Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Mass Spectrometry 238 (2004) 189–196
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Improved protein identification using automated high mass measurement accuracy MALDI FT-ICR MS peptide mass fingerprinting

David M. Horn[a], Eric C. Peters[a], Heath Klock[a], Andrew Meyers[b], Ansgar Brock[a,*]

[a] *Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA*
[b] *The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA*

Received 17 November 2003; accepted 21 March 2004
Available online 25 September 2004

Abstract

A comparison between automated peptide mass fingerprinting systems using MALDI-TOF and MALDI FT-ICR MS is presented using 86 overexpressed proteins from *Thermotoga maritima*. The high mass measurement accuracy of FT-ICR MS greatly reduces the probability of an incorrect assignment of a protein in peptide mass fingerprinting by significantly decreasing the score and peptide sequence coverage of the highest ranked random protein match from the database. This improved mass accuracy led to the identification of all 86 proteins with the FT-ICR data versus 84 proteins using the TOF data against the *T. maritima* database. The beneficial effect of mass accuracy becomes much more evident with the addition of variable modifications and an increase in the size of the database used in the search. A search of the same data against the *T. maritima* database with the addition of a variable modification resulted in 77 identifications using MALDI-TOF and 84 identifications using MALDI FT-ICR MS. When searching the NCBInr database, the FT-ICR based system identified 82 of 86 proteins while the TOF based system could only identify 73. The MALDI FT-ICR based system has the further advantage of producing fewer unassigned masses in each peptide mass fingerprint, resulting in greatly reduced sequence coverage and score for the highest ranked random match and improving confidence in the correctly assigned top scoring protein. Finally, the use of rms error as a measure for instrumental mass accuracy is discussed.

Keywords: Peptide mass fingerprinting; MALDI; Mass accuracy; FT-ICR

1. Introduction

Protein identification using mass spectrometry (MS) [1] has become a cornerstone of the burgeoning field of proteomics. The two most popular methods for MS based protein identification are tandem mass spectrometry (MS/MS) [2], primarily used with multidimensional separation of complex peptide mixtures [3], and peptide mass fingerprinting [4–8], generally used with digests of proteins first separated by two-dimensional gel electrophoresis [9]. Both of these methods use algorithms [10–12] that correlate each entry in a protein or genomic sequence database with mass spectral data. Generally, the database entry with the highest correlation "score" is assigned by the program as the identity of the unknown protein in the sample. Successful protein identification depends on several factors such as the sequence coverage obtained for each peptide or protein in the sample, inclusion of one or more variable post-translational modifications in the search, protein/genomic sequence database size, the choice of proteolytic enzyme, the number of extraneous masses in a given spectrum [13], and the allowed mass deviation between experimentally-determined masses and those calculated from the database sequences.

Peptide mass fingerprinting (PMF), also known as peptide mapping or peptide mass mapping, is the primary method for identification of purified proteins, such as those from excised 1D or 2D gel bands. Most PMF experiments are performed

* Corresponding author. Tel.: +1 858 812 1549; fax: +1 858 812 1918.
E-mail address: abrock@gnf.org (A. Brock).

1387-3806/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2004.03.016

using matrix-assisted laser desorption/ionization (MALDI) [14] time-of-flight (TOF) mass spectrometry, which is well suited for PMF because it is easily automated, tolerant of salts and detergents, highly sensitive, and demonstrates good resolution (up to 10,000) and mass accuracy (10–20 ppm rms). Even with such high performance instruments, there are still several issues that can dramatically decrease the chance of successful protein identification. For example, the presence of multiple proteins in a given sample significantly increases the number of masses observed in the spectrum and thus increases the likelihood of an incorrect assignment to other proteins in the database [15]. Furthermore, digests of low molecular weight or low abundance proteins typically exhibit a limited number of matched peptides resulting in low correlation scores. This is compounded by the fact that, in general, a significant percentage of the masses observed in PMF experiments cannot be assigned to a peptide from the identified protein [13,16,17]. Unexpected masses can result from natural post-translational modifications, unintentional modifications incurred during sample processing such as carbamylation, oxidation, deamidation, and nonspecific carbamidomethylation [18], nonspecific cleavages by the proteolytic enzyme, and common protein contaminants such as keratin and protease autolysis peaks. The exact amount of sequence information required for an unambiguous identification is still a matter of debate, but a typical standard requires that at least five peptides match within 30 ppm tolerance, and the score for the next most probable match be significantly lower [19]. These requirements often lead to difficulty in the unambiguous identification of a sample, requiring further analysis using much more time-consuming LC/MS/MS approaches.

To overcome these limitations, a number of techniques have been developed that utilize chemical modification to obtain sequence or chemical information for individual peptides in a fingerprint. This additional information increases the likelihood for successful identification by enhancing the specificity of the peptide assignments. Examples of such an approach include incorporation of stable isotopes for the determination of the number of a specific amino acid contained in a given peptide [20–23], the addition of chemicals that selectively modify a specific amino acid residue such as hydrogen peroxide-mediated oxidation of methionine residues [24], and hydrogen/deuterium exchange reactions for the determination of the number of exchangeable hydrogens in a given peptide [25]. Additionally, derivatization with *O*-methylisourea [26–29] or similarly-tailored compounds [23] may also improve PMF results by increasing the number of observable peptides.

A logically more straightforward approach to enhance the specificity of PMF experiments is to improve the accuracy of the mass measurements. It is undisputed that accurate mass measurements are highly advantageous for protein identification using PMF [30,31], yet the most accurate technology available remains underutilized for this purpose. Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry [32,33] can perform highly accurate mass measurements under appropriately controlled conditions [34], and thus has tremendous potential for applications using PMF as the primary method for protein identification.

Given the superior mass accuracy of FT-ICR MS, there have been surprisingly few published reports of its use in PMF experiments. The most prominent albeit unconventional work involves the analysis of highly complex proteolytic digests of cerebrospinal fluid both with and without chromatographic separation [35–38]. Another study identifies abundant cryoglobulins in blood after their extraction from two-dimensional gels using MALDI FT-ICR MS [39]. Additionally, several studies confirm the advantages of high mass accuracy for conventional PMF experiments using protein standards [31,40]. However, the true value of such studies remains unclear, because they were performed using a single purified known protein and do not describe samples whose mass, identity, and concentration are unknown before the study was undertaken. Further, no study has yet been performed that compares the data from FT-ICR MS to that of TOF MS to obtain experimental verification of the advantages of higher mass accuracy.

Recently, we described a new MALDI FT-ICR mass spectrometer for high throughput analysis of complex mixtures [34]. This instrument achieves low ppm mass accuracy even for highly complex peptide mixtures by mixing the sample and internal calibrants in the gas phase rather than on the sample plate. Custom designed software has been created to completely automate MS experiments, requiring only 5–10 s per sample for each step, including spectral acquisition, automated data reduction, and protein identification by database searching.

Here we describe the application of this instrument towards PMF experiments. Mass spectra for the tryptic digests of 86 *Thermotoga maritima* proteins, individually overexpressed and purified for high throughput structural biology studies at the Joint Center for Structural Genomics [41], were acquired using both an automated MALDI-TOF system and our automated MALDI FT-ICR mass spectrometer. We clearly show that the greatly improved mass accuracy obtainable with our system is highly advantageous for PMF experiments, and that MALDI-TOF may not be optimally suited for PMF experiments in many cases.

2. Experimental

2.1. Materials

Bradykinin, substance P, neurotensin, ACTH fragment 1–17, ACTH fragment 18–39, melittin, oxidized insulin B chain, dihydroxybenzoic acid (DHB), α-cyano-4-hydroxycinnamic acid (CHCA), dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), acetonitrile, ammonium bicarbonate, and diammonium citrate were all purchased from Sigma–Aldrich (St. Louis, MO). Hexanes were purchased from Fisher Scientific (Hampton, NH).

Sequencing grade modified trypsin was purchased from Promega (Madison, WI).

2.2. Preparation of protein digests for FT-ICR and TOF MS experiments

Eighty-six *T. maritima* proteins were overexpressed and purified as described previously [41]. Individual 6 His-tagged isolated proteins were further purified by SDS-PAGE. Coomassie blue-stained protein bands were manually excised from the gel and automatically processed using a Waters MassPREP^{TM} station (Waters Corp., Beverly, MA) using manufacturer-specified protocols. Briefly, gel spots were destained (2× with 50% acetonitrile/50 mM ammonium bicarbonate), dehydrated with acetonitrile, reduced with dithiothreitol, alkylated with iodoacetamide, and digested with sequencing grade trypsin overnight. The resulting peptides were extracted with 30 μL of 1% formic acid and deposited into individual wells on a 96-well microtiter format plate.

2.3. Automated analysis of proteins using MALDI-TOF

Two microliters of each peptide mixture was automatically mixed with 1.5 μL of a 10 mg/mL solution of α-cyano-4-hydroxycinnamic acid and spotted directly onto MALDI target plates. MALDI mass spectra were acquired automatically using a Waters M@LDI-R TOF MS, with ACTH 18–39 as a lock mass reference. Mass spectra were reduced to files containing peak lists using MassLynx^{TM}.

2.4. Automated analysis using MALDI FT-ICR

One microliter of each peptide mixture was manually loaded onto adjacent positions of a Bruker 384-spot 400 μm diameter AnchorPlate^{TM}, followed by 300 nL of a matrix solution (16 mg/mL DHB, 0.1 mg/mL diammonium citrate, 1% trifluoroacetic acid). A calibration mixture containing 5 μg bradykinin, 5 μg substance P, 5 μg neurotensin, 8 μg ACTH 1–17, 8 μg ACTH 18–39, 12 μg melittin, and 12 μg insulin B chain and 918 mg DHB was dissolved in 3 mL of 50% acetonitrile:50% water:0.1% TFA, lyophilized using a SpeedVac (Thermo Savant, Holbrook, NY), crushed with steel beads for several minutes using a vortex mixer, and resuspended in 2 mL hexanes. Approximately 50 μL of this slurry was applied across each side of the MALDI target [34].

The Tcl/Tk scripting capabilities embedded in XMASS^{TM} (Bruker Daltonics, Billerica, MA) were used to create scripts for on-the-fly database searching using a method that optimizes the protein identification score by changing the number of laser shots. A flowchart summarizing the method is shown in Fig. 1. Briefly, a mass spectrum is acquired for a given sample with a user-set minimum number of laser shots (usually 1). The spectrum is then automatically reduced to a set of monoisotopic masses using THRASH [42], and the resulting mass list is automatically submitted for protein identification to Mascot [11] by a C program called by the Tcl/Tk script.

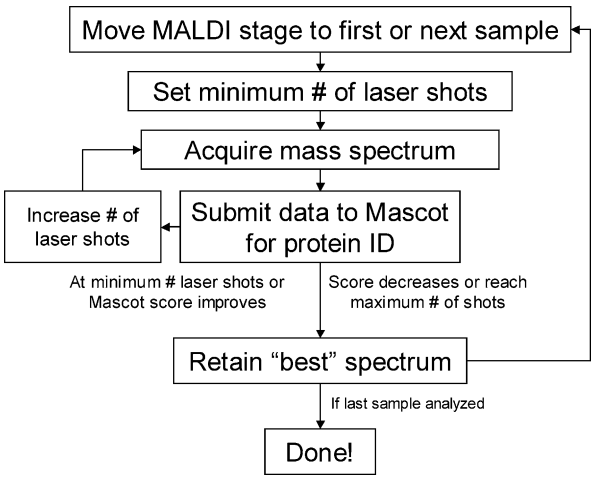

Fig. 1. Schematic for automated peptide mass fingerprinting using FT-ICR MS.

Next, the number of laser shots is doubled and the previous steps are repeated. If the score returned by Mascot is higher for the second search, the number of laser shots is increased again. This repeats until either the Mascot score decreases, generally from the reduction in resolution and mass accuracy due to excessive space charge in the analyzer cell, or until a user-defined maximum number of laser shots is reached (set at 32 for this study), indicating there is little or no protein in the sample. The MALDI stage is then instructed to move to the next sample, the number of laser shots is reset to the minimum value, and the above steps are repeated.

2.5. Protein identification

The peak lists from both instruments were submitted to Mascot for protein identification. All FT-ICR mass spectra were apodized and zerofilled prior to data reduction. The masses for each sample from the MALDI-TOF and MALDI FT-ICR data were searched against both a *T. maritima* database (1846 protein sequences) and the July 18, 2003 NCBInr database (1,472,604 sequences). All searches allowed up to two missed tryptic cleavages, and included fixed carbamidomethyl modification of cysteine and variable oxidation of methionine. Further searches were performed with an additional variable carbamyl modification of lysine. A Mascot search that returns the expected overexpressed protein as the highest score is determined to be correctly identified, regardless of whether or not the score is above or below the 95% confidence threshold returned by Mascot.

3. Results

3.1. Results of search using the T. maritima database

The results of the searches using the *T. maritima* database (1846 entries) are summarized in Table 1. Automated

Table 1
Number of identified proteins for FT-ICR (10 ppm) and TOF (50 ppm) data

Database	Variable modifications	Instrument	No. correct	Average score	Average no. of peptides	Average top random score	No. of random peptides
T. maritima	Oxidation (M)	TOF	84	155.5	19.6	35.5	7.0
	Oxidation (M)	FT-ICR	86	148.8	15.8	17.7	2.6
	Oxidation (M) + carbamyl (K)	TOF	77	120.9	21.9	37.5	10.4
	Oxidation (M) + carbamyl (K)	FT-ICR	84	107.3	16.2	16.6	3.4
NCBInr	Oxidation (M)	TOF	73	172.0	21.2	64.8	8.6
	Oxidation (M)	FT-ICR	82	154.7	16.3	39.6	3.7

MALDI FT-ICR MS was able to correctly identify all 86 proteins using 10 ppm peptide tolerance, while automated MALDI-TOF MS failed to identify two of the proteins using 50 ppm peptide tolerance. An unsuccessful lockmass correction leading to 200 ppm mass measurement errors was the cause of one of the incorrect identifications, and resubmission of the data to Mascot with 300 ppm peptide tolerance successfully identified the protein. The other unidentified protein was the 8.5 kDa conserved hypothetical protein (gi|4982136). As seen in Fig. 2, both instruments detected the same four peptides from the expected protein. It might be concluded that the greatly improved mass accuracy (∼1 ppm rms) from the internally-calibrated MALDI FT-ICR MS data versus that of the lockmass-corrected MALDI-TOF data (∼18 ppm rms) is the reason for the successful identification using the MALDI FT-ICR instrument. However, resubmission of the MALDI FT-ICR data to Mascot using 50 ppm peptide tolerance still returned the correct protein as the highest scoring hit. Upon further review of the data, it was found that the MALDI-TOF spectrum contained 40 more unassigned masses than the MALDI FT-ICR data. These additional peaks resulted in the assignment of another protein from the database rather than the expected protein, which was assigned by Mascot as the second highest scoring match.

3.2. Results of search using the T. maritima database with variable carbamyl modification

The data was resubmitted to Mascot using the *T. maritima* database allowing for the variable carbamylation of lysine (see Table 1), which can occur during proteolysis at elevated temperatures when using urea as the denaturant [18]. While the Mascot scores for all samples decreased despite the fact that the average sequence coverage of the correctly identified proteins increased, the scores of the average highest ranked random protein matches remained roughly the same. For the MALDI FT-ICR MS data, two proteins were no longer correctly identified due to the reduction of their Mascot scores

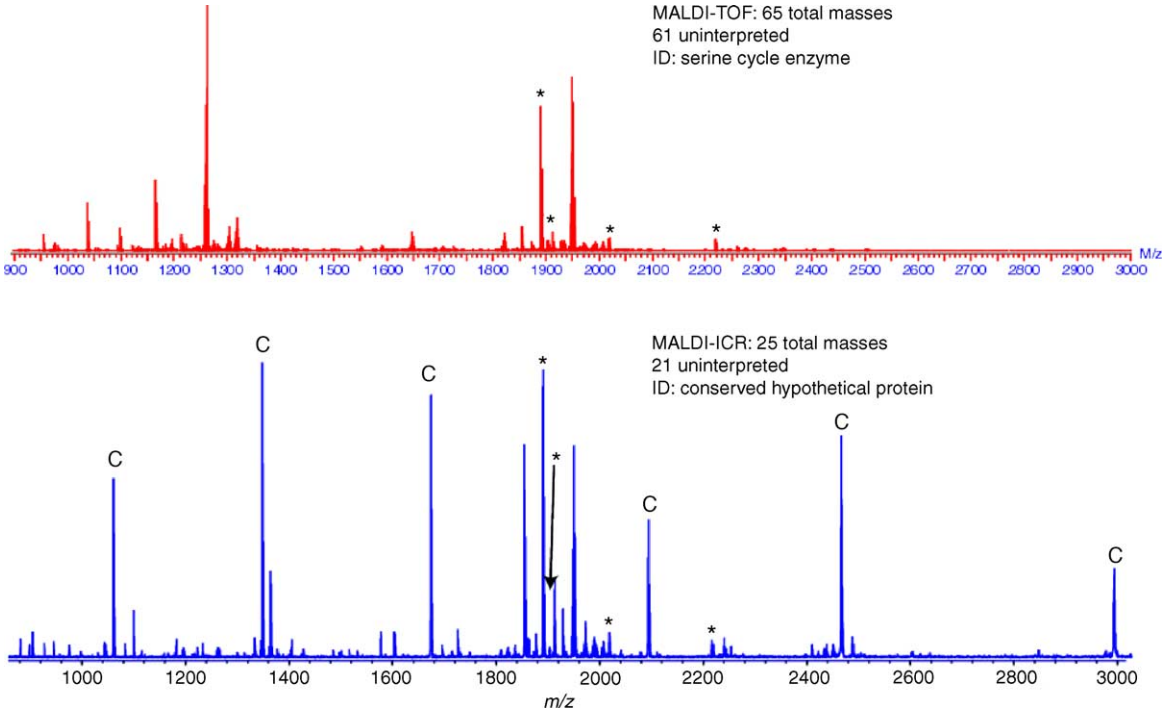

Fig. 2. Comparison of peptide mass fingerprint spectra for same sample using (a) MALDI-TOF and (b) MALDI FT-ICR MS. Peaks marked with "C" indicate internal calibrants.

below those of other random protein matches. These proteins returned scores of 24 and 37 in the original *T. maritima* search. With the addition of the variable modification, the scores for these proteins decreased below those from other proteins in the database. For the TOF data, seven additional proteins were no longer correctly identified. These had an average Mascot score of 47 with nine peptides matched in the original *T. maritima* search, but their scores decreased below those of other proteins in the database when variable carbamylation was selected.

3.3. Results of search using the NCBInr database

Finally, the data for both instruments was submitted to Mascot and searched against the full NCBInr database. The MALDI FT-ICR MS system was able to correctly identify 82 of the 86 proteins, while the TOF system was able to identify 73. For the FT-ICR MS data, the four unidentified proteins returned top scores of 24, 31, 37, and 37 in the *T. maritima* search, but were obscured by random matches from within the larger database. All of the proteins that were not correctly identified using the TOF MS data had scores of 68 or below in the original *T. maritima* searches.

4. Discussion

4.1. Effect of mass accuracy on highest ranking random match

The primary effect of improving mass accuracy in PMF is the greatly decreased score and sequence coverage obtained for the highest ranking random match (HRRM), defined here to be the highest scoring incorrectly assigned protein returned from the database search. As can be seen in Table 1, the average HRRM score was much larger under all conditions for the MALDI-TOF MS data than for the MALDI FT-ICR MS data. For the *T. maritima* database searches, the average HRRM for the TOF MS data was assigned a score and number of matching peptides that were both twice as large as those found for the FT-ICR MS data. With the addition of the variable carbamyl modification, the average number of matching peptides in the HRRM was over 10 for the TOF data, while it remained at 3.4 for the FT-ICR MS data. This shows that higher mass accuracy measurements are far less affected by a significant increase in database complexity. Furthermore, the number of peptides matched for the highest scoring hit from the FT-ICR MS data increased by an average of 0.4 per protein versus 2.3 for the TOF MS data, implying that these additional peptide assignments from the TOF MS data are likely to be random rather than true matches. For the NCBInr database searches, the HRRM averaged only 3.7 assigned peptides and a score of 39.8 for the FT-ICR MS data versus 8.7 peptides and a score of 64.6 for the TOF MS data. Even though the average sequence coverages and Mascot scores of the correct proteins returned for the MALDI-TOF MS data were slightly higher than those from the FT-ICR MS data, the greatly decreased scores for the HRRM using FT-ICR MS more than compensates and is the primary reason that MALDI FT-ICR MS identifies more proteins than the MALDI-TOF system.

Table 2 shows a more complete set of statistics for the searches of both sets of data against the NCBInr database. These data give a good indication of the mass accuracy of both instruments, with the highest number of proteins identified using 10 ppm peptide tolerance for the MALDI FT-ICR MS data and 50 ppm peptide tolerance for the MALDI-TOF MS data. Interestingly, most of the proteins are still identified with the FT-ICR MS data using only a 1 ppm peptide tolerance due to the significant proportion of the mass measurements that fall within this value (∼50%) as well as the substantially reduced HRRM scores. As the peptide tolerance increases, the average HRRM score appears to approach a limit that is roughly equivalent to the Mascot-defined 95% reliability threshold computed to be 74 for the NCBInr database. However, as can be seen in the 1 ppm peptide tolerance searches of the FT-ICR MS data, the average identified protein returns a score that is significantly less (∼62) than this value. Nevertheless, these proteins are easily identified due to the very low HRRM scores. Since the true requirement for a positive identification is simply that the correct protein returns a higher score than any potential random assignment, this constant reliability threshold employed by Mascot is unrealistic, especially for data sets with very high mass measurement accuracy. There are clearly many factors such as database size, mass accuracy, number of species found in a given mass spectrum, and inclusion of variable modifications that must be considered in order to provide a realistic threshold for the acceptance of a given protein assignment.

4.2. Unassignable masses

As illustrated in Fig. 2, the number of unassignable species in a mass spectrum can also affect the accuracy of protein identification by PMF. While the spectra obtained by MALDI-TOF MS contained on average 67 masses, those obtained using MALDI FT-ICR MS contained an average of only 40 masses. In the search against *T. maritima*, this larger number of masses in the TOF MS data results in ∼4 more matching peptides per successfully identified protein corresponding to 6% greater overall sequence coverage compared to that observed for the FT-ICR MS data. However, 72% of the masses in the TOF MS data cannot be assigned to any peptide sequence that should result from the expected protein compared to 62% of the masses for the FT-ICR MS data. Thus, the additional signals observed using TOF MS combined with its poorer mass accuracy resulted in random protein matches with higher sequence coverages than those seen using FT-ICR MS.

A comparison of average HRRM scores for several peptide tolerances with both the TOF and FT-ICR MS data is seen in Table 2. Surprisingly, the average background scores are roughly the same for the two instruments, and thus Mascot

Table 2
Average top score for random matches using FT-ICR and TOF data with NCBInr database

ppm error	FT-ICR					TOF				
	No. correct ID	Average correct ID score	No. of peptides/correct ID	Average HRRM score	No. peptides/HRRM	No. correct ID	Average correct ID score	No. of peptides/correct ID	Average HRRM score	No. peptides/HRRM
1	72	62.6	8.4	21.6 ± 3.5	2.2 ± 0.4	1	33.0	6.0	19.0 ± 4.2	2.0 ± 0.5
5	80	110.1	12.4	29.6 ± 7.9	2.7 ± 0.8	25	57.1	9.8	29.9 ± 6.4	3.0 ± 1.1
10	82	154.7	16.3	39.6 ± 6.0	3.7 ± 1.0	44	94.3	14.0	37.6 ± 9.1	3.6 ± 1.2
30	77	164.9	17.4	55.7 ± 7.6	6.1 ± 2.6	71	154.6	19.4	53.9 ± 8.8	6.7 ± 2.4
50	74	169.4	18.1	64.8 ± 10.7	8.0 ± 3.4	73	172.0	21.2	64.8 ± 9.6	8.6 ± 4.6
100	73	164	18.7	72.6 ± 9.8	10.2 ± 5.2	65	186.4	23.6	76.3 ± 9.1	14.5 ± 8.2
200	73	139.5	18.6	63.5 ± 10.3	9.3 ± 4.9	67	158.8	23.5	68.8 ± 8.7	12.2 ± 5.9

appears to adjust scores of potential identifications downwards in response to an increasing number of unassigned masses. However, the average number of assigned peptides for the HRRMs and the standard deviation of their scores are substantially larger for the TOF data. Thus, these additional masses do indeed increase the probability of an unexpected protein in the database to be returned as the highest scoring potential match.

4.3. Samples requiring high mass accuracy

From the data in Tables 1 and 2, it is clear that higher mass accuracy measurements are advantageous for searches involving variable modifications and large protein databases. As the mass accuracy decreases, the chance of the HRRM scoring higher than the expected protein increases significantly. This trend is particularly pronounced for proteins under 20 kDa that may produce too few peptides upon proteolysis to enable an unambiguous identification using lower mass accuracy data. In the NCBInr searches, eight proteins that were identified using FT-ICR MS but not by TOF MS have molecular weights of less than 20 kDa, with two of these being below 10 kDa. These samples returned an average of six assignable peptides covering ∼40% of the protein sequence upon searching, which is comparable to the coverage observed with larger proteins. However, due to the limited number of peptides, the 50 ppm mass tolerance used in the searches with the MALDI-TOF data led to an average HRRM with almost nine peptides compared to less than four peptides for the FT-ICR data at 10 ppm. The other proteins not identified by MALDI-TOF MS are larger proteins (∼30 kDa) that likely did not digest efficiently under the conditions employed, and therefore yielded low sequence coverage.

Another situation where high mass accuracy measurements prove to be critical involves samples that contain more than one protein. Digestion of several proteins simultaneously significantly increases the number of masses in the resulting spectrum, leading to an increase in the score of the HRRM. In order to determine the effect that an increasing number of masses in a spectrum has on the ability to identify an individual protein, a plot of the number of peptides assigned to the HRRM is compared to the number of masses

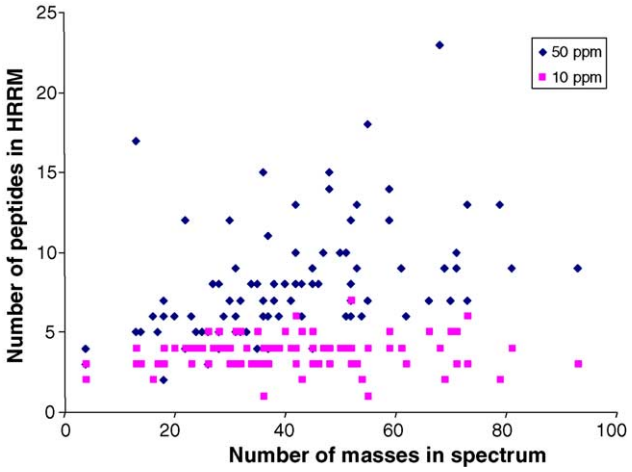

Fig. 3. Plot of peptide sequence coverage of HRRM vs. number of masses in spectra using 10 and 50 ppm mass tolerance.

observed in each spectrum. Fig. 3 shows a general trend of an increasing number of peptides assigned for the HRRM as the number of masses in a spectrum increases, with this effect being more pronounced at 50 ppm mass accuracy than at 10 ppm. While it is difficult to extrapolate to the hundreds of masses that would be expected from the digestion of multiple proteins in a single sample, it is clear that an instrument with higher mass accuracy is better suited to handle such complexity. Preliminary results show that our MALDI FT-ICR instrument can identify up to four proteins per band from a sample of mouse proteins separated on a 2D gel (data not shown), but more work is required to fully understand the maximum number of proteins that can readily be identified from such samples.

4.4. Performance of automated FT-ICR MS system

In order to successfully implement automated PMF on a given mass spectrometer, two difficulties must be overcome. The first issue arises from the presence of “sweet spots” routinely encountered using MALDI MS. For TOF instruments, the area irradiated by the laser is generally much smaller than the sample spot in order to minimize the amount of desorbed charge. This prevents both saturation of the microchannel

plate detectors and also the loss of resolution and mass accuracy due to space charge effects in the ion source region. Unfortunately, the signals obtained vary greatly across the dimensions of the crystallized sample [43] and, as a result, numerous regions of the sample must be queried to increase the probability that a location yielding strong signal is analyzed. The second challenge involves accounting for the potentially large range of sample concentrations. Current automated systems are designed to either change the number of laser shots or alter the laser power [44] until a desired total signal level is obtained. While these methods obtain a mass spectrum with an optimal balance of resolution, mass accuracy, and number of masses observed, this by no means guarantees that the highest quality peptide mass finger print is acquired. Rather, it is more likely that a mass spectrum with either more or less signal could result in a higher score from a database query due to either an increase in sequence coverage or a reduction of the number of unassigned masses respectively.

The described automated PMF system employs novel methods to address these challenges. We have already shown that the sweet spot issue can be mitigated by a combination of high laser power, a large area of illumination, and the use of hydrophobic/hydrophilic surface patterned plates [34]. In addition, this system also automatically adjusts the number of laser shots to account for potential variations in sample concentration. However, unlike other automated PMF systems, the actual Mascot score generated from the acquired data is used to gauge the quality of the spectrum rather than signal intensity. This effectively balances protein sequence coverage, the number of unassigned masses, and mass accuracy, resulting in the best possible peptide mass finger print rather than simply the "best looking" mass spectrum. This method also successfully accounts for the widely differing concentrations of the *T. maritima* samples, as the number of laser shots required to return the optimal Mascot scores for the samples spanned the full range (1–32) allowed for this experiment. There were more spectra that maximized at one laser shot (27 of 86 total) than at any other number, indicating that these samples were highly concentrated. A decrease in laser power and a corresponding increase in the maximum number of laser shots could even further improve the results from these samples, although they already yielded sufficiently high Mascot scores for unambiguous identification.

Data acquisition times ranged from 30 s to 2 min per sample for this system, depending on the database used as well as the number of method iterations performed. The database searches required the majority of this time, especially for larger databases. Although the throughput of this MALDI FT-ICR MS based system may appear to be a little lower than that claimed for other automated PMF systems, in reality, this system substantially improves the performance of a complete proteomics platform. Each additional protein that this system identifies saves roughly 1 h due to the elimination of the extra ESI-LC/MS/MS analysis of the unidentified samples. Further, the intrinsic MS/MS capabilities of FT-ICR MS, although not used here, could be used to identify proteins that are not successfully identified using PMF, also precluding the need for ESI-LC/MS/MS. Thus, this system should save a substantial amount of time where peptide mass fingerprinting is a vital part of a proteomics platform, such as those that use 2D gel electrophoresis to separate complex protein mixtures.

4.5. A note about root-mean-squared (rms) ppm mass accuracy

The optimal mass tolerances found for both instruments (10 ppm for MALDI FT-ICR MS and 50 ppm for MALDI-TOF MS) seem rather high considering the mass accuracies generally attributed to FT-ICR MS (∼1–2 ppm rms) and TOF MS (∼10–20 ppm rms). However, most reports convey the mass accuracy of their instruments as rms ppm error rather than absolute mass tolerance. For the FT-ICR MS data presented here, the average rms ppm error was 1.7 using 10 ppm mass tolerance in the database search, while for the TOF MS data the average rms ppm error was 17 ppm using 50 ppm tolerance. Therefore, the mass accuracies reported here are consistent with the mass accuracies often reported in the literature. However, a drop in mass tolerance in the database search to 5 ppm using the FT-ICR data led to the loss of two correct protein identifications, a drop in the average assigned score of ∼10, and a decrease in the average number of assigned peptides from 16.3 to 12.4 peptides. Since these data show that ∼25% of all data returned mass measurement errors between 5 and 10 ppm, the distribution of mass errors is clearly not Gaussian. Furthermore, when the mass tolerance in the database search using the TOF MS data was decreased to 30 ppm, two correct protein identifications were lost, the average score of correctly identified proteins dropped 18, and the average number of assigned peptides for the top scoring hit dropped from 21.2 to 19.4. Thus, ∼10% of all mass measurement errors are between 30 and 50 ppm, indicating again a non-Gaussian distribution. Without knowledge of the true distribution of mass errors on an instrument, it is difficult to determine the optimal absolute mass tolerance that should be employed for database searches. Thus, rms ppm mass measurement accuracy does not truly indicate the mass accuracy that can be confidently used to describe an instrument's performance in a proteomics experiment.

5. Conclusions

It is clear that the better mass accuracy of MALDI FT-ICR MS provides substantially higher confidence in protein identifications by PMF. While FT-ICR MS has not yet gained widespread use in PMF, the mass accuracy and dynamic range of this instrument are well suited for this technique. This work shows that it is possible to create an automated system based on FT-ICR MS and demonstrates the promise of FT-ICR MS for the routine analysis of samples extracted from gels.

References

[1] K. Gevaert, J. Vandekerckhove, Electrophoresis 21 (2000) 1145.
[2] T. Keough, M.P. Lacey, A.M. Fieno, R.A. Grant, Y. Sun, M.D. Bauer, K.B. Begley, Electrophoresis 21 (2000) 2252.
[3] M.P. Washburn, D. Wolters, J.R. Yates III, Nat. Biotechnol. 19 (2001) 242.
[4] W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, Electrophoresis 90 (1993) 5011.
[5] P. James, M. Quadroni, E. Carafoli, G. Gonnet, Biochem. Biophys. Res. Commun. 195 (1993) 58.
[6] M. Mann, P. Hojrup, P. Roepstorff, Biol. Mass Spectrom. 22 (1993) 338.
[7] D.J.C. Pappin, P. Hojrup, A. Bleasby, Curr. Opin. Biotechnol. 3 (1993) 327.
[8] J.R. Yates III, S. Speicher, P.R. Griffin, T. Hunkapiller, Anal. Biochem. 214 (1993) 397.
[9] H.W. Lahm, H. Langen, Electrophoresis 21 (2000) 2105.
[10] J.K. Eng, A.L. McCormack, J.R. Yates III, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[11] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.
[12] H.I. Field, D. Fenyo, R.C. Beavis, Proteomics 2 (2002) 36.
[13] J.A. Karty, M.M. Ireland, Y.V. Brun, J.P. Reilly, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 782 (2002) 363.
[14] M. Karas, F. Hillenkamp, Anal. Chem. 60 (1988) 2299.
[15] J.A. Karty, M.M. Ireland, Y.V. Brun, J.P. Reilly, J. Proteome Res. 1 (2002) 325.
[16] A. Gattiker, W.V. Bienvenut, A. Bairoch, E. Gasteiger, Proteomics 2 (2002) 1435.
[17] Q. Ding, L. Xiao, S. Xiong, Y. Jia, H. Que, Y. Guo, S. Liu, Proteomics 3 (2003) 1313.
[18] J. McCarthy, F. Hopwood, D. Oxley, M. Laver, A. Castagna, P.G. Righetti, K. Williams, B. Herbert, J. Proteome Res. 2 (2003) 239.
[19] J. Rappsilber, M. Moniatte, M.L. Nielsen, A.V. Podtelejnikov, M. Mann, Int. J. Mass Spectrom. 226 (2003) 223.
[20] S. Gu, S. Pan, E.M. Bradbury, X. Chen, Anal. Chem. 74 (2002) 5774.
[21] J.M. Pratt, D.H. Robertson, S.J. Gaskell, I. Riba-Garcia, S.J. Hubbard, K. Sidhu, S.G. Oliver, P. Butler, A. Hayes, J. Petty, R.J. Beynon, Proteomics 2 (2002) 157.
[22] T.C. Hunter, L. Yang, H. Zhu, V. Majidi, E.M. Bradbury, X. Chen, Anal. Chem. 73 (2001) 4891.
[23] E.C. Peters, D.M. Horn, D.C. Tully, A. Brock, Rapid Commun. Mass Spectrom. 15 (2001) 2387.
[24] S. Corless, R. Cramer, Rapid Commun. Mass Spectrom. 17 (2003) 1212.
[25] W.V. Bienvenut, C. Hoogland, A. Greco, M. Heller, E. Gasteiger, R.D. Appel, J.J. Diaz, J.C. Sanchez, D.F. Hochstrasser, Rapid Commun. Mass Spectrom. 16 (2002) 616.
[26] F.L. Brancia, S.G. Oliver, S.J. Gaskell, Rapid Commun. Mass Spectrom. 14 (2000) 2070.
[27] J.E. Hale, J.P. Butler, M.D. Knierman, G.W. Becker, Anal. Biochem. 287 (2000) 110.
[28] T. Keough, M.P. Lacey, R.S. Youngquist, Rapid Commun. Mass Spectrom. 14 (2000) 2348.
[29] R.L. Beardsley, J.P. Reilly, Anal. Chem. 74 (2002) 1884.
[30] K.R. Clauser, P. Baker, A.L. Burlingame, Anal. Chem. 71 (1999) 2871.
[31] M.K. Green, M.V. Johnston, B.S. Larsen, Anal. Biochem. 275 (1999) 39.
[32] A.G. Marshall, C.L. Hendrickson, S.D. Shi, Anal. Chem. 74 (2002) 252A.
[33] A.G. Marshall, C.L. Hendrickson, G.S. Jackson, Mass Spectrom. Rev. 17 (1998) 1.
[34] A. Brock, D.M. Horn, E.C. Peters, C.M. Shaw, C. Ericson, Q.T. Phung, A.R. Salomon, Anal. Chem. 75 (2003) 3419.
[35] M. Palmblad, M. Wetterhall, K. Markides, P. Hakansson, J. Bergquist, Rapid Commun. Mass Spectrom. 14 (2000) 1029.
[36] M. Wetterhall, M. Palmblad, P. Hakansson, K.E. Markides, J. Bergquist, J. Proteome Res. 1 (2002) 361.
[37] M. Ramstrom, M. Palmblad, K.E. Markides, P. Hakansson, J. Bergquist, Proteomics 3 (2003) 184.
[38] J. Bergquist, M. Palmblad, M. Wetterhall, P. Hakansson, K.E. Markides, Mass Spectrom. Rev. 21 (2002) 2.
[39] E. Damoc, N. Youhnovski, D. Crettaz, J.D. Tissot, M. Przybylski, Proteomics 3 (2003) 1425.
[40] M. Witt, J. Fuchser, G. Baykut, J. Am. Soc. Mass Spectrom. 14 (2003) 553.
[41] S.A. Lesley, P. Kuhn, A. Godzik, A.M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H.E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L.S. Brinen, M.D. Miller, T.M. McPhillips, M.A. Miller, D. Scheibe, J.M. Canaves, C. Guda, L. Jaroszewski, T.L. Selby, M.A. Elsliger, J. Wooley, S.S. Taylor, K.O. Hodgson, I.A. Wilson, P.G. Schultz, R.C. Stevens, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 11664.
[42] D.M. Horn, R.A. Zubarev, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 11 (2000) 320.
[43] R.W. Garden, J.V. Sweedler, Anal. Chem. 72 (2000) 30.
[44] O.N. Jensen, P. Mortensen, O. Vorm, M. Mann, Anal. Chem. 69 (1997) 1706.