Proteome Analysis in Hematology Using Capillary Electrophoresis Coupled On-Line to Mass Spectrometry

E.M. Weissinger[1,*], H. Mischak[2], J. Kontsendorn[1], A. Hahn[1], N. Hahn[1], M. Morgan[1] and A. Ganser[1]

[1]*Hannover Medical School, Department of Hematology, Hemostasis, Oncology, and Stem Cell Transplantation, 30625 Hannover;* [2]*Mosaiques Diagnostics & Therapeutics AG, 30625 Hannover, Germany*

Abstract: Proteome analysis, the key technology for biomarker discovery, continues to gain importance in clinical diagnosis and follow-up. In this review we describe proteome analysis in the context of allogeneic, hematopoietic stem cell transplantation concentrating on capillary electrophoresis coupled on-line to mass spectrometry.

Key Words: Proteomics, clinical trial, diagnostics, stem cell transplantation, biomarker, mass spectrometry, capillary electrophoresis.

INTRODUCTION

Proteome analysis is emerging as an important tool for deciphering biological processes and for the discovery of biomarkers for diseases from tissues and / or body fluids. Proteins and peptides (polypeptides) in body fluids are responsible for the flow of information to cells and tissues, and changes in protein/peptide abundance reflect changes relevant for cells, tissues and even the whole organism. A deeper insight into the functional relevance of these polypeptides under different physiological and pathophysiological conditions is one of the main challenges in proteome research [1-3]. The choice of the accurate technological platforms, well defined patient groups and reasonable controls, other than healthy individuals are required to reach the high standard necessary for clinical application [4]. Major abundant proteins such as albumin and globulin may obscure the identification of less abundant peptides, plus there are other ways of looking for these large molecules, like western blot, ELISA, immunoblotting and others. Thus, removal of the large analytes from the samples will greatly enhance the spectrum of smaller and less abundant peptides. Pre-fractionation, improved separation, higher sensitivity of identification, and depletion of albumin, globulin, transferrin, and others prior to analysis are necessities results on the smaller molecules. The application of different proteomic techniques and the proper use of biostatistics and bioinformatics will lead to the goal of "unraveling the proteome of diseases".

Although it is of great value to know that the particular disease patterns consist of molecules that are important known key players in the pathophysiology of the disease, proteomic screening opens a much broader view of normal and disease states – a cornucopia of new possibilities and yet unknown and thus undefined biomarkers for certain diseases.

*Address correspondence to this author at the Department of Hematology, Hemostasis, Oncology and Stem Cell Transplantation, Hannover Medical School, 30625 Hannover, Germany; Tel: (511)-532-9518; Fax: (511)-532-6843; E-mail: mischak-weissinger.eva@mh-hannover.de

The complexity and wide dynamic range of the proteins/peptides in the sample pose an enormous challenge to both protein/peptide separation as well as subsequent identification tools [5]. To enable an in-depth analysis of most of the polypeptides present in complex biological samples, robust high resolution separation must be coupled to high resolution mass analyzers. Clinical proteomics is a newly evolving field. Several recent developments have driven the transition of proteome research to the more complex challenge of screening patient samples. In order to apply proteomic screening in a meaningful way to clinical use, it is of utmost importance that clinicians raise a defined problem and ask specific, well defined questions, with respect to what benefit proteomic screening would hold for the patients studied. The choice of sample, the subsequent sample preparation and the technology best suited to answer the question should be discussed together with proteome researchers and statisticians. Sample preparation should be kept to a minimum to avoid introducing artifacts [6]. Body fluids obtained from patients are particularly interesting for detecting disease [7], as well as monitoring disease progression and response to therapy [8, 9]. Bioinformatics must be of extremely high standard in clinical proteomics; the limited numbers of samples compared to the enormous amount of information that can be obtained with most proteomic screening tools pose an enormous challenge to the software tools used to interpret inter- and intra-patient variability. In the following we shall briefly summarize the technology platforms currently used for clinical proteomics, focusing later on the application of capillary electrophoresis (CE) coupled on-line to mass spectrometry (MS) to the follow-up of patients undergoing allogeneic, hematopoietic stem cell transplantation (allo-HSCT).

1. CAPPILLARY ELECTROPHORESIS AND MASS SPECTROMETRY

1.1. General Observations

All modern MS techniques currently require appropriate sample preparation, as well as fractionation/separation steps prior to the MS analysis. Fig. (**1**) summarizes the steps common for proteome analysis of body fluids in a clinical

1389-5575/09 $55.00+.00

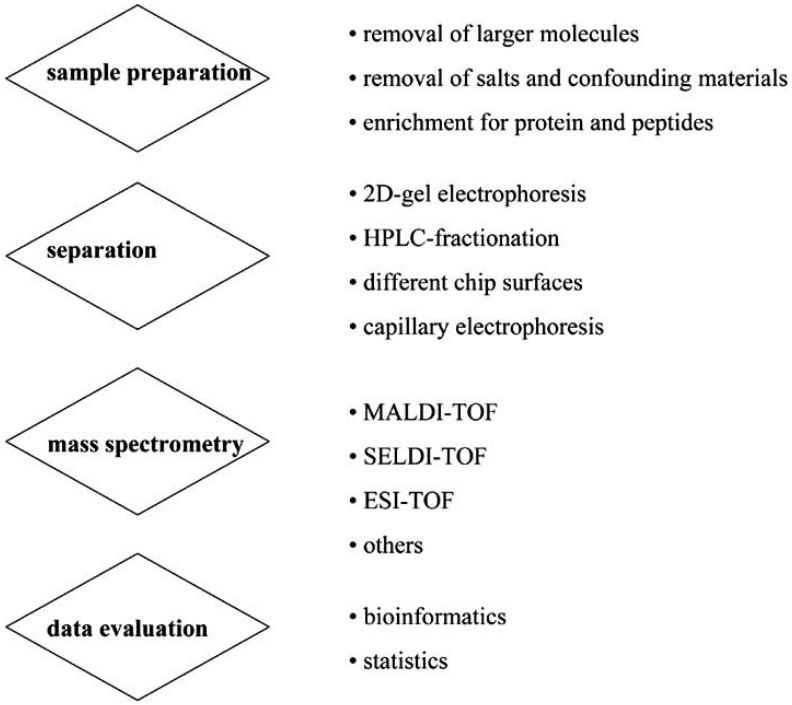


Fig. (1). Summary of common steps and methods to accomplish proteomic screening of body fluids.

Sample preparation: Once the proper sample is chosen (blood, urine, other), the sample must be prepared for analysis. Purification and concentration of the samples is necessary, but should be limited to as few steps as possible, such as removal of large molecules and enrichment for polypeptides and peptides within the samples.

Separation: The high complexity of biological samples also necessitates a separation step, prior to MS-analysis. Gel electrophoresis (1 or 2 dimensions), fractionation using high performance liquid chromatography (HPLC) and capillary electrophoresis (CE) are the most commonly used methods to separate analytes according to their size, isoelectric point and charge, respectively prior to analysis in the MS.

Ionization: Electrospray ionization (ESI) and matrix-assisted laser desorption / ionization (MALDI) are the two techniques most commonly used to volatize and ionize proteins or peptides for mass spectrometry. ESI ionizes the analytes from a solution and is therefore readily coupled to liquid-based (for example, chromatographic like HPLC or electrophoretic like capillary electrophoresis (CE) separation tools. MALDI sublimates and ionizes the samples out of a dry, crystalline matrix *via* laser pulses. MALDI-MS is gnerally used to analyze relatively simple peptide mixtures, whereas integrated liquid-chromatography (LC-MS) or ESI-MS systems are preferred for the analysis of complex samples.

Mass spectrometry: A mass spectrometer consist of a ion source (ionization) in combination with mass analyzers, which measure the mass/charge (m/z) ratio and detectors registering the number of ions at each m/z value (mass spectrometry) yielding to the signal intensity.

Data evaluation: Most proteomic studies accomplished in clinical settings showed that a pattern of different biomarkers may be more useful then a single marker for differential diagnosis of diseases. Thus, statistical analysis and applications of tools like support vector machines [59] have become increasingly important for clinically oriented proteomic analysis.

setting. We used capillary electrophoresis (CE) for separation of the analytes present in complex biological samples, coupled on line to an electrospray-ionization time-of-flight mass spectrometer for analyses (ESI-TOF). For measuring in a TOF-instrument the ions are accelerated to high kinetic energy and are separated along a flight tube as a result of their different velocities. TOF are most commonly used in pre-clinical and clinical proteomics approaches. Sequencing as well as information on post-translational modification (PTM) requires sequential use of mass spectrometers, termed tandem mass spectrometry [10], (MS/MS). In general, the first MS instrument serves as a mass filter, selecting only ions with the mass of interest ("parent ions"), and the second MS instrument analyzes the fragmentation products ("daughter ions") which may be generated by collision with other molecules (CID) [11] or transfer of electrons (electron transfer dissociation. Clinical proteomics can be seen as a comparative analysis of multidimensional datasets, which is further complicated by biological variability. It cannot be overemphasized that any experiment must include assessment of all variable parameters in order to accurately evaluate the data. Furthermore, appropriate use of the correct statistical methods (e.g. adjustment for multiple variables) is of the utmost importance [12]. In this review we focus on the use of CE [13] coupled on line to an ESI-TOF-MS (CE-MS) [14] for clinical diagnostic and follow-up of disease [15-18].

1.2. Capillary Electrophoresis Coupled On-Line to Mass Spectrometry (CE-MS)

CE separates proteins based on migration time in an electrical field (300-500 V/cm) with high resolution. CE-MS (Fig. **2**) has the advantage of providing fast and robust separation at high resolution [19], is compatible with most buffers and analytes [20], and uses inexpensive capillaries instead of expensive LC columns [2]. Since CE is compatible

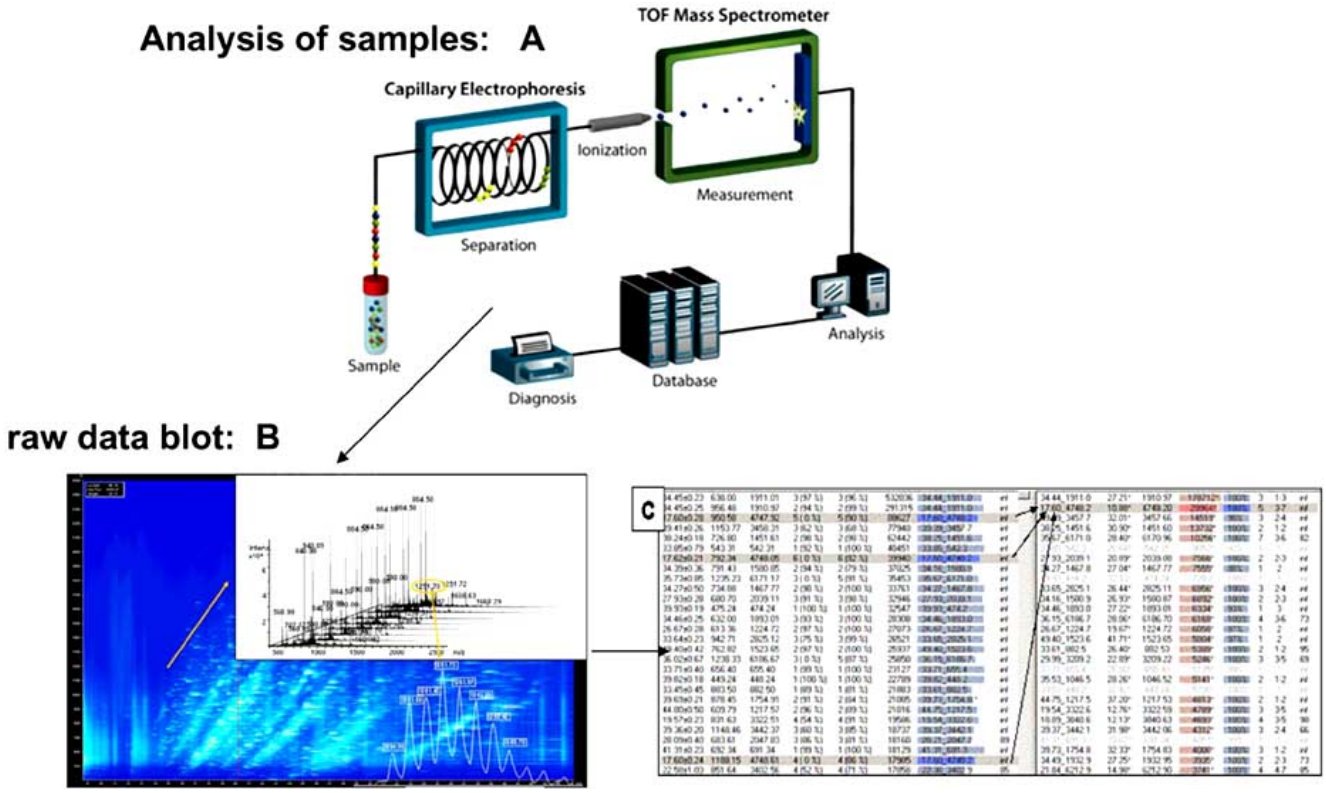


Fig. (2). Capillary electrophoresis coupled on-line to mass spectrometry.

A schematic drawing of the on-line coupling of CE-MS showing the work-flow from the sample preparation to final data processing and identification of the unique peptide pattern is shown (**2A**). The CE-MS spectra obtained every 3 seconds are shown in Fig (**2B,** insert) and the resulting 3-dimensional raw data blot is shown (lower). The m/z (Y-axis) is plotted against the migration time (X-axis, in minutes) in the CE, while the signal amplitude is depicted as a color code ranging from black to white with increasing intensity of the signal. All data are stored in a database as peak list (**2C**) as shown on the right.

with most buffers and analytes [21], it provides a stable constant flow, thus avoiding elution gradients that may otherwise interfere with MS detection [22]. Similar to LC, CE can be interfaced with most mass spectrometers, technical considerations that must be taken into account for such coupling have been reviewed [23, 18]. Limitations include the difficulty to apply CE to the analysis of high-molecular-weight proteins, due to the acidity of the buffer generally used for CE-MS analysis, which results in precipitation of larger proteins. However, this limitation is less pronounced than that observed in LC. Another limitation of CE is the relatively small sample volume that can be loaded onto the capillary (less than 1μl), leading to a lower sensitivity of detection in comparison to LC. Improved methods of ionization by micro- or nano-ion spray and improvements in the detection limits of mass spectrometers enabling detection in the low- or sub-femtomolar range have overcome these problems [24-26]. In addition, improved delivery of the separated proteins from the tip of a capillary to the MS instrument in a small stream of liquid by nano-ion spray has also increased sensitivity. Consequently, CE-MS has become a viable alternative to the commonly used proteomic technologies and has recently been successfully applied in several clinical studies [27-30]. For MS/MS sequencing, CE is currently not the method of choice, due to its low loading capacity. However, CE-MS data can be matched to liquid chromatography (LC)-MS/MS data, using migration time as a second identifying parameter [12]. Hence, a combination of CE-MS analysis, enabling higher throughput and consequently the generation of data of higher statistical quality, with LC-MS/MS for subsequent sequencing of the potential biomarkers may be a promising approach for biomarker discovery and characterization.

1.3. Direct Comparison of CE-MS and LC-MS

As already outlined in a previous review [2], CE holds several advantages over LC. Those advantages were very recently described in detail in several reviews [31, 32], and are especially beneficial when analyzing large numbers of heterogeneous samples that contain interfering compounds, such as lipids, precipitates, and others. CE's main advantages are it's robustness, ability to recondition quickly with NaOH, simple separating principle with high reproducibility, and, with respect to MS interfacing, use of a buffer system that does not change its composition during analysis, as no buffer-gradient is applied. There are currently several reviews on different CE, thus we shall not elaborate on these here, but concentrate on what we are using in clinical diagnosis and CE-MS [2, 33].

A disadvantage of CE is the limited loading capacity. Whereas ml quantities can be loaded onto an LC column, a

CE can be filled with a maximum of about 1 µl; and in general only 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, corresponding to 0.25 – 1 µL when using 50 or 75 µm inner diameter capillaries with 80 – 100 cm length. The limited loading capacity does not seem to present a significant problem in CE-MS coupling, because the amount of information in the sample is generally extremely high [33].

2. PROTEOME ANALYSIS USING CE-MS

2.1. Choice of the Sample Collection, Storage and Preparation for Proteomic Screening

Success for the search for biomarkers and for the development of diagnostic tools is dependent upon the choice of sample for any research or diagnostic study. Two basic sources of material are available for proteomic studies: body fluids (e.g. urine, blood) and tissue. While examination of tissue may be advantageous at least in a scientific setting, several issues like accessibility and high variability due to sample heterogeneity (e.g. presence of several different cell types) have greatly hindered progress in this field. Proteome analysis requires pure, selected cell types in order to find biomarkers for disease or activation markers, which are not due to different cell types present in a biopsy for example. None the less it has also been outlined recently that biomarkers discovered in tissue could subsequently be detected in body fluids [34]. For clinical application of proteomic screening body fluids hold several advantages over biopsies as outlined below. In contrast to polypeptides in tissues and most cell types, the polypeptides in body fluids are relatively easily accessible and changes within the circulating peptides/proteins can be readily detected. Comprehensive profiling of peptides and proteins in body fluids such as plasma or urine has advantages over the analysis of proteins and peptides expressed in particular cells or tissues.

2.2. Body Fluids in Proteome Analyses

Among various body fluids, urine and blood (plasma and/or serum) are most extensively studied. While blood-derived samples appear as the first choice, more in depth investigation has revealed several, yet-unresolved problems. Activation of proteases and consequently generation of an array of proteolytic breakdown products is associated with both, the blood collection and clotting procedures [35]. Standardization of collection protocols and storage is therefore necessary, since different pre-analytical handling of the samples yield to activation of proteases and this appears to be major causes for difficulties to compare data between centers. Furthermore, the wide dynamic range (10^{12}) of proteins poses yet another major problem [36, 37]. Serum has been analyzed and found essentially unsuitable for the screening with CE –MS [15]. The proteolytic break down products are detected readily by the highly sensitive MS and this leads to less comparable data and high variability between samples collected at different centers. Plasma frozen immediately after collection is a more reliable source of information, but the strict requirement for accurate pre-analytical handling is a substantial problem in a routine clinical setting [38, 36].

Urine, on the other hand, is an attractive source of information [2]. One of the first attempts to define the urinary proteome was published by Spahr and collegues [39, 40]. Tryptic peptides of pooled urine samples were analyzed using LC-MS, and 124 proteins were identified. While this study did not attempt to define any urinary biomarkers for a disease, it clearly highlighted the potential information in the urinary proteome and also indicated a possible approach towards its mining. More recently, Adachi *et al.* identified more than 1,500 proteins or their fragments in urine of healthy individuals, further underlining the complexity of the human urinary proteome [41, 33]. One of the greatest advantages of urine over serum/plasma is the stability of this body fluid. Urinary polypeptides are stable and generally do not undergo significant proteolysis within several hours after the collection. Urine can be stored for up to 3 days at 4°C [42] or up to 6 hours at room temperature [43] and still yields comparable results without significant degradation. Thus, urine seems to hold potential promise for the detection of biomarkers significant for different diseases or disease progression.

2.3. Sample Preparation for CE-MS Analysis

A critical issue in clinical proteomics is the sample preparation. Ideally, the crude samples should be analyzed, Thus, allowing assessment of the analytes in the without any manipulation. Unfortunately, this ideal situation cannot be accomplished, and samples must undergo pre-analytical manipulation to remove salts or other confounding materials. To enable a subsequent comparison, this step should be robust and highly reproducible.

In our set up an ultrafiltration step in the presence of urea and SDS, followed by a desalting step on PD-10 columns [44] is used in the sample preparation. As outlined in detail by Theodorescu *et al*., the presence of detergent and chaotropic agent efficiently inhibits protein-protein interaction to limit loss of analytes due to association with other proteins (*e.g.,* albumin). This sample preparation protocol allows the preparation of samples containing low-molecular mass proteins and peptides, and results in a good comparability of data from individual patients or samples collected at different time points.

2.3. Analysis of Body Fluids Using CE-MS

The CE-MS technology platform is shown in Fig. (**2**), the schematic drawing gives an overview of sample processing from sample preparation to diagnosis. Urine samples obtained from healthy volunteers and/or patients with different diseases, or at different stages of disease, are prepared and analyzed within 45 to 60min (Fig. **2A**), [15, 45]. CE-MS analysis is performed as described [16, 29, 46] using a P/ACE MDQ (Beckman Coulter, Fullerton, USA) system coupled on-line to an ESI-TOF MS (Mariner Biospectrometry Workstation, Applied Biosystems, Farmington, CT) using sheath flow coupling (30% methanol, 0.5% formic acid in H_2O). The potential of the ESI sprayer (Agilent Technologies, Palo Alto, CA, USA) is set between 3 and 4kV. Data acquisition and MS acquisition methods are automatically controlled by the CE *via* contact-close-relays. Spectra are

accumulated every 3sec, over a range of m/z from 400 to 3000 (Fig. **2B**, insert).

The MS-spectra are transformed into the so called "raw data plots" (Fig. **2B**) by application of a specialized software package, MosaiquesVisu (accessible at www.proteomiques.com), the only soft ware currently available for depiction of MS-spectra as a 3-dimensional blot. The raw data blot contains information on mass/charge (m/z) and signal intensity (shown as color code ranging from black (0 MS counts) to white (25000 MS counts)) blotted against the migration time in the CE. All detected polypeptides are deposited, matched, and annotated in a Microsoft SQL database (Fig. **2C**), allowing further analysis and comparison of multiple samples and/or patient groups.

2.4. Data Processing and Evaluation

The evaluation of an entire CE-MS run consisting of about 1000 – 1500 single spectra can be performed in about 5 minutes using MosaiquesVisu, with an error rate below 1% as judged by the presence of 200 internal polypeptide standards present in each urinary sample analyzed to date [44]. The reproducibility and comparability of the data is of utmost importance. Thus, high mass accuracy is an absolute requirement for such analyses. Mass spectral ion peaks representing identical molecules at different charge states are deconvoluted into single masses using the MosaiquesVisu 2.1.0 software [29]. In order to ensure comparability between different measurements, the migration time and ion signal intensity (amplitude) are normalized using the internal polypeptide standards present in every urine sample analyzed to date [44]. The resulting peak list characterizes each polypeptide by its molecular mass [Da], normalized migration time [min] and normalized signal intensity. Polypeptides within different samples are considered identical, if the mass deviation was less than 200ppm and the migration time deviation is less than 2min. The reproducibility of the data obtained with CE-MS allows using normalization algorithms and subsequent compilation of data obtained from different patients or different patient groups, resulting in the definition of potential biomarkers for certain patient groups and diseases. Comparison of the compiled data in the database is performed to identify potential biomarkers for disease. Given the large number of polypeptides, pre-selection is necessary. To avoid artifacts due to diet, exercise, medication and other causes, the second spot urine of the day is collected and only polypeptides present in at least 40% of the samples in a given group (healthy individuals, specific disease controls or specific disease) are considered. This value was set empirically, allowing discrimination of patient groups with different complications or diseases [45]. In general, between 10 and 100 polypeptides showing statistically significant differences in signal intensity and/or frequency between groups are found. The best indicative single biomarker allows distinguishing between a certain disease and controls with 70 – 90% accuracy. By combining several of these potential biomarkers, yielding a diagnostic pattern of several biomarkers, classification accuracy can be improved. This is accomplished by using support vector machines (SVM) [47-49]. SVM's non-linearly map their n-dimensional input space into a high dimensional feature space. The features are the selected polypeptides, each of which representing one dimension. The support vectors are then used to construct an optimal hyperplane in this particular feature space. The optimal hyperplane is then used for classification and is given as classification factor (CF). In a first step, a list of predefined polypeptides is obtained, considering all available data sets of the sample groups compared (e.g. aGvHD *vs*. controls). AS described previously, only polypeptides showing a difference in frequency of >0.4 or a difference in signal amplitude of >2-fold between the compared groups are predefined. The pre-defined set of polypeptides is further vali-

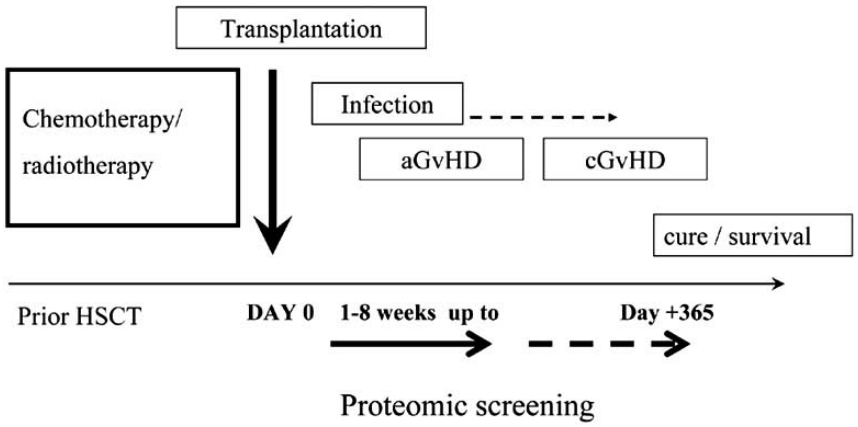


Fig. (3). Typical time-line of treatment prior to and follow-up after allogeneic hematopoietic stem cell transplantation (HSCT). Prior to HSCT the patient receives conditioning treatment, consisting of radio- and/or chemotherapy to eradicate leukemic cells and an immunosuppressive treatment to ensure survival of the allogenic graft. During this time and for first 1-2 weeks after HSCT is the time of aplasia, when the patient has no functioning immune system and needs particular prophylaxis against bacterial, viral and fungal infections. Despite the prophylaxis, most patients develop infectious complications within these first few weeks. By day 14 after HSCT, engraftment takes place and now one of the most feared complications, namely the graft-versus-host disease (GvHD) becomes of major importance. The acute form (aGvHD) develops rapidly and is seen as early as 7 days after transplantation and is a continuous problem for about the first 100 days. The chronic (cGvHD) form is differentiated mainly by the organ manifestation and the histological presentation of the sample biopsies. Most cases of cGvHD are diagnosed after day +100, but sometimes cGvHD may develop earlier. Grading of aGvHD and cGvHD have been revised according to the NIH-consensus conferences [60-62].

dated by randomly excluding 30% of available samples. This kind of bootstrapping procedure is repeated up to 10 times and yields markers of high statistical significance. Discriminatory polypeptides are included in the SVM derived prediction model using the MosaiquesCluster software [27].

3. CE-MS ANALYSIS APPLIED TO CLINICAL DIAGNOSTICS

3.1. Allogeneic Hematopoietic Stem Cell Transplantation (allo-HSCT)

Allogeneic hematopoietic stem cell transplantation (allo-HSCT) is applied with success to the treatment of malignant hematological diseases, such as leukemia and lymphomas, as well as to hematological dysfunction syndromes, e.g. severe aplastic anemia. Although allo-HSCT is currently the only curative treatment for patients with hematologic malignancies, the application of allo HSCT is not only limited by the availability of a donor, but also by severe complications occurring after allo-HSCT, such as concomitant infections or severe graft-versus-host disease (GvHD). Fig. (**3**) summarizes the major steps during treatment and recovery phase of patients undergoing allo-HSCT. Prior to HSCT, the patient receives radio- and/or chemotherapy as a conditioning treatment to reduce the number of leukemic cells and as an immunosuppressive treatment to ensure survival of the allogeneic graft. Within the first 2 weeks after allo-HSCT the immune system of the host is exchanged, resulting in aplasia of

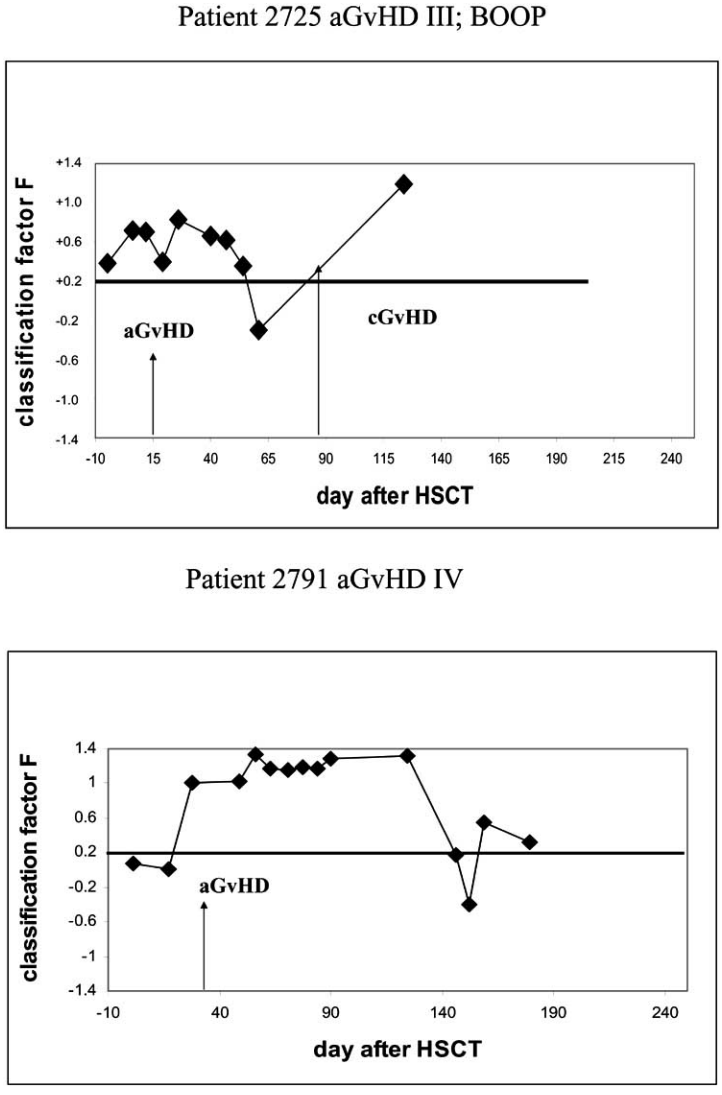


Fig. (4). Graphs of the proteomic pattern diagnostic classification factor F over several time-points post HSCT:

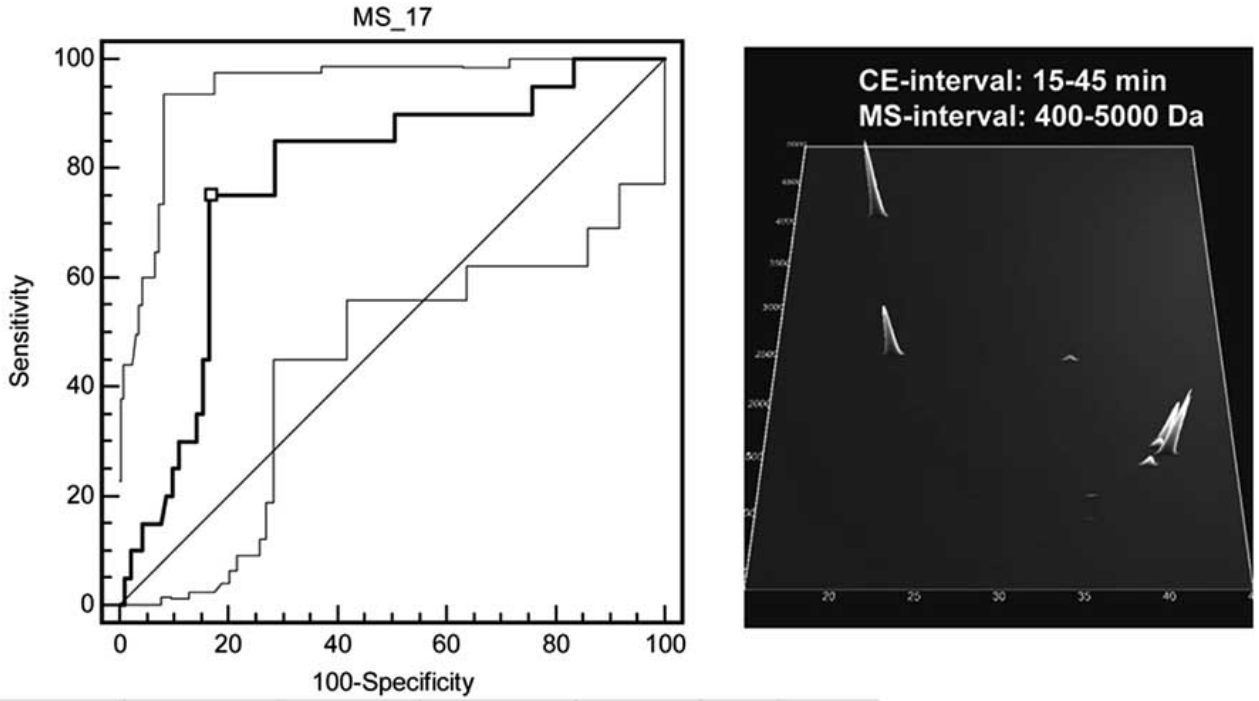
Fig. (4a). Graphs of 2 calculation models classification factor F (y- axis) plotted over the time (in days) after HSCT (days after HSCT X-axis) for patient 2725. This patient developed clinical signs of aGvHD grade III by day +29 post allo-HSCT, but the aGvHD proteomic pattern became positive as early as day +7. After responding to therapy the patient developed lung problems and was diagnosed for cGvHD by day +85, at this time the proteomic pattern became positive again and BOOP (bronchiolitis obliterans organizing pneumonia) was diagnosed clinically.

Fig. (4b). The same data are shown for patient 2791. Severe aGvHD grade IV was diagnosed by clinical manifestation of severe diarrhea and biopsy of the intestine by day +29. The proteomic pattern became positive earlier (day +15 and day +20). Despite severe immunosuppressive therapy, the aGvHD continued and finally led to the death of the patient due to infectious complications (fungal infection of the brain).

the marrow. During this time concomitant infectious complications, such as bacterial, viral and fungal infections bare the greatest risk for the patients. Around day +14 after allo-HSCT, engraftment of the donor cells occurs and is clinically defined by the development of more than 1000 leukocytes/μL. With the engraftment of donor cells a new complication impacts on morbidity and mortality after allo-HSCT, namely the graft-versus-host disease (GvHD). Donor-T-cells recognize the tissues and organs of the recipient as foreign and this may result in severe organ damage in some cases. The acute form (aGvHD) develops rapidly and may occur as early as 7 days after transplantation and poses a continuous problem for about the first 100 days. The chronic form is differentiated mainly by the different organ manifestation and the histological presentation of the sample biopsies taken. In general cGvHD arises after day +100, but sometimes cGvHD may develop earlier and is diagnosed due to its characteristic features. Differential diagnosis of aGvHD and infectious complications may be difficult; diagnosis of aGvHD is mainly based on clinical examination or on the histopathological examination of organ biopsies. Recently a proteomic pattern allowing early diagnosis of aGvHD was developed by our group [45]. Screening urine samples obtained from patients after allo-HSCT revealed that changes in the urinary proteome of these patients' could be linked to the development of aGvHD.

3.2. Diagnosis of Acute Graft-Versus-Host Disease (aGvHD) Based on CE-MS Spectra

To date, CE-MS analysis of urinary samples has been applied to several different diseases. Initially, chronic nephropathy was examined using the CE-MS as renal diseases were thought to result in indicative changes in the urinary proteome [15, 17, 19, 50, 51]. Screening urinary samples of more than 200 patients after allo-HSCT revealed that urine of these patients also held information about other diseases in addition to acute renal failure: diseases like septic complications and acute graft-versus-host disease (aGvHD) [52] could be displayed by indicative polypeptides. Previously we showed that the generated proteomic patterns could predict aGvHD development about 1 week prior to clinical manifestation [45, 52]. Sensitivity and specificity of the proteomic patterns have been evaluated in a multicenter fashion and were around 75-85% respectively [45]. Fig. (**4a,** upper part) shows the training sets used to establish the proteomic pattern specific for the diagnosis of aGvHD: Samples from 10



Criterion	Sensitivity	95% CI	Specificity	95% CI	+LR	-LR
>0.288 *	75.0	50.9 - 91.2	83.5	74.3 - 90.5	4.55	0.30

Area under the ROC curve	0.776
Standard error	0.065
95% Confidence interval	0.687 to 0.850
Significance level P (Area=0.5)	0.0001

Fig. (5). Differential diagnosis of acute graft-versus-host-disease based on proteomic patterns: The receiver operated curve (ROC) for a set of patients (n=110) transplanted at Hannover Medical School is shown here. The retrospective anlysis of prospectively screened patients revealed correct diagnosis of aGvHD with a sensitivity of 75% (CI 95%: 50.9 – 91.4) and a specificity of 83.5% (CI 95% 74.3 – 90.5). Criterion: cut off for diagnosis of aGvHD according to classification factor (CF) >0.288.

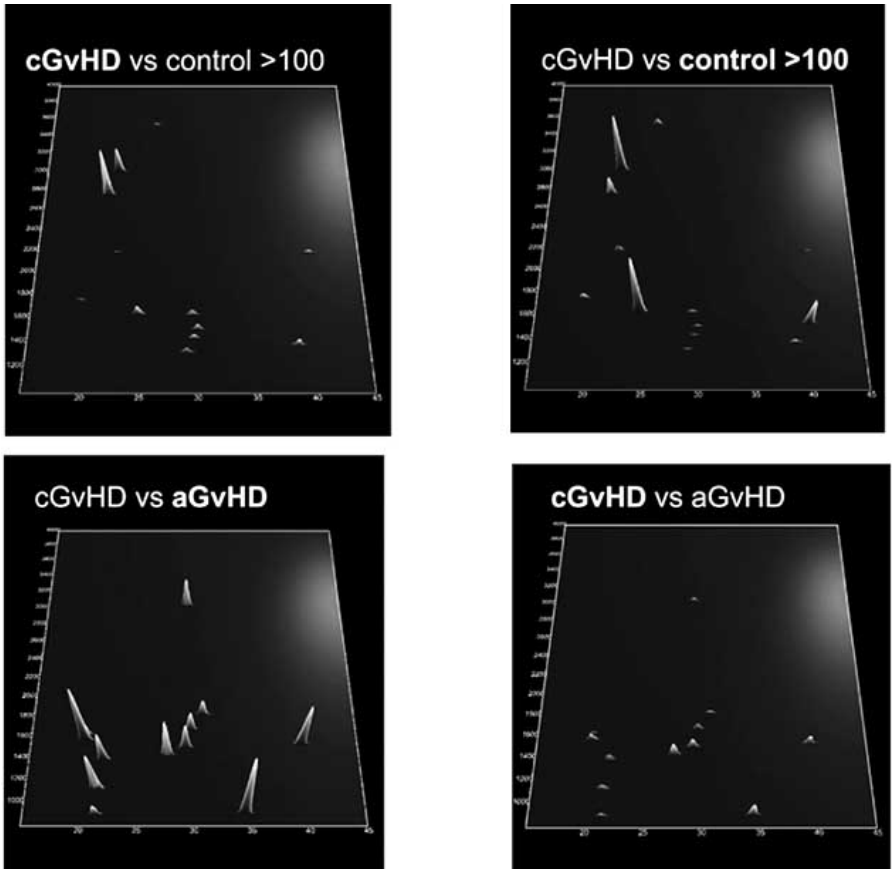


Fig. (6). Proteomic patterns for diagnosis of chronic GvHD: the left side shows polypeptides forming the pattern specific for diagnosis of cGvHD compared against the controls collected after day +100 (**control** "bold print") The right side represents the polypeptides that differentiate the controls > day +100 from patients with cGvHD (**cGvHD** "bold print"). The lower panel shows the polypeptide pattern differentiation **aGvHD** from cGvHD and vice versa. The X-axis shows the migration time in the CE, the Y-axis shows the deconvoluted mass in kDa.

patients with aGvHD were chosen to establish the aGvHD-pattern. The control set for the aGvHD patients consisted of 23 patients undergoing allo-HSCT without complications at early time points (prior day +100) post HSCT [45]. In a similar way we proceeded to generate a pattern for the early diagnosis of development of chronic GvHD. In order to define changes in patients with cGvHD early on, we used samples from 10 patients diagnosed with cGvHD. Patients without acute or chronic GvHD post HSCT were sampled at a time when these patients are off immunosuppression (more than 100 days post allo-HSCT) and those served as a control set. At the time of sample collection they had neither relapse of the original disease nor infectious complications. Application of MosaiquesCluster to the generated patterns yield classification factors (CF) for patients and the progress or the response to therapy are depicted very well by plotting the CF over time after HSCT. Fig. (**4a)** and (**4b**) show application of the classification model for aGvHD, which can both be used for diagnosis and follow up of patients with acute or chronic GvHD. The cut off for diagnosis of aGvHD is a CF greater than +0.19.

Response to therapy is depicted by the decline of the CF to values below. In addition, receiver operated curves (ROC) for diagnosis of aGvHD on more then 200 patients transplanted at Hannover Medical School are shown. Sensitivity and specificity for diagnosis of aGvHD are shown (Fig. **5**).

3.3. Biomarkers for Diagnosis of cGvHD

A proteomic pattern consisting of polypeptides that allow differentiation of aGvHD from cGvHD as well as patients without complications (control day +100) is shown in Fig. (**7**). The pre-defined set of polypeptides was further validated by randomly excluding 30% of available samples [45, 52]. Discriminatory polypeptides were included to a support vector machine (SVM) derived prediction model using MosaiquesCluster software [27]. Polypeptides identifying a GvHD can also be excreted by patients with cGvHD, but due to the differences between acute T-cell response and delayed T-cell activation, different markers can be identified, allowing discrimination of aGvHD and cGvHD. Fig. (**6**) shows the 3-dimensional depiction of the discriminatory marker peptides allowing differential diagnosis of cGvHD from controls (day +100) as well as from patients developing acute or late acute GvHD.

3.4. MS/MS Sequencing of Peptides Forming the aGvHD-Specific Pattern

Obtaining accurate sequence information is important to obtain evidence that the polypeptides defined as identical (based on mass and migration time) are really identical –and to subsequently identify disease-related polypeptides to gain deeper insights into the pathophysiology of the disease. However, it should be noted that sequencing is not a prereq-

uisite for the use of a particular biomarker pattern for diagnostic purposes. Furthermore, as recently outlined in detail, [53-55] the mere knowledge of sequence without knowledge of post-translational-modifications, often results in erroneous data interpretation (e.g. if a specific albumin fragment represents a marker for diabetic nephropathy, then merely detecting albumin will not allow diagnostic assessment). CE can be used coupled off-line to a MALDI-TOF-TOF [56], but this can lead to unavoidable signal suppression in MALDI, thus yielding few sequences. We use LC for separation instead of CE for sequencing, because of the higher loading capacity and combine LC with an ESI-TOF-TOF. Further improvements can be accomplished by the use of FT-ICR instruments coupled to CE [57]. Polypeptides up to 10 kDa that contain posttranslational modifications can be identified based on accurate sequence-tags that can be generated by using the high FT-ICR instruments. Also orbitrap instruments equipped with electron-transfer dissociation (ETD) [53, 55] can be used for sequencing, increasing the number of sequences obtained. This approach allowed identification of >500 polypeptides [58]. Based on these data we have established a map of defined urinary polypeptides yielding the "urinary proteome" [33] with information on sequence and posttranslational modifications or changes in frequency or signal intensity in disease for each polypeptide. Sequencing of polypeptides forming a particular diagnostic pattern will lead to a more in depth understanding of the pathophysiology of hematologic disease or complications after HSCT, for example GvHD [30, 45].

CONCLUSIONS AND PERSPECTIVES

The data presented here indicate that the CE-MS technology holds significant potential to define biomarkers for a wide variety of diseases. Using this technology, it is possible to analyze hundreds of samples within a relatively short time, namely weeks, and also to process and compare the data to appropriate controls. The sample preparation and analyses techniques developed will allow undertaking large scale studies to identify e.g. clinically useful biomarker patterns or follow-up screenings for patients after allogeneic hematologic stem cell transplantation with emphasis on complications. In addition, efforts will also be undertaken to identify all polypeptides within the pattern using FT-ICR-MS or similar approaches in order to obtain further insight into the pathophysiology of the different diseases analyzed.

ACKNOWLEDGEMENTS

This work was sponsored in part by the Deutsche Jose Carreras Leukämie Stiftung DJCLS R 05/08 to EMW and Bernd Herstenstein and by the "Deutsche Forschungsgesellschaft" DFG Mi-658/01 and SFB 738 (MHH; project A2) to EMW and AG.

Conflict of interest: HM is the founder and co-owner of mosaiques diagnostics and therapeutics AG, whose potential product has been investigated here. EMW is married to HM.

ABBREVIATIONS

aGvHD	=	Acute graft-versus-host disease
allo-HSCT	=	Allogeneic, hematopoietic stem cell transplantation
CE	=	Capillary electrophoresis
CI 95%:	=	Confidence interval (95%)
cGvHD	=	Chronic graft-versus-host disease
CF	=	Classification factor
ESI	=	Electrospray ionization
FT-ICR	=	Fourier transformation ion cyclotron résonance
GvHD	=	Graft-versus-host disease
HPLC	=	High performance liquid Chromatography
LC	=	Liquid chromatography
Q	=	Quadrupole
MS	=	Mass spectrometry
MS/MS	=	Tandem mass spectrometry
m/z	=	Mass-to-charge ratio
MALDI	=	Matrix assisted laser desorption/ionization
ROC	=	Receiver operated curves
SELDI	=	Surface enhanced laser desorption ionization
SVM	=	Support vector machines
TOF	=	Time of flight
2DE-MS	=	Two dimensional gel electrophoresis

REFERENCE

[1] Vlahou, A.; Schellhammer, P. F.; Mendrinos, S.; Patel, K.; Kondylis, F. I.; Gong, L.; Nasim, S.; Wright, J. G., Jr. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. *Am. J. Pathol.*, **2001**, *158*, 491-1502.

[2] Kolch, W.; Neususs, 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.

[3] Issaq, H. J.; Conrads, T. P.; Prieto, D. A.; Tirumalai, R.; Veenstra, T. D. SELDI-TOF MS for diagnostic proteomics. *Anal. Chem.*, **2003**, *75*, 148A-55A.

[4] Weissinger, E. M.;Mischak, H. Application of proteomics to post-transplantational follow-up. *Methods Mol. Med.*, **2007**, *134*, 217-28.

[5] Kolch, W.; Mischak, H.; Pitt, A. R. The molecular make-up of a tumour: proteomics in cancer research. *Clin. Sci. (Lond)*, **2005**, *108*, 369-83.

[6] Theodorescu, D.;Mischak, H. Mass spectrometry based proteomics in urine biomarker discovery. *World J. Urol.*, **2007**, *25*, 435-43.

[7] Haubitz, M.; Wittke, S.; Weissinger, E. M.; Walden, M.; Rupprecht, H. D.; Floege, J.; Haller, H.; Mischak, H. Urine protein patterns can serve as diagnostic tools in patients with IgA nephropathy. *Kidney Int.*, **2005**, *67*, 2313-20.

[8] Rossing, K.; Christensen, P. K.; Hansen, B. V.; Carstensen, B.; Parving, H. H. Optimal dose of candesartan for renoprotection in type 2 diabetic patients with nephropathy: a double-blind randomized cross-over study. *Diabetes Care*, **2003**, *26*, 150-5.

[9] Rossing, K.; Mischak, H.; Parving, H. H.; Christensen, P. K.; 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.

[10] Mohammed, S.; Chalmers, M. J.; Gielbert, J.; Ferro, M.; Gora, L.; Smith, D. C.; Gaskell, S. J. A novel tandem quadrupole mass spec-

trometer allowing gaseous collisional activation and surface induced dissociation. *J. Mass Spectrom.*, **2001**, *36*, 1260-8.

[11] Sleno L.;Dietrich A.Volmer Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.*, **2004**, 1091-112.

[12] Zurbig, P.; Renfrow, M. B.; Schiffer, E.; Novak, J.; Walden, M.; Wittke, S.; Just, I.; Pelzing, M.; Neususs, C.; Theodorescu, D.; Root, K. E.; Ross, M. M.; Mischak, H. Biomarker discovery by CE-MS enables sequence analysis *via* MS/MS with platform-independent separation. *Electrophoresis*, **2006**, *27*, 2111-25.

[13] Figeys, D.; van, O., I; Ducret, A.; Aebersold, R. Protein identification by capillary zone electrophoresis/microelectrospray ionization-tandem mass spectrometry at the subfemtomole level. *Anal. Chem.*, **1996**, *68*, 1822-8.

[14] Ingendo, A.; Kiehne, A.; Greiner, M. CE-MS-MS of Peptides Using a Novel Orthogonal CE ESI Sprayer and Ion trap MS. *Chromatographia*, **1999**, *49*, 87-92.

[15] Kaiser, T.; Wittke, S.; Just, I.; Krebs, R.; Bartel, S.; Fliser, D.; Mischak, H.; Weissinger, E. M. Capillary electrophoresis coupled to mass spectrometer for automated and robust polypeptide determination in body fluids for clinical use. *Electrophoresis*, **2004**, *25*, 2044-55.

[16] Weissinger, E. M.; Wittke, S.; Kaiser, T.; Haller, H.; Bartel, S.; Krebs, R.; Golovko, I.; Rupprecht, H. D.; Haubitz, M.; Hecker, H.; Mischak, H.; Fliser, D. Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes. *Kidney Int.*, **2004**, *65*, 2426-34.

[17] Weissinger, E. M.; Kaiser, T.; Meert, N.; De Smet, R.; Walden, M.; Mischak, H.; Vanholder, R. C. Proteomics: a novel tool to unravel the patho-physiology of uraemia. *Nephrol. Dial. Transplant.*, **2004**, *19*, 3068-77.

[18] Weissinger, E. M.; 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.

[19] Wittke, S.; Fliser, D.; Haubitz, M.; Bartel, S.; Krebs, R.; Hausadel, F.; Hillmann, M.; Golovko, I.; Koester, P.; Haller, H.; Kaiser, T.; Mischak, H.; Weissinger, E. M. Determination of peptides and proteins in human urine with capillary electrophoresis–mass spectrometry, a suitable tool for the establishment of new diagnostic markers. *J. Chromatogr. A*, **2003**, *1013*, 173-81.

[20] Johannesson, N.; Wetterhall, M.; Markides, K. E.; Bergquist, J. Monomer surface modifications for rapid peptide analysis by capillary electrophoresis and capillary electrochromatography coupled to electrospray ionization-mass spectrometry. *Electrophoresis*, **2004**, *25*, 809-16.

[21] Hernandez-Borges, J.; Neususs, C.; Cifuentes, A.; Pelzing, M. On-line capillary electrophoresis-mass spectrometry for the analysis of biomolecules. *Electrophoresis*, **2004**, *25*, 2257-81.

[22] Neususs, C.; Pelzing, M.; Macht, M. A robust approach for the analysis of peptides in the low femtomole range by capillary electrophoresis-tandem mass spectrometry. *Electrophoresis*, **2002**, *23*, 3149-59.

[23] Petersson, M. A.; Hulthe, G.; Fogelqvist, E. New sheathless interface for coupling capillary electrophoresis to electrospray mass spectrometry evaluated by the analysis of fatty acids and prostaglandins. *J. Chromatogr. A*, **1999**, *854*, 141-54.

[24] Sassi A.P.; Frank Andel III; Hans-Marcus L.Bitter1; Michael P.S.Brown1; Robert G.Chapman; Jeraldine Espiritu; Alfred C.Greenquist; Isabelle Guyon; Mariana Horchi-Alegre; Kathy L.Stults; Ann Wainright; Jonathan C.Heller; John T.Stults An automated, sheathless capillary electrophoresis-mass spectrometry platform fordiscovery of biomarkers in human serum. *Electrophoresis*, **2005**, *26*, 1500-12.

[25] Ullsten, S.; Zuberovic, A.; Wetterhall, M.; Hardenborg, E.; Markides, K. E.; Bergquist, J. A polyamine coating for enhanced capillary electrophoresis-electrospray ionization-mass spectrometry of proteins and peptides. *Electrophoresis*, **2004**, *25*, 2090-9.

[26] Klampfl, C. W. Recent advances in the application of capillary electrophoresis with mass spectrometric detection. *Electrophoresis*, **2006**, *27*, 3-34.

[27] Decramer, S.; Wittke, S.; Mischak, H.; Zurbig, P.; Walden, M.; Bouissou, F.; Bascands, J. L.; Schanstra, J. P. Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat. Med.*, **2006**, *12*, 398-400.

[28] Kaiser, T.; Hermann, A.; Kielstein J.T.; Wittke, S.; Bartel, S.; Krebs, R.; Hausadel, F.; Hillmann, M.; Golovko, I.; Koester, P.; Haller, H.; Weissinger, E. M.; Fliser, D.; Mischak, H. Capillary electrophoresis coupled to mass spectrometry to establish polypeptide patterns in dialysis fluids. *J. Chromatogr. A*, **2003**, *1013*, 157-71.

[29] Theodorescu, D.; Wittke, S.; Ross, M. M.; Walden, M.; Conaway, M.; Just, I.; Mischak, H.; Frierson, H. F. Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol.*, **2006**, *7*, 230-40.

[30] Weissinger, E. M.; Mischak, H.; Ganser, A.; Hertenstein, B. Value of proteomics applied to the follow-up in stem cell transplantation. *Ann. Hematol.*, **2006**, *85*, 205-11.

[31] Song, E. J.; Babar, S. M.; Oh, E.; Hasan, M. N.; Hong, H. M.; Yoo, Y. S. CE at the omics level: towards systems biology--an update *Electrophoresis*, **2008**, *29*, 129-42.

[32] Bakry, R.; Huck, C. W.; Najam-ul-Haq, M.; Rainer, M.; Bonn, G. K. Recent advances in capillary electrophoresis for biomarker discovery. *J. Sep. Sci.*, **2007**, *30*, 192-01.

[33] Metzger, J.; Schanstra, J. P.; Mischak, H. Capillary electrophoresis-mass spectrometry in urinary proteome analysis: current applications and future developments. *Anal. Bioanal. Chem.*, **2008**,

[34] Shiwa M; Nishimura Y; Wakatabe R; Fukawa A; Arikuni H; Ota H; Kato Y; Yamori T Rapid discovery and identification of a tissue-specific tumor biomarker from 39 human cancer cell lines using the SELDI ProteinChip platform. *Biochem. Biophys. Res. Commun.*, **2003**, *309*, 18-25.

[35] Marshall, J.; Kupchak, P.; Zhu, W.; Yantha, J.; Vrees, T.; Furesz, S.; Jacks, K.; Smith, C.; Kireeva, I.; Zhang, R.; Takahashi, M.; Stanton, E.; Jackowski, G. Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. *J. Proteome. Res.*, **2003**, *2*, 361-72.

[36] Rai, A. J.; Gelfand, C. A.; Haywood, B. C.; Warunek, D. J.; Yi, J.; Schuchard, M. D.; Mehigh, R. J.; Cockrill, S. L.; Scott, G. B.; Tammen, H.; Schulz-Knappe, P.; Speicher, D. W.; Vitzthum, F.; Haab, B. B.; Siest, G.; Chan, D. W. HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome sample**s**. *Proteomics*, **2005**, *5*, 3467-74.

[37] Tammen, H.; Schulte, I.; Hess, R.; Menzel, C.; Kellmann, M.; Mohring, T.; Schulz-Knappe, P. Peptidomic analysis of human blood specimens: Comparison between plasma specimens and serum by differential peptide display. *Proteomics*, **2004**, *5*, 3414-22.

[38] Hanash S. HUPO Initiatives Relevant to Clinical Proteomics. *Mol. Cell. Proteomics,* **2004**, *4*, 298-301.

[39] Spahr, C. S.; Davis, M. T.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Courchesne, P. L.; Chen, K.; Wahl, R. C.; Yu, W.; Luethy, R.; Patterson, S. D. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. *Proteomics*, **2001**, *1*, 93-107.

[40] Davis, M. T.; Spahr, C. S.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Yu, W.; Luethy, R.; Patterson, S. D. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. II. Limitations of complex mixture analyses. *Proteomics*, **2001**, *1*, 108-17.

[41] Timperman AT;Aebersold R. Peptide electroextraction for direct coupling of in-gel digests with capillary LC-MS/MS for protein identification and sequencing. *Anal. Chem.*, **2000**, *72*, 4115-21.

[42] 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.

[43] Theodorescu, D.; Wittke, S.; Ross, M. M.; Walden, M.; Conaway, M.; Just, I.; Mischak, H.; Frierson, H. F. Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol.*, **2006**, *7*, 230-40.

[44] Theodorescu, D.; Fliser, D.; Wittke S.; Mischak, H.; Krebs, R.; Walden, M.; Ross, M.; Eltze, E.; Bettendorf, O.; Wulfing, C.; Semjonow, A. Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis*, **2005**, *26*, 2797-808.

[45] Weissinger, E. M.; Schiffer, E.; Hertenstein, B.; Ferrara, J. L.; Holler, E.; Stadler, M.; Kolb, H. J.; Zander, A.; Zurbig, P.; Kellmann, M.; Ganser, A. Proteomic patterns predict acute graft-

versus-host disease after allogeneic hematopoietic stem cell transplantation. *Blood*, **2007**, *109*, 5511-9.

[46] Wittke, S.; Haubitz, M.; Walden, M.; Rohde, F.; Schwarz, A.; Mengel, M.; Mischak, H.; Haller, H.; Gwinner, W. Detection of acute tubulointerstitial rejection by proteomic analysis of urinary samples in renal transplant recipients. *Am. J. Transplant.*, **2005**, *5*, 2479-88.

[47] Peng S.; Xu Q; Ling X.B; Peng X.; Du W; Chen L Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. *FEBS Lett.*, **2003**, *555*, 358-62.

[48] Brown, M. P.; Grundy, W. N.; Lin, D.; Cristianini, N.; Sugnet, C. W.; Furey, T. S.; Ares, M., Jr.; Haussler, D. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 262-7.

[49] Raudys, S. How good are support vector machines? *Neural Netw.*, **2000**, *13*, 17-9.

[50] Mischak, H.; Kaiser, T.; Walden, M.; Hillmann, M.; Wittke, S.; Herrmann, A.; Knueppel, S.; Haller, H.; Fliser, D. Proteomic analysis for the assessment of diabetic renal damage in humans. *Clin. Sci. (Lond)*, **2004**, *107*, 485-95.

[51] Wittke, S.; Kaiser, T.; Mischak, H. Differential polypeptide display: the search for the elusive target. *J Chromatogr. B Anal. Technol. Biomed. Life Sci.*, **2004**, *803*, 17-26.

[52] Kaiser, T.; Kamal, H.; Rank, A.; Kolb, H. J.; Holler, E.; Ganser, A.; Hertenstein, B.; Mischak, H.; Weissinger, E. M. Proteomics applied to the clinical follow-up of patients after allogeneic hematopoietic stem cell transplantation. *Blood*, **2004** , *104*, 340-9.

[53] Good, D. M.; Thongboonkerd, V.; Novak, J.; Bascands, J. L.; Schanstra, J. P.; Coon, J. J.; Dominiczak, A.; Mischak, H. 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.

Received: 28 January, 2009 Revised: 10 February, 2009 Accepted: 25 February, 2009

[54] Zimmerli, L. U.; Schiffer, E.; Zurbig, P.; Good, D. M.; Kellmann, M.; Mouls, L.; Pitt, A. R.; Coon, J. J.; Schmieder, R. E.; Peter, K. H.; Mischak, H.; Kolch, W.; Delles, C.; Dominiczak, A. F. Urinary proteomic biomarkers in coronary artery disease. *Mol. Cell Proteomics*, **2008**, *7*, 290-8.

[55] Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell Proteomics*, **2007**, *6*, 1942-51.

[56] Villanueva, J.; Philip, J.; Entenberg, D.; Chaparro, C. A.; Tanwar, M. K.; Holland, E. C.; Tempst, P. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal. Chem.*, **2004**, *76*, 1560-70.

[57] Chalmers, M. J.; Mackay, C. L.; Hendrickson, C. L.; Wittke, S.; Walden, M.; Mischak, H.; Fliser, D.; Just, I.; Marshall, A. G. Combined top-down and bottom-up mass spectrometric approach to characterization of biomarkers for renal disease 1. *Anal. Chem.*, **2005**, *77*, 7163-71.

[58] Coon, D.; Beriwal, S.; Heron, D. E.; Kelley, J. L.; Edwards, R. P.; Sukumvanich, P.; Zorn, K. K.; Krivak, T. C. High-dose-rate rotte "Y" applicator brachytherapy for definitive treatment of medically inoperable endometrial cancer: 10-year results. *Int. J. Radiat. Oncol. Biol. Phys.*, **2008**, pre-published ahead of print.

[59] Leslie, C.; Eskin, E.; Weston, J.; Noble, W. S. Mismatch string kernels for SVM protein classification. *Bioinformatics*, **2004**, *20*, 467-76.

[60] Jacobsohn, D. A.;Vogelsang, G. B. Acute graft versus host disease 3. *Orphanet J. Rare Dis.*, **2007**, *2*, 35-40

[61] Flowers, M. E.; Lee, S.; Vogelsang, G. An update on how to treat chronic GVHD 1. *Blood*, **2003**, *102*, 2312-7

[62] Lee, S. J.; Vogelsang, G.; Gilman, A.; Weisdorf, D. J.; Pavletic, S.; Antin, J. H.; Horowitz, M. M.; Akpek, G.; Flowers, M. E.; Couriel, D.; Martin, P. J. A survey of diagnosis, management, and grading of chronic GVHD. *Biol. Blood Marrow Transplant.*, **2002**, *8*, 32-9.