Identification of 2D-Gel Proteins: A Comparison of MALDI/TOF Peptide Mass Mapping to µ LC-ESI Tandem Mass Spectrometry

Hanjo Lim,* Jimmy Eng,† and John R. Yates, III
Department of Cell Biology, The Scripps Research Institute, La Jolla, California, USA

Sandra L. Tollaksen and Carol S. Giometti
Biosciences Division, Argonne Laboratory, Argonne, Illinois, USA

James F. Holden and Michael W. W. Adams
Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia, USA

Claudia I. Reich and Gary J. Olsen
Department of Microbiology, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

Lara G. Hays
Diversa, Inc., San Diego, California, USA

A comparative analysis of protein identification for a total of 162 protein spots separated by two-dimensional gel electrophoresis from two fully sequenced archaea, *Methanococcus jannaschii* and *Pyrococcus furiosus*, using MALDI-TOF peptide mass mapping (PMM) and µLC-MS/MS is presented. 100% of the gel spots analyzed were successfully matched to the predicted proteins in the two corresponding open reading frame databases by µLC-MS/MS while 97% of them were identified by MALDI-TOF PMM. The high success rate from the PMM resulted from sample desalting/concentrating with ZipTip$_{C18}$ and optimization of several PMM search parameters including a 25 ppm average mass tolerance and the application of two different protein molecular weight search windows. By using this strategy, low-molecular weight (<23 kDa) proteins could be identified unambiguously with less than 5 peptide matches. Nine percent of spots were identified as containing multiple proteins. By using µLC-MS/MS, 50% of the spots analyzed were identified as containing multiple proteins. µLC-MS/MS demonstrated better protein sequence coverage than MALDI-TOF PMM over the entire mass range of proteins identified. MALDI-TOF and PMM produced unique peptide molecular weight matches that were not identified by µLC-MS/MS. By incorporating amino acid sequence modifications into database searches, combined sequence coverage obtained from these two complimentary ionization methods exceeded 50% for ~70% of the 162 spots analyzed. This improved sequence coverage in combination with enzymatic digestions of different specificity is proposed as a method for analysis of post-translational modification from 2D-gel separated proteins. (J Am Soc Mass Spectrom 2003, 14, 957–970) © 2003 American Society for Mass Spectrometry

Since the introduction of matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), mass spectrometry has revolutionized the structural analysis of biomolecules [1, 2]. For example, tandem mass spectrometry (MS/MS) is now routinely used for protein sequencing instead of the traditional Edman degradation method because of its flexibility, sensitivity, speed, reliability, and accu-

racy. Biological mass spectrometry led to a birth of proteomics by enabling large-scale protein analysis (for review, see references [3, 4]). The combination of high-resolution protein separation by two-dimensional gel electrophoresis (2DE) and mass spectrometry has proven to be an essential tool for proteomics to identify proteins [5–7], and post-translational modifications (PTM) [8, 9]. New methods employing on-line multidimensional liquid chromatography separations of protein or peptide mixtures greatly extended the breadth and depth of proteome analysis relative to those performed solely by liquid chromatography and mass spectrometry [10–16]. In addition to qualitative protein characterization, measurement of relative and absolute protein expression between two different sample states is also important. Relative quantification can be obtained by imaging the intensity of fluorescent dye-labeled [17] or stained proteins separated by 2DE. Methods to measure relative expression of proteins labeled with stable isotopes have emerged that create m/z differences for peptides and proteins that can be measured in the mass spectrometer. Stable isotopes can be incorporated into proteins by metabolic, covalent, or enzymatic labeling and expression ratios are measured by comparing peak areas for the protein or peptide ions measured in the mass spectrometer [18–23]. The role of mass spectrometry is important in almost all areas of proteomics.

Protein identification is an essential step to understand the function and roles of proteins in the cell. Although new methods using multidimensional liquid chromatography to identify proteins in mixtures have appeared, protein separation by 2DE and subsequent protein identification by mass spectrometry is a widely used strategy in proteomics. Among the available mass spectrometric methods for protein identification of proteins separated by 2DE proteins, peptide mass mapping (PMM) by matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS) [24] (hereafter PMM for MALDI-TOF PMM) and electrospray tandem mass spectrometry (ESI MS/MS) [25] have been almost exclusively used for this purpose. An initial screen by PMM and subsequent sequence analysis of unidentified proteins from the first analysis by MS/MS is a common strategy for identification of large numbers of 2D-gel separated proteins [5]. In particular, on-line micro capillary reversed phase liquid chromatography interfaced to a tandem mass spectrometer (μLC-MS/MS) (hereafter LC-MS/MS for μLC-MS/MS) and data-dependent MS/MS acquisition [26] has made the second stage of MS/MS analysis more comprehensive and higher throughput.

Although proteomes of many different organisms have been analyzed using 2DE/mass spectrometry [5–7], there has not been a comparative study on the two different approaches for protein identification of 2D-gel separated proteins. Recently, Burlingame and coworkers compared the MALDI and ESI methods for peptide analysis in the identification of proteins isolated from rabbit intestine and purified by one-dimensional sodium dodecylpolyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie staining [27]. Two proteins were identified by MALDI-TOF post source decay (PSD) and two additional proteins were identified by nanospray MS/MS from an HPLC fraction generated by off-line reversed phase HPLC. The two approaches for protein identification were found to be complementary in the analysis of a single band from SDS-PAGE. However, the methods employed were not commonly used for the identification of 2D-gel proteins, and the comparison was conducted with a band from a 1D gel. In this paper, we report a comparative analysis of protein identification of proteins separated by 2DE from two fully sequenced archaea, *Methanococcus jannaschii* and *Pyrococcus furiosus,* using MALDI-TOF MS and LC-MS/MS to perform PMM and MS/MS database searching, respectively. More than 160 protein spots isolated by using 2DE from the two organisms were used for the comparison. Since the catalog of the proteome of the two organisms has been or will be reported in separate publications [28], this study compared several aspects of the two methods including the number of matched peptides, sequence coverage, effect of modifications for protein identification and sequence coverage, and the utility of combining the two methods. Also an in-house developed PMM search engine PRO-QUEST will be introduced and its optimized use will be described.

Experimental and Methods

Materials

Deionized water from a Milli-Q ultrapure water system (Bedford, MA), HPLC grade acetonitrile from Fisher Scientific (Fair Lawn, NJ), and glacial acetic acid from J. T. Baker (Phillipsburg, NJ) were used for HPLC. Modified trypsins for in-gel digestion were purchased from Promega (Madison, WI) and Boeringer Mannheim (Mannheim, Germany). Ammonium bicarbonate, dithiothreitol (DTT), iodoacetamide (IAA), α-cyano-4-hydroxy cinnamic acid (α-CHCA) were obtained from Sigma (St. Louis, MO) and used without further purification. Formic acid was obtained from Aldrich (Milwaukee, IL). Tris(2-carboxyethyl)-phosphine (TCEP) hydrochloride from Pierce (Rockford, IL) was also used for reduction during in-gel digestion. *M. jannaschii* and *P. furiosus* cells were grown in the laboratories of Gary Olson (University of Illinois) and Michael Adams (University of Georgia), respectively. The proteins were separated by 2DE in the laboratory of Carol Giometti at Argonne National Laboratory.

Sample Preparation and 2DE of Archaeal Proteins

Whole cell extracts, soluble fractions, and membrane fractions from cells grown in minimal nutrient media were mixed with equal volumes of a solution contain-

ing 9M urea, 2% 2-mercaptoethanol, 2% ampholytes (pH 8–10, BioRad), and 2% Nonidet P40 (a nonionic detergent). The soluble, denatured proteins were recovered by centrifugation at 435,000 × g for 10 min using a Beckman TL100 tabletop ultracentrifuge. Protein (400 μg for Coomassie Blue staining of whole cell extracts and soluble fractions and 200 μg for membrane fractions) was loaded onto isoelectric focusing gels containing 50% pH 5–7 with 50% pH 3–10 (*M. jannaschii*) or 12% pH 3–10 and 88% pH 5–7 (*P. furiosus*) ampholytes [29]. After 14,000 V-hours, the gels were equilibrated with SDS and the proteins were separated by SDS-PAGE using a linear gradient of 10–17% acrylamide [30]. Proteins were then detected by staining with Coomassie Blue R250 [31].

In-Gel Trypsin Digestion

Spots were collected from 2–4 replicate gels to insure sufficient amount of sample available for both PMM and LC-MS/MS analysis. A slightly modified procedure that was originally developed by Shevchenko et al. [32] was used for in-gel digestion. Briefly, Coomassie-stained spots were destained and washed with 100 mM ammonium bicarbonate and acetonitrile, reduced with either TCEP [33] at room temperature for 20 min or DTT at 60 °C for 40 min., and then alkylated by IAA in a dark place for 30 min. The gel was incubated in 50 μL of a 12 ng/μl modified trypsin solution in 50 mM ammonium bicarbonate, pH 8.6, and incubated at 37 °C overnight. The resulting peptides were extracted first with a 1:1 solution of 25 mM ammonium bicarbonate and acetonitrile and then twice with a 1:1 solution of 5% formic acid and acetonitrile. The extracted tryptic peptides were lyophilized and resuspended with 15–20 μl of 5% formic acid for mass spectrometric analysis. Approximately one-third each of the resuspended tryptic digests was used for PMM and LC-MS/MS, respectively, while the remaining one-third was reserved for another analysis in case the first analyses failed. Based upon the relative intensities of peaks from samples and externally added internal standards (~500 femtomoles/each) in the obtained MALDI spectra, the sample quantity used for each mass spectrometric analysis was estimated to be in the high femtomole to low picomole range.

Mass Spectrometry

MALDI-TOF MS. All samples were desalted and concentrated with a 10 μL ZipTip$_{C18}$ (Millipore, Bedford, MA) [34], following the instructions provided by the manufacturer. Peptides were eluted in a volume of 1.5 μl using a concentrated solution of α-CHCA in 50% acetonitrile and 0.3% trifluoroacetic acid in water and deposited onto the MALDI target plate. Before the sample/matrix solution dried, 0.5 μl (~500 femtomoles/each calibrant) of calibration mixture 2 of the Sequazyme peptide standard kit (Applied Biosystems,

Framingham, MA) was added on top of the spot as internal calibrants. The calibration mixture consisted of angiotensin I, ACTH 1–17 clip, ACTH 18–39 clip, ACTH 7–38 clip, and bovine insulin. Once the spots dried on the target plate, the plate was introduced into the Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems, Framingham, MA) for analysis. The MALDI-generated ions were extracted with a 135 ns delay and accelerated to 25 kV. The TOF was operated in the reflectron mode. Each spectrum was an average of 64–128 laser shots, depending on the signal-to-noise (S/N) ratio reflected on the oscilloscope. The resulting mass spectra were calibrated by a two-point internal calibration with one of the trypsin autolysis peaks appearing at *m/z* 842.5100 and the ACTH 18–39 clip peak at *m/z* 2465.1989 in the calibration mixture solution. Calibrated spectra were submitted to database searches using the PROQUEST peptide mass mapping software. Analysis of 15–20 samples per day was achieved from sample cleanup by ZipTip$_{C18}$ to manual inspection of each spectra to searches and researches with PROQUEST.

μLC-MS/MS. The tryptic peptide samples were separated and analyzed using a LCQ classic ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) using a homemade fritless capillary column as described previously [35]. Briefly, a 365 × 100 μm fused silica tubing (Polymicro Technologies, Phoenix, TX) was pulled with a P-2000 laser puller (Sutter Instrument, Novato, CA) to create a 5 micron tip, and then packed with POROS 10 R2 10 μm hydrophobic packing material (Applied Biosystems, Framingham, MA) to a bed length of 10 to 15 cm using a high-pressure helium vessel. Samples were then loaded onto the column by placing an eppendorf tube containing the sample solution into a high pressure vessel, sealing the vessel, and then inserting the blunt end of the column through a swagelock fitting into the eppendorf tube. The high-pressure vessel is then pressurized to 400–500 psi to force the sample solution onto the column. To minimize column clogging, the resuspended tryptic digest solution was centrifuged first, and only the supernatant was forced onto the column. The bound tryptic peptides were then separated by a 30-min linear gradient of 0–60% buffer B, where buffer A was 0.5% acetic acid in water and buffer B was 80% acetonitrile and 0.5% acetic acid in water. The flow rate was reduced from 150 μl/min to ~300 nl/min by using a 75 μm pre-column restriction capillary tubing to split the flow. An HP 1100 pump (Agilent Technologies, Palo Alto, CA) was used to create the flow and gradient. A short 15-min gradient of 0–80% buffer B was used to remove peptides left on the column from the previous run. Each column was used for approximately 50 analyses.

Data-dependent MS/MS was employed to generate tandem mass spectra during LC analysis. The top three most intense ions were selected from the full MS scan. A three-min dynamic exclusion was also applied to min-

imize acquisition of redundant MS/MS data. Tandem mass spectra were directly submitted to SEQUEST-PVM [36, 37] searches for protein identification. Approximately eight samples were analyzed per day by manual LC-MS/MS from sample loading onto a column to column washing between sample runs by 15 min short gradient to SEQUEST-PVM database searches.

Database searching.   Open reading frame (ORF) databases of *M. jannaschii* (1.7 Mbp) and *P. furiosus* (2.0 Mbp) were downloaded in FASTA format via file transfer protocol (FTP) from the website of The Institute for Genome Research (TIGR) and from the Center of Marine Biotechnology (COMB) at University of Maryland, respectively. These databases were stored locally for database searches. The *M. jannaschii* database was fully annotated at the time of data collection while that of *P. furiosus* was not.

Results and Discussion

MALDI-TOF Peptide Mass Mapping by PROQUEST

A total of 162 spots from *M. jannaschii* and *P. furiosus*, 100 from *M. jannaschii* and 62 from *P. furiosus*, were analyzed using MALDI-TOF MS. PROQUEST, an in-house developed web-based PMM database search program, was used to search the data. Default search parameters used in this study were: (1) Protein molecular weight search window of 0 to 150,000 Dalton (Da); (2) peptide mass tolerance of 0.1 Da; (3) minimum number of peptide match of 2; (4) minimum sequence coverage of 10%; (5) upper limit of mass tolerance of 25 parts per million (ppm); (6) cysteine modification by iodoacetamide (+57 Da) and methionine oxidation (+16 Da).

Two layers of search criteria were used to sort the search results. Briefly, proteins were ranked by a first criterion and a second criterion was activated if multiple proteins were matched by the first criterion. There are several options to choose from for each criterion, e.g., number of peptide matches, sequence coverage, or *m/z* deviation in ppm. In this study, the number of peptides matched and *m/z* deviation in ppm were used as the first and second criteria, respectively. If a protein ranked number 1 from a search, it was considered identified if it met the following criteria: (1) A minimum of 3 peptide *m/z* values was required to match values predicted from the theoretical digestion, and (2) the search result must be consistent with another database search with a second data set obtained from the same sample spot. Once a protein was identified from the PROQUEST search, a second search was performed using unmatched molecular weight values from the initial search. The same reproducibility requirements were applied in identifying the second and third proteins in the spot. Collecting multiple sets of data from

the same spot ensured the peptide maps were reproducible. Various search parameters were optimized for successful PROQUEST PMM searches as described below.

Optimization of PROQUEST Search Parameters

Internal calibration.   Before importing *m/z* values from a MALDI-TOF spectrum, all *m/z* were calibrated using a two-point internal calibration and thus a tight peptide mass tolerance could be used in the database search. Internal calibration was particularly critical for successful PROQUEST PMM protein identification. In order to obtain the best two-point internal calibration peptide peaks, three different pairs of internal calibrants were initially tested with approximately 10 different samples. The three pairs were des-Arg[1]-Bradykinin (*m/z* 904.4681)/Glu[1]-Fibrinopeptide B (*m/z* 1570.6774), angiotensin I (*m/z* 1296.6853)/ACTH clip 18–39 (*m/z* 2465.1989), and a trypsin autolysis fragment (*m/z* 842.5100)/ACTH clip 18–39 (*m/z* 2465.1989). We concluded that a combination of peaks at *m/z* 842.5100 and *m/z* 2465.1989 was found to provide the best PROQUEST search results at 20–30 ppm and this two-point curve covered the *m/z* range for most of the expected tryptic peptides. These two *m/z* values were used to correct *m/z* values throughout this study.

Manual inspection of mass spectra.   Before importing the calibrated peaks into PROQUEST, S/N of the peaks in the MALDI mass spectrum was inspected. Poor quality *m/z* values (S/N <5/1) were deleted, and *m/z* values with S/N greater than 5/1, but not labeled by the software program were labeled by manual manipulation of the software. This manual inspection was necessary due to inconsistent peak picking by the instrument's software. When attempts are made to automatically generate MALDI-TOF spectra for batch processing by any PMM software, this inconsistent automatic peak picking frequently makes an overall success rate for protein identification from the first round batch PMM searches much less than 50% (data not shown). Therefore unidentified spectra have to be manually inspected and re-searched, or data re-collected for the still unidentified spots. The time-consuming manual inspection of mass spectral data is often necessary to identify the maximum number of proteins when using PMM. This bottleneck makes it difficult for PMM to be a fully automated protein identification method. Although there have been many efforts, including a regression algorithm [38], an isotope-fit algorithm [39], and a Poisson peak harvesting algorithm [40], to improve the accuracy of automated peak picking from MALDI-TOF data for more reliable automation, it is still a challenging task.

Number of peaks for database searching.   Using the appropriate number of *m/z* values is one of the most important factors for successful protein identification by

Table 1. 2DE spots identified with less than 5 peptides from ProQuest PMM search

Spot ID	Identified protein	Molecular weight (Da)	Theoretical tryptic peptides[a]	Peptide matches	Matched peptides with >25ppm	Rank with 0– 150 kDa search window	Rank with 0–50 kDa search window
mj47	MJ0891	22700	8 (5)	4	1	1	1
mj75	MJ0892	22577	9 (6)	4	1	1	1
mj51	MJ1203	12686	8 (7)	4	0	1	1
mj82	MJ0312	22367	12 (11)	4	2	1	1
mj83	MJ0312	22367	12 (11)	4	1	1	1
mj96	MJ1333	9871	5 (5)	4	0	1	1
mj27	MJ0892	22577	9 (6)	3	0	2	1
mj46	MJ0892	22577	9 (6)	3	1	2	1
mj48	MJ0508	10363	5 (4)	2	0	2	1

[a]Number of complete trypsin cleavage at m/z 600 cutoff. Numbers in parentheses are the number of complete trypsin cleavage between m/z 600 and 3000.

PMM. It has been known from a previous study that using too many m/z values negatively affects the search results with a large database, such as NCBI non-redundant database due to an increasing probability of random matches from very complex theoretical tryptic peptide distributions [41]. However, this issue is not a serious problem with small databases. From initial comparisons of search results using different numbers of m/z values with the two databases from the two organisms used in this study (~2000 ORFs/each), we found that the use of as many peaks as possible provided the best PROQUEST search results. Therefore, all m/z values above a S/N of 5/1 were used for the database searches. The number of m/z values used for PROQUEST database searches in this study varied from 4 to 52, but about two thirds of the spots identified by PROQUEST (108/157, 69%) were identified with 10–30 m/z values. A recent report also suggests 20–30 peaks are likely to be the optimum number of peaks for reliable automated PMM protein identification [39].

Mass tolerance for database searching. In order to determine the optimized peptide mass tolerance from the two-point internal calibration for a PROQUEST PMM search, five different mass tolerances (10, 20, 30, 40, and 50 ppm) were tested with ~10 samples. PROQUEST search results with each mass tolerance were compared with database search results using MS/MS data. Twenty and 30 ppm were found to produce the best protein identification results, but 25 ppm was chosen as a default setting for the entire sample set. This 20–30 ppm mass tolerance range was previously shown to work well with internally calibrated tryptic peptides up to m/z 3000 [41, 42]. It should be noted here that 25 ppm was not used as an upper limit of mass tolerance for each individual peptide to match to a protein, but as an average value of mass tolerance for all peptides matches to a protein. Therefore, some peptides were matched to a protein at a value greater than 25 ppm of their theoretical masses. A peptide with an m/z deviation of up to 96 ppm was matched to a corresponding protein, but 92% of matched peptides were within 30 ppm. One

major reason for incorporating the average ppm rather than an individual absolute upper limit was that PROQUEST could match more peptides to proteins than would be matched with less than 5 peptide hits. It was found from the analyses of the entire sample set that this average ppm feature did not negatively affect search results for proteins matched with a sufficient number of peptides (usually >5 peptides) and that it did improve the performance of PROQUEST in identifying multiple proteins and small proteins in a spot. For example, among the 168 identified proteins by PROQUEST PMM in this study, 11 proteins could be successfully identified by this feature. Among them, 6 proteins were identified as the second most abundant protein in spots from second pass searches, and 5 other proteins were small proteins of <23 kDa.

Searching for small proteins. In general, at least 4–5 matched peptides for a protein are required to be confident of protein identification using PMM [41]. This is especially true if a search engine, which ranks proteins based on the number of peptide matches, such as PROQUEST, is used. Even with the recent developments of more sophisticated PMM search engines incorporating statistical scoring schemes [43, 44], small proteins with insufficient trypsin cleavage sites are still challenging for unambiguous identification by PMM. In an effort to overcome this problem of identification with small proteins, a narrow mass search window (0–50 kDa) option was incorporated into the database search. Table 1 shows the spots identified with less than 5 peptides from the PROQUEST search. The 9 proteins listed in Table 1 are all small proteins with molecular weights (MW) less than 23 kDa. As shown in the two rightmost columns in Table 1, the first 6 spots matched with 4 peptides, mj47, 75, 51, 82, 83, and 96, were not negatively affected by searches with the two different search windows. However, the next 2 spots identified with 3 peptides (mj27 and 46) and the bottom spot identified with 2 peptides (mj48) were ranked from second to first and thus unambiguously matched to the corresponding proteins only when the narrow search

Table 2. 2DE spots that were not identified by PROQUEST PMM but by LC-MS/MS

Spot ID	Identified proteins by LC-MS/MS	Molecular weight (Da)	Matched peptides by LC-MS/MS
mj89	MJ0316	40294	1
mj94	MJ1129	10583	2
	MJ0742	11703	2
pf06	Pf_1817171	48244	1
pf25	Pf_1314579	65719	1
pf6b	Pf_882732	43345	1

window was employed. Spot mj48 could be confidently identified with only 2 peptide matches because no other proteins were matched with the narrow mass range search from duplicate experiments. Identification of all proteins in Table 1 was confirmed by LC-MS/MS.

The column with the number of matched peptides with >25 ppm shows the effect of the average upper limit of 25 ppm in small protein identification. By using this average feature, spots mj47, 75, 82, 83, and 46, could be identified with higher confidence with the 1–2 additional m/z values.

Another column shows the number of theoretical trypsin cleavage sites for each matched protein. Numbers prior to parentheses indicate the number of complete cleavage sites >m/z 600 and numbers in parentheses indicate those between m/z 600 and 3000, in which most ions were matched in these experiments. The numbers are all less than 10 except for spots mj82 and 83. Thus, assuming the overall efficiency of protein digestion and peptide extraction to produce peptides within an m/z range of 600–3000 Daltons is less than 50%, then proteins with less than 10 cleavage sites would be expected to yield about five peptides. The fact that all the proteins shown in Table 1 were matched with less than 5 peptides by PMM supports this simple hypothesis. Note that these matches were obtained for peptides present in amounts of at least 500 femtomoles, and they have been concentrated and desalted using a ZipTip$_{C18}$. Thus, as the amount of protein decreases to the low femtomole ranges, it will be even more difficult to identify small proteins using PMM.

Interestingly, the same protein appeared at the same location on 2D gels from three different sample fractions; a whole cell lysate, a soluble fraction, and a membrane enriched fraction. Spots mj75 from soluble fraction, mj27 from whole cell extract, and mj46 from membrane fraction were consistently matched to MJ0892 with the three same peptides matched from spots mj27 and 46. This identification consistency to a protein, which migrated to same location on different gels, provides added confidence to the identification.

Five spots out of a total of 162 spots analyzed in this study were not identified by PMM and are listed in Table 2. All of these spots were successfully identified by LC-MS/MS. As shown in Table 2, one of the 5 spots (mj94) was identified as two small proteins and the other 4 proteins were identified as medium-size pro-

teins by LC-MS/MS with only 1 or 2 peptide matches. No m/z values or only single m/z values were observed to match to the proteins in Table 2 by PROQUEST search. This observation suggests the tryptic peptides of the proteins listed in Table 2 may be difficult to ionize and/or detect using MALDI-TOF mass spectrometry and thus no or too few peptide m/z values were observed. Duplicate experiments with another tryptic digest were performed, but the same results were obtained. Further measurements on these peptides were not pursued.

Comparison with Other Search Engines

In order to compare the performance of PROQUEST search results with two widely used PMM algorithms, Profound [44] and MS-Fit [45], our data set was re-searched using these programs. Identifying an abundant protein with a large number of peptide m/z values is relatively straightforward for any PMM software. For example, 157/162 spots analyzed in this study were already successfully identified by the PROQUEST search. Thus, the comparison focused more on how well the two PMM software programs performed in more challenging areas, such as identification of small proteins and multiple proteins in a single spot. One hundred *M. jannaschii* spots were chosen for comparison because many small proteins and multiple proteins in a single spot were identified by PROQUEST compared with data from *P. furiosus*. For comparison purposes, the same search parameters used in PROQUEST searches were used in searches by Profound and MS-Fit (using the MOWSE scoring system), which included a molecular weight range of 0–150 kDa, static Cys-modification by IAA, and 25 ppm as the upper limit of mass tolerance. The 25 ppm upper limit was the absolute upper limit for individual matched peptides for both Profound and MS-fit. Database searches were performed against the same *M. jannaschii* species-specific database as used in the PROQUEST search. Criteria for protein identification required a Z score >0.95 and P <1 for Profound and MOWSE score >100 for an MS-Fit search.

The most abundant proteins in a spot were identified by both Profound and MS-Fit with exactly the same results as that of PROQUEST. The 98 most abundant proteins identified by PROQUEST were also identified by both programs, and the 2 spots that failed to be identified by PROQUEST were not identified by either of the two. In both cases no m/z values or only single m/z values were observed by MALDI-TOF for peptides from proteins identified solely by LC-MS/MS. The improved specificity of MS/MS data allowed the proteins to be identified in these two cases. Several spots that contained multiple proteins in a single spot that were not identified as such by PROQUEST were identified by both Profound and MS-Fit to contain multiple proteins. Both Profound and MS-Fit identified 15 spots as containing multiple proteins, while PROQUEST
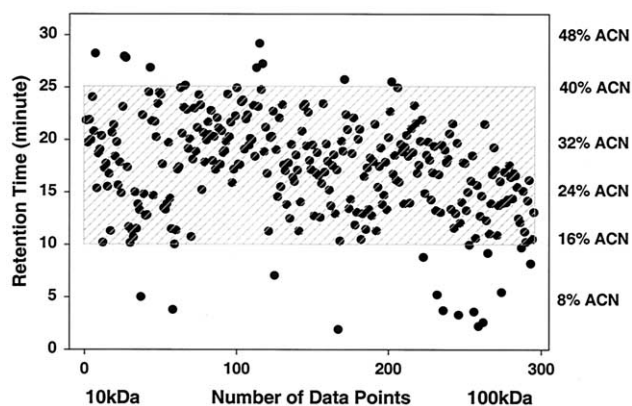
Figure 1. Distribution of retention times of 295 unique peptides matched to 30 *P. furiosus* ORFs from linear 30 min LC gradient. Retention times are plotted as a function of protein molecular weight for the proteins identified. The 273/295 (92%) peptides in the shaded area successfully matched to the 30 *P. furiosus* proteins indicating the 15 min span of the 30 min LC gradient is sufficient to identify 2D-gel proteins.

found 11 spots. Both Profound and MS-Fit identified the same set of proteins, which may indicate a similar sensitivity for the two scoring algorithms used in the two programs. No single program demonstrated better performance over the others for the identification of small proteins. For example, all of the 9 small proteins identified by PROQUEST shown in Table 1 were initially ranked first by both Profound and MS-Fit. However, the statistical confidence in the identifications using the above criteria would not have been high enough for acceptance of some of the proteins. This problem is due to the smaller number of peptide *m/z* values obtained for each small protein and thus a lower statistical significance is obtained. This situation also suggests a limitation to the PMM approach for confident identification of small proteins.

Towards High-Throughput µLC-MS/MS

In this study, the same linear 30-min gradient scheme was consistently used for identification of 2D-gel separated proteins. A total of 1875 peptides were matched to identify a total of 315 ORFs by the SEQUEST-PVM search. The distribution of all matched peptides across the 30-min gradient, 295 unique peptides matched to 30 selected *P. furiosus* proteins, were plotted as a function of LC retention time and the size of the protein to which the peptides were matched (Figure 1). shows more than 90% of the matched peptides were detected with retention times between 10 and 25 min, which corresponded to 16–40% of acetonitrile in mobile phase buffer. In other words, almost all peptides used to identify proteins were detected in a 15 min span of the 30-min gradient. Therefore, it was believed that utilizing a non-linear gradient of ~15– 40% of acetonitrile would reduce data collection time by about half without sacrificing HPLC resolution and the ability to acquire tandem mass spectra for the protein identification pro-

cess. Ultimately the gradient should preserve the dynamic range advantage of a high-resolution separation while decreasing the time required to perform the separation. To verify this observation a non-linear gradient was successfully used to identify other 2D-gel separated proteins prepared from several different organisms (data not shown). A combination of this separation strategy and automated LC-MS/MS using precolumn sample focusing [46] would be expected to provide higher throughput processing of 2D-gel separated proteins maintaining high identification success rate.

Protein Identification

Success rate. From the analysis of 100 *M. jannaschii* and 62 *P. furiosus* spots by both PMM and LC-MS/MS, 100% of the analyzed spots from both organisms were successfully matched to the predicted proteins in the two corresponding ORF databases by LC-MS/MS, while success rates of 98% (98/100) and 95% (59/62) were obtained from *M. jannaschii* and *P. furiosus*, respectively, by PMM. This relatively high success rate (≥95%) for protein identification by PMM was enhanced by using sample clean up and concentration by ZipTip$_{C18}$ preconcentrator. According to a test analysis of 5 spots, including both dark and faint spots, the quality of MALDI spectra from faint spots were greatly improved with a ZipTip$_{C18}$ while those from dense spots did not change appreciably regardless of whether a ZipTip$_{C18}$ was used or not (data not shown). An additional advantage of using a ZipTip$_{C18}$ to clean up the sample was demonstrated in the extended MALDI-TOF MS scan range. Typically, a MALDI mass spectrum without ZipTip$_{C18}$ concentration has high α-CHCA matrix interference signal up to *m/z* 900, which makes it difficult to distinguish peptide peaks from the interfering matrix signal up to the *m/z* region. However, the desalting effect by ZipTip$_{C18}$ significantly reduced the α-CHCA matrix signal in that *m/z* range and made it possible to acquire usable data down to an *m/z* of 600. This enabled the use of *m/z* values between an *m/z* 900 and 600 to be used in the data searching process and contributed to the high success rate of ≥95%. For example, of the 157 MALDI spectra that successfully matched to proteins, 71% (111/157) of them matched peptides between *m/z* 600 and 900. For some proteins, up to 9 peptides were matched in this region. Moreover, among the 9 small protein spots in Table 1, 3 proteins (MJ0891, MJ0892, and MJ0312) from the 6 spots (mj47, 75, 82, 83, 27, and 46) could be more confidently matched with 1–2 matches in this *m/z* region. This strongly suggests that sample clean up and concentration is essential for successful identification of small proteins when using PMM.

Specificity of LC-MS/MS. From this 162 spot analysis, it was found that all the proteins identified by PMM were always identified by LC-MS/MS. Except for a few cases,

proteins that matched with the most number of peptides by LC-MS/MS were also the proteins exclusively identified as the number 1 ranked proteins in the first round of a PROQUEST PMM search. No unique proteins were identified exclusively by PMM. This result shows that MALDI-TOF is complementary to MS/MS identification of 2D-gel proteins, but lacks the dynamic range of LC-MS/MS. Improved dynamic range is obtained through the LC separation process to enable the acquisition of tandem mass spectra for proteins present in a spot in less abundance. MS/MS database searching uses fragmentation patterns indicative of specific amino acid sequence to match a protein and thus has a higher level of specificity than PMM. Thus, only one peptide MS/MS spectrum with good signal to noise, fragmentation, and length can identify a protein in the database. Another drawback of the PMM approach in protein identification can be found when judging a search result. If the score of the first ranked protein is around or below the boundary of the criteria set by the software, the user cannot be confident with the result and has to manually inspect the data to obtain additional information (e.g., sequence) to increase the confidence of the result. In most cases, the protein has to be reanalyzed preferably by MS/MS. An example of this situation was described earlier in evaluating the identification of small proteins by Profound and MS-Fit.

Number of identified proteins in a single spot.   The specificity of each mass spectrometric method in protein identification is well illustrated in Figure 2, which shows the number of identified proteins in a single spot from each method for each organism. As shown in the pie charts, up to 7 proteins were identified by LC-MS/MS; up to 3 proteins by PMM. Of the identified spots by each method, 60% (60/100) and 34% (21/62) of the spots were identified as multiple proteins in *M. jannaschii* and *P. furiosus,* respectively by LC-MS/MS, while 12% (12/98) and 3% (2/59) were identified as multiple proteins in the spots by PMM. This dramatic difference in identification of the number of multiple proteins from a single spot clearly demonstrates the improved dynamic range of LC-MS/MS over PMM in protein identification from 2DE.

This improved dynamic range of LC-MS/MS in protein identification can be an important issue in drug discovery. For example, the presence of multiple proteins in a spot when trying to measure a differential expression ratio of a protein between normal and disease cells is important in biology and drug target identification. Currently, 2D-gel image analysis, especially differential in-gel electrophoresis (DIGE) technology via fluorescent dye labeling [17] provides a reliable measure of protein expression changes in two different cell states by running two different samples on the same 2D gel. The relative change of a protein is obtained by comparing fluorescent signals at two different wavelengths from a 2DE image obtained after the combined separation of fluorescently labeled proteins from the
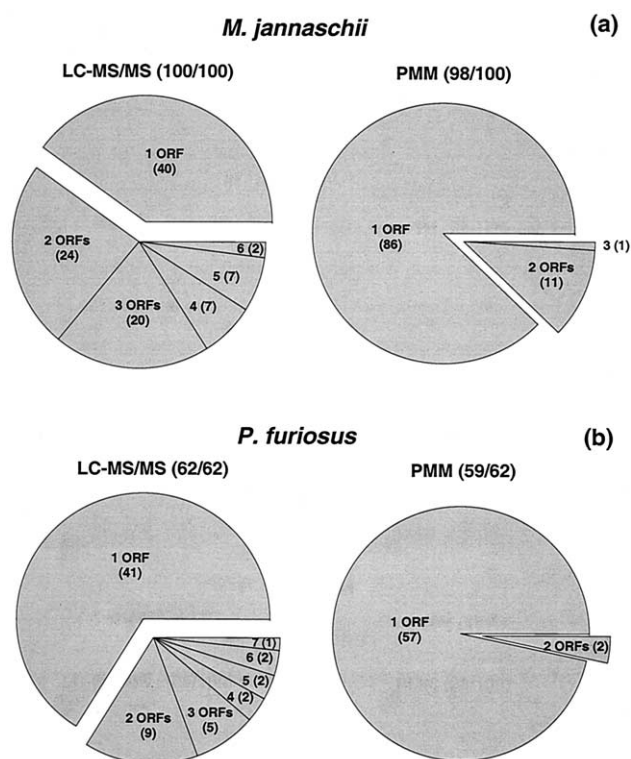

Figure 2. Number of ORFs identified from a single spots by LC-MS/MS and PMM for (a) *M. jannaschii* and for (b) *P. furiosus.* Numbers in parentheses are the number of spots identified with the corresponding number of ORFs.

two different cell states. After proteins are identified by mass spectrometry, each protein ID is linked to a spot ID by the software. This means no matter how many proteins are identified from a single spot, only one difference ratio will be assigned to the proteins identified from the spot. Since the most widely used pH range for isoelectric focusing (ISF) for 2DE is currently 4–7, it is likely that more than 20% of the analyzed spots will contain multiple proteins [6]. For some organisms, the number can approach up to 6 or 7 proteins as shown in Figure 2. Thus, reporting just 1 or 2 proteins by PMM for a differentially expressed spot with more than 2 proteins would be incorrect. Although employing LC-MS/MS will not help solve this problem, it is more important to know how many and what proteins actually exist in the spot to minimize incorrect assessment of up- or down regulation of protein expression. Presumably co-elution can be minimized by increasing the resolution of protein separation by using narrow pH range strips for isoelectric focusing (ISF) prior to the second-dimension separation.

Protein sequence coverage.   Figure 3a and b show the number of matched peptides and sequence coverage from both organisms over the entire mass ranges obtained by LC-MS/MS and PMM respectively. In order to measure sequence coverage for proteins of different sizes, the168 proteins identified from the two organisms
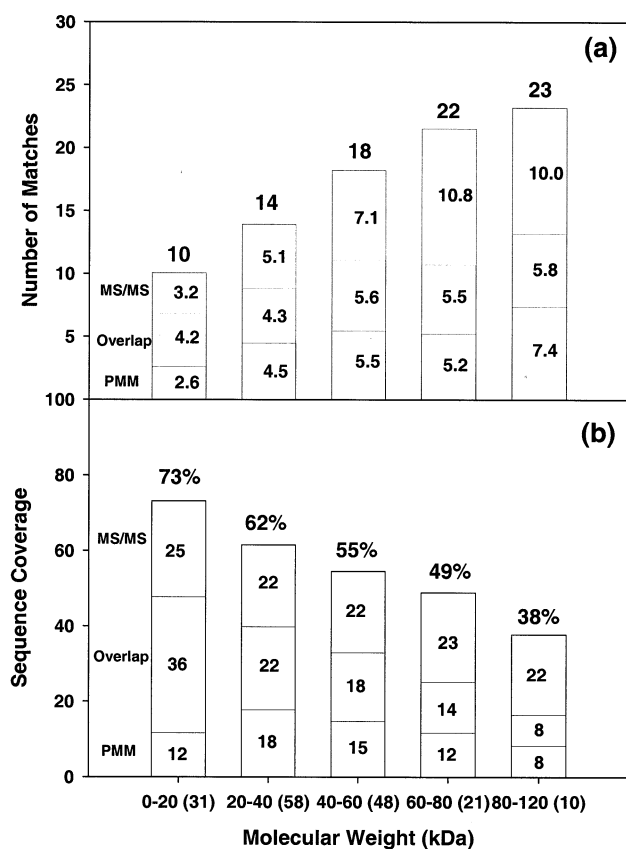
Figure 3. Average values of (**a**) combined matched peptides and (**b**) combined sequence coverage from LC-MS/MS and PMM for a total of 168 ORFs identified by both methods from *M. jannaschii* and *P. furiosus*. All 168 identified ORFs were grouped in the following molecular weight bins: 0–20 kDa, 20–40 kDa, 40–60 kDa, 60–80 kDa, and 80–120 kDa. The numbers in parentheses on the x-axis are the number of identified proteins used to obtain the average value for the number of matched peptides and sequence coverage for each group, respectively. MS/MS indicates the unique peptides obtained using LC-MS/MS to identify peptides and calculate protein sequence coverage; overlap indicates peptides matched in common between the two methods, overlapped matched peptides and sequence coverage from both methods, and PMM indicates the unique peptides obtained using of MALDI-TOF and PMM to identify peptides and sequence coverage. The numbers above each stacked bars indicate combined number of matched peptides and sequence coverage. Rounded values of integers are used except for the number of matched peptides for MS/MS, overlap, and PMM to show subtle differences in numbers.

by each method were grouped into five different sizes: 0–20 kDa, 20–40 kDa, 40–60 kDa, 60–80 kDa, and 80–120 kDa. The numbers in parentheses on the x-axis indicate the number of identified proteins used to obtain the average value of the number of matched peptides for each group. For example, 31 proteins were used to determine the average number of matched peptides for proteins of 0–20 kDa, 58 for 20–40 kDa, 48 for 40–60 kDa, 21 for 60–80 kDa, and 10 for 80–120 kDa. The bottom part of the stacked bar in Figure 3a is the average number of PMM-unique matched peptides, the middle section is the average number of overlap-

ping peptides from both PMM and LC-MS/MS, and the top section is the average number of unique peptides identified by LC-MS/MS. Thus, the sum of all three parts is the average number of combined matched peptides. Integers close to the actual average values were used for easy viewing and the numbers are 10, 14, 18, 22, and 23, respectively, as displayed on top of each stacked bar. Figure 3b shows corresponding sequence coverage for each group, and the same scheme as used in Figure 3a was again applied.

As the size of the protein increases, the number of trypsin cleavage sites increases, and thus the theoretical number of peptides to be potentially matched increases. The actual number of peptides detected will be a function of site-specific digestion efficiency and recovery from the gel extraction procedure. The efficiency of cleavage and extraction produces a range of peptide abundances expected for a particular protein. Coupled together with different ionization efficiencies for each peptide, a variety of signal intensities can be expected. Thus, if LC-MS/MS has better dynamic range than MALDI-TOF MS for protein identification it is expected that sequence coverage for the identified proteins should be better using LC-MS/MS. Figure 3a shows that both LC-MS/MS and PMM follow this simple hypothesis. It is shown that the relative proportion of unique peptides identified by LC-MS/MS increases at a steeper rate than that of PMM as a function of protein size. This slope approached a plateau for proteins >80 kDa. Overall, the numbers of LC-MS/MS-unique peptides were higher than that of PMM over the whole range of protein size and this effect was more substantial for medium to large proteins. However, the difference for relatively small proteins (0–40 kDa) was shown to be minimal. This trend of a higher number of LC-MS/MS-unique peptides for all five groups of proteins in Figure 3a is well reflected in its higher sequence coverage for all five groups of proteins in Figure 3b. All of these results indicate several interesting points. First, the contribution of peptide separation by LC was not significant for relatively small proteins. A possible explanation for this is that not many theoretical tryptic peptides were available and thus the dynamic range of PMM was sufficient to acquire peptide *m/z* data for useful peptide matches. However, the limited dynamic range of PMM is clearly seen for medium to large proteins. Second, the combined efficiency of data-dependent MS/MS acquisition and LC separation from a 30-min linear gradient used in this study seems to reach its maximum for proteins of >80 kDa. This limited data acquisition efficiency for the proteins of >80 kDa was also illustrated in Figure 3b to negatively affect the sequence coverage of large proteins more substantially than other group of proteins. Employing a gradient longer than 30 min would be expected to improve sequence coverage for these large proteins at the cost of higher throughput. As shown in Figure 3b sequence coverage decreases substantially with respect to the increasing molecular weight of proteins. In general,

LC-MS/MS demonstrated higher sequence coverage than PMM over the entire molecular weight range of proteins identified.

Combined sequence coverage.   Obtaining more than 50% of sequence coverage from a tryptic digest of a 2D-gel separated protein by either PMM or LC-MS/MS is difficult [47]. This was also found in our study as shown in Figure 3b, but it was also shown that relatively small proteins (0–40 kDa) have a greater chance of obtaining sequence coverage of ~50% than medium to large proteins. In fact, protein sequence coverage of >60% was readily obtained by LC-MS/MS for very small proteins in the group of 0-20 kDa. As shown in Figure 3b, average sequence coverage from PMM decreased 48%, 40%, 33%, 26%, and 16% with increasing molecular weight of proteins, while those by LC-MS/MS varied 61%, 44%, 40%, 37%, and 30%. However, when the matched peptides were combined from these two different ionization methods, the average combined sequence coverage increased to 73%, 62%, 55%, 49%, and 38% with increasing molecular weight of proteins. Among the total of 168 identified proteins, 112 proteins (67%) achieved greater than 50% sequence coverage.

One important area in proteomics is the identification of post-translational modifications (PTM). It has been reported that more than 200 PTMs occur in nature [48]. 2DE can display different forms of proteins that can indicate the state of modified proteins. Thus mapping modified sites of these 2DE-separated proteins by mass spectrometry is important to understand their function. In order to obtain maximum information of a protein, ideally 100% sequence coverage should be obtained from experiments. The limitations of digestion and extraction efficiency in gels make it difficult to consistently obtain 100% sequence coverage. Combining the sequence coverage of both PMM and LC-MS/MS presents an approach to increase the amount of sequence coverage obtained for a protein. As described above, about two thirds of the identified proteins had >50% sequence coverage. In some cases, greater than 80% sequence coverage was observed. In order to increase the chance of identifying modified sites, at least 80% sequence coverage is desired [47]. Use of other proteases with different cleavage specificity in addition to trypsin is a good way of increasing sequence coverage [14, 49, 50]. Jungblut and coworkers reported that they could achieve up to 80% sequence coverage for medium-quantity spots by PMM by employing two additional proteases, Asp-N and Glu-C, in addition to trypsin [47]. Therefore, by employing other proteases in addition to trypsin and by combining the matched peptides from two complementary protein identification methods, PMM and LC-MS/MS, maximum sequence coverage from a gel spot could be obtained, increasing the probability of identifying a PTM.

Effects of modification searches.   Two widely used modification searches, cysteine modification by iodoacet-amide (+57 Da) (C + 57) and methionine oxidation (+16 Da) (M + 16), were included for both PROQUEST and SEQUEST database searches. Both (C + 57) and (M + 16) were differentially searched during SEQUEST MS/MS searches. However, since PROQUEST did not have options of such differential modification searches at the time of data collection, two searches, modification and non-modification, were conducted separately, and the results were combined. It was found from the (C + 57) searches that cysteine-modified peptides were dominantly matched over the non-modified form from both PROQUEST and SEQUEST database searches, which indicates that most of the cysteine-containing peptides were alkylated during sample processing, and thus that there is very little value in searching for unmodified Cys. However, the degree of contribution to sequence coverage enhancement from (C + 57) was minimal compared with results from non-modification searches. This seems to be due to the relatively low frequency of cysteine residues in proteins in microorganisms.

On the contrary, differential (M + 16) searches demonstrated noticeable sequence coverage enhancement. Because the effect was much higher from SEQUEST database search than from PROQUEST PMM search, only its effect on MS/MS data will be discussed. Oxidation of methionine is common modification occurring in proteins, but most of the modification detected by mass spectrometry is thought to come from adventitious oxidation of the residues when samples are exposed to acidic conditions and oxygen during protein processing and in-gel digestion [14]. Methionine can be oxidized to methionine-sulfoxide (+16 Da) and further oxidized to methionine-sulfone (+32 Da). Since methionine-sulfoxide is more dominant than its sulfone derivative, only modification to methionine-sulfoxide was considered in the database search. Table 3 shows an example of the most dramatic effect of (M + 16) in the SEQUEST search. This spot was initially matched to MJ0784 (MW 38689) with 10 peptide matches at 37% sequence coverage when only (C + 57) search was included. However, with the additional (M + 16) incorporation during the database search, 13 more peptides were matched and its sequence coverage was increased from 37 to 60%. Among these 12 additionally matched peptides, the underlined 4 peptides were newly identified sequences that contributed to increasing the sequence coverage to 60%. The other 10 peptides were previously matched from (C + 16) search, but they were matched again as sequences containing oxidized methionine.

Incorporating (M + 16) into the search also demonstrated its usefulness in identifying additional proteins by improving the confidence level of protein identification. Table 4 shows that this (M + 16) search could identify additional proteins by two different contributions. First, the top 5 proteins in Table 4 (MJ0784 from spot mj15, MJ0891 from spot mj46, MJ0784 from spot mj57, MJ0318 from spot mj72, and PF_1346821 from spot pf12) were not identified from SEQUEST search

Table 3. Sequence coverage change of a 2DE spot by SEQUEST MS/MS data search with (C + 57) and (M + 16) differential modifications

Number	Sequence of matched peptides[a]	Matched peptides by (C + 57)	Additional matches by (M + 16)
1	R.EEAVEGADIVITWLPK.G	O	
2	K.EVGKPEIALTHSSITYGAELLHLVPDVK.E	O	
3	K.EVMEAHLSGNPESIMPK.I	O	
4	K.EVM*EAHLSGNPESIMPK.I		O
5	K.EVMEAHLSGNPESIM*PK.I		O
6	K.EVM*EAHLSGNPESIM*PK.i		O
7	K.FADAIPEGAIVTHAC*.T	O	
8	K.FADAIPEGAIVTHAC*TIPTTK.F	O	
9	K.GIANM*EEALDPAALLGTADSM*C*FGPLAEILPTAL.V		O
10	K.GIANM*EEALDPAALLGTADSMC*FGPLAEILPTAL.V		O
11	K.GQVYIAEGYASEEAVNK.L	O	
12	K.IAILGAGC*YR.T	O	
13	K.ILGAPADFAQM*M*ADEALTQIHNLM*K.E		O
14	K.ILGAPADFAQM*M*ADEALTQIHNLMK.E		O
15	K.ILGAPADFAQM*MADEALTQIHNLM*K.E		O
16	K.ILGAPADFAQMM*ADEALTQIHNLM*K.E		O
17	K.ILGAPADFAQMMADEALTQIHNLM*K.E		O
18	K.LYEIGK.I	O	
19	R.THAAAGITNFM*R.A		O
20	R.THAAAGITNFMR.A	O	
21	K.VTSDDREAVEGADIVITWLPK.G	O	
22	K.DLGREDLNITSYHPGC*VPEM*K.G		O
23	R.EDLNITSYHPGC*VPEM*I.G		O
	Obtained Sequence Coverage	37%	60%

[a]C* and M* indicate modified Cys residue by iodacetamide and oxidized Met residue, respectively.

when only (C + 57) was included. However, with inclusion of (M + 16), these proteins could be confidently identified by SEQUEST with 1 to 3 newly identified peptides containing a modified methionine residue. Second, the remaining three proteins in Table 4 (MJ0212 from spot mj94, PF_1861160 from spot pf19, and PF_263488 from spot pf40) were initially matched with only 1 peptide by SEQUEST search incorporating (C + 57) only, but their identification was not accepted because the SEQUEST cross-correlation (Xcorr) scores and the quality of their MS/MS spectra were not convincing enough to accept. However, with inclusion of (M + 16), these three protein identifications were confidently recovered with 2–3 matched peptides with good XCorr values and spectral quality. SEQUEST searches incorporating (M + 16) differential modification of Met is being routinely used in our laboratory for protein identification searches. The only disadvantage of this (M + 16) differential modification search is the fact that it requires sufficient computing power to obtain reasonable search speed since the search algorithm has to consider all possible permutation of modifications on peptides. This problem is more important especially when dealing with a large number of MS/MS spectra. All these differential modification searches could be finished at a relatively fast speed by SEQUEST-PVM on our Linux cluster of 12 nodes. Since PTM analysis is getting more and more important in proteomics, utilizing a high computing power for high-speed MS/MS database searches will be more and more important in the future.

Conclusions and Perspectives

Using ZipTip$_{C18}$ sample clean up and optimizing several search parameters for PMM searches with the search program, PROQUEST, a success rate exceeding 95% was achieved from the 162 Coomassie-stained archaea 2D-gel spots. It was demonstrated that some of the low molecular weight (<23 kDa) proteins could be successfully identified with less than 5 peptide matches by applying a narrow protein mass search window. Additional peptide matches between m/z 600 and 900 could be obtained by using ZipTip$_{C18}$ sample clean up

Table 4. Identified 2DE spots by SEQUEST MS/MS data search with (M + 16) modification

Spot ID	Identified proteins	Peptide matches with (C + 57)	Peptide matches with (C + 57) & (M + 16)
mj15	MJ0784	0	1
mj46	MJ0891	0	3
mj57	MJ0784	0	3
mj72	MJ0318	0	1
pf12	PF_1346821	0	2
mj94	MJ0212	1	3
pf19	PF_1861160	1	2
pf40	Pf_263488	1	2

and by incorporating a tight 25 ppm <u>average</u> mass tolerance during database searches. The additional low molecular peptide *m/z* values contributed to the improved identification of small proteins. However, by comparison with other PMM search engines, it was found that the confident identification of small proteins with limited trypsin cleavage sites was still difficult no matter what search engine was used.

MALDI PMM is very useful in high-throughput protein identification. In favorable cases up to 2 proteins in a single spot can be confidently identified by PMM. The comparisons shown in this paper demonstrate the ability of PMM to identify multiple proteins in a single spot lags behind that of LC-MS/MS, presumably because of dynamic range differences between the methods for data acquisition. This dynamic range differences may be an issue in drug target identification if a spot that is differentially expressed between two different cell states has more than 2 proteins in it. By not fully identifying the proteins present in a spot can lead to a wrong conclusion on the spot's ID and its relative expression changes. Therefore, if there is no need for very high-throughput analysis, e.g., analysis of several hundred spots per day, the use of LC-MS/MS should be preferred to obtain the most accurate protein identifications for differentially expressed spots. Ideally the LC-MS/MS method should employ an autosampler to meet a medium throughput capability and to have a success rate of protein identification comparable to that from manual loading of samples. An automated LC-MS/MS system utilizing pre-column sample focusing with a fast non-linear LC gradient strategy described in this paper will be a very useful tool for analysis of differential 2D-gel proteins in drug discovery.

The combination of LC-MS/MS and SEQUEST database searching showed better protein sequence coverage than MALDI PMM across a range of protein molecular weights. However, the PMM method did produce unique peptide *m/z* values that were not found by LC-MS/MS. Combined sequence coverage obtained from these two complimentary methods produced better then 50% sequence coverage for ~70% of the spots identified by both methods from a single trypsin digestion alone. Improvement in sequence coverage would be expected if multiple digestions of different specificity were applied to a protein and the results pooled. Higher sequence coverage will not contribute to better confidence in protein identification, but will be very useful to identify post-translational modifications.

Acknowledgments

References

1. Karas, M.; Hillenkemp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, *60*, 2299–2301.
2. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246*, 64–71.
3. Yates, J. R., III. Mass Spectrometry and the Age of the Proteome. *J. Mass Spectrom.* **1998**, *33*, 1–19.
4. Aebersold, R.; Goodlett, D. R. Mass Spectrometry in Proteomics. *Chem. Rev.* **2001**, *101*, 269–295.
5. Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. Linking Genome and Proteome by Mass Spectrometry: Large-Scale Identification of Yeast Proteins from Two Dimensional Gels. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14440–14445.
6. Link, A. J.; Hays, L. G.; Carmack, E. B.; Yates, J. R., III. Identifying the Major Proteome Components of Haemophilus influenzae Type Strain NCTC 8143. *Electrophoresis* **1997**, *18*, 1314–1334.
7. Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* **2002**, *1*, 845–867.
8. Lisacek, F. C.; Traini, M. D.; Sexton, D.; Harry, J. L.; Wilkins, M. R. Strategy for Protein Isoform Identification from Expressed Sequence Tags and Its Application to Peptide Mass Fingerprinting. *Proteomics* **2001**, *1*, 186–193.
9. Claverol, S.; Burlet-Schiltz, O.; Girbal-Neuhauser, E.; Gairin, J. E.; Monsarrat, B. Mapping and Structural Dissection of Human 20 S Proteasome Using Proteomic Approaches. *Mol. Cell. Proteomics* **2002**, *1*, 567–578.
10. McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., III. Direct Analysis and Identification of Proteins in Mixtures by LC/MS/MS and Database Searching at the Low-Femtomole Level. *Anal. Chem.* **1997**, *69*, 767–776.
11. Link, A. J.; Eng. J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. Direct Analysis of Protein Complexes Using Mass Spectrometry. *Nat. Biotech.* **1999**, *17*, 676–682.
12. Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotech.* **2001**, *19*, 242–247.
13. Florens, L.; Washburn, M. P.; Raine, J. D.; Anthony, R. M.; Grainger, M.; Haynes, J. D.; Moch, J. K.; Muster, N.; Sacci, J. B.; Tabb, D. L.; Witney, A. A.; Wolters, D.; Wu, Y.; Gardner, M. J.; Holder, A. A.; Sinden, R. E.; Yates, J. R., III; Carucci, D. J. A Proteomic View of the *Plasmodium falciparum* Life Cycle. *Nature* **2002**, *419*, 520–526.
14. MacCoss, M. J.; McDonald, W. H.; Saraf, A.; Sadygov, R.; Clark, J. M.; Tasto, J. J.; Gould, K. L.; Wolters, D.; Washburn, M.; Weiss, A.; Clark, J. I.; Yates, J. R., III. Shotgun Identification of Protein Modifications from Protein Complexes and Lens Tissue. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7900–7905.
15. Wall, D. B.; Kachman, M. T.; Gong, S.; Hinderer, R.; Parus, S.; Misek, D. E.; Hanash, S. M.; Lubman, D. M. Isoelectric Focusing Nonporous RP HPLC: A Two-Dimensional Liquid-Phase Separation Method for Mapping of Cellular Proteins with Identification Using MALDI-TOF Mass Spectrometry. *Anal. Chem.* **2000**, *72*, 1099–1111.
16. Kachman, M. T.; Wang, H.; Schwartz, D. R.; Cho, K. R.; Lubman, D. M. A 2-D Liquid Separations/Mass Mapping Method for Interlysate Comparison of Ovarian Cancers. *Anal. Chem.* **2002**, *74*, 1779–1791.

17. Unlu, M.; Morgan, M. E.; Minden, J. S. Difference Gel Electrophoresis: A Single Gel Method for Detecting Changes in Protein Extracts. *Electrophoresis* **1997,** *18,* 2071–2077.

18. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotech.* **1999,** *17,* 994–999.

19. Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. Quantitative Proteome Analysis by Solid-Phase Isotope Tagging and Mass Spectrometry. *Nat. Biotech.* **2002,** *19,* 512–515.

20. Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. Quantitation and Facilitated de Novo Sequencing of Proteins by Isotopic N-terminal Labeling of Peptides with a Fragmentation-Directing Moiety. *Anal. Chem.* **2000,** *72,* 4047–4057.

21. Goshe, M. B.; Conrads, T. P.; Panisko, E. A.; Angell, N. H.; Veentra, T. D.; Smith, R. D. Phosphoprotein Isotope-Coded Affinity Tag Approach for Isolating and Quantitating Phosphopeptides in Proteome-Wide Analyses. *Anal. Chem.* **2001,** *73,* 2578–2586.

22. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. Accurate Quantitation of Protein Expression and Site-Specific Phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96,* 6591–6596.

23. Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic O18 Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **2001,** *73,* 2836–2842.

24. Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. Rapid Identification of Proteins by Peptide Mass Fingerpriting. *Curr. Biol.* **1993,** *3,* 327–332.

25. Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. Femtomole Sequencing of Proteins from Polyacrylamide Gels by Nano-Electrospray Mass Spectrometry. *Nature* **1996,** *379,* 466–469.

26. Link, A. J.; Carmack, E.; Yates, J. R. III. A Strategy for the Identification of Proteins Localized to Subcellular Spaces: Application to *E. coli* Periplasmic Proteins. *Int. J. Mass Spectrom. Ion Processes* **1997,** *160,* 303–316.

27. Medzihradszky, K. F.; Leffler, H.; Baldwin, M. A.; Burlingame, A. L. Protein Identification by In-Gel Digestion, High-Performance Liquid Chromatography, and Mass Spectrometry: Peptide Analysis by Complementary Ionization Techniques. *J. Am. Soc. Mass Spectrom.* **2001,** *12,* 215–221.

28. Giometti, C. S.; Reich, C.; Tollaksen, S.; Babnigg, G.; Lim, H.; Zhu, W.; Yates, J. R., III; Olsen, G. Global Analysis of a "Simple" Proteome: Methanococcus janaschii. *J. Chromatogr. B* **2002,** *782,* 227–243.

29. Anderson, N. G.; Anderson, N. L. Analytical Techniques for Cell Fractions. XXI. Two-Dimensional Analysis of Serum and Tissue Proteins: Multiple Isoelectric Focusing. *Anal. Biochem.* **1978,** *85,* 331–340.

30. Anderson, N. G.; Anderson, N. L. Analytical Techniques for Cell Fractions. XXI. Two-Dimensional Analysis of Serum and Tissue Proteins: Multiple Gradient Slab-Gel Electrophoresis. *Anal. Biochem.* **1978,** *85,* 341–354.

31. Anderson, N. L.; Nance, S. L.; Tollaksen, S. L.; Giere, F. A.; Anderson, N. A. Quantitative Reproducibility of Measurements from Coomassie Blue-Stained Two-Dimensional Gels: Analysis of Mouse Liver Protein Patterns and a Comparison of BALB/c and C57 Strains. *Electrophoresis* **1985,** *6,* 592–599.

32. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **1996,** *68,* 850–858.

33. Han, J. C.; Han, G. Y. A Procedure for Quantitative Determination of Tris(2-Carboxyethyl)Phosphine, an Odorless Reduc-

ing Agent More Stable and Effective Than Dithiothreitol. *Anal. Biochem.* **1994,** *220,* 5–10.

34. Erdjument-Bromage, H.; Lui, M.; Lacomis, L.; Grewal, A.; Annan, R. S.; McNulty, D. E.; Carr, S. A.; Tempst, P. Examination of Micro-Tip Reversed-Phase Liquid Chromatographic Extraction of Peptide Pools for Mass Spectrometric Analysis. *J. Chromatogr. A.* **1998,** *826,* 167–181.

35. Gatlin, C. L.; Kleemann, G. R.; Hays, L. G.; Link, A. J.; Yates, J. R. III. Protein Identification at the Low Femtomole Level from Silver-Stained Gels Using a New Fritless Electrospray Interface for Liquid Chromatography-Microspray and Nanospray Mass Spectrometry. *Anal. Biochem.* **1998,** *263,* 93–101.

36. Eng, J. K.; McCormack, A. L.; Yates, J. R. III. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994,** *5,* 976–989.

37. Sadygov, R. G.; Eng, J. K.; Durr, E.; Saraf, A.; McDonald, W. H.; MacCoss, M. J.; Yates, J. R., III. Code Developments to Improve the Efficiency of Automated MS/MS Spectra Interpretation. *J. Proteome Res.* **2002,** *1,* 211–215.

38. Gras, R.; Muller, M.; Gasteiger, E.; Gay, S.; Binz, P-A.; Bienvenut, W.; Hoogland, C.; Sanchez, J-C.; Bairoch, A.; Hochstrasser, D. F.; Appel, R. D. Improving Protein Identification from Peptide Mass Fingerprinting Through a Parameterized Multi-Level Scoring Algorithm and an Optimized Peak Detection. *Electrophoresis* **1999,** *20,* 3535–3550.

39. Berndt, P.; Hobohm, U.; Langen, H. Reliable Automatic Protein Identification from Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Peptide Fingerprints. *Electrophoresis* **1999,** *20,* 3521–3526.

40. Breen, E. J.; Hopwood, F. G.; Williams, K. L.; Wilkins, M. R. Automatic Poisson Peak Harvesting for High Throughput Protein Identification. *Electrophoresis* **2000,** *21,* 2243–2251.

41. Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. Identification of the Components of Simple Protein Mixtures by High-Accuracy Peptide Mass Mapping and Database Searching. *Anal. Chem.* **1997,** *69,* 4741–4750.

42. Krause, E.; Wenschuh, H.; Jungblut, P. R. The Dominance of Arginine-Containing Peptides in MALDI-Derived Tryptic Mass Fingerprints of Proteins. *Anal. Chem.* **1999,** *71,* 4160–4165.

43. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999,** *20,* 3551–3567.

44. Zhang, W.; Chait, B. T. Profound: An Expert System for Protein Identification Using Mass Spectrometric Peptide Mass Mapping Information. *Anal. Chem.* **2000,** *72,* 2482–2489.

45. Clauser, K. R.; Baker, P.; Burlingame, A. L. Role of Accurate Mass Measurement (±10 ppm) in Protein Identification Strategies Employing MS or MS/MS and Database Searching. *Anal. Chem.* **1999,** *71,* 2871–2882.

46. Swart, R.; Koivisto, P.; Markides, K. E. Column Switching in Capillary Liquid Chromatography-Tandem Mass Spectrometry for the Quantitation of pg/ml Concentrations of the Free Basic Drug Tolterodine and its Active 5-Hydroxymethyl Metabolite in Microliter Volumes of Plasma. *J. Chromatogr. A* **1998,** *828,* 209–218.

47. Scheler, C.; Lamer, S.; Pan, Z.; Li, X-P; Salinikow, J.; Jungblut, P. Peptide Mass Fingerprint Sequence Coverage from Differently Stained Proteins on Two-Dimensional Electrophoresis Patterns by Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF). *Electrophoresis* **1998,** *19,* 918–927.

48. Wold, K. R. Identification of Common Post-Translational Modifications. *Protein Structure—A Practical Approach;* In

Creighton, T, Ed.; Oxford University Press: New York, 1997; 91–116.

49. Anderson, J. S.; Sogaard, M.; Svensson, B.; Roepstorff, P. Localization of an O-Glycosylated Site in the Recombinant Barley α-Amylase 1 Produced in Yeast and Correction of the Amino Acid Sequence Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Peptide Mixtures. *Biol. Mass Spectrom.* **1994,** *23,* 547–554.

50. Gatlin, C. L.; Eng, J. K.; Cross, S. T.; Detter, J. C.; Yates, J. R. III. Automated Identification of Amino Acid Sequence Variations in Proteins by HPLC/Microspray Tandem Mass Spectrometry. *Anal. Chem.* **2000,** *72,* 757–763.