CE coupled MS (CE-MS) has become an increasingly employed technology in proteome analysis with focus on the identification of biomarker peptides in clinical proteomics. In this review, we will cover technical aspects of CE-MS coupling and highlight the improvements made in the last few years. We examine CE-MS from an application point of view, and evaluate its merits and vices for biomarker discovery and clinical applications. We discuss the principal theoretical and practical obstacles encountered when employing CE-MS (and most other proteomic technologies) for the analysis of body fluids for biomarker discovery. We will present several examples of a successful application of CE-MS for biomarker discovery, implications for disease diagnosis, prognosis, and therapy evaluation, and will discuss current challenges and possible future improvements.

Keywords: Biomarker / CE / MS / Peptides / Urine

1 Introduction

The interfacing of CE with its high resolving power in separation, with high-resolution MS has become increasingly employed in peptide analysis, especially when reproducibility and analysis of complex mixtures are an issue. In Fig. 1A, the main components of CE-MS are shown and their typical characteristics are described. Most frequently in proteomics, the sample containing peptides is separated in acidic environment in a 50–100 cm fused silica capillary of 50–75 μM diameter, at field strengths in the range of 250–500 V/cm. Resolution in the range of 100 000 theoretical plates can be routinely achieved, making it an ideal analytical platform to analyze complex mixtures containing several thousand different analytes (peptides) in one single analytical procedure.

Body fluids such as urine, plasma, cerebrospinal fluid, synovial fluid, saliva, or others contain thousands of peptides and small proteins, and several of those are specific biomarkers for disease, disease progression, or response to therapy. As outlined in detail below, CE-MS, due to its high resolving power, robustness and reproducibility, and acceptable time required for routine analysis, has been successfully used to identify such biomarkers in urine suitable for use in the clinic, and we anticipate that it will see even much broader use in this area in the near future, due to its apparent success.

2 Technical aspects of CE-MS

2.1 CE

Mostly, CZE has been utilized in MS coupling, and CE is often and also here, used synonymously for CZE. Other
initially quite promising approaches like capillary IEF appear to be less widely used, mostly due to sophisticated technology that requires exceptional experts to perform such analysis, and also due to technical limitations (e.g., the problem of background ampholytes that interfere with MS detection). Although initial manuscripts indicated that proteome analysis might be possible on a large scale with capillary IEF-MS [1], initial optimism unfortunately has not yet been substantiated with additional reports. The different types of CE modes that can be applied toward proteome analysis have recently been described in detail in excellent reviews [2–5].

2.2 Coating

Several types of internal capillary coating are described to reduce interaction of proteins and peptides with the capillary wall, as well as the electroosmotic flow. Those types of coatings and their potential advantages were described in detail in recent manuscripts and reviews [6–12]. However, both phenomena appear to be of little or no consideration at the very acidic pH of 2–2.5 that is typically used in peptide separation. In addition, the application of coatings can be a very time-intensive process, and the stability of the coating at high pH is generally poor. The latter results in the requirement to renew coating frequently if high-pH rinses are being employed for e.g. efficient cleaning of the capillary after each run [13]. As a consequence, several groups, including ours, were unable to determine a clear overall benefit of coating tested.

2.3 Coupling/interfacing

An excellent and comprehensive overview on the different methods of coupling (both ESI and MALDI) is given in a recent review [14]. CE can be coupled off-line to MALDI targets, as described in several recent manuscripts [15, 16]. However, although coupling to MALDI appears to be less technically challenging and interpretation of the data is more straightforward [13], this approach also results in loss of resolution, and higher variability of signals due to matrix
effects. Consequently, coupling to ESI appears to be the preferred option. As outlined in several reviews, sheathless and sheath-flow interfacing can be used [10, 17]. These two different types of interfacing are graphically shown in Fig. 1B. Although sheathless coupling has the benefit of improved detection limit (due to lower flow rates), it also shows reduced stability, a major disadvantage when comparing large number of samples. Consequently, the majority of reports on CE-MS in fact utilize sheath-flow interfacing [10], which represents a sensitive detection device (in the high amol range), as outlined in a number of recent articles and reviews [10, 13, 18–21]. In general, CE can be interfaced with any type of MS, similar to LC.

2.4 Calculation of migration time

A hallmark of CE-separation of biological fluids is the appearance of “streaks” of peptides, when migration time is plotted against mass (Fig. 2). These “streaks” are the result of the separation principle used. Separation is accomplished by applying electrical force onto the ions. That force in turn is dependent on the charge and on the flow resistance, which is dependent on the cross-section area of the ion. At acidic pH, all amino groups are protonated, and protons, in general, are the sole source of charge under these conditions. The position of each peptide in a CE separation can, therefore, be calculated with good accuracy and confidence if its mass and the number of basic amino acids are known [16, 22, 23].

2.5 CE-MS in comparison to LC-MS

In general, every sample that can be analyzed using LC-MS can also be analyzed using CE-MS. CE, as an orthogonal technology, holds several advantages, but also disadvantages in comparison to LC [13, 24, 25]. Among CE’s main advantages are the robustness, short run times, and ability to recondition fast using high-pH treatment (e.g. 0.1 M NaOH), resulting a total time requirement of about 60 min per run when using a 90 cm capillary; 10 min stacking, 35 min separation, and 15 min for a high pH wash, and reconditioning with running buffer. The simple separating principle with high reproducibility, and, with respect to MS interfacing, a buffer that does not change its composition during analysis, as no buffer gradient is applied are other advantages. Such changes in buffer composition require a ramping of ionization parameters to maintain optimal ESI. Another advantage is the absence of “void”. Although small or well-charged peptides can usually not be detected by LC-MS, as these do not bind to the generally used reverse-phase material, all peptides migrate in the CE at distinct positions, in general, different from salt, and hence can be detected in the absence of interfering compounds. This is also evident in the recent comparison by the European kidney and urine proteomics (EuroKUP) group (www.eurokup.org), where standard urine samples were generated and analyzed using different MS-based instruments [26]. The advantages of CE, which were very recently outlined in detail in several reviews [27, 28], are especially beneficial when analyzing a large number of

![Figure 2. Compiled CE-MS data from healthy volunteers. Contour plot of the entire urine peptidome. The molecular mass (logarithmic scale) on the y-axis is plotted against normalized CE migration time on the x-axis. The arrangement of the analyzed peptides in distinct lines is obvious. The yellow arrows indicate the “streaks” resulting from 4, 3, 2, and 1 amino group present in the peptide.](image-url)
heterogeneous samples that contain interfering compounds, such as lipids, precipitates, etc.

A disadvantage of CE is the limited loading capacity. Although milliliter quantities can be loaded onto an LC column, a CE can be filled with a maximum of ca. 1 μL; in general, only with 10–100 nL. Although pH stacking can be used very effectively, a maximum of 30–50% of the total capillary volume can be filled with sample; that volume corresponds to 0.5–2 μL when using 50 or 75 μm id capillaries with 80–100 cm length. However, the limited loading capacity does not seem to of major concern in CE-MS coupling, as the concentration of peptides in samples is relatively high, taking the detection limits of modern MS instrument into account. In a typical urine or plasma sample, several thousand peptides and small proteins can routinely be detected using CE-MS. Loading capacity appears to be of relevance in MS/MS applications, where more material is required, due to the generation of fragment ions. Recently, several approaches to on-line sample preconcentration (more than 1000-fold values are reported) have been described, these are extensively and well reviewed in [29].

2.6 Microfluidic devices

Currently efficient CE separation of a biological fluid requires a turn-around time of ~1 h. A promising approach to greatly decrease the time required for a single CE-MS analysis appears to be the application of microchips for CE separation. This approach might reduce analysis times to less than 5 min [10, 30]. Although no routine applications have been reported yet, we anticipate that such devices will further spur the CE-MS applications in clinical proteomics, since these miniaturized devices hold the promise of high speed, combined with high resolution, and excellent sensitivity [31]. However, issues such as the demand for higher sensitivity of the MS and very high resolution of the CE (due to the large number of analytes in the sample) still must be addressed.

2.7 Targeted material

The body fluid most extensively examined using CE-MS is certainly urine. Although CE-MS has been used for the analysis of other body fluids, urine appears to be particularly well suited for clinical proteomics.

Although urine was in the past considered an unstable body fluid that contained only a low amount of information, it has gained considerable interest, and some of the previously thought obstacles turned out to be more of a myth than actual facts. Urine can be obtained in large quantities, and medically trained personnel is not required for collection. The urinary proteome is quite stable, due to the fact that urine is “stored” in the bladder for a considerable amount of time before collection. This provides sufficient time for complete proteolytic processing by endogenous proteases present. Stability was well demonstrated initially by Schaub et al. [32] and Theodorescu et al. [33], who showed that the urinary proteome does not undergo any significant changes for 3 days at 4°C, or even for 6 h at room temperature. However, it is important to note that other issues will influence quality and comparability of proteomic data from urine samples. Among those are protein concentration and variability due to diet and exercise. The variation in concentration (due to, e.g., liquid intake) can be compensated for by adjustments based on urinary creatinine or abundant “urinary housekeeping peptides”; i.e., peptides present in almost every human urine sample [34]. Variability due to diet or exercise can, in part, be avoided by collection of the second urine of the day, to produce consistent proteome/peptidome data [35]. That observation is most likely due to the fact that changes due to exercise and diet display in the urine with several hours delay. Urine collected in the afternoon/evening was found to show the highest degree of variability.

Urine likely primarily reflects information on diseases of organs in direct contact with urine, such as the kidney and bladder, but also the vascular system. On the contrary, information from essentially every organ is deposited in blood. Although that potential wealth of information in the blood proteome appears as an advantage at first sight, it might turn out to be a large problem in addition to other obstacles such as, e.g., the large dynamic range, highly abundant proteins, and the activation of proteases [36].

Given the urinary proteome’s complexity that we can currently only estimate, relevant changes associated with differences among samples due to variations in procedures for collection, storage and, of course, processing are to be expected. Those issues must be taken into account. Therefore, standardized protocols for urine sampling and standard urine reference samples have been developed [26].

2.8 Sample preparation

A critical issue in (clinical) proteomics is sample preparation. Ideally, the sample should be collected in a setting compatible with the clinical situation and not be manipulated at all to directly assess analytes in the sample without any interference. Unfortunately, analysis without preanalytical manipulation can generally not be accomplished. To avoid introduction of significant artefacts, this step should be robust, highly reproducible, and kept to a minimum.

A protocol employing an ultrafiltration step in the presence of urea and SDS, followed by a desalting step on PD-10 columns, was found to be very well suited for CE-MS analysis [37]. The presence of detergent and chaotropic...
agent inhibits protein–protein interaction and limits loss of analytes due to association with other proteins. This protocol resulted in a higher comparability of data from patients with and without proteinuria. Furthermore, this protocol also gave excellent reproducibility and comparability when analyzing urine samples of other species like rodents [38], as shown in Fig. 3.

2.9 Sequencing

Although CE can be interfaced with an MS/MS instrument, direct sequencing off the CE does represent a challenge, due to the limited amount of sample that can be loaded onto the capillary. Low-intensity peaks in the MS often result in no significant signal in the subsequent MS/MS analysis.

As the theoretical migration time can be used to calculate the position of a peptide in CE-MS, sequences can be obtained using LC-MS/MS and subsequently assigned to a position in the CE-MS analysis. This approach, in combination with highly accurate precursor-ion mass determination with CE-FT-ICR analysis, has proven quite successful [39]. However, a large number of peptides that produced high-quality spectra could not be identified. We attribute this failure to the inability to correctly interpret the spectra of peptides that contain unknown PTMs.

CE fractions can be collected and spotted off-line onto a MALDI target plate. Subsequently, the polypeptides of interest can be analyzed and sequenced with MALDI-TOF/TOF [24, 40]. That approach has the advantage that the signal of interest can be located in the MS mode, and optimal fragmentation conditions can be determined without repeated separation. However, in our hands sequencing of native peptides with MALDI-TOF/TOF was unsuccessful in most cases, apparently mostly due to low sensitivity and insufficient mass accuracy.

2.10 Data evaluation, classifiers, and statistics

Several software solutions that enable the extraction of relevant information, the generation of a list of compounds defined by mass, migration time, and if possible, sequence, have been described in recent manuscripts [13, 41, 42].

Migration time shows variability due mostly to the amount of ions present in a sample. However, the relative migration time (in correlation to the other peptides present in the sample) does not change considerably. Hence, internal standards can be used to subsequently calibrate migration time. Those processes enable assignment of unique and reproducible identifying parameters to each peptide: mass and migration time, with amplitude (ion counting) as a measure of relative abundance [34].

A limitation of proteomic methods is the lack of possibilities to directly quantify analytes, due mainly to different ionization properties of the peptides that make comparison of ion signals challenging. As a consequence, chemically synthesized marker peptides with stable isotope labels were introduced as tools for absolute quantification [43]. Due to the isotope-specific mass differences, the synthetic peptides can be unambiguously identified in the mass spectra, even in the case of extremely complex protein samples. This approach is time consuming and expensive. As a consequence, substantial efforts were undertaken to develop and improve strategies based on ion counting [44], which is not suitable for absolute quantification. However, that approach can be optimized in such a way that it allows the relative quantification of peptides with deviation characteristics of approximately 10% [45]. To establish an optimal standardization method for CE-MS generated urinary peptide profiles, we used peptides that can be found with high probability and signal intensity in a urine sample, generally independent from disease. Linear regression of the average values of 29 of these peptides can be used to calibrate the

Figure 3. Reproducibility of urinary rat polypeptide evaluation. Examples of electropherograms from four (of 19) consecutive measurements of a single urine sample. The m/z values of the 2-D raw data plots (upper panel) and the molecular mass (logarithmic scale) of the deconvoluted 3-D plots (lower panel) on the y-axis are plotted against CE migration time on the x-axis. Reprinted with permission from [38].
obtained ion counts, which can serve as a measure for relative abundance. Our data suggest that ion intensities expressed as peak counts can be used with good confidence for the quantification of urinary polypeptide levels, and that absolute quantification by the addition of isotope-labeled marker peptides offered no additional benefit [34].

A main goal in clinical proteomics is the definition of differentially present proteins/peptides, requiring establishment of identity. Since identity is based on spectral information, certain deviation must be taken into account, compromising the accuracy of protein assignment. When using TOF-MS, a mass deviation of 10–50 ppm appears as an acceptable compromise between the need to assess identity with high accuracy, and at the same time avoid assigning a different identity to the same peptide. Similarly, a deviation of 1% was found to be the optimal compromise for the CE migration time.

After initial processing of peak spectra to identify protein or peptide targets, the next step is to use the data sets to conduct comparative studies on the basis of multivariate statistical analyses. Several classifiers can be employed with good success. In general, high-dimensional algorithms like support vector machines outperform linear discriminant analysis (Dakna et al., submitted). Independent of the method chosen, statistical evaluation of the different peptides appears highly beneficial to reduce the high dimensionality of the original data set. It is worth noting that a given biomarker that shows statistical significance does not automatically perform well as a class-discriminating item.

Considering the high dimensionality of the data set, the statistical analysis must correct for any multiple testing artifacts that are inherent to such an analysis, as outlined recently [13, 42]. Bonferroni corrections, and their relatives such as the Holm procedure, are the most widespread approach to control the experiment-wide false-positive rate [46]. Distribution-free resampling methods, such as those from Westfall and Young [47], are also very powerful and strict methods to control for the experimental error rate. A major drawback of those procedures is that they may lack sufficient statistical power (no significant biomarker can be identified), especially when a limited number of data sets is available. This drawback has lead Benjamini and Hochberg [46]. Distribution-free resampling methods, such as those from Westfall and Young [47], are also very powerful and strict methods to control for the experimental error rate. A major drawback of those procedures is that they may lack sufficient statistical power (no significant biomarker can be identified), especially when a limited number of data sets is available. This drawback has lead Benjamini and Hochberg [46].

3 Application of CE-MS

The field of clinical proteomics has produced a large number of reports proposing potential clinically useful biomarkers of disease, but the majority of those studies were without followup. This is most probably caused by the absence of at least three requisites in these studies: (i) use of validated preanalytical procedures and analytical platforms, (ii) use of appropriate statistics (e.g., correction for multiple testing) and sufficient patient numbers, and (iii) use of independent validation cohorts. These problems and solutions are outlined in a recent consensus article [49]. As described above, CE-MS has been identified as a validated biomarker discovery platform for urine fulfilling, the first requisite for successful clinical useful biomarker discovery. Below, we will describe CE-MS-based urinary biomarker discovery studies that used appropriate statistics, relatively large patient cohorts for discovery and independent validation cohorts. The section will focus on the detection and validation of these biomarkers potentially useful in the clinic and will not comment pathophysiological insight obtained in those studies. These will be separately discussed in the next section. CE-MS has been mostly employed in the examination of urine samples. There are no technical restrictions in the application toward other body fluids, but these are either, not as easily collected and available (e.g., cerebrospinal fluid, bronchial lavage, or synovial fluid), or show very high variability in its low-molecular-weight proteomic content (e.g., plasma or serum), or even a combination of both. As a consequence until to date, most clinical proteomic studies on the low-molecular-weight proteome are focused on urine.

3.1 Urinary biomarkers for renal diseases

The analysis of urine has, as would be expected, directed research to the identification of urinary biomarkers of diseases of the kidney and the urinary tract although recent examples show that urine can also be an excellent reservoir of biomarkers for sites more distant from the kidney.

3.1.1 Diabetic nephropathy and chronic kidney disease

In a recent manuscript, Good et al. [50] demonstrate that an array of urinary biomarkers can be identified that enable assessment of renal disease with very high accuracy. Urine samples from 230 patients harbouring chronic renal disease of different etiologies were compared with samples from 379 apparently healthy controls, resulting in the identification of 634 peptides that differed significantly between these two cohorts, 273 of these potential biomarkers could be sequenced. The 273 biomarkers were combined using support vector machines to a chronic kidney disease-specific urinary biomarker model. The biomarkers and the biomarker model were subsequently validated in an independent blinded test set of 144 samples, with a specificity of 100% and a sensitivity of 86%, significantly outperforming serum creatinine, the currently accepted biomarker for chronic renal disease. Collagen fragments represent a large
that may even prevent the development of clinically relevant disease. These data indicate that a key factor in chronic renal disease is a reduction of collagen degradation, resulting in an increase in insoluble collagen fibres, and fibrosis. Further analysis of longitudinally collected samples indicate that this chronic kidney disease-specific urinary biomarker model enables detection of disease at a very early stage (when no clinical symptoms are detectable), opening the avenue toward early, targeted therapeutic intervention that may even prevent the development of clinically relevant disease [51].

Diabetic nephropathy (DN) has become the most prevalent cause of end-stage renal disease and is the most common and serious complication of both type I and type II diabetes, affecting up to 40% of all diabetic patients [52]. Early detection of DN, before persistent microalbuminuria, is becoming of utmost importance as inhibition of the renin–angiotensin system been shown to efficiently slow down or even block the progression of DN and thus will reduce the number of patients developing this serious DN complication [51].

In a first study employing 305 individuals, biomarkers for DN were defined and validated in blinded data sets using CE-MS [53]. A panel of 40 urinary biomarkers was identified that distinguished patients with diabetes from healthy individuals with 89% sensitivity and 91% specificity. Among the patients with diabetes, 102 urinary biomarkers differed significantly between patients with normoalbuminuria and nephropathy, and allowed to construct a model that correctly identified DN with 97% sensitivity and specificity in the blinded validation. Interestingly, these biomarkers, albeit currently only studied on a reduced number of patients, also identified patients with microalbuminuria and diabetes at risk for progression. It allowed sorting patients that progressed toward overt DN over 3 years. In addition, these biomarkers allowed to differentiate between DN and other chronic renal diseases with 81% sensitivity and 91% specificity more closely mimicking the actual clinical situation where only rarely patients need to be distinguished from healthy controls. The data were recently confirmed in [54] and in a multicentre (145 type II diabetes) study with a 93.8% sensitivity and 91.4% specificity and an area under the receiver operating curve (AUC) of 0.948 (95% CI 0.898–0.978) (Alkhalaf et al., submitted).

### 3.1.2 ANCA-associated vasculitis

Renal activity is difficult to assess in anti-neutrophil cytoplasmic antibody-associated vasculitis (AAV) because of renal scarring. Even repeated biopsies do not suffice to clearly identify patients with active renal AVV. CE-MS was used to identify and validate urinary biomarkers that enable differential diagnosis of disease and assessment of disease activity [55]. The data were compared with healthy individuals, patients with other renal and nonrenal diseases, and patients with AAV in remission leading to the identification of 113 potential biomarkers. Sensitivity and specificity of models based on 18 sequenced biomarkers were validated using blinded urine samples of 40 patients with different renal diseases. Discrimination of AAV from other renal diseases in blinded samples was possible with 90% sensitivity and 86.7–90% specificity depending on the model. Interestingly, ten patients with active AAV were followed for 6 months after initiation of treatment. Immunosuppressive therapy led to a change of the proteome toward “remission.”

### 3.1.3 Polycystic kidney disease

To gain insight into early disease progression in human autosomal dominant polycystic kidney disease (ADPKD), Kistler et al. [56] analyzed the urine proteome of 41 patients with ADPKD using CE-MS with relatively well-preserved renal function. Those were compared with age-matched healthy controls and patients with other renal diseases. This resulted in the selection of 38 biomarkers for ADPKD. Validation of those 38 sequenced markers in an independent cohort yielded (24 cases and 35 controls) revealed a sensitivity of 87.5% and a specificity of 97.5% (AUC: 0.95). The data could further be validated in samples from the independent consortium for radiologic imaging studies of polycystic kidney disease cohort (Kistler et al., in preparation). To date, identification of biomarkers that enable prognosis of disease progression failed, in part likely due to the lack of samples that were prospectively collected, where patients by now reached hard clinical endpoint (kidney failure). As also observed in DN, the identification of potential biomarkers based on surrogate endpoints may fail, since surrogate endpoints are frequently of only moderate accuracy [57].

### 3.1.4 Paediatric renal disease

Noninvasive analysis not only allows easy followup of chronic renal disease but is also warmly welcomed in pediatric disease where it has the potential to replace invasive followup of frequently encountered infant diseases.

With this in mind, Decramer et al. [58] used CE-MS-based urinary proteome analysis to define specific urinary biomarker patterns for different grades of ureteropelvic junction (UPJ) obstruction, a frequently encountered pathology in newborns that currently needs invasive followup to determine the severity of disease and the necessity for surgical intervention. In total, 53 urinary biomarkers were identified and validated in a blinded prospective study that predicted the clinical outcome of newborns with 95% accuracy 9 months in advance. The accuracy was increased even further to 97% after 12 months [59]. Recently, an independent but small-scale study (n = 19), with patients recruited in another country using
3.1.5 Acute kidney disease

Acute kidney injury (AKI) is frequent in critically diseased patients. In order to prevent progression and to improve outcome, early and accurate detection of AKI appears crucial. CE-MS was used for the identification of potential biomarkers of AKI [62]. A training set of 87 urine samples prospectively collected at five consecutive days from intensive care unit patients, who later developed AKI (n = 16) or maintained normal renal function (n = 14) were used. Twenty sequence-identified marker peptides were combined to a diagnostic pattern. The polypeptides were specific degradation products of six proteins: albumin, α-1-antitrypsin, and β-2-microglobulin (upregulated fragments), fibrinogen α and the collagens 1 α (I) and 1 α (III) (down-regulated fragments). Good diagnostic performance of the marker pattern was indicated by an AUC of 0.91 for the training set after total cross-validation. This was confirmed in a blinded intensive care unit validation set (n = 20, AUC: 0.84) and in allogeneic hematopoietic stem cell transplantation (HSCT) patients with (n = 13), or without (n = 18) AKI episode (AUC: 0.90). Comparison of the AKI marker pattern with other potential biomarkers of AKI including serum cystatin C and urinary levels of kidney injury molecule-1, interleukin-18, and neutrophil gelatinase associated-lipocalin revealed superior prognostic value of the proteomic marker pattern in detecting AKI with a maximum of 5 days in advance to the rise of serum creatinine.

3.2 Urinary biomarkers for urological disorders

The application of CE-MS in tumor biomarker research has recently been extensively reviewed by Simionato et al. [63]. Theodorescu et al. [33] described the CE-MS detection and validation of biomarkers of urothelial carcinoma. A bladder cancer-specific biomarker pattern was established by an initial definition in a training set composed of 46 patients with urothelial carcinoma and 33 healthy subjects, and further refinement with CE-MS spectra of 366 urine samples from healthy volunteers and patients with malignant and nonmalignant genitourinary diseases. With this two-step biomarker discovery approach, the authors could establish a prediction model composed of 22 urinary peptides. This model correctly classified all urothelial carcinoma patients and all healthy controls, when applied to a blinded test set that contained 31 urothelial carcinoma patients, 11 healthy individuals, and 138 nonmalignant genitourinary disease patients. Differentiation between bladder cancer and other malignant and nonmalignant diseases (such as renal nephrolithiasis) was accomplished with at least 86–100% sensitivity. In a further extension of this study, the same group identified urinary biomarkers that could distinguish between invasive and noninvasive urothelial carcinomas, indicating the potential of this approach to deliver information beyond the mere detection of a tumor, which may guide patient management [64].

In a pilot study [37], CE-MS techniques were used to define potential urinary peptide biomarkers for prostate cancer (PCa). However, the biomarkers could not be validated in a subsequent blinded assessment; once more, the importance of a blinded test set was underlined. In a subsequent study, first-void urine was found to be a more appropriate sample to define of PCa-specific biomarkers. To enable validation of first-void urine, biomarkers specific of first-void urine were determined. The results of that study indicated that the identified biomarkers originate from secretions of the prostate into urine. After refinement of the PCa-specific biomarker pattern, using urine samples from 54 PCa and 62 BPH patients, a model with ten potential biomarkers resulted in the prediction of 88.9% (32/36) of the PCa and of 66.7% (16/24) of the BPH patients in a second blinded set of patient samples [65]. These data could further be validated in an independent cohort of 211 samples collected under “real life conditions,” by resident
urologists, revealing a sensitivity of 88% and a specificity of 59%.

3.3 Application of urinary proteome analysis to nonrenal and urological diseases

Body fluids are suspected to be highly informative of the tissues with which they are directly in contact. That assumption is one of the reasons why urine was used in studies on diseases of the urogenital tract. However, the fact that plasma is filtered by the kidney also stimulated researchers to examine urine for biomarkers of disease from more distant organs. This has been exploited for the identification of biomarkers of cardiovascular disease and for the clinical followup of patients after allogeneic HSCT. CE-MS was applied to the clinical followup of patients after (HSCT [66, 67]). Urine samples from 40 patients after HSCT (35 allogeneic and 5 autologous) and 5 patients with sepsis were collected during a period of 100 days (a maximum of ten samples per patient) for CE-MS analysis. A pattern that consisted of 16 differentially excreted polypeptides discriminated patients with early graft versus host disease from patients without complications with 82% specificity and 100% sensitivity. In a subsequent blinded multicenter validation study on 100 patients with more than 600 samples collected prospectively, the initial results were confirmed, although with reduced specificity and sensitivity [68]. Initial results of preemptive therapy based on proteome profiling clearly indicate a benefit for the yet limited number of patients (Weissinger et al., submitted).

3.3.1 Coronary artery disease

Coronary artery disease (CAD) is a leading cause of morbidity and mortality worldwide. Despite multiple clinical, electrographic, and biochemical characteristics, there are subgroups of patients who progress to severe, life-threatening CAD without clinically overt symptoms and signs [69]. For example, patients with type II diabetes and the elderly frequently suffer from silent myocardial infarctions with significantly increased risk of complications [70]. Early diagnosis of CAD in its presymptomatic stage would allow for better targeted and hence more effective primary prevention as compared with current clinical recommendations. Recently, urinary biomarkers for CAD were defined and validated in an independent population [71]. Urine from 88 CAD patients and 282 controls was examined by CE-MS. This resulted in the identification of 15 peptides that defined a characteristic CAD signature panel. In a second step, this panel was evaluated in a blinded study on 47 CAD patients and 12 healthy individuals. CAD patients were identified with 90% sensitivity and specificity. In addition, the polypeptide CAD signature panel significantly changed after therapeutic intervention toward the polypeptide signature of healthy humans. Recent data show that patients with CAD could be distinguished from patients presenting symptoms of CAD, but without clinical evidence in the coronary angiography [72]. The prospective value of the urinary proteomics for CAD could further be validated in prospectively collected samples from patients with type I diabetes [54]. That study also highlighted another feature of the CE-MS analysis: one analysis can be investigated for several different biomarker panels that are indicative for different pathological conditions, as shown in Fig. 4. In a further extension of this study, Delles et al. [73] were able to define a pattern of 238 CAD-specific polypeptides by comparing 586 spot urine samples from 408 subjects. This pattern identified patients with CAD in a blinded cohort of 138 urine samples (71 patients with CAD and 67 healthy individuals) with high sensitivity and specificity (AUC: 0.87). The sequences of the discriminatory polypeptides include fragments of α-1-antitrypsin, collagen types 1 and 3, granin-like neuroendocrine peptide precursor, membrane-associated progesterone receptor component 1, sodium/potassium-transporting ATPase γ chain, and fibrinogen-α-chain. Several biomarkers changed significantly toward the healthy signature following 2 years treatment with irbesartan, whereas short-term treatment with irbesartan did not significantly affect the polypeptide pattern.

Although there is no apparent clinical need for biomarkers for diabetes, investigation of proteomic changes associated with diabetes may help in further understanding pathophysiology, especially of consequences of diabetes. Based on the comparison of 30 cases and controls, each, a panel of 40 potential biomarkers was identified by Rossing et al. [53]. In addition, an extended panel of 261 potential biomarkers was defined by the authors, employing all 205 samples available in the study. These panels were further validated by Snell-Bergeon et al. [54] in a cohort of 38 prospectively collected samples from the coronary artery calcification in type I diabetes study. The results showed a clear benefit of using the 261-biomarker panel, indicating that a larger number of significant biomarkers is beneficial, resulting in higher accuracy in assessment of disease. In a recent large study on 902 subjects from ten different clinical centers, this biomarker pattern could be further validated, resulting in 94% accuracy in this multicentric assessment [74]. What is more, the authors could also demonstrate that significant differences between diabetes type I and type II exist in the urinary proteome. These may help to pinpoint potential pathophysiological changes early in disease that are specific for type I or type II diabetes.

3.4 Aging

Aging induces morphological changes of the kidney, and leads to a reduction of renal function. In order to gain insight into the processes of renal aging, [75] examined urine samples collected from 324 healthy individuals (aged
2–73 years) with CE-MS. A total of 325 peptides displayed statistically significant age-related changes. Most of the markers changed significant during puberty, and coincided with the completion of renal development. However, 49 peptides could be correlated with aging in adults. A striking observation was that some of these peptides were also found to be differentially secreted in chronic renal diseases, including DN, focal, and segmental glomerulosclerosis, membranous glomerulonephritis, and vasculitis. Association of renal aging and chronic renal disease was confirmed in a blinded evaluation of samples from healthy individuals and DN patients. Sequence information of some of those aging markers suggested that one prominent mechanisms of human aging is a reduced turnover of extracellular matrix (ECM), which results in increased fibrosis. After further refinement, the age-related polypeptide marker patterns might allow the noninvasive detection of renal lesions in healthy persons, and the testing of an individuals' suitability for kidney donation.

4 Pathophysiological aspects of biomarkers

Although the majority of potential CE-MS-identified urinary biomarkers described to date have not been sequenced, sequences are available for more than 500 different urinary peptides [39]. Not unexpectedly, most of those peptides are derived from the most-abundant proteins in the body: collagen – mainly type I, II, and III, albumin, β 2-macro-globulin, and uromodulin. Consequently, a valid question is whether urinary peptidomics in renal disease is not just another way to measure glomerular injury, that could probably be assessed with similar precision, but less effort, by measuring albuminuria [76]. Although that question cannot be answered with absolute confidence, it is certain, however, that differential diagnosis based on urinary proteome analysis is possible [77–80]. The fact that patients in complete remission without albuminuria still exhibit apparently disease-specific changes in urinary polypeptides [35] strongly suggests that those peptides contain clues about the pathogenesis and are not merely degradation products.

Although the evidence is still scarce, it is an attractive hypothesis that urinary peptides of diagnostic value are not merely degradation products of abundant larger proteins, but a result of distinct, disease-specific processes; in many cases, due to significant changes in the activity of proteases as suggested by Haubitz [77]. That hypothesis is further strengthened by the detection of specific collagen fragments that correlated with the disease-specific activity of matrix metalloproteases. That assumption is supported by various findings: (i) the increase of collagen and ECM is observed in patients with diabetes and DN, (ii) collagen fragments are significantly reduced in diabetic urine [80, 81], and (iii) reduced activity of proteases and protection of the ECM from proteolysis by advanced glycation end products, proposed key pathological changes in diabetes mellitus [82].

A similar scenario might 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. Once a substantial number of additional peptides are sequenced, a thorough examination of the sequences of the urinary peptides and comparison with protease specificities might provide additional support for the above hypothesis, and could lead to a better insight into the regulation and pathophysiological role of specific proteases in many diseases.

A related hypothesis can be proposed, on the urinary peptidome displaying, to a large degree, the turnover of the ECM. That hypothesis has been generated as a result of the observation that the major urinary peptides are not, as expected, the “usual suspects” such as albumin or uromodulin, but rather specific collagen degradation products, and that several of those products are significantly reduced in diseases where an increase of ECM has been reported [83]. Consequently, those peptides might be derived from ECM turnover. Changes in that turnover also result in indicative changes in urinary peptides, which serve as a very specific, noninvasive indicator for alterations in ECM turnover, which in turn is likely to be disease specific. Such changes in the ECM turnover might be due to, e.g., an invasion of tumors (ECM must be “dissolved” in order to make room for the growing tumor), fibrosis (reduced ECM degradation), increased arterial stiffness (change in ECM composition), or changes in endothelium.

5 Concluding remarks and outlook

CE-MS fulfills the requirements for broad application in routine clinical practice, as indicated by the validation of a variety of different pathological conditions and controls that might represent with similar symptoms, as e.g., shown recently by Haubitz et al. [55] (Fig. 6). Hence, we are confident that publicly available databases that contain thousands of data sets from individual samples will greatly ease and expedite the definition of relevant biomarkers. This

Figure 5. Graphic depiction of comparable data sets (analyzed with identical preanalytical preparation, instruments, and analytical parameters) obtained from subjects with different diseases/(pathological) conditions currently represented in the human urinary proteome database. Reprinted with permission from [39].
is one of the main aims of the EuroKUP COST action (www.eurokup.org), as outlined recently [84]. As a first step in this direction and to ease the comparability between data sets, a standard protocol for urine collection, and two standard urine samples (male and female) that have been extensively characterized using an array of different technologies, have been generated and established [26].

H. M. was supported in part by EU Funding through InGenious HyperCare (LSHM-C7-2006-037093) and SysKID (HEALTH–F2–2009–241544). The work of JPS was supported by Inserm, the “Direction Régional Clinique” (CHU de Toulouse, France) under the Interface program, by the Fondation pour la Recherche Médicale and by the ANR grant: ANR-07-PHYSI0-004-01).

Figure 6. Compiled protein patterns of the CE-MS analysis of urine samples from patients and controls examined in [55]. Shown are compiled patterns consisting of all samples from patients with active vasculitis and each of the six control groups. (AAV, ANCA associated vasculitis; NC, apparently healthy normal control; MNGN, membranous glomerulonephritis; DN, diabetic nephropathy; FSGS, focal segmental glomerular sclerosis; IgAN, IgA nephropathy; MCD, minimal change disease). The molecular mass on a logarithmic scale (0.8–25 kDa, indicated on the left) is plotted against normalized migration time (18–45 min, indicated on the bottom). Signal intensity is encoded by peak height and color. Although the difference in peptides present between apparently healthy controls and the six different chronic renal diseases is evident, distinct biomarkers that can distinguish between AAV and the other chronic renal diseases are less obvious. Reprinted with permission from [55].
The authors have declared the following conflict of interest: H. M. is the founder and co-owner of Mosaiques Diagnostics, who developed the CE-MS technology for clinical application.

6 References


