

URINARY PROTEOME ANALYSIS USING CAPILLARY ELECTROPHORESIS COUPLED TO MASS SPECTROMETRY: A POWERFUL TOOL IN CLINICAL DIAGNOSIS, PROGNOSIS AND THERAPY EVALUATION

ANALIZA PROTEOMA U URINU PUTEŃ KAPILARNE ELEKTROFOREZE UDRUŐENE S MASENOM SPEKTROMETRIJOM: MOĆNO SREDSTVO U KLIŃIČKOJ DIJAGNOSTICI, PROGNOŠTICI I PROCENI TERAPIJSKOG UĆINKA

Harald Mischak, Eric Schiffer, Petra Zűrbig, Mohammed Dakna, Jochen Metzger

Mosaïques diagnostics & therapeutics, Hannover, Germany

Summary: Proteome analysis has emerged as a powerful tool to decipher (patho)physiological processes, resulting in the establishment of the field of clinical proteomics. One of the main goals is to discover biomarkers for diseases from tissues and body fluids. Due to the enormous complexity of the proteome, a separation step is required for mass spectrometry (MS)-based proteome analysis. In this review, the advantages and limitations of protein separation by two-dimensional gel electrophoresis, liquid chromatography, surface-enhanced laser desorption/ionization and capillary electrophoresis (CE) for proteomic analysis are described, focusing on CE-MS. CE-MS enables separation and detection of the small molecular weight proteome in biological fluids with high reproducibility and accuracy in one single processing step and in a short time. As sensitive and specific single biomarkers generally may not exist, a strategy to overcome this diagnostic void is shifting from single analyte detection to simultaneous analysis of multiple analytes that together form a disease-specific pattern. Such approaches, however, are accompanied with additional challenges, which we will outline in this review. Besides the choice of adequate technological platforms, a high level of standardization of proteomic measurements and data processing is also necessary to establish proteomic profiling. In this regard, demands concerning study design, choice of specimens, sample preparation, proteomic data mining, and clinical evaluation should be considered before performing a proteomic study.

Keywords: clinical proteomics, capillary electrophoresis, mass spectrometry, biomarker, urine

Kratak sadržaj: Analiza proteoma je postala je moćno sredstvo za dešifrovanje (pato)fizioloških procesa, što je za rezultat imalo uspostavljanje oblasti kliničke proteomike. Jedan od glavnih ciljeva je otkrivanje biomarkera oboljenja iz tkiva i telesnih tećnosti. Zbog ogromne složenosti proteoma, pri proteomskoj analizi zasnovanoj na masenoj spektrometriji potrebno je izvršiti separaciju. U radu su opisane prednosti i ograničenja proteomske analize pri separaciji proteina putem dvodimenzionalne gel elektroforeze, tećne hromatografije, SELDI i kapilarne elektroforeze (KE), sa fokusom na KE-MS. KE-MS omogućava separaciju i detekciju proteoma male molekularne tećžine u biološkim tećnostima uz visoku reproducibilnost i preciznost u samo jednom koraku radnog postupka i za kratko vreme. Pošto pojedinaćni senzitivni i specifićni biomarkeri moćda i ne postoje, strategija za premošćivanje te dijagnostićeke praznine pomera se sa detekcije pojedinaćnog analita na simultanu analizu više analita koji zajedno ćine obrazac specifićan za dato oboljenje. Takvi pristupi, mećdutim, nose sa sobom dodatne izazove, koje ćemo predstaviti u ovom radu. Pored izbora odgovarajućih tehnoloških platformi, neophodan je visok nivo standardizacije proteomskih merenja i obrade podataka kako bi se vršilo profilisanje proteoma. U tom pogledu, zahtevi koji se tiću nacрта studija, izbora primeraka uzoraka, analize proteomskih podataka i klinićeke evaluacije trebalo bi da budu razmotreni pre izvođenja proteomske studije.

Ključne reći: klinićeke proteomika, kapilarna elektroforeza, masena spektrometrija, biomarker, urin

Nonstandard Abbreviations: PSA, prostate specific antigen; GFR, glomerular filtration rate; SDS-PAGE, sodium dodecylsulfate polyacrylamide gel electrophoresis; 2DE, two-dimensional electrophoresis; LC, liquid chromatography; SELDI, surface-enhanced laser desorption/ionization; CE, capillary electrophoresis; ESI, electrospray ionization; TOF, time-of-flight; (MS)/MS, (tandem) mass spectrometry; 2D-DIGE, two-dimensional difference gel electrophoresis; ROC, Receiver Operating Characteristic; HSP, heat shock protein; CKD, chronic kidney disease; DN, diabetic nephropathy; PCa, prostate cancer; HSCT, allogeneic hematopoietic stem cell transplantation; GVHD, graft-versus-host-disease; CAD, coronary artery disease.

Address for correspondence:

Harald Mischak
Mosaïques diagnostics & therapeutics AG
Mellendorfer Str. 7-9, 30625 Hannover
Phone +49-511-554744-13
Fax +49-511-554744-31
email: metzger@mosaïques-diagnostics.com

Requirements for clinical proteomic studies

In initial clinical proteomic studies applying relatively crude methods and analyzing a limited number of study subjects, differences between patients and controls could be observed (1, 2). However, it soon became evident that the dynamics of the human proteome are far more complex than expected, and differences of a similar order of magnitude can be found, when comparing controls with controls. As presented schematically in *Figure 1*, many organizational, technical, and procedural aspects should be taken into consideration for the successful completion of a clinical proteomic study. These were outlined in detail recently (3), and will be discussed shortly here.

Study design

Study design must involve proper selection of patient and control groups, choice of sample material, and the selection of a proteomic platform that fulfills the methodological requirements. Further, cooperation between clinicians, statisticians, clinical chemists and basic scientists is required (3). It is generally obsolete to use samples from healthy individuals as the sole control group; patients with similar clinical characteristics/comorbidities must be included as additional controls. Individuals included in the study have to be selected carefully in order to exclude the possibility for marker identification to be

influenced by drug administration and other therapeutic regimens.

For reproducibility and comparability of clinical proteome analyses, it is of major importance to minimize variability of sample collection, handling, and storage. Another challenging task is to establish a uniform sampling protocol that is equally applicable to any sample, irrespective of its physical and biochemical properties. The options to control processes of sample collection and preparation are limited, particularly when multiple centers are involved and samples are transferred from the site of collection to the site of analysis. The best strategy to overcome these uncertainties of sample acquisition appears to perform large studies providing sufficient information to decide, which marker lies within or exceeds the normal range of analytical and biological variability.

This also argues for the establishment of a public database, where essential information, such as the patient's proteome profile and all clinical information, can be deposited. An example for the organization of proteomic and clinical data using a database system is shown in *Figure 2*. Ideally, with the help of appropriate software, patient groups can be selected retrospectively for inclusion in a comparative study on the basis of the recorded clinical data and by a direct linkage to the patient's proteomic profile. First step in this direction is the recently reported human urinary proteome database (4).

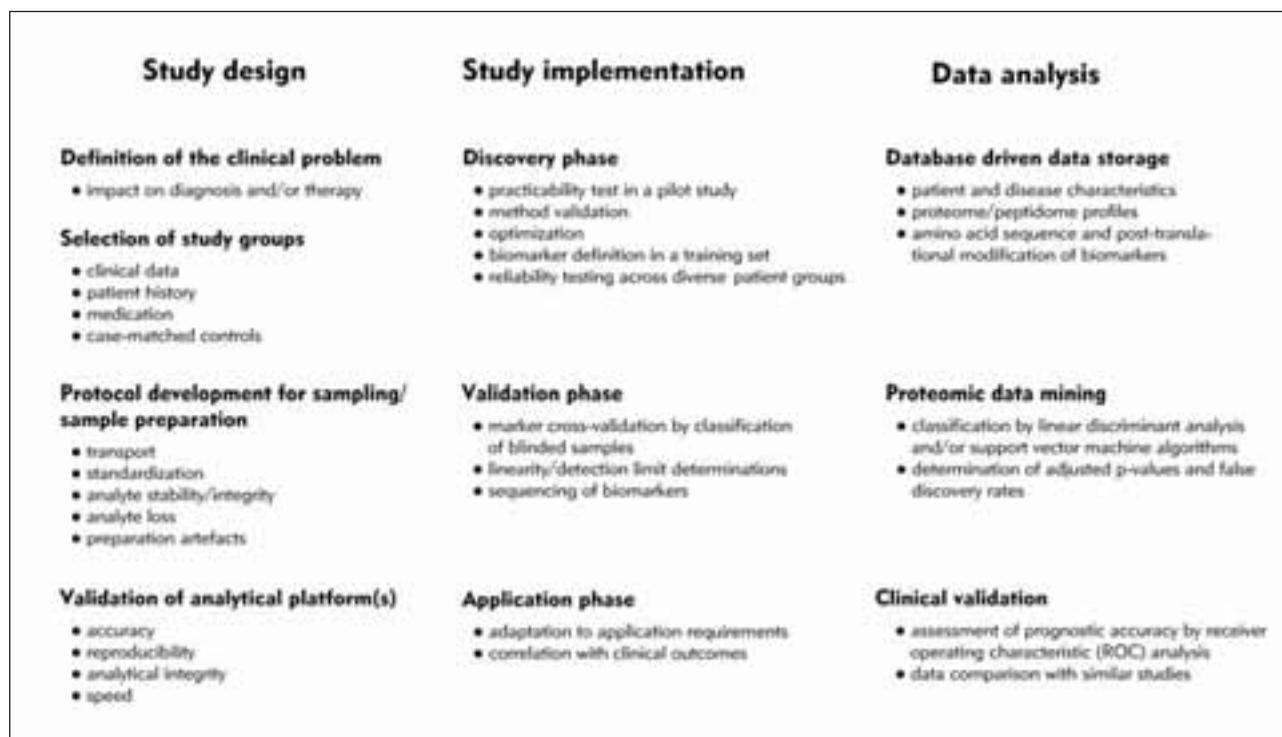


Figure 1 Areas to be considered when performing a clinical proteomic study.

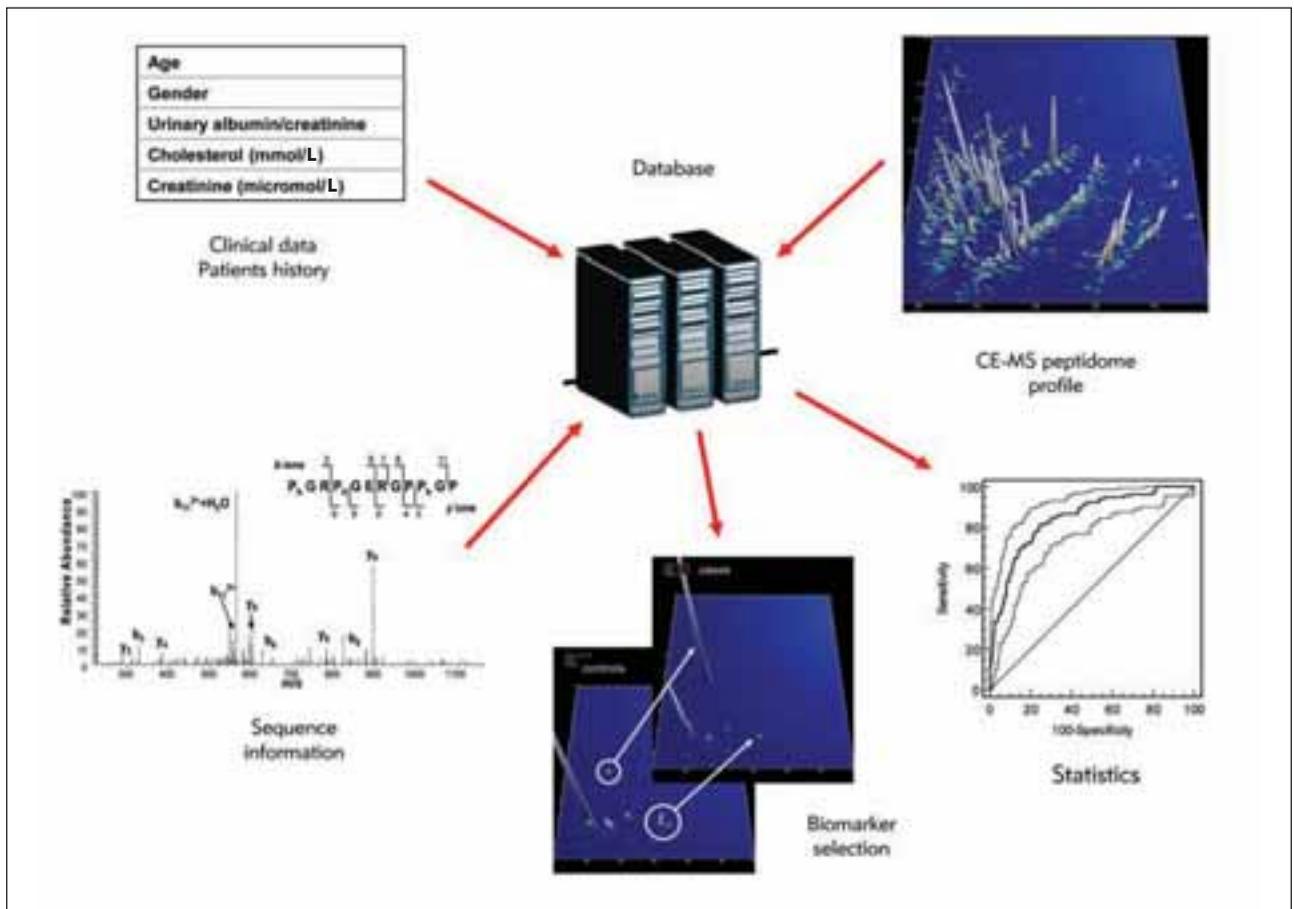


Figure 2 Composition of a database for storage and retrieval of protein/peptidome profiles, protein/peptide sequences, and patient clinical records allowing sample selection and differential proteomic profiling for biomarker discovery.

Source of specimen

Generally, proteomic profiling can be performed with tissue extracts, cell lysates, or body fluids (such as blood, urine, cerebrospinal fluid, and saliva) as biological materials. Due to proximity to the primary site of the disease, affected cells and tissues should contain the highest amounts of biomarkers. This property makes them an ideal source for biomarker definition. Unfortunately, these are not easily accessible and require invasive or surgical methods for sampling. This also applies to samples from cerebrospinal, synovial, and body cavity fluids. The poor accessibility of tissue is reflected by the fact that to date proteomic studies using tissue samples are rare (5, 6). However, as outlined recently by Lescuyer et al. (7), such approaches may be more successful than serum (or plasma)-based approaches, which, as the authors point out, have not resulted in measurable diagnostic success to date.

Malignant transformation and increased death of cells in disease-affected tissues and organs is associated with diffusion of tissue- and organ-specific proteins into the extracellular space and into blood

circulation. The role of blood as a transporter for molecules to and from tissues, together with the ease of sample collection, makes blood an attractive source for biomarker discovery. Unfortunately, there are a number of problems to overcome before blood can be effectively used for clinical proteomic research (8–10). Complete identification of the human plasma proteome is currently impossible due to the high dynamic range between low- and high-abundant proteins. By removal of highly abundant proteins, most notably albumin and IgG, this dynamic range can be reduced, but this also creates new challenges, like loss of low-abundance proteins by their binding to albumin or to the resin of the subtraction column. As reported by Kolch et al. (11), persistent proteolytic activity in the blood sample is another source of experimental variability, making meaningful comparison of serum proteome data between individual samples even more challenging. These reports and considerations suggest that, while blood is certainly the richest source for biomarkers, it is highly unlikely that these can actually be identified using today's technologies.

Via proteolytic degradation, stable, protease-resistant peptide fragments are generated, traversing epithelial barriers and passing into the urine by glomerular filtration. Although peptides and small proteins of the blood, which enter the lumen of the renal tubule, are mostly reabsorbed by proximal tubular cells, a small quantity escapes this process and is excreted into urine. As a consequence, urine can be used as a biomarker source for various diseases (10, 15).

Urine as a sample matrix provides several advantages. First, it is non-invasively accessible and can be obtained in large quantities. Second, urine is relatively stable in its composition if handled properly. As reported by Schaub et al. (12), this is also the case after long storage times. Third, since urine represents the ultrafiltrate of plasma, the composition of the urinary peptidome is highly susceptible to changes caused not only by renal but also by a wide range of non-renal diseases including cardiovascular, auto-immune and infectious diseases as well as certain types of cancer.

Sample preparation

Since biological fluids are complex in their composition, their preanalytical processing is a prerequisite for an efficient MS-based analysis. Currently, a plethora of different protocols for sample preparation for different tissues and body fluids exist and it is beyond the scope of this manuscript to review them. Highly standardized and reproducible preparation protocols with the ability to eliminate interfering compounds are required to ensure analytical reproducibility. Furthermore, it should be kept in mind that each additional step in sample preparation will introduce new, additional artifacts. Hence, preanalytical manipulation should be robust and kept to a minimum.

Biomarker quantification

A limitation of proteomic methods in respect to their clinical applicability was the inability to specify the amount of a particular protein or peptide in a given sample. To solve this problem, several MS-based quantification strategies have been developed, many of which are based on the use of isotope-coded labels (13). These, however, are time-consuming and expensive. As a consequence, further efforts were made to develop more accurate and simple quantification strategies based on signal intensity/ion counting (14). While these measures, in contrast to stable isotope labelled internal standards (15), do not permit absolute analyte quantification, they were specially adapted to perform relative quantification with acceptable deviation characteristics (+/- 10%, (16) and own observation). In recent experiments,

both absolute and relative quantification gave highly similar results, with clear methodological advantages of the ion counting approach (17) as it avoids direct manipulation of the specimens. In addition, ion counting procedures are not limited to the analysis of biomarkers with known peptide sequence, making it an ideal tool for biomarker discovery.

Proteomic data mining

Due to the large amount of information provided by a single proteome analysis, adequate software solutions are required to correctly interpret and process proteomic data. Mandatory features include the ability to determine the charge of a particular peak, to identify and combine peaks of the same analyte at different charge states, and to perform efficient standardization/normalization to compensate for differences between individual measurements.

Biomarkers are defined from these datasets, generally based on multivariate statistical analyses of datasets. The subsequent classification algorithm may be based among others on linear discriminate analysis (18) or support vector machine (19). As with any classification procedure, these methods have their own advantages and drawbacks. It should be noted, however, that neither of these supervised learning methods do include a variable selection procedure per se. Therefore, statistical evaluation of the different peptides appears mandatory. Nevertheless, a given biomarker showing statistical significance does not automatically perform well as a class discriminating item. Considering the high dimensionality of the statistical problems, the stochastic analysis must correct for multiple testing artifacts inherent to such analyses. An example may illustrate, why this is of utmost importance: The presumption is that n independent tests using 0.05 as the critical significance level are performed. The probability for a single test to come to a non-significant result (that is a correct conclusion) is hence $1-0.05 = 0.95$ (95%). Since the n tests are independent from each other, the probability that all n tests correctly reject the n null hypothesis is determined by the product of the single results: $0.95 \times \dots \times 0.95 = 0.95^n$. Hence, the probability to at least wrongly reject one of the n null hypotheses is given by $1-0.95^n$. If our experiment involves 200 tests on 200 biomarkers, the experimental error probability is $1-0.95^{200} = 0.999965$. In other words, it is almost certain that when performing 200 tests on 200 biomarkers several of the declared significant findings are false positive. Because of the test independence, the probability of k such false positives among n biomarkers is given by the binomial distribution with the significance level α as the probability of »success« (i.e. having a false positive). In the example of 200 biomarkers tested at the significance level of 0.05,

this probability amounts for $k = 5$ to 0.97355. Even for $k = 8$, the probability that the findings are false positives is still 0.78669. Bonferroni corrections, and their relatives such as the Holm procedure, are the most widespread approach for controlling the experiment-wide false positive rate (20). Distribution-free re-sampling methods, like from Westfall and Young (21), are excellent methods to control the experimental error rate. A major drawback of these procedures is that they may lack sufficient statistical power. This has led Benjamini and Hochberg to introduce the elegant approach of false discovery rate, which conserves more statistical power (22).

These reports and the application of statistical methods on a theoretical example shown above clearly underline the importance of using proper statistics. If not strictly observed, the data obtained will likely hold no value, and will be proven invalid in the next set of experiments.

Clinical validation

The performance of the biomarker(s) defined as the ability to correctly classify samples into healthy and diseased subjects may best be expressed as a Receiver Operating Characteristic (ROC) plot, since this will indicate the degree of overlap between the two groups by plotting the sensitivity against 1-specificity at each level (23). ROC plot analysis is the method of choice, since it has the property to be independent of the prevalence of the disease in the sample cohort.

After identification of candidate biomarkers in a training set of well-defined samples, their discriminating and prognostic value must be validated in a second independent sample collective to prevent overfitting of training set data. Additionally, the validation process should test the biomarker's ability to discriminate between the disease of interest and other diseases and health conditions. For example, heat shock proteins, in particular HSP70, which were described as biomarkers for certain types of cancer (24–26), are frequently released by affected cells after the onset of other, non-cancer associated stimuli, such as oxidative stress, heavy metals, tobacco smoke, and metabolic poisons (27). Therefore, it is difficult to utilize HSP up-regulation as a specific tumor marker.

MS-based proteome analysis

Two-dimensional separation of proteins according to their isoelectric point and molecular weight in sodium dodecylsulfate polyacrylamide gels (SDS-PAGE), first described by O'Farrell (28), provided the basis for proteomic research. It soon became clear that high resolution protein separation

is only one part of the solution. The other is unambiguous identification of the proteins. Initially, proteins excised from the gel were examined by Edman degradation, a tedious and often unsuccessful procedure. Therefore, whenever possible, detection was performed by nitrocellulose transfer and staining with specific antibodies (29). Subsequently, the implementation of mass spectrometry led to a step-by-step identification of hundreds of proteins based on a proteolytic in-gel digest, gel extraction, and MS analysis of the resultant peptide fragments (30).

In recent years, two major technologies emerged, which allowed high-throughput screening of proteomes. These two methods are proteomic microarrays and separation technologies coupled to mass spectrometry. Protein-detecting microarrays rely on the development of antibody engineering technologies and automated spotting techniques for biomolecule immobilization onto solid supports and will not be covered here [For more details see (31)]. Electrophoretic and chromatographic separation technologies, such as two-dimensional electrophoresis (2DE), liquid chromatography (LC), surface-enhanced laser desorption/ionization (SELDI) and capillary electrophoresis (CE), were developed concurrently with the microarray technology, and were coupled to mass spectrometers with different ionization sources, i.e. matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI), and analyzing systems, i.e. quadrupole, time-of-flight (TOF) or Fourier transform ion cyclotron resonance. We will briefly outline the advantages and shortcomings of the different technological platforms, and subsequently focus on CE-MS. For a more detailed review of the technologies, we refer to (32).

Two-dimensional gel electrophoresis coupled to mass spectrometry (2DE-MS)

In 2DE, proteins are resolved by electro-focusing according to their isoelectric point followed by orthogonal separation by SDS-PAGE. MS detection of proteolytic digests has become the method of choice for protein identification from the 2D-gels: tryptic digestion of excised protein spots, extraction of the proteolytic fragments, and, as soon as mass information of at least three fragment ions is available, comparison with public databases provided by e.g. the National Center for Biotechnology Information. Matches can subsequently be verified by tandem mass spectrometric (MS/MS) sequencing or immunoblotting.

2DE-MS is a time-consuming, technically challenging approach with high analytical (gel-to-gel) variability, compromising comparison between samples. While the latter is to an extent solved by the use of two-dimensional difference gel electrophoresis (2D-DIGE) (33), enabling simultaneous resolution of two differentially labelled samples within the same 2D gel,

comparison of several different experiments still remains challenging. Furthermore, the approach is by far too time-consuming to be applied in clinical routine laboratories. Hence, potential biomarkers defined by 2DE-MS have to be transferred to an application platform, where they need to be validated. Despite these limitations, 2DE-MS remains a commonly used technique for comparative analysis of large proteins and definition of potential biomarkers >20 kDa (34).

Liquid chromatography coupled to mass spectrometry (LC-MS)

LC represents a very versatile high resolution separation method. A great variety of LC columns (e.g. reversed phase, ion exchange, size exclusion), are available that allow separation of large amounts of analyte. LC can be combined with any mass spectrometer. LC-MS was further extended to multi-dimensional separation with different separation media. Most important in this field is multidimensional protein identification technology (MudPIT), which uses cation exchange pre-fractionation followed by reversed phase separation and MS/MS detection (35). Disadvantageous is the considerably long period of time for analysis, making its use for high-throughput screening of hundreds of samples difficult. Further, comparative examination of datasets obtained from multi-dimensionally separated samples still represents an unresolved challenge. Other challenges of the LC-method are lipids and deter-

gents in the sample interfering with separation and detection, and analytes precipitating on the column material. As 2DE, the approach is currently not suited for routine clinical application.

Surface-enhanced laser desorption/ionization coupled to mass spectrometry (SELDI-MS)

SELDI uses selective adsorption of proteins to different active surfaces, e.g. hydrophilic, reversed-phase, or affinity reagents as lectins or antibodies, to reduce the complexity of a given biological sample. Application of matrix material followed by laser desorption allows soft laser ionization for MS detection. SELDI has been used for the tentative identification of biomarkers for a variety of diseases. Unfortunately, data in general could subsequently not be validated (e.g. (36)). The advantages of SELDI are its ease of operation and high-throughput capabilities, and low sample volume requirements. Several limitations have essentially precluded the use of SELDI for clinical diagnostic purposes. These include difficulties with inter-laboratory comparison of datasets, since the proteome profiles generated by SELDI are influenced by factors, such as the type of surface coating, pH and salt conditions, and protein concentration of the sample. Different conditions and chip surfaces lead to different datasets from the same sample, resulting in a frequently observed complete lack of comparability of datasets. Other concerns raised are lack of mass precision.

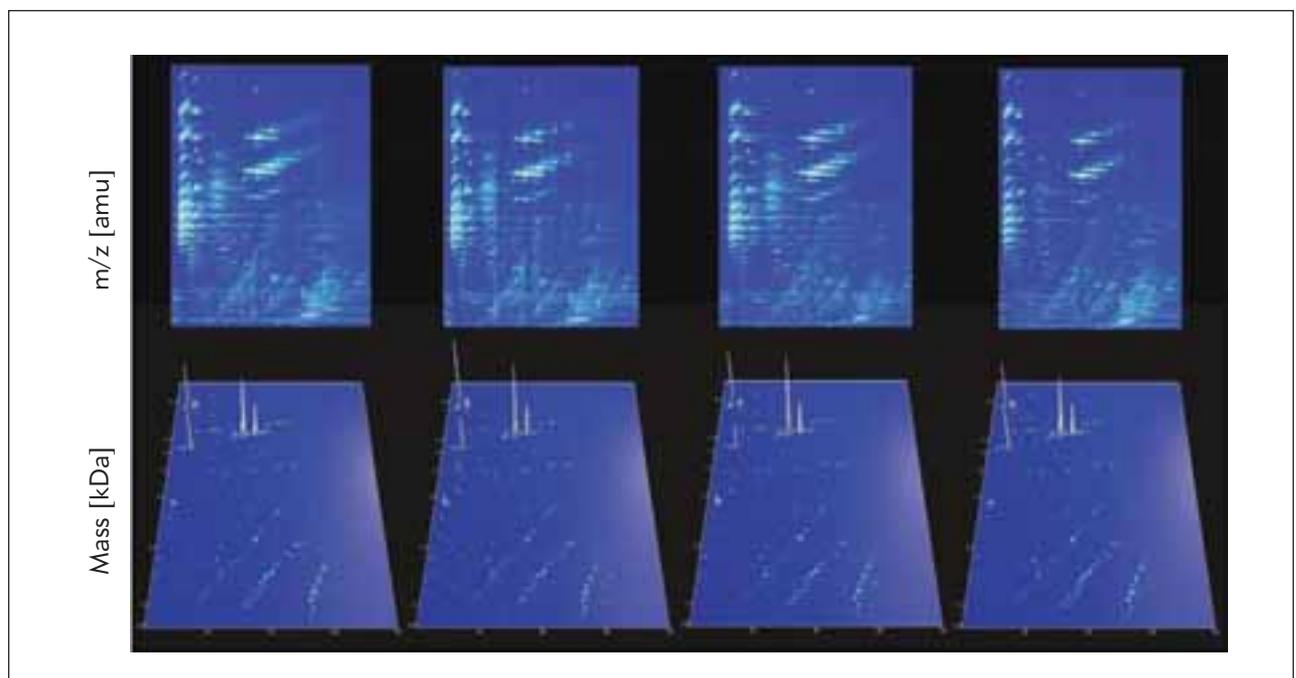


Figure 3 Reproducibility of CE-MS. Peptide profiles of four different measurements of a single rat urine sample. Representation of peptide CE-MS patterns by plotting m/z values (upper panels) or log molecular weights (lower panels) against CE-migration times (reprinted with permission from (69)).

Capillary electrophoresis coupled to mass spectrometry (CE-MS)

CE separates analytes with high resolution based on differential migration through a liquid-filled capillary in an electric field. By online coupling of CE to an electrospray ESI-TOF-MS, analysis of thousands of polypeptides within a time range of 45 to 60 min can be performed (11). CE and LC are similar with respect to resolution and compatibility with mass spectrometers. Advantages of CE are absence of buffer gradients and speed (37). Furthermore, CE is generally insensitive towards precipitating proteins and peptides that often interfere with LC-separation, and the capillary can be reconditioned with NaOH, enabling efficient cleaning after each run. Due to these capabilities, CE-MS can be applied for the separation of virtually all naturally occurring peptides and small proteins. This enables reproducible separation and detection of the low molecular weight proteome of any sample in one single step. As an example for the reproducibility of CE-MS, four peptide profiles of the same urine sample are presented in *Figure 3*. CE-MS has been successfully used for the identification and validation of specific biomarker patterns in several studies (38).

CE is limited in the separation of larger proteins (>20 kDa) due to the generally low pH used, although to a lesser extent than LC. Another limitation is the small sample volume that can be applied onto the capillary, which greatly hampers CE-MS/MS applications, but, due to the high sensitivity of mass spectrometers in the low fmol range, is of little concern in CE-MS. Today, CE-MS has been proven a stable and versatile platform for low molecular weight proteomic profiling.

Bottom-up and top-down proteomics

In a top-down experiment, intact proteins and peptides are subjected to analysis by mass spectrometry. In bottom-up analyses, proteins are digested by proteases and the generated peptides are analyzed by MS/MS. While the latter method is best suited for the identification of large proteins, the former is more applicable for peptides and small proteins and gives a more accurate definition of the potential biomarkers. As outlined recently, any biomarker should be defined by its accurate molecular composition (39). The identification of a theoretical protein based on a few tryptic fragments may be quite misleading, as it generally does not allow accounting for posttranslational modifications, which may in fact confer »biomarker quality« to the protein: glycated albumin may serve as a biomarker for diabetes, while albumin precursor, which would be defined as biomarker based on several tryptic peptides, certainly does not (40). If albumin precursor was defined as biomarker based on a top-down 2DE- or LC-MS/MS approach,

the subsequent validation of the results using an alternative technology for clinical application would have failed, due to the inaccurate definition of the original biomarker.

With the exploitation of top-down strategies for CE-MS, LC-MS, and SELDI-MS, the low molecular mass range of the proteome, also termed peptidome, came into focus as a source of information. The peptidome consists of all naturally occurring peptides, many of which are the products of proteolytic degradation. The rationale behind performing peptidomic analysis is provided by the finding that many disease states are reflected by changes in the peptide composition of biological fluids. As suggested by Haubitz et al. (41), Villanueva et al. (42), or Rossing et al. (43), some of these peptides are produced by (disease-)specific proteases. Altered activity of proteases may be more readily assessed by analysis of reaction products, the proteolytic fragments generated, than by direct assessment of the proteases themselves.

It appears that the top-down approaches are currently better suited for clinical applications than the bottom-up methods. This is in part due to the more accurate definition of biomarkers including posttranslational modifications and the apparently higher resolution obtained in the top-down approach (while distinguishing between e.g. 4700 and 4701 Da can easily be accomplished, it is almost impossible to distinguish between 60000 and 60001 Da). In addition, the higher throughput, producing larger numbers of independent samples for statistical analysis, appears to be a prerequisite in a multiple parameter setting. Hence, while the high-molecular weight proteome, generally analyzed via the bottom-up approach, may hold more information than the low molecular weight proteome/peptidome, its information is largely not accessible in a statistically sound approach. In contrast, the low molecular weight proteome/peptidome can be assessed in a statistically meaningful way.

Single versus multi-marker applications

Biomarkers are molecules (generally assessed in body fluids) used as indicators for the detection of pathological changes or disease states, drug response, etc. Many of these are proteins and peptides, the exclusive focus of this review. It is important to note that a biomarker is not only defined by its molecular structure, but also by its intended use. It can only be utilized for the intended use, but not beyond (e.g. a biomarker that indicates disease at an advanced state cannot be used as early predictor) until its value has been proven for this purpose (44).

The potential of a protein to serve as a biomarker depends on how selective and sensitive an assessment it enables of a (patho)physiological situa-

tion. Most of the analytes currently used for screening and diagnostic purposes have been discovered after extensive physiological and biochemical characterization of the disease. This laborious procedure resulted in the identification of single markers with often moderate diagnostic value, mostly due to low specificity. As a prominent example, prostate specific antigen (PSA) is currently widely used as a marker for prostate cancer. Its prognostic relevance, however, is the subject of ongoing debates due to a lack of specificity. This shortcoming not only results in unnecessary biopsies, but also in higher rates of false positive diagnosis (for review see (45)). Early detection of renal impairment, which is of vast importance for the initiation of renoprotective intervention, mainly relies on the detection of microalbuminuria. The significance of this marker is underlined by the observation of Mogensen (46) that glomerular filtration rate (GFR) and serum creatinine levels, as alternative estimates of kidney function, did not change within the microalbuminuria range. However, microalbuminuria is also found in apparently healthy individuals, and cannot be utilized as a predictive marker of renal disease (47). These two examples underline the need for other more accurate biomarkers to overcome the limitations of today's diagnostics.

It appears questionable, if screening for a single marker enables reliable, early detection of disease, unambiguously distinguishing it from other pathological conditions, and/or monitoring the efficacy of therapy. An alternative strategy is the identification of several markers, which may not be »optimal« for use alone, but work in concert, and combining them to a disease-specific pattern (32). Recently, a number of different proteomic technologies have been introduced to establish disease-specific marker patterns for clinical diagnosis and therapeutic monitoring as an alternative to single biomarker based approaches. This multi-marker approach is now widely accepted, but it comes with several challenges, which were unknown or not fully appreciated and therefore disregarded in the beginning of the proteomics era. As outlined recently in suggestions for guidelines for clinical proteomics, the general criteria that are applied onto a biomarker (e.g. known identity, reproducible detection, known deviation) also apply for the single biomarkers in a multi-marker panel (3). The initial enthusiasm and subsequent failure to deliver valid results were mostly a consequence of ignoring these principles. It is now evident that an ill defined »pattern« does not constitute a multi-marker panel and cannot serve as a clinical diagnostic tool.

Application of CE-MS technology for urinary biomarker discovery

Urinary biomarkers for renal diseases

Chronic kidney disease (CKD) is characterized by a slow, but progressive loss of renal function. Renal

biopsy, which is the standard method to differentiate between glomerular, tubular and vascular renal diseases, entails the risk of procedural complications. For differential diagnosis and continuous monitoring of disease progression, a non-invasive alternative to renal biopsy is highly desirable. For this reason, proteomic studies were initiated to identify urinary biomarkers for all types of CKD.

In an attempt to define urinary polypeptide markers specific for membranous glomerulonephritis, Neuhoff et al. used SELDI and CE-MS as proteomic platforms (48). In two clinical studies (49, 50), CE-MS profiles from patients with diabetes type I or II with/without macroalbuminuria and healthy volunteers were analyzed to create stage-specific polypeptide patterns. In patients with type II diabetes mellitus and unchanged albumin excretion rate, the detected peptide pattern differed significantly from that in patients with high-grade albuminuria. Comparable results were obtained for patients with diabetes type I, suggesting that the urinary proteome contains a much greater variety of polypeptides than previously demonstrated. Further, the effects of therapeutic intervention could be clearly demonstrated by CE-MS in an independent study on microalbuminuric subjects after Candesartan treatment (51). The results were recently confirmed in a study on 500 case and control samples (52). The authors demonstrate the efficiency of urinary peptidome profiling by CE-MS to detect both diabetes and DN, and to predict development of DN in a blinded, prospectively collected population.

Prediction of disease development was also demonstrated by Decramer et al. (53), who applied CE-MS-based urinary proteome analysis to define specific biomarker patterns for different grades of ureteropelvic junction obstruction, a frequently encountered pathology in newborns. In their blinded prospective study, the biomarker patterns could predict the clinical outcome of newborns without signs of proteinuria with 95% accuracy nine months in advance. The accuracy was increased even further to 97% after 12 months (54). Those data not only indicated the potential of urinary proteomics to enable the diagnosis of renal disease, but also suggested the potential to gauge the prognosis.

Urinary biomarkers for urological disorders

Theodorescu et al. (55) described the detection and validation of biomarkers of urothelial carcinoma using CE-MS. A specific biomarker pattern was established in a training set composed of 46 patients with urothelial carcinoma and 33 healthy subjects and refined further using CE-MS profiles of 366 urine samples from healthy volunteers and patients with malignant and non-malignant genitourinary diseases. By this two-step biomarker discovery approach, a

biomarker model composed of 22 urinary peptides was established, which, when applied to a blinded test set containing 31 urothelial carcinoma patients, 11 healthy individuals and 138 non-malignant genitourinary disease patients, correctly classified all urothelial carcinoma patients and all healthy controls. Differentiation between bladder cancer and other malignant and non-malignant diseases was accomplished with 86–100% specificity.

In a study on prostate cancer (PCa), the importance of proper sampling was underlined (56). While initially investigated midstream urine did not enable identification of valid biomarkers, first void urine served as a source for PCa-specific biomarkers, indicating that the identified biomarkers may originate from secretions of the prostate into urine. After refinement of the PCa-specific biomarker pattern using urine samples from 51 PCa patients and 35 patients with benign hyperplasia as well as 184 controls, a model based on 12 potential biomarkers was established and validated in a second blinded set of 264 patient samples. In combination with PSA, it enabled correct identification of 90% of the PCa samples with 61% specificity.

Application of urinary proteome analysis to non-renal diseases

Urine samples from 40 patients after allogeneic or autologous hematopoietic stem cell transplantation and 5 patients with sepsis were examined with CE-MS analysis (57, 58). A pattern consisting of 16 polypeptides indicating early graft-versus-host-disease (GVHD) discriminated patients with GVHD from those without GVHD with 82% specificity and 100% sensitivity. A subsequent blinded multicenter validation study of more than 100 patients with 599 samples collected prospectively confirmed these results (59). Currently an intervention trial comparing preemptive therapy based on proteome profiling compared to standard treatment is being conducted. First results indicate a benefit for the yet limited number of patients (Weissinger and Metzger, unpublished).

Zimmerli et al. (60) examined patients undergoing coronary artery bypass grafting or patients after acute coronary infarction. Urine samples from patients and controls were analyzed using CE-MS to identify coronary artery disease (CAD) specific biomarkers. In a blinded assessment of 59 samples, specific urinary biomarkers identified CAD patients with 98% sensitivity and 83% specificity. These data could further be validated in an independent study, where the authors could also present data on the identification of several urinary collagen fragments as biomarkers for CAD, and showed clear evidence for the relevance and abundance of different types of collagen in the arteriosclerotic plaques (61). In a further independent

study on prospectively collected samples from patients with type I diabetes, the urinary biomarker panel for CAD could be further validated (62). More importantly, it could be demonstrated in this study, that the urinary polypeptide patterns not only allowed accurate diagnosis, but also provided proof that patients with suspicious proteome profiles were at doubled risk (OR 2.2 [1.3–5.2]; $P=0.0016$) to suffer from acute vascular events on average 1.4 ± 1.3 years following the baseline visit compared to unsuspecting controls. The proteome pattern has proven its ability to identify patients at risk to undergo future CAD events.

Pathophysiological role of biomarkers

Albeit many potential urinary biomarkers defined by CE-MS profiling have not been sequenced yet, sequences are available for more than 400 different urinary peptides (4). Many of these peptides derived from abundant proteins: albumin, beta 2-macroglobulin, uromodulin, or collagen. Consequently, a valid question is whether peptidomics is not just another way to measure glomerular injury, that could be assessed by measuring albuminuria (63). The fact that differential diagnosis based on urinary proteome analysis is possible (32, 41, 64, 65) and that patients in complete remission without albuminuria still exhibit apparently disease-specific changes in urinary polypeptides (66) strongly suggests that these peptides contain information about the pathogenesis and are not mere degradation products. It is tempting to speculate that these peptides are indicators of disease-specific protease activity, as suggested by Haubitz et al. (64). This hypothesis is strengthened by findings recently published by Nemirovskiy et al. (67), where the presence of specific collagen fragments correlated with the disease-specific activity of matrix metalloproteases. As another indication, our recent findings that collagen fragments are significantly reduced in diabetic urine (23) fit with the observed increase of collagen and extracellular matrix in patients with diabetes and DN described by Cooper et al. (62). This further supports the hypothesis that reduced activity of proteases and protection of the extracellular matrix from proteolysis by advanced glycation endproducts may be key pathological changes in diabetes mellitus (43).

A similar scenario may be applicable to albuminuria. Consequently, an albumin-derived biomarker is not simply »an albumin fragment«, but rather a specific fragment, defined by its specific C- and N-terminus. Consistent with this view, the presence of urinary fragments of albumin and alpha-1-antitrypsin associated with nephrotic syndrome in chronic kidney disease have recently been described (68). A thorough examination of the sequences of urinary peptides and comparison with protease specificities may strengthen the above hypothesis and lead to better insight into regulation and the pathological role of proteases in disease.

Concluding remarks

Separation technologies coupled to MS currently appear the preferred option for a generic approach to identify and evaluate biomarker profiles in individuals. The available separation techniques 2DE, LC, SELDI, and CE differ in respect to throughput, robustness, accuracy, and reproducibility. It appears that CE-MS fulfills the requirements for broad application in routine clinical practice, as indicated by the validation of GVHD, renal disease, prostate and bladder cancer specific marker patterns in hundreds of patient samples (4, 53, 55, 56, 65). It must be stated, however, that future implementation of proteome profiling in laboratory diagnosis relies on more than just technological advancements. Of equal importance are concerted efforts in the development of global standardization procedures for the conduction and reporting of clinical proteomic studies. By the adoption of standardized methods in

the identification of disease-specific biomarkers, the information provided by proteomic platforms will bring clinical chemists a step closer to the ultimate goal to capture critical pieces of information of a particular disease in one single diagnostic step. This hopefully will result in the integration of MS-based proteomic methods into the armamentarium of clinical laboratories.

Acknowledgments. This work was supported in part by EUROTRANS-BIO grant ETB-2006-016 and EU Funding through InGeniousHyperCare (LSHM-C7-2006-037093) and Geninca (HEALTH-F2-2008-202230).

FINANCIAL DISCLOSURES: HM is the founder and co-owner of Mosaiques Diagnostics GmbH (Hanover, Germany), who developed the CE-MS profiling technology for clinical application. ES, PZ, MD, and JM are employees of Mosaiques Diagnostics GmbH.

References

- Anderson NG, Anderson NL, Tollaksen SL. Proteins of human urine. I. Concentration and analysis by two-dimensional electrophoresis. *Clin Chem* 1979; 25: 1199–210.
- Argiles A, Mourad G, Mion C, Atkins RC, Haiech J. Two-dimensional gel electrophoresis of urinary proteins in kidney diseases. *Contrib Nephrol* 1990; 83: 1–8.
- Mischak H, Apweiler R, Banks RE, Conaway M, Coon JJ, Dominiczak A, et al. Clinical Proteomics: a need to define the field and to begin to set adequate standards. *Proteomics Clin Appl* 2007; 1: 148–56.
- Coon JJ, Zürgbilg P, Dakna M, Dominiczak AF, Decramer S, Fliser D, et al. CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics. *Proteomics Clin Appl* 2008; 2: 964–73.
- Castronovo V, Kischel P, Guillonneau F, De LL, Defechereux T, De PE, et al. Identification of specific reachable molecular targets in human breast cancer using a versatile ex vivo proteomic method. *Proteomics* 2007; 7: 1188–96.
- Roessler M, Rollinger W, Mantovani-Endl L, Hagmann ML, Palme S, Berndt P, et al. Identification of PSME3 as a novel serum tumor marker for colorectal cancer by combining two-dimensional polyacrylamide gel electrophoresis with a strictly mass spectrometry-based approach for data analysis. *Mol Cell Proteomics* 2006; 5: 2092–101.
- Lescuyer P, Hochstrasser D, Rabilloud T. How Shall We Use the Proteomics Toolbox for Biomarker Discovery? *J Proteome Res* 2007; 6: 3371–6.
- Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 2005; 5: 3226–45.
- Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, et al. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* 2005; 5: 3262–77.
- Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002; 1: 845–67.
- Kolch W, Neussuss C, Pelzing M, Mischak H. Capillary electrophoresis-mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. *Mass Spectrom Rev* 2005; 24: 959–77.
- Schaub S, Wilkins J, Weiler T, Sangster K, Rush D, Nickerson P. Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry. *Kidney Int* 2004; 65: 323–32.
- Righetti PG, Campostrini N, Pascali J, Hamdan M, Astner H. Quantitative proteomics: a review of different methodologies. *Eur J Mass Spectrom* (Chichester, Eng) 2004; 10: 335–48.
- Zybilov B, Coleman MK, Florens L, Washburn MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal Chem* 2005; 77: 6218–24.
- Barnidge DR, Dratz EA, Martin T, Bonilla LE, Moran LB, Lindall A. Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Anal Chem* 2003; 75: 445–51.
- DeKeyser SS, Li L. Matrix-assisted laser desorption/ionization Fourier transform mass spectrometry quanti-

- tation via in cell combination. *Analyst* 2006; 131: 281–90.
17. Jantos-Siwy J, Schiffer E, Brand K, Schumann G, Rossing K, Delles C, et al. Quantitative Urinary Proteome Analysis for Biomarker Evaluation in Chronic Kidney Disease. *J Proteome Res* 2009; 8: 268–81.
 18. Shannon W, Culverhouse R, Duncan J. Analyzing microarray data using cluster analysis. *Pharmacogenomics* 2003; 4: 41–52.
 19. Smola AJ, Scholkopf B. A tutorial on support vector regression. *Statistics and Computing* 2004; 14: 199–222.
 20. Abdi H. Bonferroni and Sidak corrections for multiple comparisons. Thousand Oaks (CA): Sage, 2007.
 21. Westfall PH, Young SS. Resampling-based Multiple Testing: Examples and Methods for P-Value Adjustment. New York: Wiley, 1993.
 22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B (Methodological)* 1995; 57: 125–33.
 23. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993; 39: 561–77.
 24. Jaattela M. Escaping cell death: survival proteins in cancer. *Exp Cell Res* 1999; 248: 30–43.
 25. Helmbrecht K, Zeise E, Rensing L. Chaperones in cell cycle regulation and mitogenic signal transduction: a review. *Cell Prolif* 2000; 33: 341–65.
 26. Jolly C, Morimoto RI. Role of the heat shock response and molecular chaperones in oncogenesis and cell death. *J Natl Cancer Inst* 2000; 92: 1564–72.
 27. Lindquist S, Craig EA. The heat-shock proteins. *Annu Rev Genet* 1988; 22: 631–77.
 28. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975; 250: 4007–21.
 29. Burnette WN. »Western blotting«: electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem* 1981; 112: 195–203.
 30. Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chem Rev* 2001; 101: 269–95.
 31. Kozarova A, Petrinac S, Ali A, Hudson JW. Array of informatics: Applications in modern research. *J Proteome Res* 2006; 5: 1051–9.
 32. Fliser D, Novak J, Thongboonkerd V, Argiles A, Jankowski V, Girolami M, et al. Advances in urinary proteome analysis and biomarker discovery. *J Am Soc Nephrol* 2007; 18: 1057–71.
 33. Wu TL. Two-dimensional difference gel electrophoresis. *Methods Mol Biol* 2006; 328: 71–95.
 34. Thongboonkerd V. Recent progress in urinary proteomics. *Proteomics Clin Appl* 2007; 1: 780–91.
 35. Delahunty CM, Yates JR, III. MudPIT: multidimensional protein identification technology. *Biotechniques* 2007; 43: 563, 565, 567.
 36. McLerran D, Grizzle WE, Feng Z, Thompson IM, Bigbee WL, Cazares LH, et al. SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate cancer. *Clin Chem* 2008; 54: 53–60.
 37. Gaspar A, Englmann M, Fekete A, Harir M, Schmitt-Kopplin P. Trends in CE-MS 2005–2006. *Electrophoresis* 2008; 29: 66–79.
 38. Mischak H, Julian BA, Novak J. High-resolution proteome/peptidome analysis of peptides and low-molecular-weight proteins in urine. *Proteomics Clin Appl* 2007; 1: 792–804.
 39. Good DM, Thongboonkerd V, Novak J, Bascands JL, Schanstra JP, Coon JJ, et al. Body Fluid Proteomics for Biomarker Discovery: Lessons from the Past Hold the Key to Success in the Future. *J Proteome Res* 2007; 6: 4549–55.
 40. Wa C, Cerny RL, Clarke WA, Hage DS. Characterization of glycation adducts on human serum albumin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Chim Acta* 2007; 385: 48–60.
 41. Haubitz M, Fliser D, Rupperecht H, Floege J, Haller H, Rossing K, et al. Defining renal diseases based on proteome analysis. *Nephrology Dialysis Transplantation* 2005; 20: V20.
 42. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 2006; 116: 271–84.
 43. Rossing K, Mischak H, Rossing P, Schanstra JP, Wiseman A, Maahs DM. The urinary proteome in diabetes and diabetes-associated complications: new ways to assess disease progression and evaluate therapy. *Proteomics Clin Appl* 2008; 2: 997–1007.
 44. Goodsaid F, Bandow JE, Mischak H. Grand Rounds in Proteomics at the FDA. *Proteomics Clin Appl* 2007; 1: 1526–31.
 45. Hernandez J, Thompson IM. Prostate-specific antigen: a review of the validation of the most commonly used cancer biomarker. *Cancer* 2004; 101: 894–904.
 46. Mogensen CE. Systemic blood pressure and glomerular leakage with particular reference to diabetes and hypertension. *J Intern Med* 1994; 235: 297–316.
 47. Rossing K. Progression and remission of nephropathy in type 2 diabetes: new strategies of treatment and monitoring. *Dan Med Bull* 2007; 54: 79–98.
 48. Neuhoff N, Kaiser T, Wittke S, Krebs R, Pitt A, Burchard A, et al. Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry. *Rapid Communications in Mass Spectrometry* 2004; 18: 149–56.
 49. Mischak H, Kaiser T, Walden M, Hillmann M, Wittke S, Herrmann A, et al. Proteomic analysis for the assessment of diabetic renal damage in humans. *Clin Sci (Lond)* 2004; 107: 485–95.

50. Meier M, Kaiser T, Herrmann A, Knueppel S, Hillmann M, Koester P, et al. Identification of urinary protein pattern in type 1 diabetic adolescents with early diabetic nephropathy by a novel combined proteome analysis. *J Diabetes Complications* 2005; 19: 223–32.
51. Rossing K, Mischak H, Parving HH, Christensen PK, Walden M, Hillmann M, Kaiser T. Impact of diabetic nephropathy and angiotensin II receptor blockade on urinary polypeptide patterns. *Kidney Int* 2005; 68: 193–205.
52. Rossing K, Mischak H, Dakna M, Züribig P, Novak J, Julian BA, et al. Urinary Proteomics in Diabetes and CKD. *J Am Soc Nephrol* 2008; 19: 1283–90.
53. Decramer S, Wittke S, Mischak H, Züribig P, Walden M, Bouissou F, et al. Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat Med* 2006; 12: 398–400.
54. Decramer S, Bascands JL, Schanstra JP. Non-invasive markers of ureteropelvic junction obstruction. *World J Urol* 2007; 25: 457–65.
55. Theodorescu D, Wittke S, Ross MM, Walden M, Conaway M, Just I, et al. Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol* 2006; 7: 230–40.
56. Theodorescu D, Schiffer E, Bauer HW, Douwes F, Eichhorn F, Polley R, et al. Discovery and validation of urinary biomarkers for prostate cancer. *Proteomics Clin Appl* 2008; 2: 556–70.
57. Kaiser T, Kamal H, Rank A, Kolb HJ, Holler E, Ganser A, et al. Proteomics applied to the clinical follow-up of patients after allogeneic hematopoietic stem cell transplantation. *Blood* 2004; 104: 340–9.
58. Weissinger EM, Hertenstein B, Mischak H, Ganser A. Online coupling of capillary electrophoresis with mass spectrometry for the identification of biomarkers for clinical diagnosis. *Expert Rev Proteomics* 2005; 2: 639–47.
59. Weissinger EM, Schiffer E, Hertenstein B, Ferrara JL, Holler E, Stadler M, et al. Proteomic patterns predict acute graft-versus-host disease after allogeneic hematopoietic stem cell transplantation. *Blood* 2007; 109: 5511–9.
60. Zimmerli LU, Schiffer E, Züribig P, Kellmann M, Mouis L, Pitt A, et al. Urinary proteomics biomarkers in coronary artery disease. *Mol Cell Proteomics* 2008; 7: 290–8.
61. Von Zur Muhlen C, Schiffer E, Zuerbig P, Kellmann M, Brasse M, Meert N, et al. Evaluation of Urine Proteome Pattern Analysis for Its Potential To Reflect Coronary Artery Atherosclerosis in Symptomatic Patients. *J Proteome Res* 2009; 8: 335–45.
62. Snell-Bergeon JK, Maahs DM, Ogden LG, Kinney GL, Hokanson JE, Schiffer E, et al. Evaluation of urinary biomarkers for coronary artery disease, diabetes, and diabetic kidney disease. *Diabetes Technol Ther* 2009; 11: 1–9.
63. Comper WD, Osicka TM, Clark M, Maclsaac RJ, Jerums G. Earlier detection of microalbuminuria in diabetic patients using a new urinary albumin assay. *Kidney Int* 2004; 65: 1850–5.
64. Haubitz M, Wittke S, Weissinger EM, Walden M, Rupprecht HD, Floege, J et al. Urine protein patterns can serve as diagnostic tools in patients with IgA nephropathy. *Kidney Int* 2005; 67: 2313–20.
65. Rossing K, Mischak H, Dakna M, Züribig P, Novak J, Julian BA, et al. Proteomic discovery and validation of urinary biomarkers for diabetes and chronic renal disease. *J Am Soc Nephrol* 2008; 19: 1283–90.
66. Weissinger EM, Wittke S, Kaiser T, Haller H, Bartel S, Krebs R, et al. Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes. *Kidney Int* 2004; 65: 2426–34.
67. Nemirovskiy OV, Dufield DR, Sunyer T, Aggarwal P, Welsch DJ, Mathews WR. Discovery and development of a type II collagen neoepitope (TIINE) biomarker for matrix metalloproteinase activity: From in vitro to in vivo. *Anal Biochem* 2006; 361: 93–101.
68. Candiano G, Musante L, Bruschi M, Petretto A, Santucci L, Del BP, et al. Repetitive fragmentation products of albumin and alpha1-antitrypsin in glomerular diseases associated with nephrotic syndrome. *J Am Soc Nephrol* 2006; 17: 3139–48.
69. Frommberger M, Züribig P, Jantos J, Krahn T, Mischak M, Pich A, et al. Peptidomic analysis of rat urine using capillary electrophoresis coupled to mass spectrometry. *Proteomics Clin Appl* 2007; 1: 650–60.

Received: June 30, 2009

Accepted: July 20, 2009