

# Proteomics applied to the clinical follow-up of patients after allogeneic hematopoietic stem cell transplantation

Thorsten Kaiser, Haytham Kamal, Andreas Rank, Hans-Jochem Kolb, Ernst Holler, Arnold Ganser, Bernd Hertenstein, Harald Mischak, and Eva Maria Weissinger

**A phase 1 diagnostic study was performed to evaluate a novel technology for clinical proteom research based on capillary electrophoresis and mass spectrometry. Urine from 40 patients after hematopoietic stem cell transplantation (HSCT; 35 allogeneic, 5 autologous) and 5 patients with sepsis was collected for a period of 100 days and analyzed. More than 1000 different polypeptides could be detected in individual samples. Polypeptide patterns excreted in the urine of patients were significantly different from those of**

**healthy volunteers. No significant differences were detected comparing different conditioning regimens. The aim of this study was to identify polypeptide patterns functioning as early indicators of graft-versus-host disease (GVHD). Eighteen patients developed GVHD after allogeneic HSCT. Sixteen differentially excreted polypeptides formed a pattern of early GVHD markers, allowing discrimination of GVHD from patients without complications with 82% specificity and 100% sensitivity, cross-validated. Inclusion of 13 sepsis-specific polypep-**

**tides allowed us to distinguish sepsis from GVHD with a specificity of 97% and a sensitivity of 100%. Sequencing 2 prominent GVHD-indicative polypeptides led to the identification of a peptide from leukotriene A4 hydrolase and a peptide from serum albumin. The data reveal that capillary electrophoresis and mass spectrometry allow identification of biomarkers for a variety of diseases or related complications. (Blood. 2004;104:340-349)**

© 2004 by The American Society of Hematology

## Introduction

Allogeneic hematopoietic peripheral blood stem cell transplantation (allo-HSCT) is applied with success to many hematopoietic malignancies. Despite its curative potential, the application of allo-HSCT is limited by life-threatening complications such as severe acute graft-versus-host disease (GVHD).<sup>1,2</sup> Depending on the type of transplantation, the immunosuppressive treatment, and the underlying disease, between 35% and 70% of patients develop GVHD, requiring immunosuppressive treatment in more than 35% of patients.<sup>3</sup> Early diagnosis and better control of GVHD will be necessary to increase the safety of allo-HSCT, also with respect to a broader application of HSCT.

Currently, diagnosis of GVHD is based mainly on clinical parameters such as skin rash, diarrhea, elevation of serum liver enzymes, or other. Differential diagnosis of GVHD depends on organ biopsies to distinguish GVHD from other common complications that present with similar clinical symptoms (eg, reactivation of endogenous viruses or medication-induced side effects). The use of biomarkers such as differentially expressed or excreted polypeptides and proteins has the potential for improving early and accurate diagnosis of GVHD and other complications of allo-HSCT without requiring invasive procedures, such as biopsies.<sup>4</sup> Single molecules are currently described as potential markers for GVHD<sup>5-7</sup> but the data of all molecules potentially involved are not

yet reported. Due to a lack of suitable technology, the search for polypeptides or proteins involved in GVHD up to now has been naturally biased by the preferential analysis of known molecules with potential pathophysiologic importance. An analytic display of all proteins and peptides present or changed after allo-HSCT might allow the gain of significantly more insight into the development of and processes involved in GVHD. Proteomics are being developed to characterize and identify the molecules significant for different diseases, enabling early identification of biomarkers and early intervention in cancer.<sup>8-10</sup> A technique allowing the reproducible analysis of all polypeptides present in complex biologic samples within a short time and suitable for high throughput analysis was recently developed by our group.<sup>11-13</sup> The stable on-line coupling of capillary electrophoresis and mass spectrometry (CE-MS), together with advances in software analysis, led to the display of more than 1000 polypeptides present in individual samples, identified via their particular migration time in the CE and their actual mass.<sup>14</sup> All data generated from individual samples are stored in a database, allowing intraindividual comparison of the samples taken at different time points, yielding a patient-specific pattern, as well as comparison of patient groups and controls.

Screening of urine from healthy volunteers led to the establishment of a "normal urine polypeptide pattern" consisting of more

From Mosaiques Diagnostics & Therapeutics and Hannover Medical School, Departments of Hematology/Oncology and Nephrology, Hannover, Germany; Jose-Carreras-Unit, Klinikum Großhadern, München, Germany; and the Department of Hematology, University of Regensburg, Regensburg, Germany.

Submitted February 11, 2004; accepted March 21, 2004. Prepublished online as *Blood* First Edition Paper, March 30, 2004; DOI 10.1182/blood-2004-02-0518.

Supported in part by grant no. 0312939 from BioProfile "Funktionelle Genomanalyse" and by grant no. 203.19-32329-5-461 from the Lower Saxony Ministry of Economy (H.M.).

T.K., H.M., and E.M.W. are employed by Mosaiques Diagnostics & Therapeutics, whose potential product was studied in the present work.

**Reprints:** Eva M. Weissinger, Feodor-Lynen Str 21, 30625 Hannover, Germany; e-mail: eviweissinger@gmx.de.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2004 by The American Society of Hematology

than 500 polypeptides.<sup>15</sup> This allowed a comparison with patterns obtained from patients with different diseases and led to the detection of polypeptide patterns indicative of the health status of individuals. Here we report the application of CE-MS as a phase 1 clinical diagnosis trial to screen patients with hematologic malignancies after allo-HSCT, to evaluate the feasibility of identifying marker peptide patterns for GVHD or other complications related to allo-HSCT. Thus, samples were collected from 45 patients (35 after allo-HSCT, 5 after autologous transplantation, and 5 with septic complications) and analyzed by CE-MS over a period of about 100 days after HSCT.

## Patients, materials, and methods

### Patients

The protocol for this study was approved by the local ethics committees and informed consent was obtained from all participants. There were 40 patients who underwent transplantation at Hannover Medical School, the University of Regensburg, and the University of Munich who were included in the analysis. There were 35 patients (26 with acute myeloid leukemia [AML], 4 with acute lymphocytic leukemia [ALL], one with high-risk chronic myeloid leukemia [CML], one with non-Hodgkin lymphoma [NHL], one with follicular lymphoma, one with multiple myeloma [MM], and one with myelodysplastic syndrome [MDS]) who received transplants from allogeneic donors, whereas 5 patients (1 with AML, 2 with MM, 2 with NHL) received autologous stem cells. Urine samples were collected prior to conditioning and then twice a week from each patient over 20 to 100 days after HSCT.

In addition, samples from 5 patients in the intensive care unit (ICU) with severe septic complications were included.

### Conditioning and transplantation

Nineteen patients were treated with reduced-intensity conditioning regimens<sup>16</sup> consisting of low-dose total body irradiation (TBI) and fludarabine (FAra) in the majority of the protocols; 10 of these patients were treated according to the FLAMSA (fludarabine, amsacrine) protocol.<sup>17</sup> Standard-intensity protocols were used to treat 16 patients: 8 received TBI ( $6 \times 2$  Gy over 3 days) and cyclophosphamide (60 mg/kg for 2 days), and 8 received busulfan (4 mg/kg for 4 days) followed by cyclophosphamide (120 mg/kg for an additional 2 days). There were 5 patients who received additional radioimmunotherapy (RIT).

There were 20 patients who received transplants from unrelated donors (19 from a matched unrelated donor [MUD], 1 from a mismatch), 12 patients who received stem cells from HLA-identical family donors, and 3 patients who received stem cells from haploidentical family donors. Stem cell sources were peripheral blood stem cells in 33 patients and bone marrow in 4 patients. There were 2 patients with haploidentical donors who received bone marrow plus peripheral blood stem cells.

GVHD prophylaxis was methotrexate (MTX) or mycophenolate mofetil (MMF) and cyclosporin A (CSA) in 32 patients, and T-cell depletion in 3 patients.

There were 5 patients (1 with MM, 1 with AML, and 3 with NHL) who received transplants of autologous peripheral stem cells and who served as controls for the allo-response and GVHD patterns in this setting. The clinical data of the patients after HSCT are summarized in Table 1. In addition, samples were obtained from 5 patients with severe septic complications from the ICU. Patterns of these patients and from one patient after HSCT developing sepsis were compared with the polypeptide patterns "significant" for GVHD.

### Sample preparation for capillary electrophoresis

Spot urine samples were collected twice a week starting before conditioning until discharge from the ward and stored at  $-20^{\circ}\text{C}$  until analysis. Aliquots of 2 mL were adjusted to pH 10.0 using ammonia and cleared by centrifugation for 10 minutes at 13 000g as previously described.<sup>11</sup> In brief,

2 mL was applied onto a Pharmacia C2 column (Pharmacia, Uppsala, Sweden) to enrich for proteins and peptides and to remove urea, salts, and other confounding material. Polypeptides were eluted with 50% acetonitrile in  $\text{H}_2\text{O}$  containing 0.5% formic acid, frozen, and lyophilized overnight in a Christ Speed-Vac RVC 2-18/Alpha 1-2 (Christ, Osterode, Germany). The samples were resuspended in 20  $\mu\text{L}$  high performance liquid chromatography (HPLC)-grade water, sonicated for one minute, and centrifuged for 10 minutes (13 000g at  $4^{\circ}\text{C}$ ). The CE was injected with 100 nL, and the remaining material was stored at  $-80^{\circ}\text{C}$  for further evaluation like repeating runs or sequencing. All chemicals were purchased from Merck KGaA (Darmstadt, Germany).

### Capillary electrophoresis and mass spectrometry

About 100 nL of the prepared sample was injected into the CE system, a P/ACE MDQ (Beckman Coulter, Fullerton, CA) system, and separation was performed with 30 kV on the injection site. Upon application of high voltage the ions (polypeptides) in the sample were initially focused and subsequently separated by electrophoresis. For detection and characterization of the polypeptides, the CE was coupled on-line with an electrospray ionization time-of-flight mass spectrometer (ESI-TOF-MS). CE-ESI-MS coupling was accomplished using a CE-ESI-MS sprayer kit (Agilent Technologies, Palo Alto, CA). On-line TOF detection and data acquisition was performed on a Mariner Biospectrometry Workstation (Applied Biosystems, Farmington, CT). The data acquisition and the MS run were automatically controlled by the CE program via contact-close relays. To achieve highest signal intensities, the sheath flow rate was set to a minimum (100-1000 nL/min), while the nebulizer gas was turned off during acquisition. Under these conditions, 50 fmol of a set of different standard proteins and peptides resulted in signals with signal-to-noise ratios between 50 and 500.<sup>14</sup> The used TOF-MS delivers the data with mass accuracy better than 100 parts per million (ppm) under the conditions applied. This setup enables the analysis of amounts of less than 10 pg of polypeptides and can potentially yield the display of thousands of different polypeptides present in one individual sample without the need for any specific reagents.<sup>14,15</sup>

### Data processing

The enormous amount of information obtained in one single CE-MS run required the development of specialized software to evaluate the data in a reproducible and automated fashion.<sup>18</sup> The software (MosaïquesVisu; Biomosaïques software, Hannover, Germany) recognizes MS peaks, determines the charge of each signal based on isotopic distribution and conjugated mass, and subsequently generates a list of polypeptides defined by mass and migration time, which is the basis for comparison with other samples and is stored for each individual sample in the database. The signal intensity of the individual molecules is shown in a color code (ranging from 0 to 25 000 MS counts) and serves as a measure for the relative abundance of particular peptides. The data generation and data processing are shown in Figures 1 and 2.

To account for run-to-run variations, the CE migration times were normalized using 104 polypeptides present with high probability in urine samples. This allowed comparison and search of conformity within different individual samples. The signal intensity was normalized to the total ion current. Polypeptides were considered identical if the mass deviation was less than 300 ppm and the CE migration-time deviation was less than 5 minutes.

### Support vector machine

Statistical analysis was performed using support vector machines. This tool has the advantage of discriminating data in high-dimensional parameter space. Its fast and stable algorithms showed good performance in

**Table 1. Clinical data of the patients after HSCT**

Patient no./disease	Age, y/sex	Transplantation date, d/mo/y	HSCT	Conditioning	GVHD	Day after HSCT	Other complications
1123/AML 1.R	61/M	19/06/03	MUD PB SCT	FLAMSA	Y, grade II skin	+14	None
1124/AML refr.	60/M	05/09/03	h-PBSC/BMT	FLAMSA	Y, grade II skin	+28	Fever day -5/+6
1125/ALL Cr2	54/M	11/09/03	MUD PB SCT	FLAMSA	Y, grades I,II skin	+13	Fever day -5/+1
1127/AML 1.R	44/F	15/11/03	MUD PB SCT	FLAMSA	Y, grade II skin	+20	None
1128/AML 1.R	36/M	13/02/03	h-PBSC/BMT	FLAMSA	N	NA	None
1129/AML Cr1	50/F	14/10/03	SIB PB SCT	FLAMSA	N	NA	None
1130/AML Cr1	56/M	22/10/03	MUD PB SCT	FLAMSA	N	NA	None
1131/AML Cr1	58/F	28/10/03	SIB PB SC	FLAMSA	Y, grade II skin	+20	None
1132/AML-refr	54/F	29/10/03	MUD PB SCT	FLAMSA	N	NA	Fever day -2/5
1133/AML 1.R	34/M	28/11/03	h-PB SCT	FLAMSA	N	NA	Fever day 0
711/T-NHL	22/F	21/05/03	MUD BMT	TBI 8 Gy FAra CY	Y, grade II skin	+75	Fever
712/AML MDS	60/M	14/07/03	MUD PB SCT	FA-BCNUMel/RIT	Y, skin liver	+35	None
713/ALL/Cr1	26/M	09/07/03	MUD PB SCT	TBI-CY	Y, grade II intestine	+57	None
714/AML	50/F	07/07/03	MUD PB SCT	FA-BCNUMel/RIT	Y, skin	+26	None
717/Foll. Ly	46/M	27/08/03	MUD PB SCT	TBI 8 Gy FAra CY	Y, grade III skin	+19	None
720/AML	49/F	17/09/03	SIB PB SCT	FA-BCNUMel/RIT	N	NA	None
721/AML	37/M	25/09/03	SIB PB SCT	FA-BCNUMel/RIT	N	NA	None
291/AML Cr2	62/M	06/02/03	MUD PB SCT	FAra-Bu	N	NA	None
292/AML Cr1	51/F	07/02/03	MUD PB SCT	RIT-Bu-CY	Y, skin III intestine IV	+32	VOD +9/ARF+13
293/ALL Cr2	24/M	11/02/03	SIB PB SCT	TBI-VP16	Y, skin; intestine IV	+15	None
294/AML Cr1	40/F	10/02/03	SIB PB SCT	TBI-CY	N	NA	None
297/AML 1.R	52/M	01/03/03	MUD PB SCT	Bu-CY	N	NA	VOD, relapse day +77
312/AML Cr2	52/F	12/03/03	SIB PB SCT	RIT-Bu-CY	Y, skin intestine II	+49	D+14 - +49; FUO
313/AML Cr2	59/F	07/03/03	MUD PB SCT	RIT-Bu-CY	N	NA	FUO +12; CMV +46
314/AML Cr1	22/F	21/03/03	SIB PB SCT	RIT-Bu-CY	N	NA	None
315/MM PR	44/F	01/04/03	MUD PB SCT	FAra/Mel	Y, grade II skin	+13	None
316/AML Cr1	52/F	24/03/03	SIB PB SCT	TBI-CY	N	NA	Sepsis day +6
317/CML CP1	42/F	1/04/03	MM PB SCT	TBI-CY	Y, grade II skin	+28	CMV-R day +28
318/AML Cr2	45/F	14/03/03	MUD PB SCT	RIT/FAra/Mel	N	NA	FUO day +6
481/ALL Cr3	56/F	7/04/03	MUD BMT	TBI-CY	N	NA	CMV-R day +15
483/AML Cr1	38/F	14/04/03	SIB PB SCT	TBI-CY	N	NA	CMV-R day +14
484/AML Cr2	44/M	24/04/03	MUD PB SCT	Bu-CY	Y, grade III skin	+19	None
485/AML Cr2	19/M	11/04/03	MUD PB SCT	TBI-CY	N	NA	None
486/AML Cr2	49/M	07/05/03	SIB PB SCT	Bu-CY	Y, grade III intestine	+85	None
910/MDS	41/M	18/07/03	SIB PB SCT	Bu-CY	Y, grade II	+22	ARF day +1
605/AML Cr1	35/F	26/06/03	Auto	Bu-CY	NA	NA	None
1085/MM PR	57/M	05/12/03	Auto	HD-Mel	NA	NA	None
1087/NHL PR	47/M	06/12/03	Auto	BEAM	NA	NA	None
1088/NHL Cr2	63/F	02/12/03	Auto	BEAM	NA	NA	None
1139/NHL PR	63/F	21/11/03	Auto	BEAM	NA	NA	None

Patients marked with "Fever" had bacteria in the blood culture in combination with fever on the indicated days and were treated with specific antibiotics. Patients with GVHD are marked "Y" and the date of the diagnosis is given in the column "Day after HSCT." HSCT indicates hematopoietic stem cell transplantation; MUD PB SCT, matched unrelated donor peripheral stem cells; SIB PB SCT, HLA-identical sibling donor peripheral blood stem cell transplantation; MM, mismatched unrelated donor; h-PBSC/BMT, haploidentical peripheral blood and bone marrow transplantation; Auto, autologous stem cell transplantation; FAra Cy, fludarabine + cyclophosphamide; FA-BCNUMel/RIT, fludarabine, 1,3 bis(2-chloroethyl)-1-nitrosourea-carmustine, melphalan/radioimmunotherapy; VP16, etoposide; FAra-Bu, fludarabine + busulfan; HD-Mel, high-dose melphalan; BEAM, dexamethasone, carmustine [BCNU], etoposide, cytarabine, melphalan; RIT, radioimmunotherapy; TBI, total body irradiation; CMV-R, cytomegalovirus-reactivation; ARF, acute renal failure; NA, not applicable; N, no GVHD; Y, GVHD; and FUO, fever of unknown origin.

the evaluation of clinical markers<sup>15,19</sup> and different areas of biologic analyses like DNA arrays.<sup>20</sup>

### Sequencing of discriminatory polypeptides using a matrix-assisted laser desorption/ionization (MALDI-tandem MS)

The MS-MS analyses were performed with a MALDI-TOF-TOF-MS (Ultraflex; Bruker Daltonik, Bremen, Germany) as previously described.<sup>14,21</sup> A complete CE run (as described in "Capillary electrophoresis and mass spectrometry") was spotted onto the MALDI target (one spot every 15 seconds) with the matrix solution (5 mg/mL sinapinic acid in 50% acetonitrile and 0.1% trifluoroacetic acid [TFA]) added as sheath liquid at 4  $\mu$ L/min. The target was subsequently examined in MS mode for the polypeptides of interest, based on the data from the CE-MS analyses. Polypeptides of interest were sequenced in MS-MS mode, without the use of collision-induced dissociation (CID) gas.

## Results

CE-MS allows the characterization of more than 1000 polypeptides present in complex biologic samples by the identification of the m/z of the molecules and their migration time in the CE. As a third dimension, the signal intensity (MS counts) serves as a measure for the relative abundance of the molecules detected. The MS spectra are generated by continuously analyzing the m/z and signal intensity of molecules present in a sample. These data are collected every 3 seconds over the CE run of about 45 minutes and this information is stored as individual spectra. Thus, about 1000 individual spectra are generated and stored, yielding the total ion current chromatogram.<sup>11,15</sup>

It is impossible to show all individual spectra and information obtained in a single CE-MS run. To depict the raw data, the information (resolution) is reduced and the data are shown as 3-dimensional contour plots (Figure 1). The contour plot of the raw data is generated by plotting all information received corresponding to mass per charge (x-axis), signal intensity (color code), and CE migration time (y-axis). Each dash shown in the contour plot corresponds to a single polypeptide with a particular charge state analyzed in the CE-MS run (Figure 1A). To display the information content of the plot, which is generated by the single spectra, a magnified view of MS data of 3 individual spectra at the corresponding parts of the plot is shown (Figure 1B-D). To reveal the high resolution of the technology used, a further magnification of MS spectra of 2 individual polypeptides is shown in Figure 1E-F.

As evident, it is impossible to “manually” analyze the data contained in about 1000 spectra within a reasonable time (less than 5 days). Hence, software (MosaiquesVisu) was developed to automatically process the enormous amount of data generated when CE-MS is applied toward the analysis of complex body fluids.

The processing of a CE-MS spectrum using MosaiquesVisu is shown in Figure 2. Signals (peaks above a certain signal-to-noise ratio, see spectra Figure 1B-F) within the individual spectra are recognized by the software and form the raw data plot (Figure 2A). In the next step, these peaks are combined to a list of “relevant signals.” Relevant signals are defined as peaks that are present in at least 3 consecutive spectra. These relevant signals are graphically depicted in Figure 2B and as part of the resulting peak list in Figure 2D (left panel). After collecting all relevant signals and assigning a charge to each peak, the actual mass is calculated based on the mass-per-charge ratio of the individual polypeptides (Figure 2C-D, right panel). As a last step, the data are normalized with respect to migration time and amplitude, using a set of internal standards present in all samples screened to date.<sup>15</sup> The data obtained for each individual sample are deposited as peak lists in a Microsoft Access database (Microsoft, Seattle, WA) containing the information on mass and CE migration time as a unique identity of the polypeptides and signal intensity, corresponding to the relative abundance of the molecules.

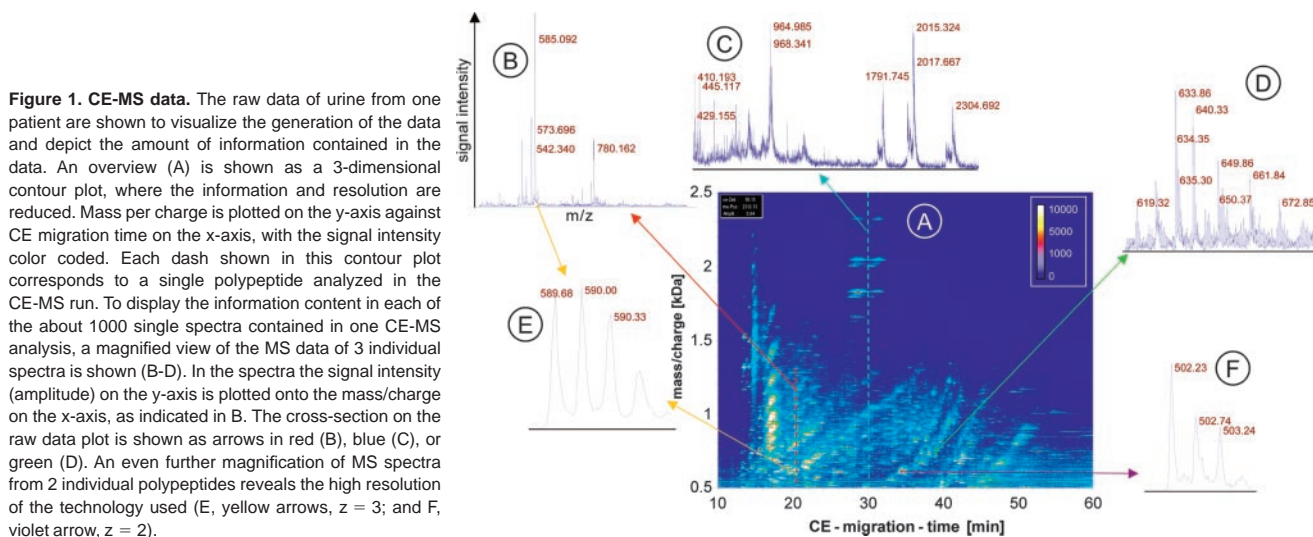
This type of analysis yields highly comparable spectra and hence enables the generation of a high-resolution polypeptide pattern, which is generally present in healthy individuals.<sup>15</sup> The information present in the individual spectra is compiled to

createtypical polypeptide patterns (Figure 3). To avoid that polypeptides present in only one or a few samples appear as “typical,” a penalty algorithm is utilized to reduce the signal intensity of these peptides. Hence, only consistently observed spots are displayed.

For this phase 1 diagnostic study, spot urine samples from 40 patients were collected after HSCT (35 allo-HSCT, 5 auto-HSCT) at different time points, ranging from the time before conditioning until 100 days after transplantation or discharge from the hospital. The clinical data of these patients showing the type of transplantation, conditioning treatment, and specific problems occurring within the observation period are summarized in Table 1. In addition, spot urine samples of 5 ICU patients with sepsis were collected. After performing the CE-MS analysis, the data were compiled and compared with healthy controls. The patient samples were grouped according to the conditioning regimens, resulting in 3 groups: those receiving reduced-intensity conditioning ( $n = 19$ ), those receiving standard-intensity conditioning ( $n = 16$ ), and those with the autologous protocols ( $n = 5$ ). The analysis of these data showed significant differences between patients and controls, as expected. Based on a list of 21 polypeptides differentially excreted in the urine (Table 2), patients could be distinguished from the controls with a specificity of 95% and a sensitivity of 100% after cross-validation using support vector machines.<sup>20</sup> As evident from Table 2, a comparison of the data from the 3 distinct groups of patients revealed no significant discrepancy between the different conditioning regimens. Hence, the differences seen between the healthy control and the patients were most likely due to the underlying disease and not due to different conditioning regimens.

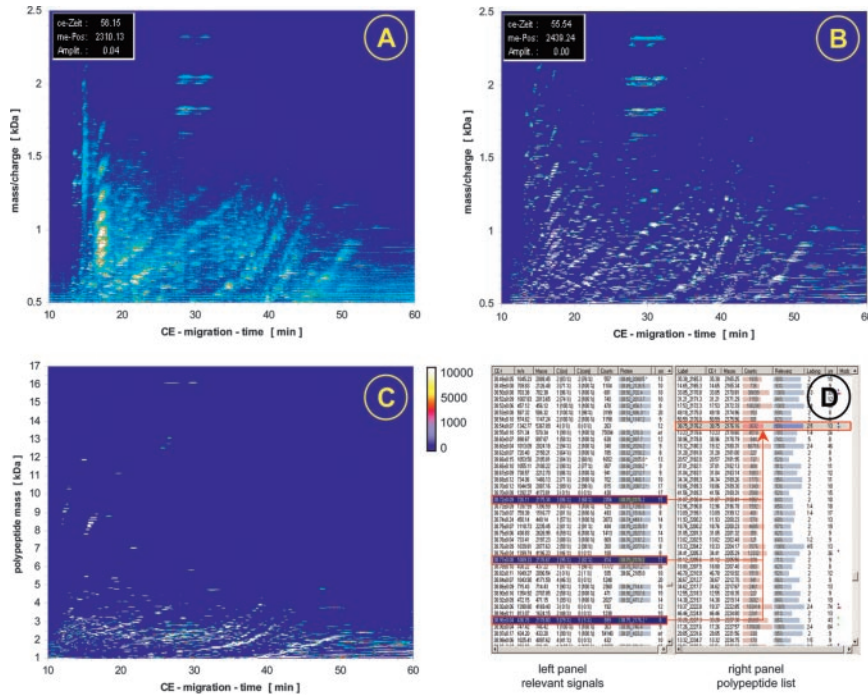
In patients who did not develop severe complications after HSCT within the observation period, the polypeptide pattern did not show any significant changes during this period. As an example, data from patient 291, who was treated according to the reduced-intensity conditioning prior to HSCT,<sup>16</sup> is shown in Figure 4A. No statistically significant change occurred in the polypeptide patterns from this patient when all the samples were compared with each other.

There were 18 patients who developed GVHD after allo-HSCT, between day 13 and day 85 (mean: day +26). The polypeptide data from the samples collected at the diagnosis of GVHD were compared with the data from the patients without complications.



**Figure 1. CE-MS data.** The raw data of urine from one patient are shown to visualize the generation of the data and depict the amount of information contained in the data. An overview (A) is shown as a 3-dimensional contour plot, where the information and resolution are reduced. Mass per charge is plotted on the y-axis against CE migration time on the x-axis, with the signal intensity color coded. Each dash shown in this contour plot corresponds to a single polypeptide analyzed in the CE-MS run. To display the information content in each of the about 1000 single spectra contained in one CE-MS analysis, a magnified view of the MS data of 3 individual spectra is shown (B-D). In the spectra the signal intensity (amplitude) on the y-axis is plotted onto the mass/charge on the x-axis, as indicated in B. The cross-section on the raw data plot is shown as arrows in red (B), blue (C), or green (D). An even further magnification of MS spectra from 2 individual polypeptides reveals the high resolution of the technology used (E, yellow arrows,  $z = 3$ ; and F, violet arrow,  $z = 2$ ).





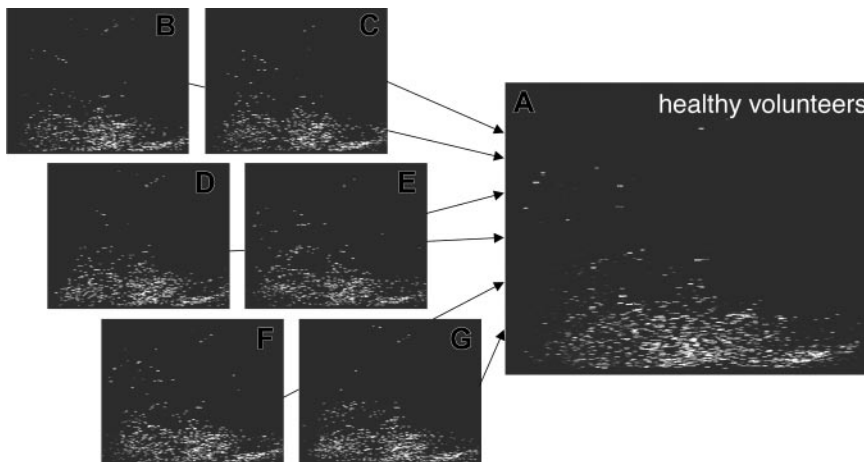
**Figure 2. Raw data contour plots and data processing.** The individual steps of the data processing are illustrated here, showing data obtained after screening urine from one patient. The initial raw data (A) are examined for signals that fit the criteria of “real polypeptide” signals, which are collected (B) and deposited in a peak list (D, left panel). Subsequently, a charge is assigned to each of these signals, conjugated peaks are combined as indicated by the red arrows, and as a result a spectrum based on mass and normalized migration time is obtained as graphically depicted in C. This spectrum contains the information on all peaks (a partial list is shown in D, right panel), including time, mass, intensity and signal-to-noise ratio.

This comparison revealed significant changes of polypeptide excretion at the time of the clinical diagnosis of GVHD. Only polypeptides present/absent with an absolute difference of more than 50% or with a more than 10-fold difference in the signal intensity between the 2 groups were accepted as significant disease markers. Under these very stringent conditions, 16 polypeptides potentially significant for GVHD could be identified. The molecular weight of some of these polypeptides and their appearance in the “GVHD,” “no complications,” or “septic” groups are shown in Table 3.

Figure 4B shows some of the polypeptide patterns from samples of patient 293 over a period of 35 days after HSCT. Grade II GVHD was diagnosed on day +14 with an increase of severity over time resulting in grade IV GVHD of the intestine. The polypeptide plots

were searched for differentially excreted peptides, defined in Table 3 as GVHD polypeptide pattern. As an example, the excretion of one of the marker peptides, defined via the actual mass of 1.85 kDa and the CE time of 34 minutes, is shown in Figure 4C (293). Peptide 1.85 kDa/34 minutes is excreted early before diagnosis of GVHD (day +1), gaining signal intensity over the progression of GVHD. The same coordinates (1.85 kDa/34 minutes) are shown for patient 291 (Figure 4C) without problems after HSCT, and this polypeptide could not be detected.

To show all data obtained from all patients in the groups, the compiled patterns and the distribution of some polypeptides with discriminatory value for all patients (GVHD indicated also as “w/” and “HSCT with no complication” indicated also as “w/o”) are shown in Figure 5. The magnified view in the lower panel of Figure



**Figure 3. Generation of patterns for groups of test persons.** The screening of samples obtained from different healthy individuals yielded the generation of several hundreds of plots to date. The storage of all data within the database allows the comparison of the individual data and the compilation of the data into a general pattern significant for health (shown here). This allows screening for common molecules excreted in different individuals and can also be applied toward the data exploitation of patients with different diseases or even different stages of disease. In analogy to the generation of the control pattern, pattern for patients undergoing HSCT were generated.

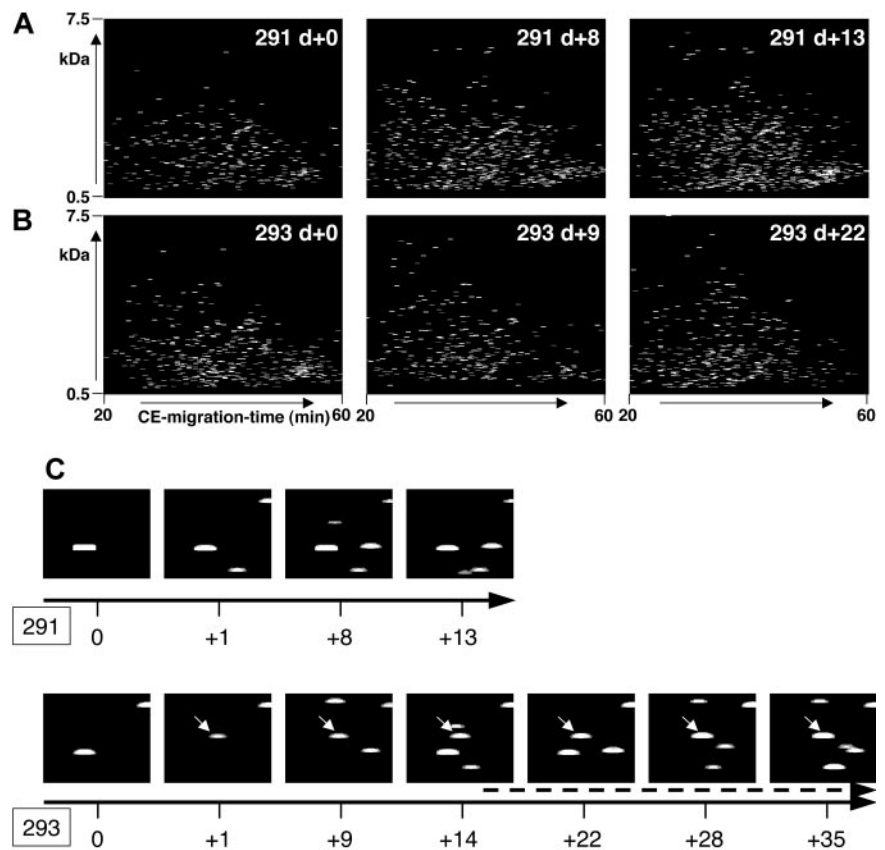
**Table 2. Polypeptide patterns for HSCT in comparison to control**

Polypeptide identification			Frequency by group				Mean amplitude, MS counts	
No.	Mass, kDa	Time, min	Auto. %	Reduced, %	Standard, %	Control, %	HSCT	Control
1	2.03	43.7	80	67	67	0	520	0
2	1.68	44.0	60	78	58	5	1 178	7
3	2.56	25.6	100	67	50	65	2 768	220
4	1.91	37.9	80	100	92	95	33 597	90 569
5	2.90	40.9	60	67	33	55	16 873	157
6	2.05	34.0	100	89	92	100	12 051	3 894
7	2.66	25.4	80	56	42	75	12 720	864
8	1.93	35.4	80	56	33	90	204	3 762
9	0.89	31.7	0	0	0	60	0	233
10	1.22	57.0	0	0	0	60	0	292
11	3.71	31.4	0	22	25	80	501	1 599
12	1.49	43.0	0	33	17	80	1 035	1 091
13	3.28	36.1	0	22	42	90	617	2 040
14	1.86	57.1	0	11	17	75	61	126
15	3.41	38.1	0	22	50	95	1 694	2 332
16	3.43	42.7	60	22	8	90	110	1 290
17	3.30	45.0	40	22	25	95	272	815
18	1.26	49.2	0	11	17	80	26	1 563
19	2.85	33.8	20	11	8	85	116	1 132
20	3.50	42.8	20	0	0	80	542	3 040
21	1.47	40.2	20	33	0	95	31	1 265

Shown here are 21 of the polypeptides differentially excreted with statistical significance in patients undergoing HSCT and in healthy volunteers (control). The frequency of appearance within each group is shown, as well as the mean amplitude of the signal intensity, which serves as a marker of relative concentration of the polypeptides in urine.

5 shows 4 differentially excreted polypeptides, defined by their particular mass and CE migration time. These changes reflect all changes in all patients without (w/o) or with (w/) GVHD. This allowed the generation of a typical polypeptide pattern that could

be used to distinguish between “no complications” and GVHD. Applying the support vector machine algorithms, classification based on the 16 polypeptides shown in Table 3 is possible with a sensitivity of 100% and a specificity of 82% after cross-validation.



**Figure 4. Polypeptide plots from patient no. 291 (no complications) and patient no. 293 (GVHD).** Polypeptide plots (polypeptide mass over CE migration time) are shown for 2 patients, no. 291 (A) and no. 294 (B), at an early posttransplantation period. While patient 291 had no complications in the observation period, patient 294 was diagnosed with GVHD on day +14. To visualize the search for biomarkers and to show the differences between these patients, panel C shows a magnified view of the plots obtained at different days after HSCT (0 to +13/+35 as shown by the black arrow) of the polypeptide 1.85 kDa/34 minutes. This polypeptide appears in the samples of patient 293, as early as day +1, showing increasing intensity over the progression of GVHD (arrow, starting at day +15, indicating the clinical diagnosis of GVHD).

**Table 3. Discriminating polypeptide patterns for GVHD and sepsis**

No.	Polypeptide identification			Frequency by group			Mean amplitude, MS counts		
	Mass, kDa	Time, min	Marker type	GVHD, %	No problems, %	Sepsis, %	GVHD	No problems	Sepsis
1	1.97	34.7	GVHD	82	18	92	801	89	253
2	1.75	36.0	GVHD	55	0	75	280	0	231
3	1.85	34.0	GVHD	64	9	33	314	28	73
4	1.07	37.8	GVHD	50	0	17	434	0	261
5	1.57	36.0	GVHD	59	9	83	246	25	650
6	1.87	46.3	GVHD	50	0	33	108	0	113
7	4.00	30.2	GVHD	50	0	8	842	0	103
8	1.83	33.2	GVHD	95	73	75	16 298	2047	5552
9	2.38	34.6	GVHD	18	73	17	53	301	89
10	3.44	44.1	GVHD	45	100	17	4581	12 182	169
11	1.35	48.1	GVHD	23	82	42	173	1572	76
12	3.21	33.6	GVHD	23	82	42	825	3605	1300
13	6.19	38.8	GVHD	23	82	8	275	2161	41
14	1.73	50.1	GVHD	32	91	25	344	799	138
15	3.09	45.9	GVHD	27	91	0	283	974	0
16	1.85	56.6	GVHD	9	82	25	5	650	19
17	1.54	50.7	Sepsis	18	9	83	33	7	267
18	2.17	42.1	Sepsis	18	9	83	114	97	604
19	1.24	32.4	Sepsis	23	27	92	141	227	549
20	2.05	46.2	Sepsis	18	18	83	38	46	169
21	2.15	41.6	Sepsis	36	36	100	877	290	1948
22	1.10	46.0	Sepsis	23	18	83	108	102	330
23	1.60	39.6	Sepsis	27	9	83	304	28	534
24	1.81	40.6	Sepsis	32	55	100	255	217	878
25	1.85	51.6	Sepsis	27	36	92	87	464	306
26	3.00	47.2	Sepsis	59	64	0	240	877	0
27	3.39	37.0	Sepsis	55	73	0	775	1348	0
28	3.84	25.8	Sepsis	59	73	0	3635	2219	0
29	4.04	29.2	Sepsis	59	73	0	1751	1405	0

Shown here are 16 polypeptides forming the diagnostic pattern (marker type: GVHD) with their frequency of appearance in groups of patients after HSCT (no "Other complications" as noted in Table 1), patients with GVHD (GVHD), and patients with sepsis. Thirteen additional polypeptides forming the "sepsis" pattern raise the discriminatory value of the GVHD panel, allowing differentiation of patients with sepsis from those with GVHD with 97% specificity and 100% sensitivity. The mean amplitude of signal intensity is shown for all groups.

Subsequent evaluation of the samples obtained prior to the clinical diagnosis of GVHD for the same pattern revealed that the "GVHD-typical" pattern could be observed between 5 and 15 days in advance of the clinical diagnosis. The value of these polypeptides became even more obvious when single molecules significant for GVHD were followed over time after HSCT and changes of the signal intensity were plotted against the days relative to diagnosis of GVHD in 4 different patients (Figure 6). The excretion of one polypeptide (1.83 kDa) is shown relative to the time of the clinical diagnosis of GVHD. This particular marker peptide appears with significant intensity between 5 and 15 days before the clinical diagnosis of GVHD in all patients.

The excretion of polypeptides in GVHD might be similar to the polypeptides observed during sepsis and in patients with bacteraemia. Since it is necessary to distinguish between these complications, we examined the urinary polypeptides in septic patients. To this end, samples from 5 ICU patients with sepsis were examined. As shown in Table 3, some of the indicative polypeptides for GVHD were, in fact, found in the septic patients. However, the majority of the GVHD-indicative polypeptides showed a distribution different from sepsis and, in addition, urine from septic patients contained several polypeptides that were not present in either "no complication" or GVHD patients. Hence, septic patients could be distinguished from patients with GVHD based solely on the polypeptide pattern. Using the panel of GVHD- and sepsis-specific polypeptides,

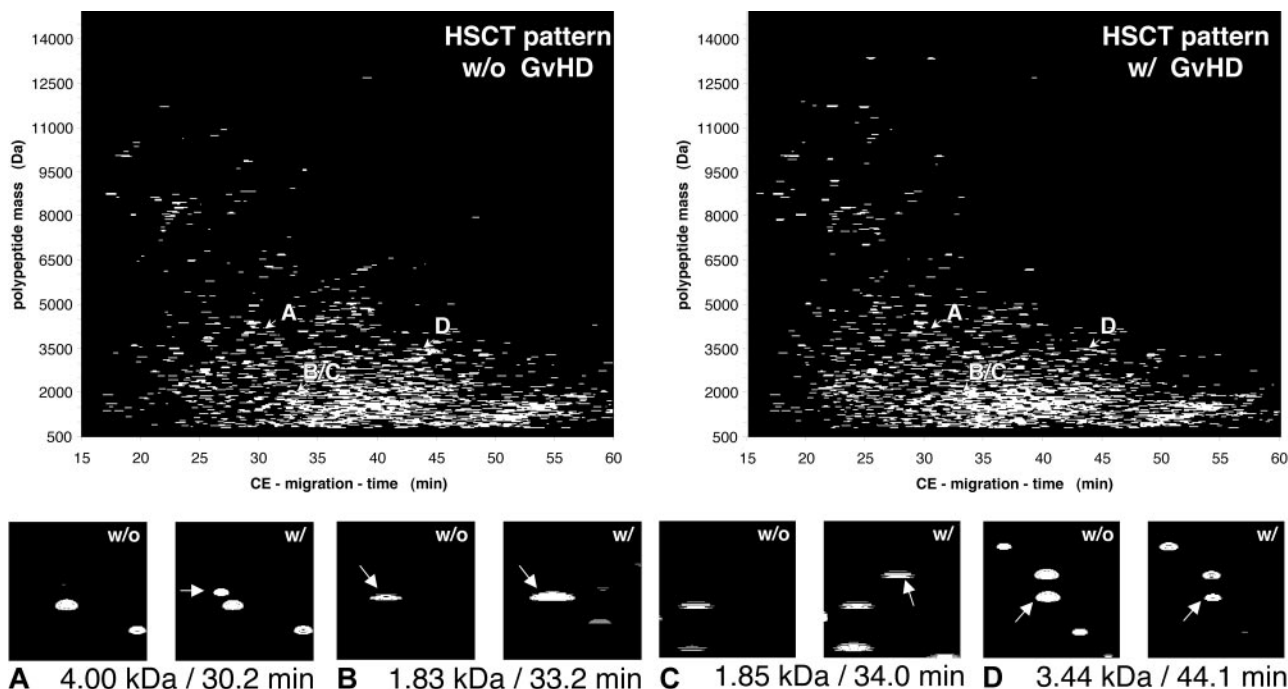
sepsis could be distinguished from GVHD with 100% sensitivity and 97% specificity after cross-validation.

To further characterize the GVHD-specific polypeptides, 2 of the differentiating polypeptides indicative for GVHD development were sequenced. The sample from patient 312 was separated by CE and spotted onto the MALDI target. Figure 7 shows the MALDI-MS-MS spectra obtained. The 1.85-kDa polypeptide (CE migration time: 34 minutes) was identified as a fragment of leukotriene A4 hydrolase, the rate-limiting enzyme in the biosynthesis of leukotriene 4B, known as a mediator of inflammation.<sup>22,23</sup> The 1.83-kDa polypeptide (CE migration time: 33.2 minutes) was a fragment of serum albumin excreted in the urine. This molecule was present with high abundance in patients with GVHD, while it was found with much lower signal intensity in healthy controls, patients with no problems after HSCT, and patients with sepsis or fever (Table 3).

## Discussion

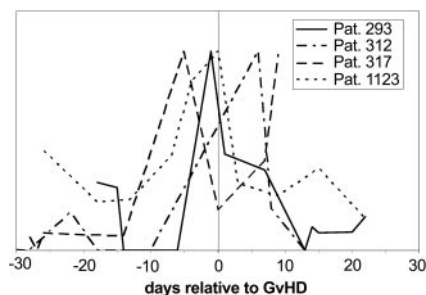
The results presented here as well as reports in the literature strongly suggest that proteomic analysis of body fluids will become increasingly important in the clinical diagnosis of various diseases.<sup>24,25</sup>

The aim of this study was to evaluate the application of CE-MS to hematopoietic stem cell transplantation, in order to



**Figure 5. Compiled plots of patients with and without GVHD.** The data obtained on the samples collected over the observation period were compiled for the patients without complications after HSCT (w/o GVHD) and compared with the polypeptide patterns obtained from patients at diagnosis of GVHD (w/GVHD). Comparison of the 2 patterns shows a number of differentially excreted polypeptides, as visualized here by inserts that magnify the particular parts of the pattern where the changes occur (marked by arrows). Shown are 4 discriminatory peptides, marked in each pattern with A to D and arrows indicating the position of each of these molecules in the compiled patterns, for patients without (w/o) or with (w) GVHD. Panels A and C show polypeptides newly appearing in all patients with GVHD. Panel B shows a more than 10-fold increase in signal intensity observed in all samples of patients with GVHD, while panel D shows a more than 10-fold reduction of the intensity of the 3.44 kDa/44.1 minutes peptide.

identify indicative differences in polypeptide patterns generated from patients developing GVHD in comparison to patterns generated from patients without problems or different complications. The patients represent a cross-section of those undergoing HSCT and were treated according to the different conditioning regimens: autologous transplantation and reduced-intensity or standard-intensity conditioning. CE-MS currently appears to be the best-suited technology to apply proteom analysis toward clinical samples. The wealth of data obtained in a single analysis is impossible to display in a figure, but can be judged by the example shown in Figures 1 and 2. The high resolution of this technology would allow the examination of more than 100 000 different polypeptides in a single analysis within about 50 minutes. This high resolution is necessary, allowing an initial



**Figure 6. Individual polypeptides relevant for GVHD in individual patients.** The signal intensity of the discriminating polypeptide (1.83 kDa/33.2 minutes CE migration time) is shown over the time course relative to GVHD development in 4 patients with GVHD relative to the time of GVHD diagnosis. The signal intensity of the molecule shown here rises at the time before diagnosis of GVHD, at least 5 days and up to 15 days before diagnosis.

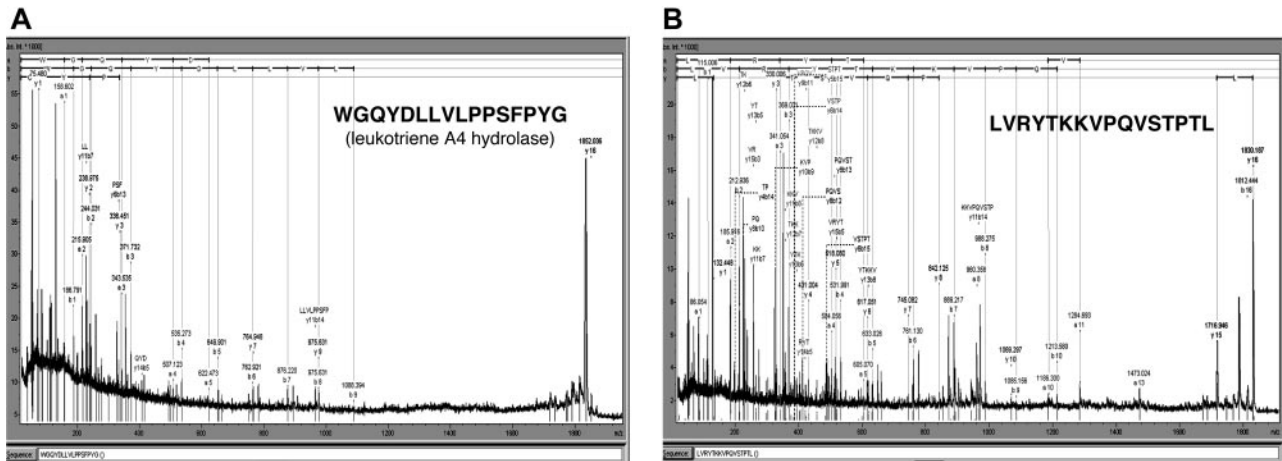
large number of different polypeptides to be examined, getting an unbiased view and leading eventually to a highly discriminatory but restricted “disease” or “complication-specific pattern.” The high resolution evidently requires software solutions for the automated extraction of the relevant data. The lack of such software has most likely hampered similar approaches in the past. New developments in the software applied for data processing now allow researchers to obtain information on polypeptides present in body fluids in a reproducible way.<sup>14</sup>

The reproducibility of this type of analysis is represented by Figure 3. In general, about 500 polypeptides are present with an absolute probability of more than 50% in urine samples, allowing compilation of the data and the generation of a typical pattern significant for health or certain disease. Randomly appearing polypeptides found in individual samples with low frequency most likely reflect individual alterations for example of diet, exercise, etc.<sup>11,15</sup> Such polypeptides are excluded from the compilation, but still saved.

Due to the speed and accuracy of CE-MS, a rapid diagnosis of complications based on the pattern recognition will be possible, without the need for specific reagents or different matrices. This is important when aiming toward an unbiased diagnosis, allowing all potential molecules to be included in the generation of differentiating patterns.

The polypeptide data from the CE-MS analysis (Table 2) clearly allowed differentiation from healthy controls solely based on the polypeptide pattern with a specificity of 95% and a sensitivity of 100%. When the 3 different conditioning groups were compared, no significant difference between these 3 groups could be identified. Patients undergoing HSCT can be distinguished from the





**Figure 7. Sequence of 2 discriminating polypeptides indicative for GVHD.** Two prominently excreted polypeptides for GVHD (Table 3) were sequenced yielding peptides shown above as MALDI-MS-MS spectra as well as sequence. Searching of the databases revealed one 1.85 kDa/34 minutes polypeptide from leukotriene A4 hydrolase (A) and one 1.83 kDa/33.2 minutes polypeptide from serum albumin (B).

controls, most likely due to the underlying disease or the general pretreatment. The different conditioning regimens might induce more subtle changes in the polypeptides excreted, but these will only be found as more patients are examined.

The application of CE-MS yielded the identification of polypeptides that enable us to clearly distinguish between patients with GVHD and those with no problems after HSCT with a high specificity (82%) and a sensitivity of 100% (Table 3). Applying very stringent parameters, 16 polypeptides could be identified that each revealed more than 50% absolute difference between patients without problems and those with GVHD. Hence, each one of these polypeptides on its own is able to distinguish more than 50% of all examined GVHD patients from the controls. This further underlines the power of this type of analysis.

Examination of urine obtained from patients with septic complications yielded a pattern that included several of the polypeptides also indicative for GVHD, but a number of significant differences from the GVHD pattern could be found. It is not surprising that sepsis and GVHD present with partially similar biomarkers excreted in the urine, since both diseases have inflammatory components. However, an additional 13 polypeptides indicative for sepsis could be identified in these patients (Table 3). This allows the distinction of patients with sepsis from those with GVHD with a specificity of 97% and a sensitivity of 100% after cross-validation. Preliminary data suggest that other complications such as viral infections or organ failure result in specific patterns that can be distinguished from the GVHD pattern. Patterns from samples of 3 patients (483, 481, and 313) with cytomegalovirus reactivation alone and those of 2 patients (292 and 910) at the time of acute renal failure did not bear resemblance to the GVHD pattern.

The aim of this study was to detect polypeptide patterns common to all patients with GVHD, thus creating a pattern that would allow recognition of developing GVHD. The number and concentration of possible markers rises with the development and progression of GVHD. The patterns described here are "early GVHD patterns." A number of additional polypeptides are excreted over the course of GVHD development and as the severity increases, more polypeptides are excreted at higher concentrations (Figure 4C, 293). Treatment of GVHD is usually

initiated with steroids and is switched to second-line treatment if there is no response. Second-line treatment of GVHD includes the use of anti-T-cell antibodies such as antithymocyte globulin or OKT3, as well as antibodies directed against the IL-2 receptor or cytokines, such as tumor necrosis factor (TNF). In general, the outcome of patients requiring second-line treatment is poor and the early identification of patients with poor response to first treatment might improve the outcome.<sup>26</sup> The decision to start with second-line treatment can be made based on the patient's polypeptide pattern. In order to obtain more insight into the nature of the polypeptides forming the GVHD pattern, sequencing using the MALDI-MS-MS was applied. To date, 2 polypeptides were identified via their sequence. The first one was a 1.85-kDa peptide from the leukotriene A4 hydrolase, the rate-limiting enzyme in the synthesis of leukotriene 4B.<sup>22</sup> This enzyme was described previously as up-regulated during GVHD and most likely represents a marker for the inflammatory processes during GVHD development.<sup>23</sup> This finding again underlines the ability of CE-MS to display polypeptides and subsequently enable the identification of peptides and proteins involved in the pathophysiology of different diseases.

The second polypeptide was a 1.83-kDa peptide from albumin, which was excreted at a high concentration in the urine of patients with GVHD (Table 3), but not in patients without problems or even in patients with sepsis. It is tempting to speculate that this albumin fragment is excreted at higher levels in patients with GVHD since increased proteolytic activity, as described for GVHD,<sup>27</sup> might be found in the kidney as well as in other organs very early in initial inflammatory processes, well before clinical diagnosis.

While not all polypeptides detected in individual patients could already be considered in the generation of the patterns significant for GVHD or sepsis, all these data are saved and available in the database. Thus, increasing the number of patients with different complications will allow additional fine tuning of the results in later studies. The power of the technique presented here is based on the simultaneous analysis of several hundred polypeptides present in individual samples, allowing the generation of an array of biomarkers to be used for differential diagnosis. In conclusion, proteomic analysis using CE-MS may provide a powerful diagnostic tool for

early identification of patients developing complications after HSCT, leading to an early and accurate diagnosis and timely and appropriate therapeutic intervention. Our data document that proteomic analysis of urine with CE-MS coupling is a fast, reproducible, and sensitive approach for the identification of complications evolving after allo-HSCT and may be helpful in reducing transplant-related mortality. The results have to be further validated in a prospective study, which is currently underway.

## Acknowledgments

The authors thank Drs A. John Barrett (NHLBI, NIH) and J. Frederick Mushinski (NCI, NIH) for critical revision of the manuscript, Elke Dammann for excellent documentation of the patient data, and Meike Hillmann and Frank Hausadel for excellent technical support.

## References

- Horowitz MM, Gale RP, Sondel BM, et al. Graft-versus-leukemia reactions after bone marrow transplantation. *Blood*. 1990;75:555-562.
- Gratwohl A, Brand R, Apperley J, et al. Graft-versus-host disease and outcome in HLA-identical sibling transplantations for chronic myeloid leukemia. *Blood*. 2002;100:3877-3886.
- Kolb HJ, Schmid C, Barrett AJ, Schendel DJ. Graft-versus-leukemia reactions in allogeneic chimeras. *Blood*. 2004;103:767-776.
- Reddy P, Ferrara JL. Immunobiology of acute graft-versus-host disease. *Blood Rev*. 2003;17:187-194.
- Jarvis M, Marzolini M, Wang XN, Jackson G, Sviland L, Dickinson AM. Heat shock protein 70: correlation of expression with degree of graft-versus-host response and clinical graft-versus-host disease. *Transplantation*. 2003;76:849-853.
- Blijlevens NM, Donnelly JP, Meis JF, De Keizer MH, De Pauw BE. Procalcitonin does not discriminate infection from inflammation after allogeneic bone marrow transplantation. *Clin Diagn Lab Immunol*. 2000;7:889-892.
- Seidel C, Ringden O, Remberger M. Increased levels of syndecan-1 in serum during acute graft-versus-host disease. *Transplantation*. 2003;76:423-426.
- Petricoin EF, Ardekani AM, Hitt BA. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*. 2002;359:572-577.
- Klade CS, Voss T, Krystek E, et al. Identification of tumor antigens in renal cell carcinoma by serological proteome analysis. *Proteomics*. 2001;1:890-898.
- Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res*. 2002;62:3609-3614.
- Wittke S, Fliser D, Haubitz M, et al. Proteomics for clinical diagnostic and establishment of new markers and therapeutic targets: determination of proteins and peptides in urine with CE-ESI-TOF-MS. *J Chromatogr A*. 2003;1013:173-181.
- Kaiser T, Hermann A, Kielstein JT, et al. Capillary electrophoresis coupled mass spectrometry to establish polypeptide patterns in dialysis fluids. *J Chromatogr A*. 2003;1013:157-171.
- Wittke S, Kaiser T, Mischak H. Differential polypeptide display: the search for the elusive target. *J Chromatogr B*. 2004;803:17-26.
- Kaiser T, Wittke S, Krebs R, Mischak H, Weissinger EM. Capillary electrophoresis coupled to mass spectrometer for automated and robust polypeptide recognition in body fluids for clinical use. *Electrophoresis*. In press.
- Weissinger EM, Wittke S, Kaiser T, et al. Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes. *Kidney International*. In press.
- Bornhauser M, Kiehl M, Siegert W, et al. Dose-reduced conditioning for allografting in 44 patients with chronic myeloid leukaemia: a retrospective analysis. *Br J Haematol*. 2001;115:119-124.
- Schmid C, Weisser M, Ledderose G, Stotzer O, Schleuning M, Kolb HJ. Dose-reduced conditioning before allogeneic stem cell transplantation: principals, clinical protocols and preliminary results. *Dtsch Med Wochenschr*. 2002;127:2186-2192.
- Neuhoff N, Kaiser T, Wittke S, et al. Mass spectrometry for the detection of differentially expressed proteins: a comparison of SELDI and CE/MS technology. *Rapid Commun Mass Spectrom*. 2004;18:149-156.
- Dieterle F, Muller-Hagedorn S, Liebich HM, Gauglitz G. Urinary nucleosides as potential tumor markers evaluated by learning vector quantization. *Artif Intell Med*. 2003;28:265-279.
- Brown MP, Grundy WN, Lin D, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A*. 2000;97:262-267.
- Zhu K, Kim J, Yoo C, Miller FR, Lubman DM. High sequence coverage of proteins isolated from liquid separations of breast cancer cells using capillary electrophoresis-time-of-flight MS and MALDI-TOF MS mapping. *Anal Chem*. 2003;75:6209-6217.
- Chen X, Li N, Wang S, Wu N, et al. Leukotriene A4 hydrolase in rat and human esophageal adenocarcinomas and inhibitory effects of bestatin. *J Natl Cancer Inst*. 2003;95:1053-1061.
- Takatsuka H, Yamada S, Okamoto T, et al. Predicting the severity of intestinal graft-versus-host disease from leukotriene B4 levels after bone marrow transplantation. *Bone Marrow Transplant*. 2000;26:1313-1316.
- Voss T, Ahorn H, Haberl P, Dohner H, Wilgenbus K. Correlation of clinical data with proteomics profiles in 24 patients with B-cell chronic lymphocytic leukemia. *Int J Cancer*. 2001;91:180-186.
- Osin P, Shipley J, Lu YJ, et al. Experimental pathology and breast cancer genetics: new technologies. *Recent Results Cancer Res*. 1998;152:35-48.
- Graziani F, Van Lint MT, Dominietto A, et al. Treatment of acute graft versus host disease with low dose-alternate day anti-thymocyte globulin. *Haematologica*. 2002;87:973-978.
- Salmela MT, Karjalainen-Lindsberg ML, Jeskanen L, Saarialho-Kere U. Overexpression of tissue inhibitor of metalloproteinases-3 in intestinal and cutaneous lesions of graft-versus-host disease. *Mod Pathol*. 2003;16:108-114.