JOURNAL OF MASS SPECTROMETRY
J. Mass Spectrom. 2007; **42**: 312–334
Published online 5 January 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jms.1163

JMS

Comprehensive analysis of proteins of pH fractionated samples using monolithic LC/MS/MS, intact MW measurement and MALDI-QIT-TOF MS

Chul Yoo,[2†] Tasneem H. Patwa,[2] Paweena Kreunin,[2] Fred R. Miller,[3] Christian G. Huber,[4] Alexey I. Nesvizhskii[5] and David M. Lubman[1,2,5,6*]

[1] Department of Surgery, The University of Michigan Medical Center, Ann Arbor, MI 48109, USA
[2] Department of Chemistry, The University of Michigan, Ann Arbor, MI 48109, USA
[3] Barbara Ann Karmanos Cancer Institute, Wayne State University, Detroit, MI 48201, USA
[4] Department of Chemistry, Instrumental Analysis and Bioanalysis, Saarland University, 66123 Saarbrucken, Germany
[5] Department of Pathology, The University of Michigan Medical Center, Ann Arbor, MI 48109, USA
[6] Comprehensive Cancer Center, The University of Michigan Medical Center, Ann Arbor, MI 48109, USA

Received 6 June 2006; Accepted 14 November 2006

A comprehensive platform that integrates information from the protein and peptide levels by combining various MS techniques has been employed for the analysis of proteins in fully malignant human breast cancer cells. The cell lysates were subjected to chromatofocusing fractionation, followed by tryptic digestion of pH fractions for on-line monolithic RP-HPLC interfaced with linear ion trap MS analysis for rapid protein identification. This unique approach of direct analysis of pH fractions resulted in the identification of large numbers of proteins from several selected pH fractions, in which approximately 1.5 μg of each of the pH fraction digests was consumed for an analysis time of *ca* 50 min. In order to combine valuable information retained at the protein level with the protein identifications obtained from the peptide level information, the same pH fraction was analyzed using nonporous (NPS)-RP-HPLC/ESI-TOF MS to obtain intact protein MW measurements. In order to further validate the protein identification procedures from the fraction digest analysis, NPS-RP-HPLC separation was performed for off-line protein collection to closely examine each protein using MALDI-TOF MS and MALDI-quadrupole ion trap (QIT)-TOF MS, and excellent agreement of protein identifications was consistently observed. It was also observed that the comparison to intact MW and other MS information was particularly useful for analyzing proteins whose identifications were suggested by one sequenced peptide from fraction digest analysis. Copyright © 2007 John Wiley & Sons, Ltd.

KEYWORDS: pH fractionation; intact protein MW; LC/MS/MS; MALDI-QIT-TOF; monolith

INTRODUCTION

Despite the tremendous development of techniques for the analysis of proteomes of various organisms, it is still challenging to reliably analyze highly complex biological mixtures such as human cancer cells in a high-throughput manner. Two-dimensional gel electrophoresis (2DE)[1] remains the most widely used method, followed by subsequent in-gel digestion for peptide mass fingerprinting (PMF) analysis typically by MALDI-TOF MS for protein identification.[2–4] This method has a number of limitations, including the difficulty of analyzing proteins of extreme size and hydrophobicity and poor run-to-run reproducibility.[5,6] Although robotic systems have been developed to assist excision of large numbers of gel spots, complete automated integration of the entire procedures involving 2DE to MS is still limited.

Shotgun proteomics, in which the whole cell lysate is digested for extended chromatographic separations for direct tandem mass spectrometric analysis, is one of the most widely used gel-free approaches for protein identifications. In this approach, sequence information gained from peptide fragment fingerprints (PFFs) and database searching with partial coverage of a protein sequence is often sufficient for identification.[7–10] A MudPIT approach, in which sequential ion exchange and reversed-phase high performance liquid chromatography (RP-HPLC) separations are required prior to MS analysis, has been widely applied to identify a large number of proteins in various organisms.[11–13] It has also been used to elucidate post-translational modifications (PTMs) of the samples of moderate complexity.[14] A recent study by Smith and co-workers[15–17] involved the analysis of whole cell lysates by on-line nanoscale RP-HPLC separation performed at a very high

*Correspondence to: David M. Lubman, University of Michigan Medical Center, Department of Surgery, MSRBI, A510B, Box 0658, Ann Arbor, MI 48109, USA. E-mail: dmlubman@umich.edu
†Amgen Inc., Chemistry Research and Discovery, Analytical Chemistry, One Amgen Center Drive, Mail Stop 25-2-A, Thousand Oaks, CA 91 320.

pressure of *ca* 20 kpsi interfaced with Fourier transform ion cyclotron resonance (FT-ICR) MS obtained chromatographic peak capacities of $>10^3$. Another interesting non-gel-based approach was attempted, in which the yeast cell lysates were subjected to pI-based rotofor fractionation, followed by enzymatic digestion for on-line analysis by HPLC/MS/MS using a C18 packed column for over 100 min of separation.[18] This study showed the chromatographic separations of large numbers of tryptic peptides resulting from proteins in pH fractions. Although shotgun proteomics performed exclusively at the peptide level is a highly effective means for rapid protein identifications in global scale studies, it is difficult to assess valuable information contained at the protein level, which becomes lost upon enzymatic digestion, such as sequence variations of proteins resulting from splice variants and truncations that add further complexity to the proteomes.

Recently, a method based on two-dimensional (2D) liquid-phase fractionations has been developed and successfully applied to the analysis of human cancer cells of various types as well as simple organisms,[19–23] in which each of the proteins was collected for protein identifications on the basis of peptide mapping. Chromatofocusing (CF)[24] as used for the first dimension separation is an effective approach for prefractionating complex samples prior to further analysis due to its reproducibility and its compatibility with RP separation for on-line MS analysis to obtain accurate intact protein molecular weight (MW) values.

In this work, the proteins in fully malignant human breast cancer cell lines were isolated into pH fractions by CF, in which several selected fractions were enzymatically digested for tandem MS analysis using on-line monolithic capillary HPLC to rapidly obtain sequencing information about large numbers of peptides for protein identification. The same pH fractions were also subjected to comprehensive analysis for intact protein MW, PMF, and both ESI- and MALDI-based PFF of digests of proteins collected off-line from RP-HPLC to further validate the protein identification procedures from fraction digest analysis. By using a number of methods together at the protein and peptide levels, the reliability of the protein identification procedures were enhanced and provided information on the mature forms of several proteins, which is difficult to achieve solely by peptide sequencing analysis.

EXPERIMENTAL

The experimental overview is illustrated in Fig. 1. Extracts from a human breast cancer cell line were separated by

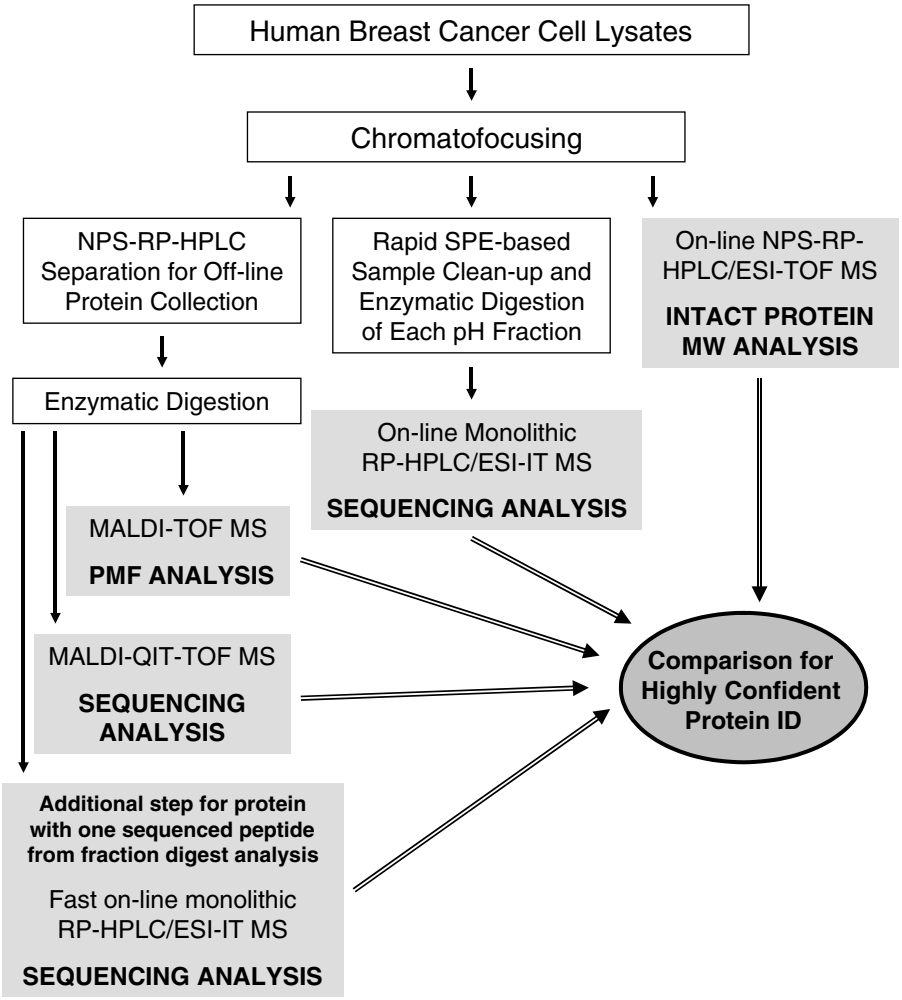

Figure 1. Overall experimental scheme of pH fraction digestion analysis and the comparison to other mass spectrometry techniques.

DOI: 10.1002/jms

CF in the range of pH 4 to 7. Each of the pH fractions was purified by solid phase extraction (SPE) and tryptically digested for on-line analysis by monolith-based capillary HPLC interfaced with a linear ion trap MS. Also, the same pH fractions were separated by nonporous (NPS)-RP HPLC for intact protein MW determination and off-line peak collection for closer examinations. Upon off-line protein collection, several proteins, for which only one sequenced peptide resulted from the analysis of fraction digests, were subjected to on-line monolithic LC/MS/MS for confirmation. The results from different MS approaches were compared.

Sample preparation

The sample used in this experiment was a fully malignant human breast cancer cell line, CA1a.cl1, prepared from a cloned variant of the MCF10 series (Barbara Ann Karmanos Cancer Institute, Wayne State University, Detroit, MI).[25] Cells were mixed with a lysis buffer containing 7 M urea, 2 M thiourea, 100 mM dithiothreitol (DTT), 2% *n*-octyl *β*-D-glucopyranoside (OG), 10% glycerol, 10 mM sodium orthovanadate, 10 mM sodium fluoride (all from Sigma, St. Louis, MO), 0.5% Biolyte ampholyte (Bio-Rad, Hercules, CA), and the protease inhibitor cocktail (Roche Diagnostics, GmBH, Mannheim, Germany) for vortexing at room temperature for 1 h. The cellular debris and other insoluble materials were removed by centrifuging the mixture at 35 000 rpm for 1 h and 15 min. The supernatant was collected for buffer exchange to replace the lysis buffer with the equilibration buffer for CF using a PD-10 G-25 column (Amersham Biosciences, Piscataway, NJ). The protein concentration was determined using the Bradford Protein Assay kit with bovine serum albumin (BSA, Bio-Rad) standard.

Chromatofocusing

The CF experiment was performed using a Beckman System Gold model 127 pump and 166 UV detector module (Beckman Coulter, Fullerton, CA) with an HPCF-1D prep column (250 mm × 4.6 mm i.d., Eprogen, Darien, IL). A linear pH gradient was generated using a combination of a start buffer (SB) composed of 6 M urea, 25 mM BisTris, and 0.2% OG and elution buffer (EB) containing 6 M urea, 0.2% OG, and 10% polybuffer 74 (Amersham Biosciences). Saturated iminodiacetic acid (Sigma) was used to adjust the pH of SB at 7.2 and EB at 3.9. The column was first equilibrated in SB until the pH of the column was the same as SB by monitoring with a postdetector on-line assembly of a pH flow cell (Lazar Research Laboratories, Los Angeles, CA). After equilibration, *ca* 3 mg of the sample was loaded onto the column at a low flow rate to allow for interactions of the proteins with the binding sites. Once a baseline was achieved, solvent flow was switched to EB and the flow rate was set to 1 ml/min for CF fraction collection at intervals of 0.2 pH units along the linear gradient, where the elution profile was recorded at 280 nm. At the end of the gradient, the column was flushed with 1 M sodium chloride (Sigma) to remove any proteins still bound to the column. All collected samples were stored at −80 °C until further analysis.

NPS-RP-HPLC protein separation and tryptic digestion

The proteins fractionated by CF were further separated by an NPS-RP-HPLC column (4.6 mm i.d. × 33 mm, Eprogen) packed with 1.5 μm C18 ODSIIIE silica beads using an HPLC system Gold (Beckman Coulter). Approximately 80 μg of protein was loaded for the separation utilizing the following gradient, in which solvents A and B comprised 0.1% trifluoroacetic acid (TFA, Sigma) in DI water and acetonitrile (ACN, Sigma), respectively; 5–15% B in 1 min, 15–25% B in 2 min, 25–31% B in 3 min, 31–41% B in 10 min, 41–47% B in 3 min, 47–67% B in 4 min, 67–100% B in 1 min, 100% B in 2 min, and 100–5% B in 1 min at a flow rate of 0.5 ml/min. The protein separation was monitored at 214 nm for off-line collection of *ca* 40 fractions from each separation using a fraction collector (SC-100, Beckman Coulter) controlled by a semiautomated acquisition program built in-house. The collected proteins were dried completely by a SpeedVac (model SC210A, Thermo Electron, Marietta, OH) and subjected to enzymatic digestion by adding 50 μl of 50 mM ammonium bicarbonate (Sigma) and 0.5 μg of TPCK-modified sequencing-grade trypsin (Promega, Madison, WI) for incubation at 37 °C for 18 h. The digests were added with 1 μl of 10% TFA to stop digestion and stored at −80 °C until further analysis.

NPS-RP-HPLC/ESI-TOF MS

NPS-RP-HPLC separation was also interfaced on-line with ESI-TOF MS (LCT, Waters-Micromass, Manchester, U.K.) for intact protein MW analysis by injecting *ca* 100 μg of proteins using the same separation condition described in the previous section, except that 0.1% TFA was replaced with 0.3% formic acid (Sigma). A flow splitter was used to deliver 40% of the eluent to the LCT with the following parameters: capillary voltage at 3200 V, sample cone voltage at 40 V, and extraction cone voltage at 3 V. The desolvation temperature was set at 300 °C with a nitrogen gas flow of 650 l/h, while the source temperature was at 120 °C. The intact MW was obtained by deconvolution utilizing MaxEnt1 of the MassLynx software version 4.0 (Waters-Micromass).

MALDI-TOF MS and data analysis

The proteins previously collected by NPS-RP-HPLC for tryptic digestion were desalted using 2 μm C18 ZipTips (Millipore, Bedford, MA) resulting in concentrated peptide mixtures in 5 μl of 60% ACN and 0.5% TFA. A matrix solution was prepared by diluting saturated *α*-cyano-4-hydroxycinnamic acid (*α*-CHCA, Sigma) solution with 60% ACN and 0.1% TFA at 1 : 4 ratio added with the internal standards, including angiotensin I, adrenocorticotropic hormone (ACTH) fragment 1–17, and ACTH 18–39 (all from Sigma). The matrix solution was spotted on the MALDI-plate so that each well contained 50 fmol of each of the internal standards, followed by layering 1 μl of the desalted sample on top.

The Micromass TofSpec 2E was used for the MALDI-TOF MS analysis in the reflectron mode with a nitrogen laser (337 nm) as the ionization source. The instrument was

operated in the positive ion mode with an operating voltage of 20 kV, an extraction voltage of 19.98 kV, and a pulse voltage of 2300 V. All spectra acquired over the mass range of 500 to 4000 Da were combined for internal calibration and post-processed using the MassLynx software to obtain monoisotopic peptide masses for submission to the MS-Fit search engine at http://prospector.ucsf.edu. The search was performed against the SwissProt database under the species of *Homo sapiens* by allowing the following parameters: one missed cleavage, mass tolerance of 50 ppm or less, no limitations set for MW and pI ranges, and possible modifications including *N*-terminal Gln to pyroGlu, oxidation of M, *N*-terminal acetylation, and phosphorylation of S, T, and Y. The search results were filtered using the following threshold: MOWSE score of $>10^3$ and sequence coverage of >20%.

MALDI-quadrupole ion trap (QIT)-TOF MS and data analysis

The MALDI-QIT-TOF MS (AXIMA-QIT, Shimadzu-Biotech, Manchester, UK) was used to perform MS/MS of selected peptides. Samples were prepared in a manner identical to those for PMF. Peptide mixtures (0.5 μl) desalted by C18 ZipTips were deposited on the MALDI target plate along with 0.5 μl of the matrix solution of 2,5-dihydroxybenzoic acid (DHB; Sigma) at a concentration of 20 mg/ml in 0.1% TFA and 60% ACN. The MALDI-QIT-TOF MS instrument was externally calibrated using a mixture of Bradykinin fragment 1–7, angiotensin II, P14R, ACTH fragment 18–39, and Insulin chain B (all from Sigma) to obtain the mass tolerance of 10 ppm. The standard instrument settings for optimum transmission at medium mass were used to record all mass spectra in this work. Data acquisition and processing were performed using the Kompact LAUNCHPAD software (Shimadzu) and the ion masses were submitted to the Mascot MS/MS ions search (http://www.matrixscience.com), where the search was performed under SwissProt database with the following parameters: (1) species: *Homo sapiens*, (2) one missed cleavage, (3) possible modifications of peptide *N*-terminal Gln to pyroGlu, oxidation of M, protein *N*-terminal acetylated and phosphorylation of S, T, and Y, (4) peptide mass tolerance of ±1 Da, (5) MS/MS tolerance of ±0.9 Da, and (6) peptide charge of +1.

pH fraction cleanup and enzymatic digestion of pH fractions

A short NPS-RP-HPLC column (14 mm × 4.6 mm i.d., Eprogen) packed with 1.5 μm silica ODS I was utilized for the removal of contaminants from each of the pH fractions obtained from CF separation, which could potentially interfere with enzymatic digestion and electrospray ionization. Approximately 20 μg of proteins from the CF was injected and the eluent was collected during a rapid gradient of 0 to 100% B in 2 min, followed by a 2 min hold at 100% B, where the solvent system and other HPLC configurations were the same as in NPS-RP-HPLC separation. The eluent was completely dried by a SpeedVac.

Trypsin (2 μg) was added to 10 μl of 10 mM DTT (Sigma) and 90 μl of 50 mM ammonium bicarbonate to each of the dried pH fractions for overnight incubation at 37 °C. The digestion mixtures were completely dried down by a SpeedVac for storage at −80 °C. Each of the samples was reconstituted in 20 μl of HPLC grade water (Fisher Scientific, Hanover Park, IL) prior to LC/MS/MS analysis.

Monolith-based ESI-LC/MS/MS and data analysis

The Ultra-Plus II MD capillary pump module (Micro-Tech Scientific, Vista, CA) was used for the separation of digested pH fractions by a monolithic capillary column. A monolithic column (360 μm o.d. × 200 μm i.d. × 60 mm) was prepared in-house by copolymerizing styrene and divinylbenzene (PS/DVB) according to the procedure described elsewhere.[26] The solvent system comprised two solvents A and B, in which 0.05% formic acid was added to HPLC grade water and ACN, respectively. The capillary column was directly connected to a micro-injector with 500 nl internal sample loop (Valco Instruments, Houston, TX) and the flow was split pre-column to generate *ca* 2.5 μl/min. The separation was controlled at 60 °C using a linear gradient of 0 to 10% B in 1 min, 10 to 50% B in 49 min, and 50 to 100% B in 10 min by loading approximately 1.5 μg of each of the pH fraction digests.

A linear ion trap MS (LTQ, Thermo Finnigan, San Jose, CA) was used to sequence the tryptic peptides from pH fraction digests. A monolithic column was connected to the LTQ with a fused silica capillary tubing of 20 μm i.d. The capillary transfer tube was set at 175 °C and the ESI voltage at +4 kV. A sheath nitrogen gas flow of 12 arbitrary units was used and ion activation was achieved with ultra-high purity helium (all from Cryogenic Gases, Detroit, MI) at a normalized collision energy of 35%. All MS/MS spectra were analyzed by the TurboSequest of Bioworks software version 3.1 SR1 (Thermo Finnigan) with SwissProt database under the species of *Homo sapiens*. The database search allowed a maximum number of missed cleavages of two, and all search results were subjected to manual inspection to consider fully tryptic peptides assigned with *X*corr values of the following: ≥1.5 for singly charged ions; ≥2.5 for doubly charged ions; and ≥3.5 for triply charged ions, while no ions at higher charged states were considered. Also, $\Delta C_n \geq 0.1$ was considered regardless of the charge states. Additionally, an X!Tandem (http://human.thegpm.org/tandem/thegpm_tandem.html) search was also performed using default parameters for ESI-IT MS. All peptides with a $\log(e)$ of < -1 were retained.[27]

Monolith-based ESI-LC/MS/MS for proteins collected off-line

Several proteins whose identifications were suggested by only one sequenced peptide from fraction digest analysis were further analyzed by off-peak collection of NPS-RP-HPLC. The tryptic digests of each individual protein were analyzed by rapid monolithic LC/MS/MS by applying a gradient of 0 to 100% B in 18 min, while all experimental platform and database search procedures were the same as those in the pH fraction digest analysis by monolithic LC/MS/MS.

DOI: 10.1002/jms

JMS

RESULTS AND DISCUSSION

Experimental platform: liquid-phase separations and combination of different MS techniques

In this experiment, the complexity of the protein mixture obtained from the human breast cancer cells was reduced by prefractionating at 0.2 pH unit intervals using CF based on weak anion exchange in the pH range of 4 to 7 prior to further analysis. A typical profile of the CF fractionation is shown in Fig. 2, in which the experimental pH is monitored in real time. It is shown that CF is an effective approach to isolate proteins, as suggested by the linearity of the pH gradient throughout separation, where a correlation coefficient, or r^2, of 0.9961 was obtained in the pH range of 4 to 7. In addition, the proteins are collected in the liquid phase and therefore more readily compatible for further RP-HPLC separation and ESI-based MS analysis than the traditional 2DE method. It is important to note that this feature can help obtain the intact protein MW values through on-line ESI-TOF MS analysis to enable proteome analysis at the protein level. Also, it has previously been shown that the comparison between experimental and theoretical pI values of proteins can provide a unique means to suggest the presence of potential modifications.[28]

As shown from the overall workflow in Fig. 1, various MS techniques have been performed in this study to comprehensively and reliably analyze proteins in each of the pH fractionated samples for comparisons, to avoid ambiguous identifications, and also to further validate the identification procedures. In addition to protein identifications obtained by off-line fraction collections from NPS-RP-HPLC separations for PMF and sequencing analyses, each of the pH fractionated samples was also digested using trypsin for monolithic capillary HPLC separation directly interfaced with linear ion trap MS to rapidly obtain protein identifications. The use of monolithic capillary columns has recently become a popular approach in separating various biological molecules[29–32] owing to its outstanding stability at extreme run conditions for the analysis of peptide mixtures.[33] The unique separation characteristics provided by polymer-based monolithic columns over conventional packed silica columns, such as high separation efficiency, high resolution, and excellent recovery, allowed for their wide use in many applications.

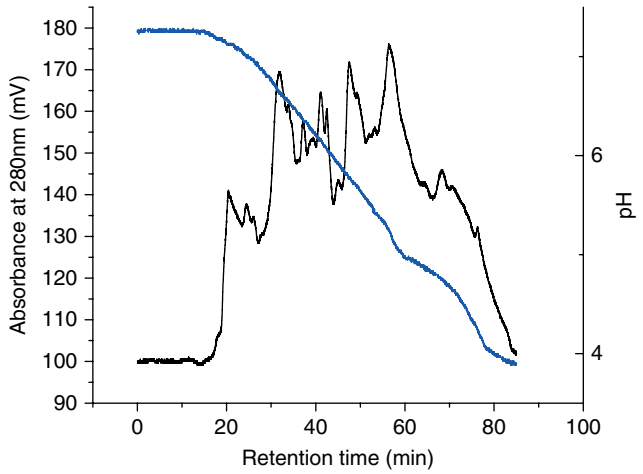

Figure 2. Chromatofocusing separation profile of human breast cancer cell line CA1a monitored at 280 nm.

Its potential usefulness in the study of peptide mapping[34] and PTMs[35] has also been described recently.

Representative TIC chromatograms obtained from tryptic digests of several pH fractions in the pH range of 4.4 to 5.2 are shown in Fig. 3, where only *ca* 1.5 μg of each of the fraction digests was required in this approach, utilizing a short length of the monolithic column. A gradient elution of 0 to 50% B in *ca* 50 min was applied and typical peak widths of a few seconds were observed. Given the complex nature of the samples studied in this experiment, the separation speed is still considered to be relatively fast, whereas the typical shotgun approach utilizing C18 packed column requires several hours of RP-HPLC separation. It is expected, though, that the tryptic peptides with less ionization efficiency might be suppressed from closely eluting peptides of higher abundance and may not be properly isolated for ion activation during this fast separation time. The problem of ion suppression may be alleviated by using shallower gradients at the cost of an increase in the analysis time. Nevertheless, it was observed that the high separation efficiency provided by PS/DVB monolithic capillary columns helped minimize this problem and allowed the detection of sufficient numbers of peptides for sequencing to identify large numbers of proteins.

Comprehensive analysis of proteins by different MS approaches

The tryptic digests of several pH fractions from CA1a cell lysates were subjected to linear ion trap MS for sequencing analysis (Fig. 3), where the same pH fractions were also subjected to protein separation by NPS-RP-HPLC of analytical scale for on-line ESI-TOF MS for intact MW determinations for direct comparison. Figure 4(A) shows NPS-RP-HPLC chromatograms obtained from the combined pH fractions of 4.8–5.0 and 5.0–5.2, where *ca* 80 μg of protein was loaded. Table 1 shows the overall summary of the comparison of the results for a selected set of proteins obtained from these four different MS techniques, where their excellent agreement suggests that identifications of these proteins are highly reliable. The proteins analyzed from fraction digests by monolithic LC/MS/MS alone are considered highly confident owing to the multiple numbers of fully tryptic peptides successfully sequenced with high *Xcorr*, but peptides sequenced with slightly lower *Xcorr* than the set criteria, but still significant, are also presented. Considering that numerous studies generally define confident protein identifications based on two or more of either fully or partially tryptic sequenced peptides,[13] many of these proteins exceeded the highly stringent criteria for reliable identification[36] with several peptides used to identify a protein.

The NPS-RP-HPLC/ESI-TOF MS analysis was also performed to obtain accurate intact protein MW values of the same pH fractions. It has been previously reported that the integration of MS information obtained at the levels of proteins and peptides successfully resulted in unambiguous identifications of proteins in several different types of human cancer cells,[19,20,22] in which intact protein MW values helped to confirm the presence of multiple

DOI: 10.1002/jms

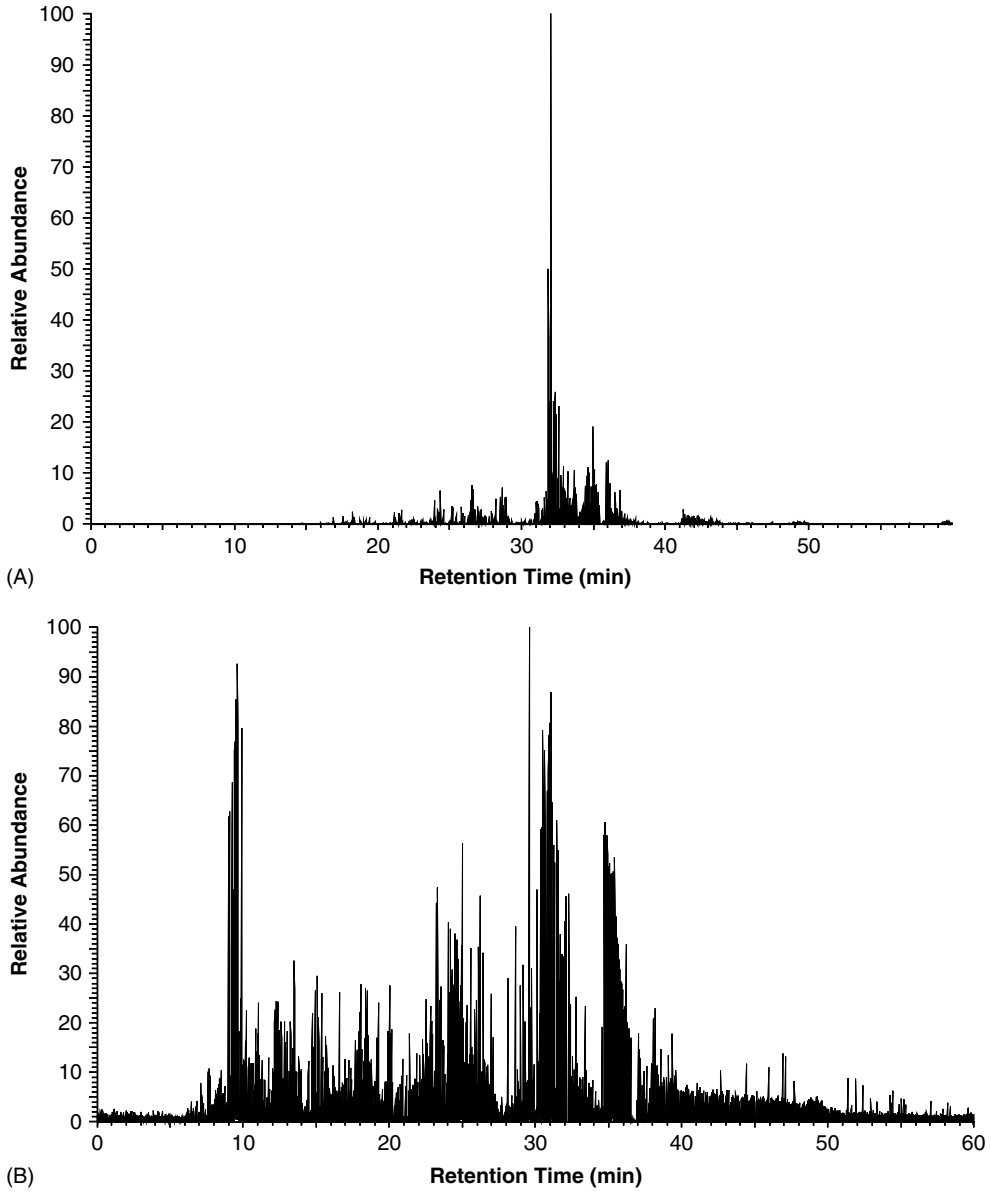

Figure 3. Monolithic capillary HPLC/MS chromatograms (TIC) for pH fraction digest of (A) 4.4–4.6, (B) 4.6–4.8, (C) 4.8–5.0, and (D) 5.0–5.2 from CA1a.

isoforms or truncated versions of a protein. It is suggested that intact MW values can provide an excellent means to further help confirm the identification of proteins and to provide complementary MS information when limited PMF or PFF information from fraction digest analysis is available. Table 1 shows that the theoretical and experimental MW values of most of the proteins match within 500 Da or better. The 500-Da window was arbitrarily chosen to account for possible modifications including minor truncations, phosphorylations, etc. on the basis of the data obtained.[19] The comparison of intact MW values is also essential for suggesting the presence of PTMs or other important sequence modifications, such as truncation, as observed for several proteins, including the stress-70 protein, protein disulfide isomerase A3 precursor, ATP synthase D chain, and heat shock protein 60 kDa, in which the experimental MW values exhibited significant deviations from their theoretical MW values. The experimental MW values of these proteins were found to closely match when the loss of transit peptides[37] was taken into account. It is of particular importance to emphasize that the confident identification of these proteins in their mature forms was possible owing to the integration of intact protein MW information. A shotgun proteomic approach, in which only partial sequence coverage is used for protein identification, cannot suggest these modifications and provides little information about the mature forms of proteins including PTMs, splice variants, truncations, and isoforms.[38]

In order to examine the validity of utilizing protein identification based on fraction digest analysis and intact protein MW values, a subset of proteins identified in the present work were collected off-line for detailed MALDI-MS and -MS/MS analyses. This facilitated a closer validation of protein identification, especially where the retention time of each protein was available for direct comparison. The results obtained from MALDI-MS and -MS/MS (Table 1) show that

DOI: 10.1002/jms

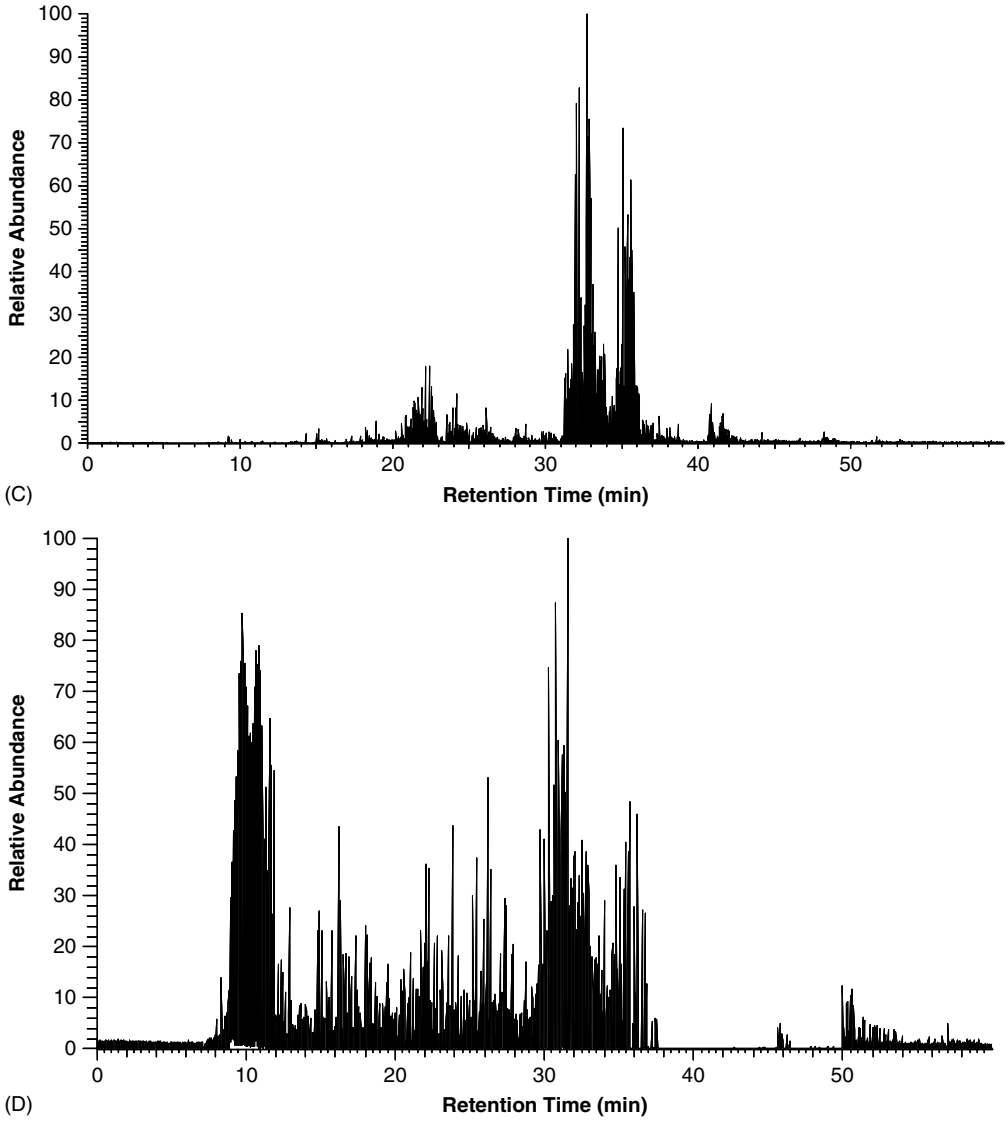

Figure 3. (*Continued*) .

the proteins identified by off-line peak collection were in excellent agreement with those identified by fraction digest analysis.

The MALDI-QIT-TOF MS analysis of proteins collected off-line from NPS-RP-HPLC resulted in PFF information for tryptic peptides that were not identified by fraction digest analysis, perhaps owing to their different ionization efficiency in the electrospray processes. For example, fraction digest analysis by ESI-MS/MS detected nine peptides for the heat shock protein 60 kDa and MALDI-MS/MS resulted in identification of eight peptides, five of which were not observed in ESI-MS/MS. Overall, they have been combined for a total of fourteen unique sequenced peptides, corresponding to a combined coverage of *ca* 45% for this large protein, resulting in a highly confident identification. In some cases, MALDI-MS/MS resulted in more sequencing information than ESI-MS/MS, as observed for cytokeratin 17, in which ten unique peptides were successfully identified, as opposed to only two found by ESI-MS/MS. Although it is difficult to compare the performance of MALDI-MS/MS and ESI-MS/MS owing to their different ionization mechanisms and the different amounts of sample consumed, it appears that the complementary nature of these two different ionization methods reported for peptide mapping applications elsewhere[34] may also apply to tandem MS analysis.

The reliability of the protein identifications by fraction digest analysis was further supported by the PMF analysis from MALDI-TOF MS. PMF analysis in Table 1 often exhibited sufficiently high sequence coverage, providing an excellent means to suggest the identification of proteins. However, less confident identifications can often result from limited sequence coverage due to many factors, including possible sample loss prior to spotting from sample cleanup, varying ionization efficiencies for different peptides with different matrices, and difficulties of detecting peptides in the low mass range. In this experiment, excellent agreement with the protein identifications provided by PFF information prevents possible false positive identifications from PMF analysis.

All the proteins identified from the pH fractions of 4.8–5.0 and 5.0–5.2 and confirmed by different MS approaches are annotated in Fig. 4(A). Likewise, the same experimental workflow was applied to the fractions of higher pH range, including 6.0–6.2 and 6.2–6.4, and the results are

DOI: 10.1002/jms

Table 1. Summary of four different MS analyses for several proteins from the pH fraction of 4.8–5.0 and 5.0–5.2 combined (only fully tryptic peptides considered for MS/MS analyses). Proteins are listed according to the elution order shown in Fig. 4(A)

	Monolithic HPLC/MS/MS of pH fraction digest					MALDI-TOF MS	MALDI-QIT-TOF MS				ESI-TOF MS		
		$[M+H]^+$								Theo.		
Protein name (Accession no.)	Peptide sequenced (charge state; amino acid no.)	Theo.	Exp.	*Xcorr*	ΔCn	PMF Cov. (%)	Peptide sequenced (amino acid no.)	M_r (exp)	Observed	Protein Score	MW[a]	pI	Exp. MW
Splicing factor, arginine/serine-rich 3 (P84103)	NPPGFAFVEFEDPR (+2; 44–57)	1621.77	1622.55	3.918	0.540	34	AFGYYGPLR (29–37)	1042.5927	1043.6000	49	19 318	11.64	20 290
							NPPGFAFVEFEDPR (44–57)	1620.2928	1621.3000				
							GPPPSWGR (91–98)	852.4927	853.5000				
Heterogeneous nuclear ribonucleoprotein (HNRNP) D0 (Q14103)	MFIGGLSWDTTK (+2; 99–110)	1355.67	1355.23	4.123	0.616	20	MFIGGLSWDTTK (99–110)	1354.6927	1355.7000	52	38 411	7.61	38 526
	GFGFVLFK (+1; 139–146)	914.51	914.26	1.822	0.371		GFGFVLFK (139–146)	913.4927	914.5000				
	IFVGGLSPDTPEEK (+2; 184–197)	1488.76	1488.54	4.226	0.524		IFVGGLSPDTPEEK (184–197)	1487.8927	1488.9000				
Keratin, cytoskeletal 8 (CK8) (P05787)	LEAELGNMQGLVEDFK (+2; 160–175)	1792.88	1792.35	5.001	0.474	48	SNMDNMFE SYINNLR (133–147)	1846.6927	1847.7000	74	53 511	5.52	53 580
	TEMENEFVLIK (+2; 186–196)	1352.68	1352.31	3.777	0.439		QLYEEEIR (225–232; Pyro-glu)	1061.4927	1062.5000				
	LEGLTDEINFLR (+2; 213–224)	1419.75	1420.16	4.075	0.524		LEGLTDEINFLRQLYEEEIR (213–232)	1418.6927	1419.7000				
	SLDMDSIIAEVK (+2; 252–263)	1320.67	1320.36	3.938	0.560		LALDIEIATYR (381–391)	1276.7928	1277.8000				
ATP synthase D chain O75947)	TIDWVAFAEIIPQNQK (+2; 9–24)	1872.99	1872.03	4.101	0.302	60	TIDWVAFAEIIPQNQK (9–24)	1871.8927	1872.9000	170	18 349	5.22	17 524
	LAALPENPPAIDWAYYK (+2; 41–57)	1931.99	1932.11	3.926	0.606		SWNETLTSR (32–40)	1092.4927	1093.5000				
	NLIPFDQMTIEDLNEAFPETK (+2; 123–143)	2465.19	2466.28	4.551	0.608		LAALPENPPAIDWAYYK (41–57)	1930.9927	1932.0000				
							AGLVDDFEK (63–71)	992.4927	993.5000				
							KYPYWPHQPIENL (148–160)	1683.8927	1684.9000				
							YPYWPHQPIENL (149–160)	1555.7928	1556.8000				

DOI: 10.1002/jms

Protein disulfide isomerase A3 precursor (P30101)	DASIVGFFDDSFSEAHSEFLK (+2; 153–173)	2348.07	2349.15	3.451	0.526	39	FVMQEEFSR (336–344)	1171.4927	1172.5000	42	56747	5.98	54307
	TFSHELSDFGLESTAGEIPVVAIR (+3; 306–329)	2575.30	2575.77	2.791	0.400		ELSDFISYLQR (472–482)	1369.5927	1370.6000		*(54214)*		
	FLQDYFDGNLK (+2; 352–362)	1359.66	1360.39	4.104	0.449								
	VVVAENFDEIVNNENK (+2; 380–395)	1832.90	1833.67	2.742	0.419								
	ELSDFISYLQR (+2; 472–482)	1370.70	1371.51	3.393	0.544								
Lamin A/C (P02545)	DLEALLNSK (+1; 136–144)	1002.55	1002.52	2.634	0.311	50	TLEGELHDLR (157–166)	1181.6927	1182.7000	91	74095	6.57	74139
	NSNLVGAAHEELQQSR (+2; 281–296)	1752.86	1753.19	5.137	0.609		NSNLVGAAHEELQQSR (281–296)	1751.7928	1752.8000				
	IDSLSAQLSQLQK (+2; 299–311)	1430.79	1430.84	3.362	0.337		LALDMEIHAYR (367–377)	1330.6927	1331.7000				
	MQQQLDEYQELLDIK (+2; 352–366)	1893.93	1894.00	5.499	0.516		SNEDQSMGNWQIK (458–470)	1536.5927	1537.6000				
	SNEDQSMGNWQIK (+2;458–470)	1536.68	1536.78	4.357	0.414		SVGGSGGGSFGDNLVTR (628–644)	1565.7928	1566.8000				
	QNGDDPLLTYR (+2; 472–482)	1291.63	1292.76	2.209	0.135								
	TALINSTGEEVAMR (+2; 528–541)	1491.75	1491.84	5.071	0.638								
	SVGGSGGGSFGDNLVTR (+2; 628–644)	1566.75	1566.83	4.302	0.568								
Alpha enolase (P06733)	GNPTVEVDLFTSK (+1; 15–27)	1406.72	1407.09	2.694	0.213	22	GNPTVEVDLFTSK (15–27)	1405.6927	1406.7000	83	47009	6.99	47374
	AAVPSGASTGIYEALELR (+2; 32–49)	1804.94	1805.99	3.286	0.444		AAVPSGASTGIYEALELR (32–49)	1802.9927	1804.0000				
	LMIEMDGTENK (+2; 92–102)	1280.59	1280.63	3.954	0.370		LAMQEFMILPVGAANFR (162–178)	1905.9927	1907.0000				

(*continued overleaf*)

Table 1. (*Continued*)

Protein name (Accession no.)	Monolithic HPLC/MS/MS of pH fraction digest: Peptide sequenced (charge state; amino acid no.)	[M + H]$^+$ Theo.	[M + H]$^+$ Exp.	*Xcorr*	Δ*Cn*	MALDI-TOF MS: PMF Cov. (%)	MALDI-QIT-TOF MS: Peptide sequenced (amino acid no.)	M_r (exp)	Observed	Protein Score	ESI-TOF MS: Theo. MW[a]	Theo. pI	Exp. MW
	DYPVVSIEDPFDQDDWGAWQK (+2; 285–305)	2510.12	2510.28	4.743	0.536		VVIGMDVAASEFFR (239–252)	1539.7928	1540.8000				
Keratin, type II cytoskeleta 7 (CK7)	PGGLGSSSLYGLGASR (+2; 30–45)	1478.76	1479.04	3.003	0.500	58	VDALNDEINFLR (214–225)	1418.6927	1419.7000	44	51 256	5.50	51 418
(P05787)	LPDIFEAQIAGLR (+2; 136–148)	1442.80	1442.92	3.130	0.450		LALDIEIATYR (382–392)	1276.6927	1277.7000				
	TLNETELTELQSQISDTSVVLSM DNSR (+2; 226–252)	3010.45	3010.76	3.413	0.626								
	SLDLDGIIAEVK (+2; 253–264)	1272.71	1272.87	3.515	0.305								
Stress-70 protein (GRP75) (P38646)	VIAVYDLGGGTFDISILEIQK (+2; 239–259)	2251.22	2252.19	6.013	0.616	34	NAVITVPAYFNDSQR (188–202)	1693.9927	1695.0000	78	73 635	5.87	68 777
	STNGDTFLGGEDFDQALLR (+2; 266–284)	2055.96	2056.63	3.484	0.489		STNGDTFLGGEDFDQALLR (266–284)	2055.8926	2056.8999		*(68 699)*		
	AQFEGIVTDLIR (+2; 349–360)	1361.74	1362.51	4.350	0.488		VQQTVQDLFGR (395–405)	1289.6927	1290.7000				
	VQQTVQDLFGR (+2; 395–405)	1290.68	1291.77	3.381	0.512		LLGQFTLIGIPPAPR (499–513)	1591.8927	1592.9000				
Keratin, type I cytoskeletal 18 (CK18)	SLGSVQAPSYGAR (+2; 14–26)	1292.66	1292.01	4.120	0.356	35	AQIFANTVDNAR (137–148)	1318.5927	1319.6000	119	47 898	5.34	47 972
(P05783)	GGMGSGGLATGIAGGLAGMGG IQNEK (+2; 55–80)	2261.10	2260.60	5.725	0.714		YETELAMR (167–174)	1011.4927	1012.5000				
	LQLETEIEALKEELLFMK (+2; 196–213)	2177.18	2176.72	5.466	0.695		AQYDELAR (253–260)	964.4927	965.5000				
	TVQSLEIDLDSMR (+2; 301–313)	1506.75	1506.15	2.550	0.334		TVQSLEIDLDSMR (303–313)	1505.6927	1506.7000				
	YALQMEQLNGILLHLESELAQTR (+3; 330–352)	2670.39	2670.75	4.049	0.472		QAQEYEALLNIK (358–369)	1401.8927	1402.9000				
							LEAEIATYR (372–380)	1064.5927	1065.6000				
Keratin, type I cytoskeletal 17 (CK17) (Q04695)	DAEDWFFSK (+1; 269–277)	1144.50	1144.14	2.137	0.438	32	IRDWYQR (115–121)	1035.5927	1036.6000	104	47 946	4.97	47 971
	LLEGEDAHLTQYK (+2; 386–398)	1516.76	1516.31	4.697	0.471		DWYQR (117–121)	766.3927	767.4000				

							DYSQYYR (129–135)	993.3927	994.4000			
							LAADDFR (163–169)	806.3927	807.4000			
							TKFETEQALR (170–179)	1221.5927	1222.6000			
							LSVEADINGLR (180–190)	1186.5927	1187.6000			
							VLDELTLAR (192–200)	1028.5927	1029.6000			
							DAEDWFFSK (269–277)	1143.4927	1144.5000			
							ASLEGNLAETENR (321–333)	1402.6927	1403.7000			
							LEQEIATYR (376–384)	1121.5927	1122.6000			
β-Actin, cytoplasmic (P60709)	VAPEEHPVLLTEAPLNPK (+3; 96–113)	1954.06	1954.32	4.240	0.600	53	IWHHTFYNELR (85–95)	1514.6927	1515.7000	126	41 737	5.29	41 664
	MTQIMFETFNTPAMYVAIQAVLSL YASGR (+3; 119–147)	3253.61	3254.48	5.260	0.627		VAPEEHPVLLTEAPLNPK (96–113)	1952.9927	1954.0000				
							GYSFTTTAER (197–206)	1131.5927	1132.6000				
							QEYDESGPSIVHR (360–372)	1498.5927	1499.6000				
Heat shock protein 60 kDa (HSP60)	TVIIEQSWGSPK (+1; 61–72)	1344.72	1344.78	2.555	0.438	56	ALMLQGVDLLADAVAVTMGPK (38–58)	2112.0928	2113.1001	94	61 017	5.70	58 308
(P10809)	LVQDVANNTNEEAGDGTTTAT VLAR (+2; 97–121)	2560.25	2560.43	4.623	0.625		TVIIEQSWGSPK (61–72)	1343.7928	1344.8000		*(57 910)*		
	GVMLAVDAVIAELK (+1; 143–156)	1428.81	1428.88	2.356	0.175		GANPVEIR (134–141)	854.4927	855.5000				
	TLNDELEIIEGMK (+2; 206–218)	1504.76	1505.03	4.689	0.511		GVMLAVDAVIAELK (143–156)	1427.7928	1428.8000				
	KPLVIIAEDVDGEALSTLVLNR (+2; 269–290)	2365.33	2365.67	3.718	0.516		ISSIQSIVPALEIANAHR (251–268)	1917.9927	1919.0000				
	DMAIATGGAVFGEEGLTLNLEDVQ PHDLGK (+3; 315–344)	3097.52	3098.60	5.086	0.587		KPLVIIAEDVDGEALSTLVLNR (269–290)	2364.2928	2365.3000				
	IQEIIEQLDVTTSEYEK (+2; 371–387)	2038.02	2038.17	6.710	0.604		APGFGDNR (302–309)	832.4927	833.5000				
	IGIEIIK (+1; 463–469)	785.51	785.46	1.784	0.064		AAVEEGIVLGGGCALLR (430–446)	1625.9927	1627.0000				
	TALLDAAGVASLLTTAEVVVTEIPK (+2; 527–551)	2482.40	2483.32	5.998	0.676								

[a] Numbers in parentheses indicate calculated MW of truncated proteins.

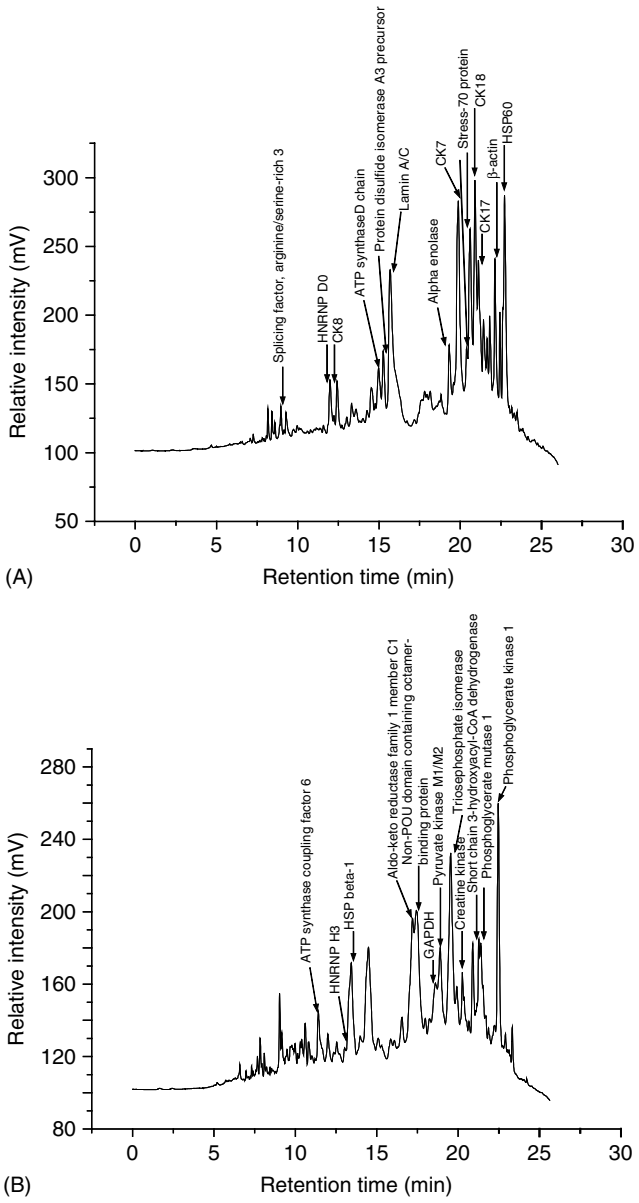

Figure 4. Chromatograms obtained from NPS-RP-HPLC for the combined pH fractions of (A) 4.8–5.0 and 5.0–5.2 and (B) 6.0–6.2 from CA1a.

summarized in Table 2 to show proteins comprehensively analyzed by these four different MS techniques and annotated in Fig. 4(B). Figure 5 shows a representative MALDI tandem mass spectrum, in which one of the tryptic peptides from short chain 3-hydroxyacyl-CoA dehydrogenase, FAGLHFFNPVPVMK (166–179), has been successfully analyzed. A significant deviation of intact MW measurement of this protein from its theoretical MW value suggested sequence truncation, where an excellent agreement was found when the loss of transit peptide sequence (1–12; 1471 Da) was taken into account. On the basis of the supporting MS information, we believe that protein identifications obtained from the analysis of pH fraction digests by monolithic capillary HPLC/MS/MS are highly reliable.

Protein identification based on fraction digest analysis and intact MW measurement

In Tables 1 and 2, it is shown that the sequencing information obtained from fraction digest analysis by monolithic capillary HPLC/MS/MS can often provide reliable protein identifications from highly complex biological mixtures, as further confirmed by other MS techniques. The intact protein MW values, in particular, provide critical information to help confirm protein identification.[39] Overall, the analysis of proteins collected off-line for parallel comparison to fraction digest analysis indicated the importance of matching the intact MW to proteins identified by fraction digest analysis.

In this study, the fraction digest analysis alone resulted in a large number of proteins identified with two or more of fully tryptic peptides sequenced from several pH fractions, as summarized in Table 3. Considering that PFF information from only two tryptic peptides is often assumed to be sufficient and accepted for confident protein identifications,[13] the protein identifications obtained in this study are generally considered reliable, where closely matching experimental intact MW values obtained for all proteins in Table 3 further support this finding. Additionally, the use of statistical data validation tools will increase the number of reliably identified proteins.[40]

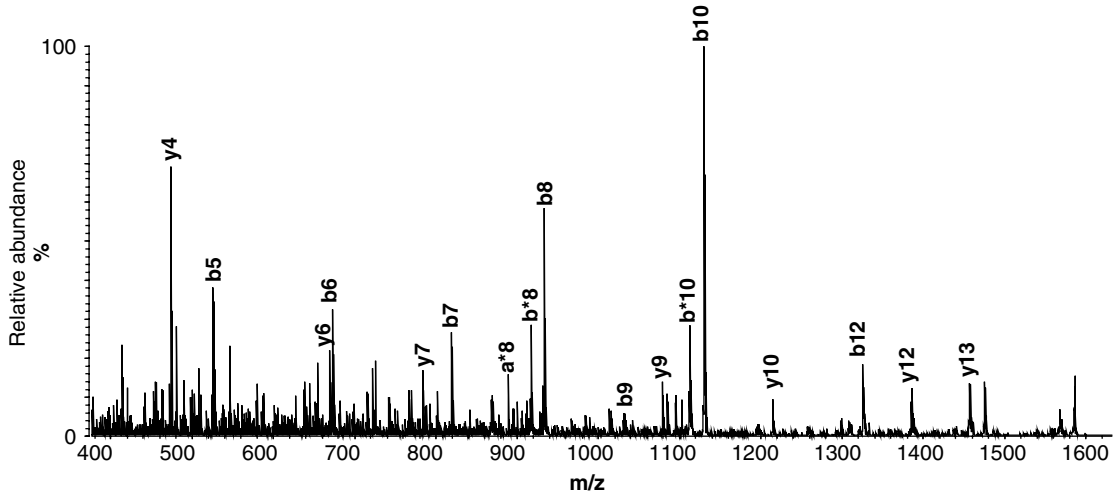

Figure 5. A MALDI tandem mass spectrum for one of the tryptic peptides from short chain 3-hydroxyacyl-CoA dehydrogenase, FAGLHFFNPVPVMK (166–179).

DOI: 10.1002/jms

Table 2. Summary of four different MS analyses for several proteins from the pH fraction of 6.0–6.2 and 6.2–6.4 combined (only fully tryptic peptides considered for MS/MS analyses). Proteins are listed according to the elution order shown in Fig. 4(B)

	Monolithic HPLC/MS/MS of pH fraction digest					MALDI-TOF MS	MALDI-QIT-TOF MS				ESI-TOF MS		
		$[M+H]^+$								Theo.		
Protein name (Accession no.)	Peptide sequenced (charge state; amino acid no.)	Theo.	Exp.	*Xcorr*	ΔCn	PMF Cov. (%)	Peptide sequenced (amino acid no.)	M_r (exp)	Observed	Protein Score	MW[a]	pI	Exp. MW
ATP synthase coupling factor 6 (P18859)	QMFGNADMNTFPTFK (+2; 80–94)	1748.78	1749.09	3.360	0.446	56	QTSGGPVDASSEYQQELER (55–73)	2062.8926	2063.8999	90	12580	9.52	12903
							QMFGNADMNTFPTFK (80–94)	1730.7928	1731.8000				
							FEVIEKPQA (100–108)	1060.5927	1061.6000				
Heterogeneous nuclear ribonucleoprotein (HNRNP) H3 (P31942)[b]	EEIVQFFQGLEIVPNGITLTMDYQGR (+2; 30–55)	3000.52	3000.46	Log (*e*) = −4.1		24	YIEIFR (85–90)	839.4927	840.5000	107	36927	6.40	36820
	ATENDIANFFSPLNPIR (+2; 206–222)	1918.18	1918.96	Log (*e*) = −4.1			ATENDIANFFSPLNPIR (206–222)	1917.8927	1918.9000				
	YIELFLNSTPGGGSGMGGSGM GGYGR (+2; 262–287)	2520.27	2523.13	Log (*e*) = −8.2			GGGGSGGYYGQGGMSGGGWR (324–343)	1803.6927	1804.7000				
Heat shock protein beta-1 (P04792)	LPEEWSQWLGGSSWPGYVR (+2; 38–56)	2234.07	2234.11	5.176	0.666	27	GPSWDPFR (13–20)	960.3927	961.4000	92	22769	5.98	22813
	YTLPPGVDPTQVSSSLSPEGTLT VEAPMPK (+2; 142–171)	3098.56	3099.39	4.850	0.671		LFDQAFGLPR (28–37)	1162.5927	1163.6000				
							QDEHGYISR (128–136)	1086.4927	1087.5000				
							LATQSNEITIPVTFESR (172–188)	1904.9927	1906.0000				
Aldo-keto reductase family 1 member C1 (Q04828)	LNDGHFMPVLGFGTYAPAEVPK (+2; 10–31)	2360.17	2359.86	4.789	0.632	58	REDIFYTSK (76–84)	1157.6927	1158.7000	56	36766	8.02	36600
	NLQLDYVDLYLIHFPVSVKPGE EVIPK (+3; 105–131)	3125.69	3125.97	6.019	0.656								
Non-POU domain containing octamer-binding protein (Q15233)	LFVGNLPPDITEEEMR (+2; 76–91)	1859.92	1859.88	4.061	0.518	57	LFVGNLPPDITEEEMR (76–91)	1858.8927	1859.9000	96	54198	9.01	54309
	NLPQYVSNELLEEAFSVFGQVER (+2; 154–176)	2668.33	2669.15	5.747	0.592		VELDNMPLR (127–135)	1085.5927	1086.6000				
	GAMPPAPVPAGTPAPPGPA TMMPDGTLGLTPPTTER (+2; 399–434)	3451.71	3451.76	3.680	0.651		FAQPGSFEYEYAMR (257–270)	1694.8927	1695.9000				
	FGQAATMEGIGGTPPAFNK (+2; 435–456)	2163.07	2163.99	6.551	0.620								

(*continued overleaf*)

Table 2. (*Continued*)

	Monolithic HPLC/MS/MS of pH fraction digest					MALDI-TOF MS	MALDI-QIT-TOF MS				ESI-TOF MS		
		$[M+H]^+$									Theo.		
Protein name (Accession no.)	Peptide sequenced (charge state; amino acid no.)	Theo.	Exp.	*Xcorr*	ΔCn	PMF Cov. (%)	Peptide sequenced (amino acid no.)	M_r (exp)	Observed	Protein Score	MWa	pI	Exp. MW
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), liver (P04406)	LVINGNPITIFQER (+2; 66–79)	1613.90	1614.33	3.093	0.473	44	LVINGNPITIFQER (66–79)	1612.8927	1613.9000	104	35 900	8.58	35 929
	WGDAGAEYVVESTGVFTTMEK (+2; 86–106)	2277.04	2277.04	5.763	0.662		GALQNIIPASTGAAK (200–214)	1409.1927	1410.2000				
	VIISAPSADAPMFVMGVNHEK (+3; 118–138)	2213.11	2212.47	3.217	0.580		LISWYDNEFGYSNR (309–322)	1762.7928	1763.8000				
	LISWYDNEFGYSNR (+2; 309–322)	1763.80	1763.35	4.553	0.585								
Pyruvate kinase M1/M2 (P14618)	TATESFASDPILYR (+2; 92–105)	1570.78	1570.59	4.428	0.546	28	LDIDSPPITAR (32–42)	1196.6927	1197.7000	38	57 770	7.95	57 899
	IYVDDGLISLQVK (+2; 173–185)	1462.82	1462.77	4.830	0.573		FGVEQDVDMVFASFIR (230–245)	1858.9927	1860.0000				
	GADFLVTEVENGGSLGSK (+2; 188–205)	1779.88	1780.44	5.185	0.593		EAEAAIYHLQLFEELR (383–398)	1930.9927	1932.0000				
	GVNLPGAAVDLPAVSEK (+2; 207–223)	1636.89	1636.59	5.578	0.698								
	LAPITSDPTEATAVGAVEASFK (+3; 400–421)	2175.12	2174.37	5.747	0.660								
	DPVQEAWAEDVDLR (+2; 475–488)	1642.77	1642.39	5.192	0.516								
Triosephosphate isomerase (P60174)	FFVGGNWK (+1; 6–13)	954.48	954.18	1.429	0.308	35	QSLGELIGTLNAAK (19–32)	1396.7928	1397.8000	81	26 522	6.51	27 150
	QSLGELIGTLNAAK (+2; 19–32)	1414.79	1415.02	4.514	0.597		DCGATWVVLGHSER (85–98)	1529.6927	1530.7000				
	VVLAYEPVWAIGTGK (+2; 160–174)	1602.89	1603.27	4.346	0.615		HVFGESDELIGQK (100–112)	1457.7928	1458.8000				
	ELASQPDVDGFLVGGASLKP EFVDIINAK (+3; 219–247)	3029.58	3030.91	2.829	0.378		VVLAYEPVWAIGTGK (160–174)	1601.8927	1602.9000				
							SNVSDAVAQSTR (194–205)	1233.6927	1234.7000				
Creatine kinase (P12532)	TVGMVAGDEETYEVFADLFDP VIQER (+2; 104–129)	2930.38	2930.03	2.647	0.611	35	LYPPSAEYPDLR (46–57)	1419.7928	1420.8000	48	47 008	8.60	42 702
							GWEFMWNER (301–309)	1253.5927	1254.6000		*(43 046)*		
							ILENLR (344–349)	756.4927	757.5000				
							GTGGVDTAATGGVFDISNLDR (354–374)	2021.9927	2023.0000				

DOI: 10.1002/jms

Short chain 3-hydroxyacyl-CoA dehydrogenase (Q16836)	TLSTIATSTDAASVVHSTDLVVEA IVENLK (+3; 96–125)	3084.63	3085.71	5.676	0.663	35	FAGLHFFNPVPVMK (166–179)	1017.5927	1018.6000	66	34 256	8.88	32 856
	LGAGYPMGPFELLDYVGLDTTK (+2; 250–271)	2357.17	2356.81	4.399	0.553		DTPGFIVNR (213–221)	1602.8927	1603.9000		*(32 785)*		
	FIVDGWHEMDAENPLHQPSPSLNK (+3; 272–295)	2761.30	2762.12	6.626	0.659								
Phosphoglycerate mutase 1 (P18669)	TLWTVLDAIDQMWLPVVR (+2; 65–82)	2156.16	2156.03	4.398	0.520	32	HGESAWNLENR (10–20)	1311.5927	1312.6000	45	28 655	6.75	29 278
	SYDVPPPPMEPDHPFYSNISK (+2; 117–137)	2417.11	2417.47	3.789	0.652		FSGWYDADLSPAGHEEAK (21–38)	1978.8927	1979.9000				
	ALPFWNEEIVPQIK (+2; 162–175)	1683.91	1683.56	3.293	0.307								
	HLEGLSEEAIMELNLPTGIPIVYELDK (+3; 195–221)	3023.57	3024.15	3.271	0.530								
Phosphoglycerate kinase 1 (P00558)	ALESPERPFLAILGGAK (+2; 199–215)	1769.00	1769.34	4.407	0.275	34	NNQITNNQR (30–38)	1100.5927	1101.6000	186	44 456	8.30	44 456
	ITLPVDFVTADK (+2; 279–290)	1318.73	1319.23	3.140	0.599		SVVLMSHLGRPDGVPMPDK (56–74)	2034.0927	2035.1000				
	VLPGVDALSNI (+1; 406 416)	1097.62	1097.28	1.665	0.463		YSLEPVAVELK (75–85)	1246.6927	1247.7000				
							ACANPAAGSVILLENLR (106–122)	1710.9927	1712.0000				
							LGDVYVNDAFGTAHR (156–170)	1633.7928	1634.8000				
							ALESPERPFLAILGGAK (199–215)	1767.9927	1769.0000				
							VLNNMEIGTSLFDEEGAK (246–263)	1964.9927	1966.0000				
							ITLPVDFVTADK (279–290)	1317.6927	1318.7000				
							QIVWNGPVGVFEWEAFAR (Pyro-glu; 332–349)	2087.1927	2088.2000				
							QIVWNGPVGVFEWEAFAR (332–349)	2104.0928	2105.1001				

[a] Numbers in parentheses indicate calculated MW of truncated proteins.
[b] This protein analyzed by the X!Tandem database search.

Table 3. List of proteins identified from several pH fractions solely on the basis of fraction digest analysis by monolithic capillary HPLC/MS/MS and intact protein MW values (only fully tryptic peptides were analyzed; proteins in Tables 1 and 2 are not included; * numbers in parentheses indicate calculated MW of truncated proteins)

Protein name	Accession no.	Theo. MW*	Theo. pI	Exp. MW	No. of peptides sequenced by monolithic HPLC/MS/MS
pH 4.4–4.6 and 4.6–4.8					
Vimentin	P08670	53 489	5.06	53 566	5
Uracil-DNA glycosylase	P13051	34 624	9.37	35 048	2
Keratin, type I cytoskeletal 15	P19012	49 138	4.71	49 080	4
Secretogranin-2-precursor	P13521	70 826	4.68	70 512	2
ATP synthase gamma chain	P36542	32 976 (30 130)	9.23	29 596	2
Keratin, type II cytoskeletal 5	P13647	62 410	8.14	62 641	2
Heterogeneous nuclear ribonucleoprotein K	P61978	50 945	5.39	50 931	3
T-complex protein 1, epsilon unit	P48643	59 633	5.45	59 196	3
Heterogeneous nuclear ribonucleoprotein A/B	Q99729	36 590	9.04	35 988	2
pH 4.8–5.0 and 5.0–5.2					
Annexin A1	P04083	38 559	6.64	38 568	7
Thioredoxin	P10599	11 599	4.82	11 606	3
Actin, aortic smooth muscle	P62736	41 982	5.24	41 817	4
Fumarate hydratase	P07954	54 603	8.85	54 307	3
Keratin, type II chtoskeletal 3	P12035	64 472	6.11	64 509	2
ATP synthase beta chain	P06576	56 525 (51 459)	5.26	51 842	2
Keratin, type I cytoskeletal 14	P02533	51 490	5.09	51 391	2
Pre-mRNA splicing factor 18	Q99633	39 836	8.19	40 273	2
Cytochrome C oxidase polypeptide VA	P20674	16 764 (12 488)	6.30	12 501	2
Heat shock-related 70 kDa protein 2	P54652	69 978	5.56	70 021	3
Heterogeneous nuclear ribonucleoprotein F	P52597	45 541	5.40	45 589	3
Keratin, type II cytoskeletal 6e	P48668	59 894	8.10	60 583	2
pH 5.6–5.8 and 5.8–6.0					
ATP synthase alpha chain	P25705	59 714 (55 158)	9.16	55 210	8
Annexin A2	P07355	38 449	7.56	38 531	5
T-complex protein 1, zeta subunit	P40227	57 857	6.25	57 616	3
Keratin, type I cytoskeletal 10	P13645	59 483	5.13	59 479	3
Annexin A1	P04083	38 559	6.64	38 568	3
Serine hydroxymethyltransferase	P34897	55 958 (52 510)	8.76	52 935	3
Dipeptidyl-peptidase I precursor	P53634	51 809	6.53	52 141	2
Splicing factor, proline- and glutamine-rich	P23246	76 102	9.45	74 719	3
S100 calcium-binding protein A16	Q96FQ6	11 795	6.28	11 732	2
Histone H2A.a	P28001	13 996	11.05	13 807	2
Heterogeneous nuclear ribonucleoprotein L	P14866	60 150	6.65	60 189	2
Cathepsin D precursor	P07339	44 524	6.10	44 648	2
26S protease regulatory subunit 8	P62195	45 598	7.11	45 058	2
39S ribosomal protein L28	Q13084	33 842	8.85	34 174	2
26S proteasome non-ATPase regulatory subunit 1	O00231	47 304	6.09	47 385	2
Transcription elongation factor B polypeptide 1	Q15369	12 466	4.74	12 190	2
pH 6.0–6.2 and 6.2–6.4					
Serine protease HTRA2	O43464	48 811 (45 354)	10.07	45 068	2
Enoyl-CoA hydratase	P30084	31 368 (28 308)	8.34	28 373	2
Histone H2A.o	P20670	13 956	10.90	13 816	3
Histone H2A.a	P28001	13 996	11.05	13 809	3
Uracil-DNA glycosylase	P13051	34 624	9.37	34 385	3

DOI: 10.1002/jms

Protein identification by one sequenced peptide from fraction digest analysis and intact MW measurement

The protein identification with PFF information from only one peptide is generally considered less reliable. At the same time, in shotgun proteomics, large numbers of proteins are identified by a single peptide. Thus, excluding all such identifications would result in a significant loss of proteins whose identification is potentially correct. Although one has to be careful when reporting the identification of proteins based on a single peptide fragment fingerprint, closer examination of several of these proteins by other MS approaches (Tables 1 and 2) suggests that they may result in correct identifications, provided that other complementary MS information, such as closely matching intact protein MW values, is available.

In Table 1, fraction digest analysis resulted in only one fully tryptic peptide with *Xcorr* exceeding the set criteria for splicing factor arginine/serine-rich 3, perhaps owing to its relatively low concentration, as observed from Fig. 4(A), where significant signal suppression by closely eluting peptides of higher abundance is expected. Assuming the typical recovery of the NPS-RP-HPLC column to be 80%,[41] it implies that the estimated amount of splicing factor arginine/serine-rich 3 protein in Fig. 4(A) is approximately 0.26 μg, or 1.28 pmol, on the basis of peak quantitation of the chromatogram using a manual baseline by the Origin software (version 6.0, Microcal software, Northampton, MA). Considering that only *ca* 1.5 μg of each of the pH fractions was consumed for fraction digest analysis, this corresponds to *ca* 24 fmol of splicing factor arginine/serine-rich 3 analyzed. However, the availability of a closely matching experimental intact MW value obtained from ESI-TOF MS strongly suggested the presence of this protein in these pH fractions. In order to examine the reliability of the identification of this protein based on one sequenced peptide, it was collected off-line from NPS-RP-HPLC for subsequent digestion, followed by rapid monolithic LC separation interfaced on-line with linear ion trap MS with a separation time of 10 min, where three unique peptides were successfully sequenced to confirm the identification, as shown in Table 4.

ATP synthase coupling factor 6 (Table 2) was also identified with only one peptide identified by PFF from fraction digest analysis, but with closely matching intact MW value available. The tandem MS analysis later resulted in three identified peptides from off-peak collection from NPS-RP-HPLC (Table 4). One protein from pH fractions of 6.0–6.2 and 6.2–6.4, delta3,5-delta2,4-dienoyl-CoA isomerase, was analyzed with one identified peptide from fraction digest analysis. This protein, observed to elute very closely with short chain 3-hydroxyacyl-CoA dehydrogenase, was collected off-line for on-line monolithic LC/MS/MS analysis, of which the TIC chromatogram is shown in Fig. 6. This protein was analyzed to obtain three unique peptides from PFF analysis with matching intact MW available. The MALDI-MS/MS analysis also confirmed the reliability of the identification of this protein with three peptides identified by PFF. Table 4 shows several other proteins whose identifications were confirmed in this manner.

As discussed, the complementary intact MW information provides an effective means of identifying proteins with only one fully tryptic peptide sequenced. It is considered particularly useful for the analysis of proteins of relatively small size, *ca* 20 kDa or less, as the PMF analysis of these remains difficult because of the nonconfident identification associated with the MALDI-MS processes.[21] Table 5 lists the small proteins analyzed by fraction digest analysis with monolithic capillary LC/MS/MS, of which database search returned only one fully tryptic peptide with high *Xcorr*. All these proteins have closely matching intact MW values. As observed from Table 4, in which one identified peptide from LC/MS/MS and closely matching intact MW suggested reliable protein identifications, the presence of the proteins shown in Table 5 in fully malignant human breast cancer cells under study is highly likely.

CONCLUSIONS

Confident identification of proteins from very complex biological mixtures is still challenging and often requires complementary information from different approaches for comparison. In this study, a comprehensive analysis that combines several different MS techniques has been successfully demonstrated to identify large numbers of proteins present in human breast cancer cells by integrating MS information from peptide and protein levels. The extreme complexity of the samples was reduced to moderate complexity by CF to make the current approach more suitable. The pH fraction digest analysis provides high speed and sensitivity due to high-resolution monolithic capillary HPLC separation for fast scanning linear ion trap mass spectrometric analysis to rapidly identify large numbers of proteins. The method has been proved to be a means to obtain reliable identifications when comprehensive analyses were performed for a subset of proteins to compare peptide mapping, additional sequencing by MALDI-MS/MS, and intact protein MW. The method also helped elucidate protein sequence variations and identify proteins that were based on a single peptide identified from a PFF. Although there has been a recent emphasis on high throughput in proteomics, this work represents an effort to obtain more detailed information and confirmation of identifications – an issue with some of the high-throughput methods. This work represents an attempt to use multiple techniques to confirm the identifications obtained by any one method alone.

Acknowledgements

This work received partial support under a number of sources including NSF-SBIR Award No. 0321763 to Syagen Technology, Inc. and a subaward to the University of Michigan, #R01CA108597 (DML), #R01CA106402 (DML), #R01CA90509 (FRM), and #R01GM49500 (DML). The MALDI-TOF MS instrument used in this work was funded by the National Science Foundation under grant DBI 99874. We thank Reiner Wintringer (Saarland University) for synthesizing the monolithic columns used in this study.

Table 4. Summary of proteins identified by one fully tryptic peptide sequenced by fraction digest analysis by monolithic LC/MS/MS, examined further by matching intact protein MW values and off-peak collection

	Monolithic HPLC/MS/MS of pH Fraction Digest							Off-peak collections from NPS-RP-HPLC for fast monolith LC/MS/MS				
		$[M+H]^+$				Intact MW			$[M+H]^+$			
Protein name (Accession no.)	Peptide sequenced (Charge state; amino acid no.)	Theo.	Exp.	*Xcorr*	Δ*Cn*	Theo.	Exp.	Peptide sequenced (Charge state; amino acid no.)	Theo.	Exp.	*Xcorr*	Δ*Cn*
pH 4.8–5.0 and 5.0–5.2												
Splicing factor, arginine/serine rich-3 (P84103)	NPPGFAFVEFEDPR (+2; 44–57)	1621.77	1622.55	3.918	0.540	19 318	20 290	VYVGNLGNNGNK (+2; 12–23)	1248.63	1248.52	2.889	0.526
								AFGYYGPLR (+1; 29–37)	1043.53	1043.40	1.423	0.360
								NPPGFAFVEFEDPR (+2; 44–57)	1621.77	1621.96	3.885	0.582
Heat shock cognate 71 kDa protein (P11142)	SENVQDLLLLDVTPLS-LGIETAGGVMTVLIK (+3; 385–415)	3238.79	3239.07	4.887	0.602	69 978	70 840	FDDAVVQSDMK (+2; 78–88)	1254.57	1254.41	4.185	0.582
								TVTNAVVTVPAYFNDSQR (+2; 138–155)	1982.00	1983.00	3.268	0.554
								TLSSSTQASIEIDSLYEGID-FYTSITR (+2; 273–299)	2997.46	2998.15	4.060	0.650
								SINPDEAVAYGAAVQAA-ILSGDK (+2; 362–384)	2260.15	2259.89	5.390	0.629
								SENVQDLLLLDVTPLSL-GIETAGGVMTVLIK (+3; 385–415)	3238.79	3240.14	5.843	0.693
								GVPQIEVTFDIDANGILN-VSAVDK (+2; 470–493)	2514.31	2515.03	4.805	0.689
Mannose-6-phosphate receptor binding protein 1 (O60664)	SVVTGGVQSVMGSR (+2; 167–180)	1363.70	1363.47	2.913	0.422	47 018	46 946	TLTAAAVSGAQPILSK (+2; 69–84)	1527.87	1528.03	4.556	0.616
								SVVTGGVQSVMGSR (+2; 167–180)	1363.70	1363.33	4.311	0.522
	* Also, 37% coverage by MALDI-TOF MS.							LGQMVLSGVDTVLGK (+2; 181–195)	1516.84	1516.75	4.229	0.472
HNRNP Q (O60506)	VADSSKGPDEAKIK (+2; 112–125)	1444.76	1445.56	3.162	0.010	69 633	69 703	TGYTLDVTTGQR (+2; 131–142)	1311.65	1311.68	3.249	0.458
								DLFEDELVPLFEK (+2; 172–184)	1593.81	1549.42	4.400	0.510
								AGPIWDLR (+1; 185–192)	927.51	927.32	1.275	0.166
								LFVGSIPK (+1; 245–252)	860.52	860.28	1.589	0.097
								VTEGLTDVILYHQPDDK (+2; 266–282)	1942.98	1942.59	4.462	0.607
								DLEGENIEIVFAKPPDQK (+2; 395–412)	2042.04	2042.91	3.992	0.563
	* Also, 19% coverage by MALDI-TOF MS.							SENQEFYQDTFGQQWK (+2; 608–623)	2034.88	2034.72	2.786	0.395

60S ribosomal protein L4 (P36578)	EAVLLLK (+1; 166–172)	785.51	785.46	1.594	0.147	47 668	48 091	IEEVPELPLVVEDK (+2; 144–157)	1608.87	1609.18	3.253	0.521
								NIPGITLLNVSK (+2; 223–234)	1268.76	1268.43	2.748	0.545
Heat shock 70 kDa protein 1L (P34931)	TTPSYVAFTDTER (+2; 39–51)	1487.70	1487.47	3.314	0.612	70 332	70 521	TTPSYVAFTDTER (+2; 39–51)	1487.70	1487.47	3.052	0.583
	* Also, 23% coverage by MALDI-TOF MS.							AKIHDIVLVGGSTR (+2; 331–344)	1465.85	1465.64	4.162	0.591
Alpha enolase, lung specific (Q05524)	YISPDQLADLYK (+2; 279–290)	1425.73	1425.87	3.873	0.582	49 447	50 550	DATNVGDEGGFAPNI-LENK (+2; 212–230)	1960.93	1960.68	6.073	0.631
								VVIGMDVAASEFYR (+2; 249–262)	1556.78	1556.55	1.878	0.174
								YISPDQLADLYK (+2; 279–290)	1425.73	1426.23	3.806	0.616
pH 6.0–6.2 and 6.2–6.4												
Delta3,5-delta2,4-dienoyl-CoA isomerase (Q13011) **Off-peak analysis with MASCOT search; observed mass *versus* Mr (exp)	MFTAGIDLMDMASD-ILQPK (+2; 113–131)	2097.01	2097.43	5.709	0.697	35 972	35 535	ISWYLR (+1; 138–143)	836.2727	837.2800	225/45	
								MMADEALGSGLVSR (+2; 232–245)	1435.3434	718.6790		
								EVDVGLAADVGTLQR (+2; 197–211)	1541.4734	771.7440		
								Additional analysis by MALDI-QIT-TOF MS resulted in identifying three unique peptide sequences at the MASCOT protein score of 50.				
Peptidyl-prolyl cis-trans isomerase B precursor (P23284)	TVDNFVALATGEK (+2; 64–76)	1364.71	1364.23	3.635	0.547	22 729	22 552	TVDNFVALATGEK (+2; 64–76)	1364.71	1364.39	3.961	0.622
								DTNGSQFFITTVK (+2; 138–150)	1457.73	1458.39	3.325	0.464
Creatine kinase (P12532)	TVGMVAGDEETYEV-FADLFDPVIQER (+2; 104–129)	2930.38	2930.03	2.647	0.611	47 008	42 702	LYPPSAEYPDLR (+1; 46–57)	1420.71	1420.58	1.782	0.298
								TVGMVAGDEETYEV-FADLFDPVIQER (+2; 104–129)	2930.38	2928.96	2.232	0.637
								VVVDALSGLK (+1; 190–199)	1000.60	1000.21	1.611	0.358
								SFLIWVNEEDHTR (+2; 257–269)	1645.80	1645.52	4.558	0.592
								LPLLSK (+1; 332–337)	670.45	670.13	1.926	0.274
								GTGGVDTAATGGVFD-ISNLDR (+2; 354–374)	2022.97	2022.71	5.445	0.646

(*continued overleaf*)

Table 4. (*Continued*)

Protein name (Accession no.)	Monolithic HPLC/MS/MS of pH Fraction Digest: Peptide sequenced (Charge state; amino acid no.)	[M + H]$^+$ Theo.	[M + H]$^+$ Exp.	*Xcorr*	Δ*Cn*	Intact MW Theo.	Intact MW Exp.	Off-peak collections from NPS-RP-HPLC for fast monolith LC/MS/MS: Peptide sequenced (Charge state; amino acid no.)	[M + H]$^+$ Theo.	[M + H]$^+$ Exp.	*Xcorr*	Δ*Cn*
HNRNP A2/B1 (P22626)	GFGFVTFDDHDPVDK (+2; 154–168)	1695.77	1695.67	4.429	0.569	37 407	37 559	DYFEEYGK (+1; 130–137)	1050.44	1050.26	1.684	0.239
								GFGFVTFDDHDPVDK (+2; 154–168)	1695.77	1695.73	4.493	0.596
	* Also, 30% coverage by MALDI-TOF MS.							GGGGNFGPGPGSNFR (+2; 214–228)	1377.63	1378.40	3.298	0.471
ATP synthase coupling factor 6	QMFGNADMNTFPTFK (+2; 80–94)	1748.78	1749.09	3.360	0.446	12 580	12 903	QTSGGPVDASSEYQQELER (+2; 55–73)	2080.94	2081.10	5.398	0.659
								QMFGNADMNTFPTFK (+2; 80–94) - its oxidized form also identified by MASCOT search (score = 247/46)	1748.78	1748.55	3.991	0.533
								FEVIEKPQA (+1; 100–108)	1060.57	1060.33	2.499	0.427
Far upstream element binding protein 1 (Q96AE4)	SVQAGNPGGPGPGGPGR (+1; 344–360)	1520.77	1522.08	1.608	0.126	67 432	67 535	IGGDAGTSLNSNDYGYGGQK (+2; 45–64)	1973.88	1974.00	4.521	0.614
								IQIAPDSGGLPER (+2; 133–145)	1352.72	1352.72	4.237	0.477
								IGGNEGIDVPIPR (+2; 271–283)	1336.72	1336.56	3.743	0.543
								SVQAGNPGGPGPGGR (+2; 344–358)	1307.65	1307.15	3.703	0.542
								TGLIIGK (+1; 387–393)	701.46	701.38	1.552	0.294
	* Also, 27% coverage by MALDI-TOF MS.											

Table 5. List of small proteins identified only with one peptide hit, but with closely matching intact protein MW values (* numbers in parentheses indicate calculated MW of truncated proteins)

Protein name	Accession no.	Theo. MW*	Theo. pI	Exp. MW
pH 4.4–4.6 and 4.6–4.8				
Interleukin-17 precursor	Q16552	17 493	8.82	17 764
U6 snRNA-associated Sm-like protein LSm7	Q9UK45	11 596	5.10	11 756
Interleukin-7 precursor	P13232	20 174	8.87	19 957
Regulator of G-protein signaling 8	P57771	20 904	9.36	20 840
Eukaryotic translation initiation factor 5A	P63241	16 691	5.08	16 873
Ig lambda chain V-IV region MOL	P06889	11 265	4.28	11 608
pH 4.8–5.0 and 5.0–5.2				
Diphosphoinositol polyphosphate phosphohydrolase	Q8NFP7	18 489	5.52	18 380
Ras-related protein Rab-7L1	O14966	23 141	6.73	23 346
Prolactin-inducible protein precursor	P12273	16 562	8.26	16 425
Lactoylglutathione lysase	Q04760	20 576	5.25	20 786
ADP-sugar pyrophosphatase	Q9UKK9	24 313	4.87	24 314
26S proteasome non-ATPase regulatory subunit 1	O75832	24 413	5.71	24 905
60S ribosomal protein L28	P46779	15 607	12.02	15 352
39S ribosomal protein L12	P52815	21 335	9.05	21 826
Stathmin-3	Q9NZ72	21 004	6.99	20 420
Histone H4	P62805	11 230	11.36	11 595
pH 5.6–5.8 and 5.8–6.0				
40S ribosomal protein S15a	P62244	14 699	10.14	14 709
Superoxide dismutase [Cu-Zn]	P00441	15 795	5.70	15 572
UMP-CMP kinase	P30085	22 209	5.44	22 335
Glutathione S-transferase P	P09211	21 233	5.44	21 575
Cytochrome C oxidase polypeptide	P12074	12 148 (9866)	9.30	9619
Histone H2B	P62807	13 767	10.32	13 777
Nucleoside diphosphate kinase A	P15531	17 138	5.83	17 212
Acylphosphatase	P14621	11 002	9.52	11 074
Mitochondrial 39S ribosomal protein L23	Q16540	17 771	9.69	17 713
T-cell leukemia/lymphoma protein 1A	P56279	13 451	4.98	13 755
40S ribosomal protein S21	P63220	9106	8.68	9160
GrpE protein homolog 1	Q9HAV7	24 264 (21 306)	8.24	21 542
Barrier-to-autointegration factor	O75531	10 053	5.81	10 054
S100 calcium-binding protein A7	P31151	11 319	6.26	11 073
pH 6.0–6.2 and 6.2–6.4				
Protein transport protein Sec61beta subunit	P60468	9838	11.57	9631
Putative RNA-binding protein 3	P98179	17 161	8.86	17 101
SH2 domain protein 1B	O14796	15 288	8.97	15 500
Prefoldin subunit 5	Q99471	17 318	5.94	17 761
Troponin I	P48788	21 194	8.88	21 503
Small nuclear ribonucleoprotein Sm D1	P62314	13 274	11.56	12 854
Peroxiredoxin 2	P32119	21 748	5.67	21 857

REFERENCES

1. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 1975; **250**: 4007.
2. Thiede B, Hohenwarter W, Krah A, Mattow J, Schmid M, Schmidt F, *et al*. Peptide mass fingerprinting. *Methods* 2005; **35**: 237.
3. Chen J, Kahne T, Rocken C, Gotze T, Yu J, Sung JJY, *et al*. Proteome analysis of gastric cancer metastasis by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionization mass spectrometry for identification of metastasis-related proteins. *J. Proteome Res.* 2004; **3**: 1009.
4. Fillet M, Crem-Olive C, Renert AF, Piette J, Vandermoere F, Rolando C, *et al*. Differential expression of proteins in response to ceramide-mediated stress signal in colon cancer cells by 2-D gel electrophoresis and MALDI-TOF MS. *J. Proteome Res.* 2005; **4**: 870.
5. Beranova-Giorgianni S. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. *Trends Anal. Chem.* 2003; **22**: 273.
6. Cooper JW, Wang Y, Lee CS. Recent advances in capillary separations for proteomics. *Electrophoresis* 2004; **25**: 3913.
7. Gan CS, Reardon KF, Wright PC. Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of Synechocystis sp. PCC 6803. *Proteomics* 2005; **5**: 2468.

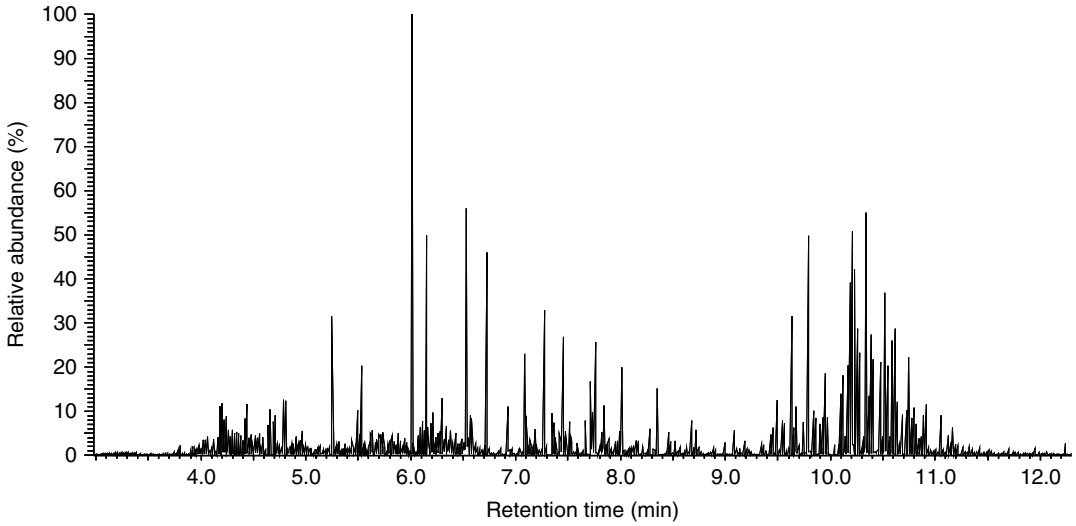

Figure 6. A representative fast monolithic capillary HPLC/MS chromatogram (TIC) for proteins collected off-line from NPS-RP-HPLC. This protein was later identified as Delta3,5-delta2,4-dienoly-CoA isomerase (see Table 4).

8. Swanson SK, Washburn MP. The continuing evolution of shotgun proteomics. *Drug Discov. Today* 2005; **10**: 719.
9. Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, *et al*. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal. Chem.* 2004; **76**: 3556.
10. Mitulovic G, Stingl C, Smoluch M, Swart R, Chervet J-P, Steinmacher I, *et al*. Automated, on-line two-dimensional nano liquid chromatography tandem mass spectrometry for rapid analysis of complex protein digests. *Proteomics* 2004; **4**: 2545.
11. Wolters DA, Washburn MP, Yates JR. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 2001; **73**: 5683.
12. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001; **19**: 242.
13. Zhu W, Reich CI, Olsen GJ, Giometti CS, Yates JR. Shotgun proteomics of Methanosoccus jannaschii and insights into methanogenesis. *J. Proteome Res.* 2004; **3**: 538.
14. MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, *et al*. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. U.S.A.* 2002; **99**: 7900.
15. Shen Y, Zhang R, Moore RJ, Kim J, Metz TO, Hixson KK, *et al*. Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000–1500 and capabilities in proteomics and metabolomics. *Anal. Chem.* 2005; **77**: 3090.
16. Shen Y, Zhao R, Berger SJ, Anderson GA, Rodriguez N, Smith RD. High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Anal. Chem.* 2002; **74**: 4235.
17. Shen Y, Tolic N, Masselon C, Pasa-Tolic L, David G, Camp I, Hixson KK, *et al*. Ultrasensitive proteomics using high-efficiency on-line micro-SPE-nanoLC-nanoESI MS and MS/MS. *Anal. Chem.* 2004; **76**: 144.
18. Janini GM, Conrads TP, Veenstra TD, Issaq HJ. Development of a two-dimensional protein-peptide separation protocol for comprehensive proteome measurements. *J. Chromatogr., B* 2003; **787**: 43.
19. Kreunin P, Urquidi V, Lubman DM, Goodison S. Identification of metastasis-associated proteins in a human tumor metastasis model using the mass-mapping technique. *Proteomics* 2004; **4**: 2754.
20. Wang H, Kachman MT, Schwartz DR, Cho KR, Lubman DM. Comprehensive proteome analysis of ovarian cancers using liquid phase separation, mass mapping, and tandem mass spectrometry: a strategy for identification of candidate cancer biomarkers. *Proteomics* 2004; **4**: 2476.
21. Zhu K, Miller FR, Barder TJ, Lubman DM. Identification of low molecular weight proteins isolated by 2-D liquid separations. *J. Mass Spectrom.* 2004; **39**: 770.
22. Hamler RL, Zhu K, Buchanan NS, Kreunin P, Kachman MT, Miller FR, *et al*. A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification. *Proteomics* 2004; **4**: 562.
23. Zheng S, Schneider KA, Barder TJ, Lubman DM. Two-dimensional liquid chromatography protein expression mapping for differential proteomic analysis of normal and O157:H78 Escherichia coli. *Biotechniques* 2003; **35**: 1202.
24. Sluyterman LAA, Elgersma O. Chromatofocusing: isoelectric focusing on ion-exchange columns. *J. Chromatogr.* 1978; **150**: 17.
25. Santner SJ, Dawson PJ, Tait L, Soule HD, Eliason J, Mohamed AN, *et al*. Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. *Breast Cancer Res. Treat.* 2001; **65**: 101.
26. Premstaller A, Oberacher H, Walcher W, Timperio AM, Zolla L, Chervet J-P, *et al*. High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Anal. Chem.* 2001; **73**: 2390.
27. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004; **20**: 1466.
28. Zhu K, Zhao J, Lubman DM, Miller FR, Barder TJ. Protein pI shifts due to posttranslational modifications in the separation and characterization of proteins. *Anal. Chem.* 2005; **77**: 2745.
29. Lee D, Svec F, Frechet JMJ. Photopolymerized monolithic capillary columns for rapid micro high-performance liquid chromatographic separation of proteins. *J. Chromatogr., A* 2004; **1051**: 53.
30. Holzl G, Oberacher H, Pitsch S, Stutz A, Huber CG. Analysis of biological and synthetic ribonucleic acids by liquid chromatography-mass spectrometry using monolithic capillary columns. *Anal. Chem.* 2005; **77**: 673.
31. Huber CG, Walcher W, Timperio A-M, Troiani S, Porceddu A, Zolla L. Multidimensional proteomic analysis of photosynthetic membrane proteins by liquid extraction-ultracentrifugation-liquid chromatography-mass spectrometry. *Proteomics* 2004; **4**: 3909.
32. Kimura H, Tanigawa T, Morisaka H, Ikegami T, Hosoya K, Ishizuka N, *et al*. Simple 2D-HPLC using a monolithic silica column for peptide separation. *J. Sep. Sci.* 2004; **27**: 897.

33. Toll H, Oberacher H, Swart R, Huber CG. Separation, detection, and identification of peptides by ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry at high and low pH. *J. Chromatogr., A* 2005; **1079**: 274.
34. Hennessy TP, Boysen RI, Huber MI, Unger KK, Hearn MTW. Peptide mapping by reversed-phase high-performance liquid chromatography employing silica rod monoliths. *J. Chromatogr., A* 2003; **1009**: 15.
35. Tholey A, Toll H, Huber CG. Separation and detection of phosphorylated and nonphosphorylated peptides in liquid chromatography-mass spectrometry using monolithic columns and acidic or alkaline mobile phases. *Anal. Chem.* 2005; **77**: 4618.
36. Cargile BJ, Bundy JL, Stephenson JL. Potential for false positive identifications from large databases through tandem mass spectrometry. *J. Proteome Res.* 2004; **3**: 1082.
37. Neupert W. Protein import into mitochondria. *Annu. Rev. Biochem.* 1997; **66**: 863.
38. Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data. *Mol. Cell Proteomics* 2005; **4**: 1419.
39. Millea KM, Krull IS, Cohen SA, Gebler JC, Berger SJ. Integration of multidimensional chromatographic protein separations with a combined ''Top-Down'' and ''Bottom-Up'' proteomic strategy. *J. Proteome Res.* 2006; **5**: 135.
40. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003; **75**: 4646.
41. Wall DB, Lubman DM, Flynn SJ. Rapid profiling of induced proteins in bacteria using MALDI-TOF mass spectrometric detection of nonporous RP HPLC-separated whole cell lysates. *Anal. Chem.* 1999; **71**: 3894.