MALDI-TOF Mass Spectrometry Analysis of Cerebrospinal Fluid Tryptic Peptide Profiles to Diagnose Leptomeningeal Metastases in Patients with Breast Cancer*

Lennard J. Dekker‡, Willem Boogerd§, Guenther Stockhammer¶, Johannes C. Dalebout‡, Ivar Siccama‖, Pingpin Zheng‡, Johannes M. Bonfrer**, Jan J. Verschuuren‡‡, Guido Jenster§§, Marcel M. Verbeek¶¶, Theo M. Luider‡, and Peter A. Sillevis Smitt‖

Leptomeningeal metastasis (LM) is a devastating complication that occurs in 5% of patients with breast cancer. Early diagnosis and initiation of treatment are essential to prevent neurological deterioration. However, early diagnosis of LM remains challenging because 25% of cerebrospinal fluid (CSF) samples produce false-negative results at first cytological examination. We developed a new, MS-based method to investigate the protein expression patterns present in the CSF from patients with breast cancer with and without LM. CSF samples from 106 patients with active breast cancer (54 with LM and 52 without LM) and 45 control subjects were digested with trypsin. The resulting peptides were measured by MALDI-TOF MS. Then, the mass spectra were analyzed and compared between patient groups using newly developed bioinformatics tools. A total of 895 possible peak positions was detected, and 164 of these peaks discriminated between the patient groups (Kruskal-Wallis, \( p < 0.01 \)). The discriminatory masses were clustered, and a classifier was built to distinguish patients with breast cancer with and without LM. After bootstrap validation, the classifier had a maximum accuracy of 77% with a sensitivity of 79% and a specificity of 76%. Direct MALDI-TOF analysis of tryptic digests of CSF gives reproducible peptide profiles that can assist in diagnosing LM in patients with breast cancer. The same method can be used to develop diagnostic assays for other neurological disorders. *Molecular & Cellular Proteomics* 4:1341–1349, 2005.

Leptomeningeal metastases (LM)\(^1\) arise when tumor cells metastasize to the cerebrospinal fluid (CSF). The flow of CSF results in widespread dissemination of the tumor cells along the surface of the central nervous system, causing symptoms by invading the brain, spinal cord, cranial nerves, and nerve roots (1).

One of the tumors most frequently associated with LM is breast cancer. During the course of the disease, \( \sim 5\% \) of patients with metastatic breast cancer will develop symptoms caused by LM. This debilitating complication’s response to therapy depends upon early treatment. However, diagnosis of LM remains challenging because 25% of samples tested are false negative at the first cytological examination of the CSF, probably because of sampling error (1).

Protein expression profiling of body fluids from patients with cancer has recently become a valuable tool for obtaining information on the state of protein circuits inside tumor cells and outside the cells at the host-tumor interface (2, 3). In serum and CSF, low molecular weight proteins and peptides that are related to this altered microenvironmental “cancerous” state can be detected.

We studied the differential tryptic peptide profiles in the CSF from patients with breast cancer with and without LM and in CSF from control subjects. Studying CSF has several advantages over studying serum. First, tumor cells in LM patients are located in the CSF and in the leptomeninges that are surrounded by CSF. Before their transport into serum, tumor-related proteins will therefore first be shed into the CSF. Second, the normal protein concentration of CSF is 100- to 400-fold lower than in serum (4). This results in a significant over-representation of LM-related proteins in CSF compared

\( \text{© 2005 by The American Society for Biochemistry and Molecular Biology, Inc.} \)

This paper is available on line at http://www.mcponline.org
with serum. The identification of protein profiles specific for
LM may be helpful in diagnosing patients with clinical suspi-
cion of LM but negative cytology. In addition, such proteins may reveal cellular mechanisms relevant to the biology of LM.

With the advent of mass spectrometry into the field of clinical proteomics, the comparison of large numbers of pro-
teins in complex biological samples such as serum and CSF has become feasible (3, 5). Until now, the most commonly used instrument was the SELDI-TOF MS. Using SELDI-TOF MS analysis of various body fluids, discriminatory protein expression profiles have been identified in various diseases (3, 6). However, SELDI-TOF MS does not allow a direct iden-
tification of the discriminatory proteins and suffers from low reproducibility and accuracy (7–10). To improve the reproduc-
ibility and accuracy and find better ways to identify relevant discriminatory proteins (11), we first digested our samples with trypsin and analyzed the resulting peptide mixtures by MALDI-TOF MS (12). The reproducibility of this type of analy-

es has been described elsewhere (13).

We analyzed CSF samples from 106 patients with active breast cancer, 54 of whom had LM, and CSF from 57 control subjects. Tryptic peptide mixtures were measured by MALDI-
TOF MS and analyzed using a newly designed bioinformatics tool. We could identify unique peptide patterns that discrimi-
nated the LM patients from the other patients with breast cancer and from control subjects.

EXPERIMENTAL PROCEDURES

Patient Selection—Using clinical databases and CSF banks, we
retrospectively identified all patients with breast cancer with available CSF samples collected in the last 7 years in four participating institu-
tions (Erasmus MC, Netherlands Cancer Institute, UMC Nijmegen, and Innsbruck Medical University). We included only patients with advanced breast cancer defined as metastatic or locally progressive disease. Patients with positive cytology or with a compatible neuro-
logical syndrome and diagnostic MRI were considered to have LM (group I, n = 54). When cytology was negative and when clinical follow-up was incompatible with LM, patients were classified as having advanced breast cancer without LM (group II, n = 52). Control subjects (group III, n = 45) were patients who were not known to have cancer and who did not suffer any known neurological disease (col-
clected at Erasmus MC, Netherlands Cancer Institute, and LUMC). We not-
et the date of diagnosis of breast cancer, the date of lumbar puncture, and the date of last follow up or death. The symptoms at the
time of lumbar puncture were scored as either central, cranial
nerve involvement, radicular, compatible with raised intracranial pres-
sure, or other. The following CSF parameters were noted: total protein concentration, white cell count, and cytology (positive or negative).
MRI and CT scans were scored as either positive, suggestive, or negative for LM. Nodular or focal linear meningeal enhancement was considered diagnostic of LM, and communicating hydrocephalus was considered suggestive of LM. All samples had been routinely centri-
fuged to discard cellular elements before storage at −80 °C.

Sample Preparation and Measurement of Samples—All samples
were blinded and analyzed in random order. From each sample, 20 μl
of CSF was put into a 96-well plate, and 20 μl of 0.2% Rapigest (Waters, Milford, MA) in 50 mM ammoniumbicarbonate buffer was added to each well. The samples were incubated for 2 min at 37 °C. 4 μl of 0.1 μg/μl gold grade trypsin (Promega, Madison, WI) in 3 mM Tris-HCl was added to each well, and the 96-well plates were incubated at 37 °C. After 2 h of incubation, 2 μl of 500 mM HCl was added to obtain a final concentration of 30–50 mM HCl, pH < 2. The 96-well plates were then incubated again for 45 min. A 96-well zip C18 microtiter plate (Millipore Corporation, Bedford, MA) was prewetted and washed twice with 200 μl of acetonitrile per well. Full vacuum was applied to the plate using a vacuum manifold (Millipore). 3 μl of acetonitrile was put on the C18 resin without vacuum to prevent it from drying. Each sample was mixed with 200 μl of water HPLC grade/TF A 0.1%. Subsequently the samples were loaded on the washed and prewetted 96-well zip C18 plate (Millipore); a pressure differential of 5 inches of Hg vacuum was used. After the wells had been cleared, the wells were washed twice with 100 μl of 0.1% TFA. Full vacuum was applied until all wells were empty. The samples were eluted in a new 96-well plate with an elution volume of 15 μl of 50% acetonitrile/water HPLC grade 0.1% TFA; a pressure differential of 5
inches of Hg vacuum was used. After elution, the samples were stored at 4 °C in the 96-well plates covered with aluminum seals. All samples were spotted on a MALDI target (300/384 anchor chip with transponder plate; Bruker Daltonik GmbH, Bremen, Germany) in trip-
licate. To do so, 2 μl of elute was mixed with 10 μl of matrix solution (2 mg of α-cyano-4 hydroxycinnamic acid; Bruker Daltonik GmbH) in 1 ml of acetonitrile for 30 min using an ultrasonic bath). Afterward, samples were automatically measured on a MALDI-TOF MS (Biflex III; Bruker Daltonik GmbH). The digestion step was repeated twice for each sample, the purified peptides were spotted in triplicate, and all the spots were measured in triplicate. This resulted in 18 spectra for each sample. The standard method for peptide measurements on the MALDI-TOF MS was used (default file Bruker “1–2kD positive” with the measurement range changed to 300–3000 Da). For the auto-
mated measurements, the settings of initial laser power of 20% and a maximum of 35% were used. The highest peak above the 750 Da had to have a signal-to-noise ratio of at least 5 and a minimum resolution of 5000. After every 30 laser shots, the sum spectrum was checked for these criteria. If the sum spectrum did not meet these criteria, it was rejected. If 13 sum spectra from 30 shots met the criteria, these were combined and saved; when 50 sum spectra from 30 shots were rejected, the measurement of that spot was then ended, and the next spot was measured.

Analysis of Spectra—First, the raw binary data files were converted
into ASCII files containing the measured intensities for all channel
indices of the spectra. We then developed a peak detection algorithm in the statistical language R (www.r-project.org). The definition of a
peak (or local maximum) in this algorithm states that the intensity of
the peak position has to be above a predefined threshold and has to be the highest intensity value in a surrounding mass window. This peak-finding algorithm was tested on a small set of spectra with different settings for the threshold and the mass window. The settings
for the peak finding were chosen such that the resulting peak list most resembled peaks that would be manually assigned, thus optimizing the trade-off between signal sensitivity and noise detection. We chose a percentile threshold of 98.5% (the intensity of the position must belong to the 1.5% highest intensity values of the spectrum) and a mass window of 0.5 Da. A quadratic fit with a number of internal calibrants was used to calibrate the channel numbers to masses. For this mass calibration, five omnipresent albumin peaks (960.5631, 1000.6043, 1149.6156, 1511.8433, and 2045.0959 m/z) were used. The accurate mass of these albumin peaks was obtained by perform-
ing a “tryptic digest” on the human albumin amino acid sequence with MS-digest (prospector.ucsf.edu/ucsfhtml4.0/msdigest.htm).

During the process of alignment and conversion, the quality of the
spectra was checked as follows. If two or more of the omnipresent albumin peaks were not detected, the spectrum was not used in the
further analysis. The peak finding algorithm was then used to create
a list of peak positions for each individual spectrum. These peak lists were combined by comparing the lists one by one. If peak positions were present in a mass window of 0.5 dalton in both spectra, these peak positions were combined. The combined peak list was then compared with a new spectrum until all peak lists had been combined. The latter peak list was used to create a matrix displaying the frequency of each peak position for each sample. Peak positions that were present in less than 5% of the spectra were deleted from the matrix to reduce the number of noise peaks. The matrix created in this way was used for statistical analysis of the data. Using a univariate analysis in R, a \( p \) value was determined for every peak position. When comparing more than two groups, we used the Kruskal-Wallis test; when comparing two groups, we used the Wilcoxon-Mann-Whitney test.

To investigate whether differences in the total CSF protein concentration of the samples affected the performance of the MALDI-TOF, we first used the Bio-Rad detergent-compatible protein assay (Bio-Rad) to determine the protein concentration of all the CSF samples. We then calculated the sum of albumin peaks detected in all seven spectra of each sample (excluding the albumin peaks that had been used for calibration). Using Prism version 4.0 (GraphPad Software, San Diego, CA), we compared the total protein concentration and the sum of the albumin peaks of the three groups. To test for statistically significant differences between the three groups, we used one-way ANOVA followed by Bonferroni’s multiple comparison test. All tests were two-sided, and \( p < 0.05 \) was considered statistically significant.

In addition, the correlation between peak frequency and protein concentration was calculated for each individual peak position. A histogram of all correlation coefficients was created and the distribution was compared, with a normal distribution using the Kolmogorov-Smirnov (SPSS Inc., Chicago, IL).

All peak positions with a frequency that was two times higher in group I than in the control groups II and III were selected. These peak values were submitted to the Mascot search engine (Matrix Science, London, UK) to search the MSDB human database using a 100-ppm tolerance.

Building a Predictive Model—A supervised multivariate analysis method was used to determine whether sample groups I and II could be separated on the basis of their peak positions. For each patient, seven mass spectra were used, and the peak positions of each of those seven spectra were combined. Therefore, the number of times that a peak was present varied between 0 and 7. To reduce noise, a minimum of two peaks was required to determine whether a peak was present (\( \geq 2, 1 \)) in a sample or not (\( < 2, 0 \)) allowing the formation of a binary data matrix. The required frequency was kept low to minimize loss of signal. To reduce the number of variables, a clustering was performed that combined peaks of similar behavior. Peptide peaks that often occurred simultaneously were grouped into the same cluster using a hierarchical clustering algorithm. The distance between each possible pair of peptide peaks was determined with the Manhattan distance measure (i.e., the sum of the absolute differences for all patients). The number of clusters was set at 50. With 50 clusters, isotope peaks were generally grouped into the same cluster. The clusters generated represented groups of peptide peaks that might be derived, at least in part, from the same protein or proteins. The clustering of the masses made it possible to compose a new data matrix. Each matrix cell contained the number of peaks present for a certain patient relative to the total number of peaks in a particular cluster. In other words, each cell in the new matrix defined the proportion of peptides that was present in a cluster for a certain patient. To further reduce the complexity of the data, we set a threshold for the presence of a cluster to obtain a binary data matrix. Using the clustered, binary variables, we constructed a non-linear predictive model that separated group I from group II. In the model thus generated, a maximum of eight clusters was allowed, and only those clusters with an area under curve (AUC) greater than 0.62 were considered. Genetic programming was used to search for the model with the highest AUC (14). To obtain an unbiased estimate of the predictive accuracy of the model, we used the bootstrapping method (15, 16). Bootstrap data sets were created by randomly selecting patients with replacements from the original data set. As an extra precaution, the clustering step was included in the bootstrapping process as well. 100 bootstrapped matrices were created from the original matrix by resampling with replacement. The clustering was repeated for each of these resampled matrices and a predictive model was constructed. The AUC of each model was measured on the bootstrap data set as well as on the original data set. The average difference between the performance on the bootstrap data set and the performance on the original data set provided a correction factor that gave an estimate of the bias of our model development process. Finally, we developed a model on the original data and corrected its AUC with the correction factor, producing a conservative estimate of the performance of the model.

RESULTS

Patients—Clinical information, CSF data, and imaging results are summarized in Table I. Forty-six percent of patients with breast cancer with LM (group I) presented with more than one neurological symptom, whereas the majority of patients with breast cancer (73%) in group II presented with a single symptom at the time of lumbar puncture. All patients in group I had LM, whereas diagnoses in group II included bone metastases (\( n = 15 \)), tension headache (\( n = 5 \)), metabolic encephalopathy (\( n = 3 \)), carpal tunnel syndrome (\( n = 3 \)), brain metastases (\( n = 2 \)), migraine (\( n = 2 \)), disseminated intravascular coagulation (\( n = 2 \)), dural metastasis (\( n = 2 \)), lumbago (\( n = 2 \)), and psychiatric problems (\( n = 2 \)). All of the following conditions were diagnosed, each in one patient: jugular vein thrombosis, herniated disk, dementia, anaplastic astrocytoma, whiplash injury, polyneuropathy, and syncope. In two patients, no cause for the symptoms was found.

The cytological examination of the CSF revealed tumor cells in all group I patients and in none of the group II patients. The CSF white cell count was increased (>4 cells/\( \mu \)l) in 56% of group I and 0% of group II patients. The total protein concentration was higher than the institutional upper limit in 85% of group I and in 30% of group II patients. Imaging studies were obtained in 39 of 41 group I patients, 35 MRIs and 4 CT scans. Imaging results were positive for LM in 23 (two CT scans) patients, suggestive in 4 patients, and negative in 10 patients (two CT scans). MRI results were available in 26 patients in group II. The MRI was considered negative in 22 patients, suggestive in 3, and positive for LM in 1 patient. The patient with a false-positive MRI had a single non-symptomatic dural enhancing lesion at T9; the CSF cytology was three times negative.

Peak Detection—We detected an average of 350 peaks per spectrum (95% CI, 250 to 450). Spectra with more than 450 detectable peaks were excluded from the analysis because these spectra consisted mainly of noise peaks (peaks without an isotopic distribution). After alignment and quality control, the number of good quality spectra per sample was counted, and samples with fewer than seven good quality spectra were
CSF Proteomics in Advanced Breast Cancer

The number of seven spectra is based on an earlier reproducibility study (13). At this threshold, the remaining number of cases per patient group was 41 in group I, 46 in group II, and 43 in group III (Table II). Of the samples with more than seven spectra, only the first seven spectra were used. The combined peak list of all these spectra contained 2006 possible peak positions. After noise reduction, the matrix created from this list contained 895 peak positions.

**Univariate Analysis**—For each peak, the significance of difference in distribution over groups I-III was tested with the Kruskal-Wallis test, and a \( p \) value was calculated. We detected 323 peak positions with \( p \leq 0.05 \) and 172 with \( p \leq 0.01 \), indicating that a considerable number of peaks correlated with diagnosis groups. In Fig. 1A, the number of peaks (frequency) is presented for each \( p \) value interval. On the same data, we performed a cross-validation by randomly assigning a group number to each CSF sample and then repeating the Kruskal-Wallis test. This scrambling procedure was repeated 10,000 times. The blue line in Fig. 1A presents the mean frequency of peaks per \( p \) value interval. The flat distribution of the \( p \) value histogram indicates that there is no correlation between peak positions and groups after scrambling. The \( p \) value histogram of the actual experiment is clearly skewed to lower \( p \) values and is significantly different from the histogram after scrambling (Fig. 1A).

We then compared the groups pair-wise with a Wilcoxon-Mann-Whitney test. The highest number of peak positions with a significant \( p \) value was found when groups I and III were compared (\( p \leq 0.05 \), 329 peak positions; \( p \leq 0.01 \), 190 peak positions) (Fig. 1B). When groups I and II were compared, the \( p \) value histogram was significantly different from a random distribution (\( p \leq 0.05 \): 326 peak positions; \( p \leq 0.01 \): 164 peak positions) (Fig. 1C). When groups II and III were compared, fewer peak positions with a significant \( p \) value were detected (\( p \leq 0.05 \): 119 peak positions; \( p \leq 0.01 \): 46 peak positions) (Fig. 1D). As a cross-check, the correlation between the different institutes and the peak occurrences were compared. The histogram of the \( p \) values did not differ significantly from a random distribution.

The CSF total protein concentration of the samples differed significantly between groups I and III (ANOVA, \( p < 0.001 \); Fig. 2A). However, the sum of albumin peaks detected per sample

### Table I

Clinical information on advanced breast cancer patients with (group I) and without leptomeningeal metastasis (group II) and control subjects (group III)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>41</td>
<td>46</td>
<td>43</td>
</tr>
<tr>
<td>Age at lumbar puncture (mean ± S.D.)</td>
<td>52 ± 10</td>
<td>52 ± 13</td>
<td>38 ± 21</td>
</tr>
<tr>
<td>Time from cancer diagnosis (months) (mean ± S.D.)</td>
<td>62 ± 65</td>
<td>69 ± 59</td>
<td></td>
</tr>
<tr>
<td>Time from LP to last follow-up or death (months) (mean ± S.D.)</td>
<td>6 ± 10</td>
<td>19 ± 19</td>
<td></td>
</tr>
<tr>
<td>Alive at last follow-up</td>
<td>1 (3%)</td>
<td>7 (16%)</td>
<td></td>
</tr>
<tr>
<td>Number of symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One symptom</td>
<td>22 (54%)</td>
<td>33 (73%)</td>
<td></td>
</tr>
<tr>
<td>Two symptoms</td>
<td>12 (29%)</td>
<td>12 (27%)</td>
<td></td>
</tr>
<tr>
<td>Three symptoms</td>
<td>7 (17%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>28 (68%)</td>
<td>14 (31%)</td>
<td></td>
</tr>
<tr>
<td>Cranial Nerve</td>
<td>11 (27%)</td>
<td>10 (22%)</td>
<td></td>
</tr>
<tr>
<td>Raised intracranial pressure</td>
<td>16 (39%)</td>
<td>5 (11%)</td>
<td></td>
</tr>
<tr>
<td>Radicular</td>
<td>12 (29%)</td>
<td>8 (18%)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>20 (44%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White cell count/μl (mean ± S.D.)</td>
<td>21.2 ± 34.4</td>
<td>0.8 ± 0.8</td>
<td>NA</td>
</tr>
<tr>
<td>Increased white cell count (&gt;4/μl)</td>
<td>23 (56%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Positive cytology</td>
<td>41 (100%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Protein concentration &gt; norm institute</td>
<td>35 (85%)</td>
<td>14 (30%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Glucose (mM , mean ± S.D.)</td>
<td>2.0 ± 1.3</td>
<td>3.8 ± 0.9</td>
<td>NA</td>
</tr>
<tr>
<td>Imaging (MRI/CT)</td>
<td>35/4a</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>23 (56%)</td>
<td>1 (2%)</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>12 (29%)</td>
<td>22 (48%)</td>
<td></td>
</tr>
<tr>
<td>Suggestive</td>
<td>4 (10%)</td>
<td>3 (7%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>2 (5%)</td>
<td>20 (43%)</td>
<td></td>
</tr>
</tbody>
</table>

* Two of four CT scans were positive for LM.

### Table II

Number of patients per group before and after quality control of spectra

<table>
<thead>
<tr>
<th></th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before quality control</td>
<td>54</td>
<td>54</td>
<td>45</td>
</tr>
<tr>
<td>After quality control</td>
<td>41</td>
<td>46</td>
<td>43</td>
</tr>
</tbody>
</table>

We then compared the groups pair-wise with a Wilcoxon-Mann-Whitney test. The highest number of peak positions with a significant \( p \) value was found when groups I and III were compared (\( p < 0.05 \), 329 peak positions; \( p < 0.01 \), 190 peak positions) (Fig. 1B). When groups I and II were compared, the \( p \) value histogram was significantly different from a random distribution (\( p < 0.05 \): 326 peak positions; \( p < 0.01 \): 164 peak positions) (Fig. 1C). When groups II and III were compared, fewer peak positions with a significant \( p \) value were detected (\( p < 0.05 \): 119 peak positions; \( p < 0.01 \): 46 peak positions) (Fig. 1D). As a cross-check, the correlation between the different institutes and the peak occurrences were compared. The histogram of the \( p \) values did not differ significantly from a random distribution.

The CSF total protein concentration of the samples differed significantly between groups I and III (ANOVA, \( p < 0.001 \); Fig. 2A). However, the sum of albumin peaks detected per sample
did not differ between the groups (ANOVA, \( p = 0.8 \); Fig. 2B). For all individual peak positions, the correlation between the peak frequency and the protein concentration was calculated and plotted in a histogram (Fig. 3). The distribution of the correlation coefficients did not significantly differ from a normal distribution (Kolmogorov-Smirnov test, \( p = 0.35 \)). The constant number of albumin peaks and the lack of effect of protein concentration on the peak frequency indicated that differences in total protein concentration had not significantly affected MALDI-TOF performance.

From the matrix file, we made a selection of all peak positions that were up-regulated two times in group I compared with both control groups II and III. This list contained 52 peak positions that were used in a mascot database search against the MSDB human database. Five of the 52 peptides matched to apolipoprotein A1 (28 kDa). These five matching peptides had an average mass error of 19 ppm compared with the MSDB database.

**Clustering Analysis**—A clustering on the masses was performed on a matrix in which the samples had been sorted on
group number. This clustering resulted in the detection of group-specific clusters (Fig. 4). In Fig. 4A, a zoom in of the dendrogram is displayed that shows peak positions that have higher frequencies in patients with breast cancer without LM (group II) and healthy control subjects (group III) than in patients with breast cancer with LM (group I). In B, peak positions with a higher frequency in patients with breast cancer with LM (group I) than in groups II and III are shown.

**Predictive Model**—Clustering of the masses resulted in a reduction of the variables from 895 peaks to 50 clusters. All albumin peaks were assigned to the same cluster (a large cluster containing 164 peaks). Table II lists 10 clusters ordered by their univariate area under the receiver operating characteristic curve (AUC), indicating the highest discriminatory value. As predicted, isotopic peaks are grouped together in the same cluster (Table III). The AUC can vary between 0.5 and 1 (0.5 for a random prediction and a value of 1 for an optimal prediction). The cluster with the highest AUC consisted of 10 peaks, all of which had $p$ values $<0.01$. When adjoining isotope peaks were combined, four distinctive peptide peaks in this cluster remained. The model with the highest predictive value, as selected with genetic programming, used six clusters (Table II). The AUC achieved by this model on the original data set was 0.836; the bootstrap-corrected AUC was 0.852. The maximum accuracy that could be achieved after bootstrapping was 77%. At this cut-off point, the corrected sensitivity was 79% and the corrected specificity was 76%.

**DISCUSSION**

We studied the value of proteomic profiling in the diagnosis of LM in patients with breast cancer. After digestion of CSF proteins with trypsin, we analyzed the resulting peptides with MALDI-TOF. In a univariate analysis, we detected that many peptides were differentially expressed among patients with breast cancer with and without LM and healthy control subjects. We then built a predictive model to diagnose LM in patients with breast cancer. After validation by bootstrapping, the model achieved a sensitivity of 79% and specificity of 76%. In current clinical practice, diagnostic tests for LM include CSF cytology and gadolinium-enhanced MRI (1, 17, 18). The sensitivity of CSF cytology (75%) and Gd MRI (76%) are comparable with our predictive model (17). The specificity of MRI (77%) is also similar, but the specificity of CSF cytology is much higher (100%) (17). We conclude that our test may be useful to support the diagnosis of LM in patients with breast cancer. It is noteworthy that the test requires only 20 μl of CSF and can therefore easily be combined with cytological examination of the CSF.

After the original highly intriguing report that the serum proteome profile can be used for the early detection of ovarian cancer (3), many researchers have applied the SELDI-TOF
technology to detect proteome profiles specific for other forms of cancer and non-malignant disease (6, 19). However, criticism has focused on the low reproducibility of the SELDI-TOF analytical tool (7, 9, 10, 20–25). Models based on SELDI-TOF protein profiling data generally performed poorly upon external validation in time (26). This lack of reproducibility may be due to variation in chip batches, mass spectrometers, sample stability, the low reproducibility of peak height, and the low number of measurements per sample (7, 20, 27). We believe that our model is less affected by these variations for several reasons. First, the sample preparation is simple, fully automated, and does not require chips or fractionations. Second, we did not include the height of the peaks in the model because quantitative measurements of peak heights with both the MALDI and SELDI methods are poorly reproducible (28). In addition, we have carefully determined before analysis the number of replicates per sample that provided the optimal reproducibility (13). The number of replicates that we used (18) was much higher than in other studies. Third, the predictors that we used in our model were clusters of peaks and not single peaks improving the robustness of the model. Changes in one peak position of a cluster, used as a predictor, will not

FIG. 4. Unsupervised clustering demonstrates peptide peaks that differentiate patients with breast cancer with and without LM. The figure illustrates close ups of an unsupervised clustering dendrogram. The clustered masses are displayed on the y-axis, whereas the x-axis represents the samples ordered by group. Colors represent the frequency of a peak position in the seven spectra of a sample (up to two spectra, increasing intensity green; three spectra, black; four to seven spectra, increasing intensity red; see key on figure. A, peak positions that are less frequently expressed in CSF from patients with breast cancer with LM (group I) than in the CSF from patients with breast cancer without LM (group II) and in CSF from control subjects (group III). B, peak positions that are more frequently expressed in group I than in groups II and III.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>AUC</th>
<th>Number of Peaks</th>
<th>Peak Masses (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.74</td>
<td>10</td>
<td>1195.60 (0.0053) 1196.61 (0.0053) 1250.65 (0.047) 1285.61 (0.00023) 1478.73 (16e-05) 1479.74 (0.0037) 1480.75 (0.00038) 1629.78 (7.5e-6) 1630.82 (0.00033) 1631.86 (3.4e-07)</td>
</tr>
<tr>
<td>2</td>
<td>0.69</td>
<td>8</td>
<td>777.26 (0.00033) 968.54 (0.0012) 1114.59 (0.13) 1395.70 (0.0099) 1753.79 (0.0052) 1754.81 (0.0052) 1787.84 (4.9e-05) 1912.00 (2e-05)</td>
</tr>
<tr>
<td>3</td>
<td>0.67</td>
<td>17</td>
<td>395.21 (0.016) 418.23 (0.079) 426.29 (0.053) 427.32 (0.0759) 519.24 (0.028) 520.28 (0.0740) 522.23 (0.047) 522.96 (0.14) 526.14 (0.014) 526.14 (0.012) 532.26 (0.12) 534.26 (0.012) 544.20 (0.0290) 550.94 (0.052) 552.04 (0.068) 567.13 (0.03) 687.30 (0.031)</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
<td>4</td>
<td>1191.62 (0.038) 1216.70 (0.110) 1433.74 (0.00012) 1632.79 (0.002)</td>
</tr>
<tr>
<td>5</td>
<td>0.65</td>
<td>5</td>
<td>1366.76 (0.12) 1367.77 (0.018) 1499.77 (0.033) 1648.76 (0.18) 1746.78 (0.0025)</td>
</tr>
<tr>
<td>6</td>
<td>0.64</td>
<td>10</td>
<td>1086.59 (0.00076) 1190.62 (0.049) 1201.68 (0.061) 1255.56 (0.052) 1288.66 (0.19) 1302.68 (0.0092) 1344.67 (0.0081) 1345.73 (0.016) 1346.73 (0.39) 1430.75 (0.160)</td>
</tr>
<tr>
<td>7</td>
<td>0.64</td>
<td>2</td>
<td>1349.71 (0.00130) 1350.71 (0.0023)</td>
</tr>
<tr>
<td>8</td>
<td>0.63</td>
<td>15</td>
<td>650.06 (0.01) 656.03 (4.5e-05) 665.97 (0.0083) 671.99 (0.00047) 825.04 (0.00052) 841.03 (9.8e-05) 843.34 (0.0052) 965.45 (0.056) 1187.61 (0.61) 1248.62 (0.0011) 1661.80 (0.077) 1785.80 (6.6e-06) 1786.83 (1.7e-05) 1910.01 (0.0024) 1911.01 (8.3e-05)</td>
</tr>
<tr>
<td>9</td>
<td>0.63</td>
<td>3</td>
<td>1034.54 (0.28) 1538.71 (0.074) 1731.77 (0.0044)</td>
</tr>
<tr>
<td>10</td>
<td>0.63</td>
<td>4</td>
<td>1837.97 (0.67) 1838.92 (0.033) 1857.98 (0.27) 1877.02 (0.19)</td>
</tr>
</tbody>
</table>

Clusters that are used in the predictive model.
have a dramatic effect on the performance of the predictive value of the entire cluster. In the future, the reliability of the method can be further improved by linking multiple peptide peaks to a single protein.

The direct identification of peptides from complex samples remains difficult because of the complexity of tryptic digests of body fluids. A direct MS/MS identification of the peptides using MALDI TOF/TOF is not possible as a result of the presence of multiple peptides, even in small mass windows. Although off line nano LC–MALDI could solve this problem, we believe that the best method to identify peptides in complex mixtures is Fourier transform MS in which the exact mass of the peptide of interest is obtained. In most cases, the detection of multiple peptides derived from a single protein will allow identification of the protein. Our database search on the up-regulated peptides has demonstrated the feasibility of this approach for apolipoprotein A1. The up-regulation of different forms of apolipoprotein has been observed before in different SELDI-TOF studies (29–31). We are currently performing Fourier transform MS to identify the other up-regulated peptides as well.

Confounding factors in the present study could be differences in sample collection and storage between institutes, differences in total protein concentration and white cell count between the groups, reproducibility of the method and patient selection bias. The number of peptides, differentially expressed between the institutes, was identical to a chance distribution, excluding potential biases introduced by differences in sample handling. The white cell count and protein concentration in CSF from patients with breast cancer without LM were increased compared with both patients with breast cancer without LM and healthy control subjects. All samples were routinely centrifuged after lumbar puncture, making contamination of the supernatant with cellular debris unlikely. The elevated protein concentration in the CSF from LM patients is well known and is caused by dysfunction of the blood-brain barrier (4), resulting in an increase of high abundance serum proteins in the CSF. A normalization on protein concentration could have been used to compensate for this difference. However, this implies that less CSF should be used from LM samples. This would result in a lower amount of CSF-specific proteins compared with the control samples. In our opinion, this would result in a bias among the three sample groups. To investigate the potential confounding effect of differences in total protein concentration, we calculated the average number of tryptic peptide digests derived from albumin in each group. The number of albumin-derived peptide peaks did not differ among the three groups. In addition, no significant negative or positive correlation between the number of peaks and the protein concentration could be detected. This provided strong evidence that the differences in protein concentration had not interfered with the analysis.

All patients with breast cancer in the present study had signs or symptoms compatible with LM, which led to the performance of a lumbar puncture. All patients in group I who were diagnosed with LM had positive cytology. All patients in group II had negative cytology combined with clinical follow up, indicating an alternative diagnosis. At this stage, we did not include a group of patients with “false-negative” CSF cytology as indicated by MRI and/or clinical follow up. It will be particularly interesting to investigate the performance of this proteome-based test also in these patients, preferably in a prospective manner.

We conclude that MALDI-TOF analysis of tryptic peptide digests derived from the CSF of patients with breast cancer can support the diagnosis of LM. We expect that the use of more accurate and sensitive measurements by Fourier transform mass spectrometry will further improve the identification of disease-specific patterns and markers from body fluids in the near future.

* This study was supported by the Netherlands Proteomics Centre, by a grant from the Erasmus MC Revolving Fund, and by the European Union grant P-mark. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

To whom correspondence should be addressed. Tel.: 31-104633327; Fax: 31-104633208; E-mail: p.sillevissmit@erasmusmc.nl

REFERENCES

resonance mass spectrometry. *Proteomics* 3, 184–190


