Journal of Neuroscience Research 81:613–621 (2005)

NeuroTechnique

Direct Profiling of the Cerebellum by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry: A Methodological Study in Postnatal and Adult Mouse

Claudine Laurent,[1] Douglas F. Levinson,[2] Sarah A. Schwartz,[3] Peter B. Harrington,[4] Sanford P. Markey,[5] Richard M. Caprioli,[3] and Pat Levitt[1]*

[1]Vanderbilt Kennedy Center for Research on Human Development and Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, Tennessee
[2]Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania
[3]Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee
[4]Ohio University Center for Intelligent Chemical Instrumentation, Athens, Ohio
[5]Laboratory of Neurotoxicology, National Institute of Mental Health, NIH, Bethesda, Maryland

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI MS) can detect substantial changes in expression of proteins in tissues, such as cancer cells. A more challenging problem is detecting the smaller changes expected in normal development or complex diseases. Here we address methodological issues regarding the acquisition and analysis of MALDI MS data from tissue sections, in a study of mouse cerebellum at different stages of development. Sections of the cerebellar cortex were analyzed at the peak of granule neuron production [postnatal day (P) 7], during synapse formation (P14), and in adults. Data were acquired (Voyager-DE™ STR Biospectrometry Workstation; seven acquisitions of 50 shots per section, 3.5–50 kDa), preprocessed (Data Explorer 4.3), and averaged. Among 846 peaks detected, in at least 50% of at least one group, 122 showed significant group differences (Kruskal-Wallis ANOVA) after Bonferroni correction. Factor analyses revealed two age-related factors, possibly reflecting gradients of expression during development. Predictive analysis of microarrays generated a model from half of the sample that correctly predicted developmental groups for the second half. Intraclass correlation coefficients, measuring within-mouse consistency of peak heights from three tissue sections, were acceptable at lower *m/z* and for larger peaks at higher *m/z*. Low mass was the best predictor of significant group differences. The analysis demonstrates that MALDI MS of normal tissue sections at different ages can detect consistent, significant group differences. Further work is needed to increase the sensitivity of the methods and to apply them reliably to brain regions and to subproteomes with relevance to diverse brain functions and diseases.

Key words: cerebellum; MALDI; mass spectrometry; proteomics

High-throughput mass spectrometry (MS) is a method for simultaneously detecting a large number of proteins in a biological tissue or extract. MS methods such as matrix-assisted laser desorption/ionization-time-of-flight MS (MALDI MS) can detect hundreds of peaks that are for the most part proteins (Karas et al., 1987; Hillenkamp et al., 1991). The protonated molecular weight for each protein is defined by the centroid of the peak at the measured mass [mass/charge (*m/z*) ratio].

MALDI MS has been used successfully for detecting patterns of substantial overexpression of proteins in cancer cells (Yanagisawa et al., 2003; Schwartz et al., 2004). There is also considerable interest in applying this method to the study of complex biological states, for example, to determine differences in protein expression between knockout and wild-type mice for a candidate gene or in brain tissue from patients with a disease such as schizophrenia compared with controls. In studies of

*Correspondence to: Pat Levitt, PhD, Box 40, Peabody, 230 Appleton Place, Nashville, TN 37203. E-mail: pat.levitt@vanderbilt.edu

Received 7 April 2005; Revised 19 May 2005; Accepted 21 May 2005

Published online 20 July 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jnr.20590

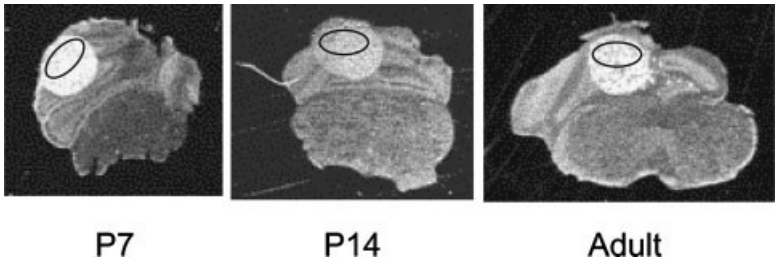

Fig. 1. Placement of matrix spots and laser impact on tissue sections. Shown are examples of coronal tissue sections at each stage of development. In each section, the bright-colored circle in the cerebellar region is the spot of matrix solution. The black oval drawn within each spot illustrates the area in which the laser impact would be targeted (average of 7 acquisitions of 50 shots each, spread within the oval area).

cellularly complex tissues, specific cell populations can be studied by acquiring data directly from tissue sections that preserve spatial resolution for the detection of uniquely expressed proteins. Through absorption of the energy from the laser pulse, surface molecules are desorbed and ionized by proton transfer processes and the resulting ions are related to their m/z through measurement of their time-of-flight to the detector (Knochenmuss and Zenobi, 2003).

Currently, it is not known whether or how reliably this method can detect differences in expression across experimental groups for a large number of proteins across the range of molecular weights and under different experimental conditions. Methodological advances are needed for optimization of sample preparation, selection and application of the matrix material necessary for the MALDI process, MS instrument settings for data acquisition, and preprocessing of digitized data for individual spectra. Perhaps most critically, there are substantial challenges for the analysis of complex data sets derived from multiple spectra across samples and experimental groups.

To address several of these methodological issues, we designed an experiment on brain tissue sections in wild-type mice comparing experimental groups in which one could expect both larger and smaller difference in protein expression across a range of molecular weights. We have studied as a model system the developing cerebellar cortex in three groups of C57Bl6 mice: pups at postnatal day (P) 7 (the peak of granule neuron production), P14 (during synapse formation), and in adults, when neural circuits are mature. There are advantages to studying the cerebellum in this type of experiment; this brain structure exhibits stereotypic development in well-defined epochs, contains only eight major cell types, and is a spatially well-defined region (Sotello, 2004), which facilitates histological localization at low magnification for accurate placing of the matrix spots in the same area of consecutive sections. These features make the developing cerebellum an ideal model for determining the extent to which high-throughput MS can be applied successfully to protein expression to detect reproducible age-related differences in the measured proteome.

Here we report a study of developing murine cerebellar cortex with MALDI MS, utilizing a combination of analytical strategies to demonstrate three patterns of protein expression that correspond to the specific age groups, with excellent internal consistency within each group of samples. This finding suggests that our methods were robust in quantifying relative levels of protein expression for a large number of proteins in a single step, but there are limitations in the degree to which the full protein complexity in the brain can be detected.

MATERIALS AND METHODS

Animal and Sample Preparation

Cerebellar specimens from 77 C57Bl6 mice (*Mus musculus domesticus*) were analyzed: 42 adults at 8 weeks of age (female n = 25, male n = 17), 19 pups at P14 (female n = 9, male n = 10), and 16 pups at P7 (sex unknown). All animal procedures were performed in accordance with the Vanderbilt University Guide for Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee. Mice were anesthetized by isofluorane and sacrificed by decapitation. The brains were dissected, immersed in liquid nitrogen for rapid freezing, and immediately stored at –80°C until sectioning on a cryostat. For each mouse, each of three consecutive frozen sections (bregma –6.5 to –6.7 mm) was collected at 14 μm thickness and was deposited and dried on gold plate, and a double spot of 0.1 μl of matrix was applied under the microscope to ensure consistency of spot areas (Fig. 1). The matrix solution was saturated in 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid; Sigma, St. Louis, MO) in acetonitrile/H_2O/trifluoroacetic acid 50/50/0.3. Tissue section thickness (10–16 μm), drying time in vacuum (10 min to 2 hr), and matrix solution applications (sinapinic acid from 10 mg/ml to saturation) were varied to identify optimized conditions for which the highest intensity and lowest noise was observed in the protein spectra (data not shown).

Mass Spectrometer Data Acquisition

Mass spectra were acquired with a Voyager-DE™ STR Biospectrometry Workstation (Applied Biosystems, Foster City, CA). This instrument was equipped with a nitrogen laser (337 nm), and data were obtained by using the linear acquisition mode under delayed extraction conditions. The laser spot size on target was approximately circular, with a diameter of 25 μm. Instrument settings were an accelerating voltage of 25 kV, 91% grid voltage, 0.05% guidewire voltage, delay time of 220 nsec, and bin size of 2 nsec. Three spectra (each an average of 7 acquisitions of 50 shots each) were acquired for each mouse,

one on each of the three sections described above. Each raw spectrum (N = 231, 3 for each of 77 mice) was preprocessed by applying noise filtering with a correlation factor of 0.7. Internal calibration standards were single-charged alpha and beta hemoglobin chains (MWs 14.982 and 15.617 kDa), thymosin beta-4 and thymosin beta-10 (MWs 4.965 kDa and 4.937 kDa), cytochrome c oxidase polypeptide VIIC (MW 5.444 kDa), ubiquitin (MW 8.565 kDa), and calmodulin (MW 16.791), which have been identified in Bl6C57.

Data Analysis

Baseline correction. By using software written in this laboratory, the chemical noise present in each spectrum was estimated by determining the minimal measured intensity value in successive 100 *m/z* windows. A function fit to these minima by least squares was defined as baseline and subtracted from the spectrum.

Binning of peaks. Data were analyzed for the MW range of 3.5–50 kDa; peaks above 50 kDa were poorly defined, and so were not included in this study. The three spectra from each mouse were averaged arithmetically (Origin Pro 7), and automated peak detection was carried out on these averaged spectra in Data Explorer 4.3 software, which determines the centroid mass (*m/z*), height, *m/z* for the 50th percentile of the lower and upper boundary of the peak, and area and signal-to-noise ratio for each peak. We used a signal-to-noise threshold of 2, a 5-kDa window for computing noise, and 50% centroid to define peak width. Then, a grand average spectrum was obtained (Fig. 2a); to give equal weight to each group, the average of the spectra for each age group was determined, and then these three averages (illustrated in Fig. 2b) were averaged. Peaks (N = 945) were detected from this grand average spectrum by using an S/N ratio of 1.5 and a noise window width (*m/z*) of 1.024 kDa. For each grand average peak, the 50th percentile upper and lower boundary was defined as a "bin." For each mouse, if there was a peak whose centroid mass fell within that range, the peak and its parameter values (height, area) were assigned to the bin, or, if not, values of 0 were assigned for height and area. We selected for further analysis the 846 peaks identified by this method in a minimum of 50% mice in at least one of the age groups.

Statistical analyses. Primary analyses considered peak heights, although analysis of peak area yielded similar results. Although areas would generally be considered more closely correlated with protein signal under conditions that were optimized for a specific protein, under high-throughput conditions without such optimization, height was considered a more reliable measure of relative quantity. Two height variables were created for the binned peaks for each mouse. 1) Ranked height was the rank from 846 (largest height) to 1 (smallest) of the heights among the 846 peaks for that mouse; identical values (such as 0 heights) were assigned the average of the range of ranks for that value. 2) Normalized height was calculated as follows: given an observed intensity, $h_{i,j}$, of peak i for mouse j; a sum, H_j, of all h_i for mouse j (for the 846 selected peaks); and an average, A of H, across all mice, then the normalized height (Ht) was calculated as

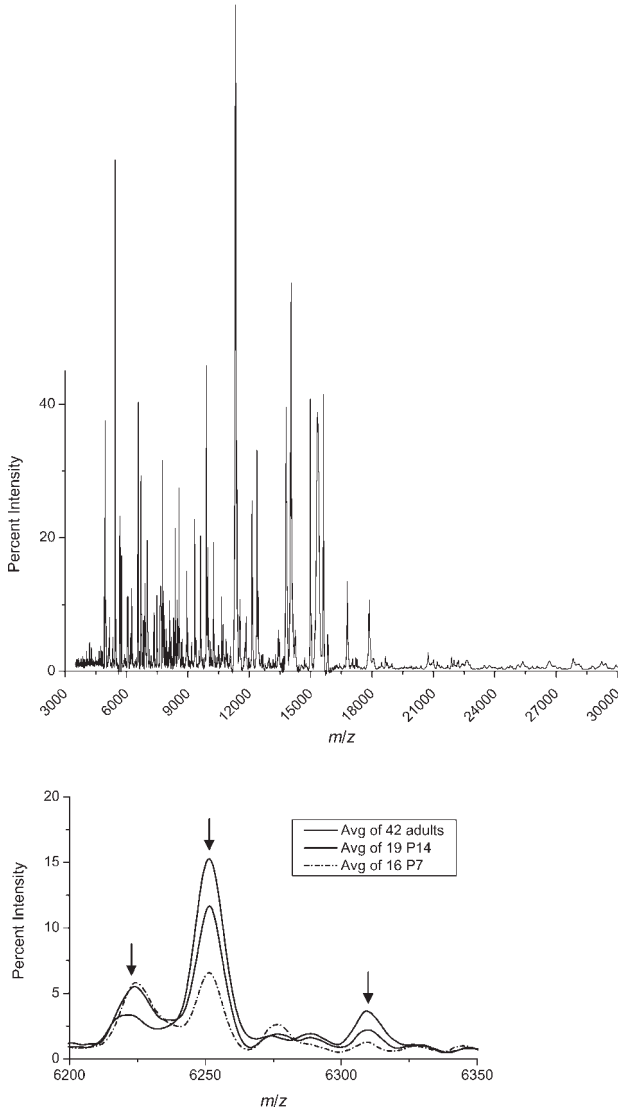

Fig. 2. Examples of averaged spectra. (a) Grand Average Spectrum (average of the 3 group averages) for the range 3.5–30 kDa. The Y-axis value is the percentage of the maximum intensity value in the spectrum. The intensity range has been truncated at 45% to permit visualization of smaller peaks. (b) Detail of the 3 group averages (6.2–6.350 kDa). The arrows mark three peaks with significant age-related differences in height, with centroid masses in the grand average spectrum at 6.224, 6.251 and 6.310 kDa.

$$Ht_{ij} = \frac{h_{ij}}{H_j} A. \quad (1)$$

Analyses of group differences were carried out by using ranked heights. This was intended as a conservative approach to analysis; i.e., ranking the heights within each mouse avoids assumptions regarding the absolute relationships among larger and smaller peak heights, permitting analysis only of changes in relative expression of proteins. Group differences in ranked peak height were analyzed by using Kruskal-Wallis analyses of variance. An alpha of 0.05 was selected as the threshold of sig-

TABLE I. Distribution of Significant Peaks by *m/z* Range and by Factor*

m/z Range (kDa)	All	3.5–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50
Total N(peaks)	846	25	89	57	22	52	76	129	137	124	135
Total significant	122	5	50	36	8	6	9	4	1	3	0
Factor 1	30	0	11	15	3	0	1	0	0	0	0
Factor 2	32	0	18	8	3	1	0	1	0	1	0
Neither	60	5	21	13	2	5	8	3	1	2	0
Mean NrmHt	709	1,114	2,925	3,229	2,216	209	186	116	102	109	86
Mean RnkHt	423.5	479	657	642	600	472	484	370	343	349	288

*Shown are the numbers of peaks detected in the grand average file in each *m/z* range, the numbers of these peaks that showed significant differences across age groups after Bonferroni correction (nominal $P < 0.00006$), the numbers of significant peaks that with loadings ≥ 0.5 or ≤ -0.5 on factors 1 or 2 or not loading this strongly on either factor. Also shown are the mean normalized heights (NrmHt, in arbitrary units; see text) and ranked heights (RnkHt) for all peaks within each range. It can be seen that the most significant group differences are detected in the range of 5–20 kDa, where the largest peak heights are detected using this methodology.

nificance, with Bonferroni correction for the 846 peaks that were analyzed. The corrected threshold of significance was therefore $1 - (1 - 0.05)^{1/846} = 0.00006$. Split-half consistency was analyzed by dividing the sample into the first and last half of males and females tested in each group, assuming that experimental conditions would vary maximally across these subsets (tested over a 6-month period). Kruskal-Wallis analyses of group differences were carried out separately in each half-sample to determine consistency of significant differences.

Analyses of variability in peak heights across the three individual spectra for each mouse were carried out by using normalized heights, with two-way random effects intraclass correlation coefficients (ICC) without interaction (McGraw and Wong, 1996). This coefficient, which is similar to Cronbach's alpha, measures consistency of individual measurements in relation to the *average* of the (three) individual values, which is the measure of interest here. It is computed by carrying out an analysis of variance for each peak with the model NrmHt = Constant + Mice + Spectra, and computing ICC as (MS[Mice] − MS[Error])/MS[Mice] (McGraw and Wong, 1996). This analysis was carried out by using 231 individual spectra from the 77 mice. Peak lists were generated and peaks were binned for each spectrum as described above, except that the signal-to-noise setting of 1.5 was used to maximize the number of spectra in which each peak was detected. Relationships among variables (log of the mean normalized height, *m/z*, ICC, and log of the Kruskal-Wallis statistic for group differences in the entire sample) were analyzed by using Pearson correlations and multiple regression analyses. Kruskal-Wallis, correlation, and analysis of variances tests were carried out in Systat 8.0.

Relationships among the 122 ranked peak heights that showed group differences were examined by subjecting them to principal-components factor analyses with varimax rotation, first for the entire sample, and then for the first and second half-samples separately. Correlations were computed between the factor loadings for the two half-samples.

Analysis of group membership prediction was carried out with prediction analysis of microarrays (PAM; Tibshirani et al., 2002). This method was developed to identify subsets of gene expression values that predict group membership, based on a form of cluster analysis using the shrunken centroids method. We used the first half-sample as a training set to determine the prediction model with the lowest prediction error with the smallest number of ranked heights and then applied the model to the second half-sample. The analysis was repeated with normalized heights, with similar results (not shown), but, with ranked heights, prediction error was lower, and few peaks were required for classification.

An additional set of analyses was carried out of reproducibility of individual spectra and of group differences, by using ANOVA-principal components analysis (Harrington et al., 2005a,b) of individual spectral points in Matlab 7.0.4 (The Mathworks Inc., Natick, MA) and a fuzzy rule-building expert system (FuRES; Harrington, 1993) to predict group membership. Details of these analyses will be reported elsewhere; we note here that these methods detected age-related group differences similar to those reported here but no sex-related or litter-related differences.

RESULTS

Group Differences

Significant group differences for ranked heights were observed for 122 peaks by Kruskal-Wallis (K-W) analysis of variance (nominal $P < 0.00006$). Among these, 87 peaks showed significant group differences at the corrected $P < 0.001$ level (nominal $P < 0.0000012$). Table I gives the distribution of analyzed peaks and of significant peaks across the range of *m/z* values. Among the 122 significant peaks, 105 (86.1%) had *m/z* values less than 25 kDa, representing 42.9% of the 245 analyzed peaks in this range; although larger numbers of peaks were detected at higher *m/z* values, few of these produced significant group differences. Factors influencing peak detection and significance are discussed further below.

To examine split-half consistency, we first determined the nominal *P* value (0.00314) that, if observed in both half-samples, would result in an overall *P* value of <0.00006 when combined using Fisher's (1954) formula. Overall, 124 peaks achieved this level of significance in the first half, 108 in the second half, and 85 in both halves (i.e., 69.8% of all peaks that were significant in the entire sample). Among the 122 peaks that were significant overall, 109 (89.3%) achieved $P = 0.00314$ in the first half, and 96 (78.0%) did so in the second half.

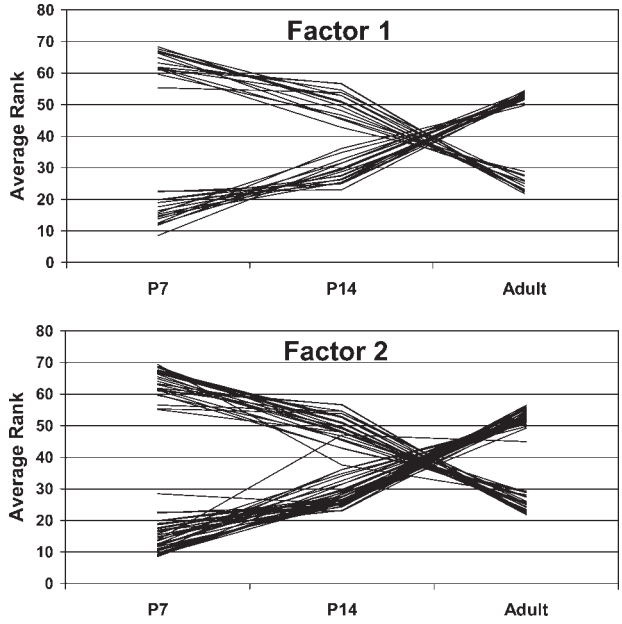

Fig. 3. Average rank by group for significant peaks. Each line represents the values, for one peak, for the average ranks (within the Kruskal-Wallis test) achieved by mice in each group. Shown are 62 of the 122 peaks with significant group differences: 30 peaks loading on Factor 1 and 32 peaks loading on Factor 2. Most peaks showed a distinct age gradient (Adult > P14 > P7, or P7 > P14 > adult).

Factor analysis revealed two major components, and, when a two-factor solution was specified, it explained 38.7% of total variance. Further details about the specific peaks loading on these factors are available upon request. Among the 122 significant peaks overall, 30 loaded at >0.50 (or <-0.50) on factor 1 of the rotated solution and 32 peaks on factor 2. The mean correlations between factor loadings for the first and second half-samples were 0.81 for factor 1 loadings and 0.85 for factor 2. Figure 3a,b illustrates group differences for the peaks loading on factors 1 and 2, respectively, showing the average Kruskal-Wallis (K-W) rank for each age group for each peak. Note that the illustrated values are not ranked heights; for each peak, the K-W test ranks each mouse from 77 to 1 (or tied values) in order of ranked peak height and determines whether the groups differ in their sums of ranks. Figure 3a,b illustrates the average of these values (i.e., if summed ranks = 425 for 19 mice, then the average rank is 22.36). The figures demonstrate that these 62 peaks were generally detected in larger quantities either in adults or at P7, with P14 typically intermediate. Thus, for each factor, there appears to be one subset of proteins whose expression increases during development, and a second set whose expression decreases during development, with varying rates of developmental change as expected.

Variability of Individual Spectra

Peak detection was also carried out for the 231 individual spectra acquired from the 77 mice. Peaks detected in these spectra were assigned to the bins representing the 846 peaks utilized in the analyses reported here. Two-way random-effects ICC values were computed for each peak by using normalized peak heights. The mean ICC was 0.325, with mean ICC values of 0.432 for peaks with m/z below 25 kDa, 0.341 between 25 kDa and 34 kDa, and 0.257 between 34 kDa and 50 kDa. For the 122 peaks showing significant group differences, mean ICC was 0.537 (considered fair to good), vs. 0.289 (poor) for all other peaks. Figure 4 shows the relationships among m/z, ICC, and whether each peak showed significant group differences. Both lower m/z and higher ICC were associated with detection of group differences. In further exploratory analyses, a number of interrelationships were observed among measures of molecular weight (m/z), statistical significance of group difference (the K-W statistic, log transformed to approximate a normal distribution more closely), consistency across individual spectra (ICC), and the log of averaged normalized peak heights. Table II lists Pearson correlations (for the 846 peaks) among these measures (Spearman correlations were almost identical). Peak height was strongly and negatively predicted by m/z, and peak height was in turn the strongest predictor of consistency (ICC) and significance (K-W statistic).

Table III shows the results of two multiple regression analyses carried out to explore these relationships further, and Figures 5 and 6 illustrate some of these relationships. In the first analysis, the significance of group differences for each of 846 peaks was strongly predicted by m/z, with low m/z associated with more significant group differences. The two other predictors interacted with m/z: for peaks with higher m/z, larger peak height (analyzed as the log of the mean normalized peak height across mice) predicted more significant group difference; whereas, for peaks with lower m/z, differences tended to be significant regardless of height; and, for peaks with lower m/z, larger ICC (consistency) predicted more significant group differences. In the second analysis, consistency (ICC) was predicted by m/z; by an interaction of m/z and height, i.e., for peaks with higher m/z, larger peak heights predicted higher ICC, whereas, for peaks with lower m/z, consistency tended to be high regardless of relative peak height (which was larger for peaks at lower m/z).

Prediction of Group Membership

Table IV shows the results of split-sample analysis using PAM; group membership was correctly predicted for all mice in the first half-sample at a threshold value of 1.73 using 68 peaks, and this model also correctly predicted group membership for all mice in the second half-sample. The peaks were generally the same as those that were significant in K-W analysis of group differences and that loaded on factors 1 and 2, particularly those for which one group had substantially lower or higher peak heights than the other groups.

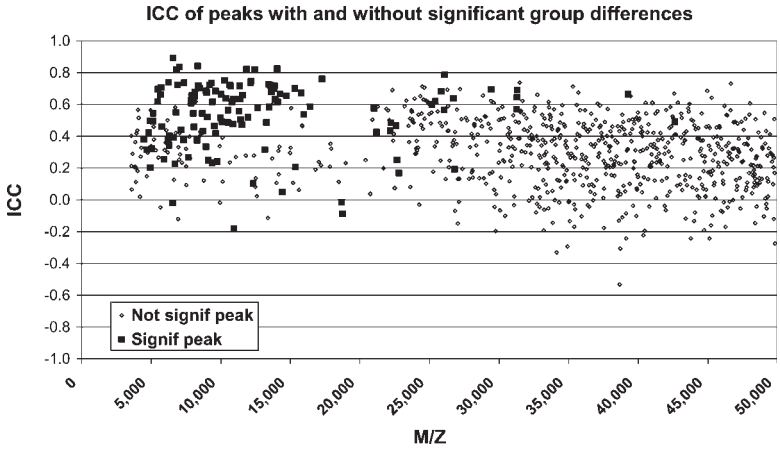

Fig. 4. ICC of peaks with and without significant group differences. ICC values for each peak are plotted against m/z. Filled squares represent peaks for which significant group differences were observed; open triangles represent peaks for which no significant group difference was observed.

TABLE II. Correlations Among Mass, Peak Height, Consistency of Individual Spectra (ICC), and Significance of Group Differences (Kruskal-Wallis Statistic)*

	Log(KW statistic)	Mass	Log(NrmHeight)	ICC
Log(KW statistic)		<0.00000001	<0.00000001	<0.00000001
Mass	−0.578		<0.00000001	<0.00000001
Log(NrmHeight)	0.615	−0.840		<0.00000001
ICC	0.378	−0.315	0.401	

*Shown are Pearson correlations (below the diagonal, with *P*-values above the diagonal) among selected variables for 846 peaks (Spearman correlations were almost identical). ICC values are the two-way random effects intraclass correlation coefficients for consistency of the average normalized peak height with the heights of the three-individual spectra for 231 spectra acquired from 77 mice. m/z Is the centroid mass for each peak in the grand average spectrum (point-by-point average across all 231 spectra after calibration and baseline correction). Peak height is the average across 231 spectra of the peak height from each spectrum after normalization (height divided by the sum of the heights of all 846 peaks, multiplied by the average sum of heights). K-W statistic is from the test of group differences for the sample of 77 mice. These relationships are further explored in Table III.

DISCUSSION

High-throughput MS was used to study protein spectra in the cerebellum at different stages of mouse development. Our goal was to determine whether reproducible data could be obtained by MALDI MS from tissue sections, after optimization of conditions for data acquisition and data preprocessing and whether differences in protein peaks could be reliably detected among defined groups of mice. Although our long-term goal is to study the more subtle changes that might be expected in complex brain diseases, we carried out this initial study with a simpler model, i.e., on mice with a uniform genetic background, maintained in the same environment (food, mouse house), and divided into groups based on the straightforward variable of age (developmental period), which could be expected to be associated with substantial differences in protein profiles.

We found that MALDI MS applied directly to tissue sections generated a large number of peaks (N = 846), among which 122 discriminated statistically among the three age groups after conservative correction for multiple testing. Factor analysis of these 122 peaks identified two major factors that accounted for 38.7% of the variance among groups, primarily because of 62 peaks that loaded strongly on one of these factors. The PAM shrunken-centroids method was also able to predict group membership based on 68 peak heights, overlapping (as would be expected) with those that contributed to the factor analysis. Thus, protein profiles generated from tissue sections detected robust differences among groups by several statistical methods.

The reproducibility of the results was examined in several ways. Three complete spectra were obtained for each mouse, from successive thin tissue sections using identical coordinates, with each spectrum generated from 350 laser shots. Normalized peak heights were reasonably consistent across these three spectra for peaks with lower molecular weights (<25 kDa), but less so for those with higher molecular weights. Not surprisingly, most of the significant group differences in peak heights were observed for peaks in this lower mass range, with mass emerging as the best single predictor of the significance of group differences. Although it has been suggested that current methods are most reliable in the lower mass range (for review see Westmacott et al., 2000), there are few empirical data sets that demonstrate it. Our data do support this conclusion. The factors responsible for this cannot be fully elucidated from the present data. Mass was strongly and inversely related to peak height, and mass and an interaction of mass and peak height predicted the significance of group differences. At lower molecular weights, peaks were larger, showed more consistent heights (high ICC) for individual spectra from the same mice, and frequently were significantly different across age groups. At higher molecular weights, peaks and ICC were

TABLE III. Peak Characteristics That Predict Detection of Group Differences and Consistency of Individual Spectra*

Dependent variable	Overall F	df	*P*	R^2	Effects	T	*p*
Log(KW)	108.96	6;839	<0.00000001	0.44	*m/z*	–4.69	<0.00001
					Height	1.02	0.31
					ICC	1.54	0.12
					m/z × height	4.75	<0.00001
					Height × ICC	–0.12	0.90
					m/z × ICC	–3.03	0.0025
ICC	88.67	3;842	<0.00000001	0.24	*m/z*	–8.71	<0.00001
					Height	0.66	0.51
					m/z × height	9.26	<0.00001

*Shown are results of multiple regression analyses of data for 846 peaks, to explore the relationships among the log of the mean normalized peak height, the centroid mass (*m/z*), the consistency (ICC) of normalized heights of the three individual spectra acquired for each mouse, and the significance of group differences in ranked peak height (log of the Kruskal-Wallis statistic) for each peak.

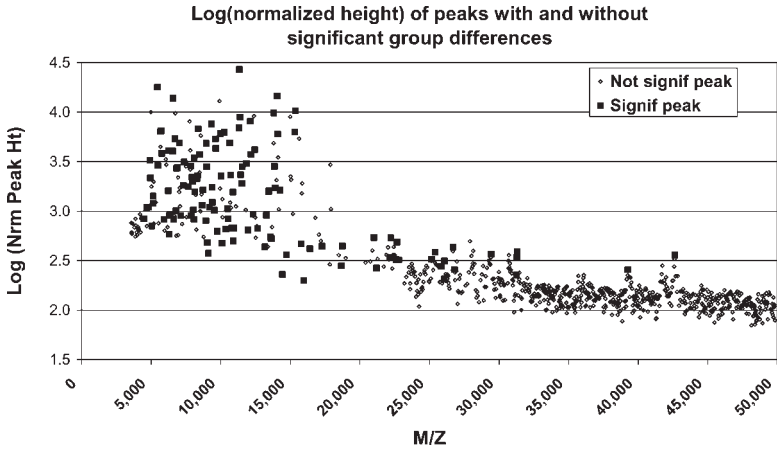

Fig. 5. Log(normalized height) of peaks with and without significant group differences. Shown for each of 846 peaks is the log of the averaged normalized peak height plotted against *m/z*. Peaks that demonstrated significant differences among age groups are designated by filled squares, and all other peaks by open diamonds.

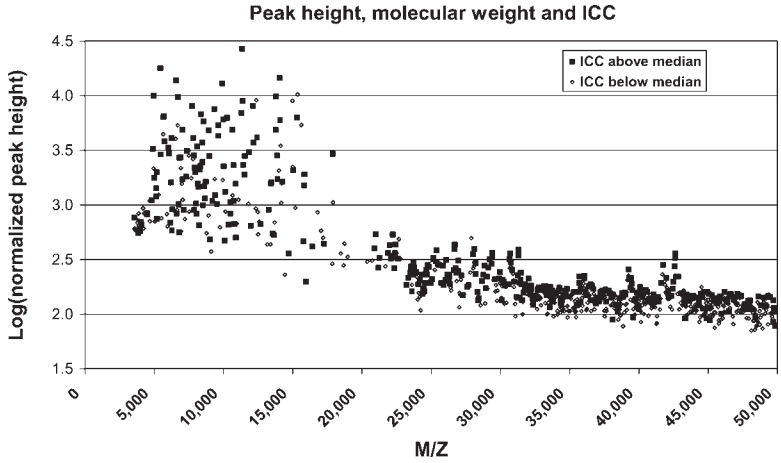

Fig. 6. Peak height, molecular weight and ICC. For each of 846 peaks, the log of the average normalized peak height is plotted against *m/z*. Black squares and open diamonds, respectively, represent peaks with values above and below the median (0.336) intra-class correlation coefficient, which measured reliability of each of the three individual spectra acquired for each mouse with the average of those spectra, in relation to variability for the same peak across mice.

smaller, but Table III and Figures 4 and 5 demonstrate that there was an even stronger positive association among peak height, ICC, and detection of significant group differences. Our data suggest that additional technical advances will be needed to detect reproducible peaks and significant group differences for proteins with higher molecular weights.

Reproducibility was also examined by carrying out tests of group differences separately in the first and second halves of the sample. Because the data were acquired over a period of 6 months, one would expect that, if there was any unknown variation in experimental conditions, this would be greatest over the longest period of time. Most of the peaks that were significant by K-W analysis also contributed to the differentiation among the three groups in each half of the sample with PAM, and the model that assigned all mice to the correct group in the first half-sample was equally successful at predicting group membership in the second half of the sample. Thus, at least for lower molecular weights,

TABLE IV. Prediction of Group Membership by PAM*

	First half-sample (training data set)			Second half-sample (test data set)		
Group	Adult	P14	P7	Adult	P14	P7
Adult	21	0	0	21	0	0
P14	0	9	0	0	10	0
P7	0	0	8	0	0	8

*Shown are the results of analyses with the predictive analysis of microarrays program (Tibshirani et al., 2002). Ranked heights for all 846 peaks were submitted for the first half of the males and females tested in each group (N = 38) (training data set). Group membership could be predicted without errors using 68 peak heights (threshold = 1.73). The same model also predicted group membership in the remaining mice (test data set).

group differences can be reliably detected from tissue sections both within and across mice. High-throughput MALDI MS of tissue sections has been used by our laboratory for analysis of different cancers (Schwartz et al., 2003, 2004; Yanagisawa et al., 2003). Both the glioma and the lung cancer studies differentiated between disease and control specimens, with reproducible results across halves of the sample. Yanagisawa et al. (2003) also reported strong correlations between the intensities of higher intensity peaks acquired from two cell lysates of human bronchial epithelial cell tissue and from two slices of the same lung tumor, as well as lower, though still substantial, correlation for two different lung tumors. Thus, analysis of a variety of tissue specimens, with intact histological organization, can yield reproducible spectra that differentiate among specimens from biological subgroups.

Most proteomic studies of high-throughput MS have involved attempts to differentiate between serum samples of patients with early and/or late stages of various cancers (brain, breast, colon, lung, ovary, and prostate cancer) compared with controls. In many of these studies, the groups of interest could be classified reproducibly with small sets of 4–25 peaks. Several studies have used SELDI-TOF methodology, including studies of prostate cancer (Adam et al., 2002; Qu et al., 2002), ovarian cancer (Petricoin et al., 2002; with reanalyses by Baggerly et al., 2004, and by Li et al., 2004), and breast cancer (Li et al., 2002). MALDI MS methods have also been used in studies of serum samples in prostate cancer (Baggerly et al., 2003; Wagner et al., 2003). In general, these studies have relied on reproducibility of group differences to establish the validity of their methods. Many statistical approaches have been successfully applied in these studies, as reviewed by Wu et al. (2003) and by Li et al. (2004). Our data suggest that, when robust group differences are detected in protein profiles, they can be quantified by diverse statistical methods.

A major issue in the cancer proteomics literature is that many of the putative biomarkers are nonspecific in their biological functions, although in some cases specific biological roles in carcinogenesis or metastasis could be hypothesized (Yanagisawa et al., 2003). Diamandis (2004) has reviewed the nonspecificity of most reported cancer biomarkers and has pointed out that the cancer literature is largely lacking in empirical studies of confounding factors such as 1) ''variability in sample collection, processing and storage''; 2) characteristics of study subjects other than diagnosis; and 3) variations in ''mass spectrometric stability and protein chip performance'' as well as statistical methods. Several groups have begun to address these issues. Cordingley et al. (2003) recently reported one of the few comprehensive empirical studies of reproducibility in this field, examining the effects of 11 experimental factors on protein spectra acquisition with the SELDI-TOF platform. They identified sample preparation and matrix preparation and application procedures that maximized reproducibility. Baggerly et al. (2004), in a reanalysis of serum data from ovarian cancer studies by Petricoin et al. (2002), addressed a number of data processing issues that could introduce artifacts into MS studies. The present study was designed to begin to address some of these methodological issues by using MALDI MS, in the context of examining tissue sections from the brain, which is likely the most complex of tissues at both the cellular and the molecular levels. The use of an experimental model, the mouse, allowed us to control for genetic and environmental variations, facilitating the examination of reproducibility both within and across subjects. Identical sample preparation, storage, and data processing procedures were applied to all specimens. Our results, demonstrating that the same protein peaks could differentiate among age groups in half-samples studied months apart, suggest that it is possible to control essential experimental variables. Because we did not systematically study the effects of each experimental variable, as was done by Cordingley et al. (2003), it remains possible that the results were confounded by undetected artifacts. We conclude, however, that a substantial number of proteins can be reliably and reproducibly detected from sections of highly complex brain tissue by using high-throughput MALDI MS, in the molecular weight range of approximately 5–25 kDa, with a well-defined set of standardized conditions. Determination of the biological relevance of the findings reported here awaits future studies to identify the proteins that differentiated the age groups.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Paula J. Woods, BS; Daniel B. Campbell, PhD; Jeremy L. Norris, PhD; Hans Aerni, MS; James A. Mobley, PhD; Dale S Cornett, PhD; and Pierre Chaurand, PhD. This study was supported in part by project 4 (P.L.) of the NIMH Conte Center for the Neuroscience of Mental Disorders (MH051456 to D. Lewis, University of Pittsburgh), NICHD Mental Retardation Developmental Disability Research Center core grant P30HD015052 (P.L.), NIGMS grant R01GM058008 (R.M.C.), and NIMH grant K24MH064197 (D.F.L.).

REFERENCES

Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer PF, Yasui Y, Feng Z, Wright GL Jr. 2002. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 62:3609–3614.

Baggerly KA, Morris JS, Wang J, Gold D, Xiao LC, Coombes KR. 2003. A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples. Proteomics 3:1667–1672.

Baggerly KA, Morris JS, Coombes KR. 2004. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 20:777–785.

Cordingley HC, Roberts SL, Tooke P, Armitage JR, Lane PW, Wu W, Wildsmith SE. 2003. Multifactorial screening design and analysis of SELDI-TOF ProteinChip array optimization experiments. Biotechniques 34:364–365, 368–373.

Diamandis EP. 2004. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. JNCI 96:353–356.

Fisher RA. 1954. Statistical methods for research workers, 12th ed. New York: Hafner.

Harrington PdB. 1993. Minimal neural networks: differentiation of classification entropy. Chemometr Intel Lab Syst 19:143–154.

Harrington PdB, Vieira NE, Espinoza J, Nien JK, Romero R, Yergey AL. 2005a. Analysis of variance-principal component analysis: a soft tool for proteomic discovery. Anal Chim Acta (in press).

Harrington PdB, Vieira NE, Espinoza J, Nien JK, Romero R, Yergey AL. 2005b. proteomic analysis of amniotic fluids using analysis of variance-principal component analysis and fuzzy rule-building expert systems applied to matrix-assisted laser desorption/ionization mass spectrometry. Anal Chim Acta (in press).

Hillenkamp F, Karas M, Beavis RC, Chait BT. 1991. Matrix-assisted laser desorption ionization mass-spectrometry of biopolymers. Anal Chem 63:A1193–A1202.

Karas M, Bachmann D, Bahr U, Hillenkamp F. 1987. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. Int J Mass Spectrom Ion Process 78:53–68.

Knochenmuss R, Zenobi R. 2003. MALDI ionization. The role of in-plume processes. Chem Rev 103:441–452.

Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. 2002. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin Chem 48:1296–1304.

Li L, Tang H, Wu Z, Gong J, Gruidl M, Zou J, Tockman M, Clark RA. 2004. Data mining techniques for cancer detection using serum proteomic profiling. Artif Intel Med 32:71–83.

McGraw KO, Wong SP. 1996. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30–46.

Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. 2002. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359:572–577.

Qu Y, Adam BL, Yasui Y, Ward MD, Cazares LH, Schellhammer PF, Feng Z, Semmes OJ, Wright GL Jr. 2002. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. Clin Chem 48:1835–1843.

Schwartz SA, Reyzer ML, Caprioli RM. 2003. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. J Mass Spectrom 38:699–708.

Schwartz SA, Weil RJ, Johnson MD, Toms SA, Caprioli RM. 2004. Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression. Clin Cancer Res 10:981–987.

Sotelo C. 2004. Cellular and genetic regulation of the development of the cerebellar system. Prog Neurobiol 72:295–339.

Tibshirani R, Hastie T, Narasimhan B, Chu G. 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99:6567–6572.

Wagner M, Naik D, Pothen A. 2003. Protocols for disease classification from mass spectrometry data. Proteomics 9:1692–1698.

Westmacott G, Frank M, Labov SE, Benner WH. 2000. Using a superconducting tunnel junction detector to measure the secondary electron emission efficiency for a microchannel plate detector bombarded by large molecular ions. Rapid Commun Mass Spectrom 14:1854–1861.

Wu B, Abbott T, Fishman D, McMurray W, Mor G, Stone K, Ward D, Williams K, Zhao H. 2003. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. Bioinformatics 19:1636–1643.

Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, Roberts JR, Edgerton M, Gonzalez A, Nadaf S, Moore JH, Caprioli RM, Carbone DP. 2003. Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet 362:433–439.