

Chul Yoo¹
Manoj Pal¹
Fred R. Miller²
Timothy J. Barber³
Christian Huber⁴
David M. Lubman^{1, 5, 6}

¹Department of Chemistry,
The University of Michigan,
Ann Arbor, MI, USA

²Barbara Ann Karmanos
Cancer Institute,
Wayne State University,
Detroit, MI, USA

³Eprogen, Inc.,
Darien, IL, USA

⁴Department of Chemistry,
Instrumental Analysis
and Bioanalysis,
Saarland University,
Saarbrücken, Germany

⁵Department of Surgery,
The University of Michigan
Medical Center,
Ann Arbor, MI, USA

⁶Comprehensive Cancer Center,
The University of Michigan
Medical Center,
Ann Arbor, MI, USA

Received September 2, 2005

Revised November 25, 2005

Accepted December 1, 2005

Research Article

Toward high sequence coverage of proteins in human breast cancer cells using on-line monolith-based HPLC-ESI-TOF MS compared to CE MS

A method is developed toward high sequence coverage of proteins isolated from human breast cancer MCF10 cell lines using a 2-D liquid separations. Monolithic-capillary columns prepared by copolymerizing styrene with divinylbenzene are used to achieve high-resolution separation of peptides from protein digests. This separation is performed with minimal sample preparation directly from the 2-D liquid fractionation of the cell lysate. The monolithic column separation is directly interfaced to ESI-TOF MS to obtain a peptide map. The protein digests were also analyzed by MALDI-TOF MS and an accurate M_r of the intact protein was obtained using an HPLC-ESI-TOF MS. The result is that these techniques provide complementary information where nearly complete sequence coverage of the protein is obtained and can be compared to the experimental M_r value. The high sequence coverage provides information on isoforms and other post-translational modifications that would not be available from methods that result in low sequence coverage. The results from the use of monolithic columns are compared to that obtained by CE-MS. The monolithic column separations provide a rugged and highly reproducible method for separating protein digests prior to MS analysis and is suited to confidently identify biomarkers associated with cancer progression.

Keywords: Electrospray ionization / High sequence coverage / Isoform / Monolith / Peptide mass fingerprinting
DOI 10.1002/elps.200500651

1 Introduction

An important problem in current proteomics is searching for biomarkers of various diseases [1]. This involves searching for proteins that undergo highly dynamic changes during the course of a disease such as cancer progression. These changes may involve sequence modifications including truncations, deletions, splice variants, and sequence substitutions, as well as post-translational modifications (PTMs), such as phosphorylation, glycosylation, acetylation, and methylation, and a host of other modifications that may affect the function

of the proteins [2–4] and play a significant role in the pathways leading to carcinogenesis [5]. In addition to this complexity, many important proteins involved in regulation are often present in low abundance, presenting further challenges in proteomic studies. Traditionally, 2-DE [6] has been used to profile protein expression and to search for changes in such expression levels of disease states, including tumor cells.

Despite the tremendous contribution the 2-DE method has made to the study of various proteomes, it has a number of fundamental limitations. These include the labor-intensive and time-consuming nature of the technique, poor reproducibility, and under-representation of certain classes of proteins, so that truly comprehensive analysis is impossible. Furthermore, it cannot provide accurate M_r information and it still remains difficult to interface 2-DE directly to MS analysis [7, 8]. More recently, a method utilizing 2-D liquid-phase separations based on pI and hydrophobicity of proteins has been developed for mapping of proteins and applied to large-

Correspondence: Dr. David M. Lubman, University of Michigan, Department of Chemistry, 930 N. University Ave., Ann Arbor, MI 48109, USA

E-mail: dmlubman@umich.edu

Fax: +1-734-615-8108

Abbreviations: DI, deionized; NPS, nonporous; PTM, post-translational modification

scale study of several different types of human cancer cells [9–11]. This method allows for direct interfacing of chromatographic separation for ESI-TOF MS analysis to obtain an accurate M_r value. Alternatively, the proteins contained in selected pH ranges can be collected, enzymatically digested, and analyzed by MALDI-TOF MS to obtain a peptide map or sequencing information by tandem mass spectrometric analysis. The protein identification can then be readily obtained by database searching.

The use of peptide mapping by MALDI-TOF MS provides very limited sequence coverage, typically, in the range of 40–50% at most, due to the ionization efficiency often depending on the choice of matrix [12], peptide composition [13, 14], ion suppression, and suppression or loss of PTMs. Thus, the issue becomes obtaining sufficiently high sequence coverage to prevent false identification of unknown proteins and then to identify the presence of important PTMs or other variations that are often particular to different stages of cancer and other diseases.

In recent work, a method combining CE/ESI-TOF and MALDI-TOF MS was used to improve the sequence coverage of peptide maps [15]. A sequence coverage of over 90% could be obtained in most cases from tryptic digests by combining the peptides detected from both techniques. The sequence could be used with the M_r value determined by ESI-TOF MS and the known M_r value from the database to determine the sequence coverage and where modifications or isoforms were present. Using this method, various isoforms of proteins found in breast cancer cells could be identified and sequence variations identified.

An alternate form of capillary separation involves the use of monolithic columns, which are formed of a single, rigid, and porous polymer. The use of these monolithic columns has been described in many studies involving biological molecules, including nucleic acids, ribonucleic acids, proteins, and peptides [16–22]. It has recently become a stationary phase of choice in LC with several advantages over conventional packed columns. Due to the lack of interstitial space, these monoliths can provide a very fast separation with high resolution. Also, as analyte does not dwell at the resin, it is capable of producing high recovery, hence increased sensitivity, and can separate very hydrophobic analytes that might otherwise be difficult to elute with conventional packed columns. Compared to CE-MS, the method is very rugged where large numbers of separations can be run and the enhanced loadability allows much improved sensitivity.

In the present work, proteins isolated from 2-D liquid separations of breast cancer cell lysates were analyzed by combining several methods to obtain high sequence cov-

erage and to search for isoforms and modified proteins. The method involved the use of capillary monolith LC separation of tryptic digests of proteins isolated from the liquid separations with on-line detection by ESI-TOF MS and linear IT MS. The sequence coverage obtained by this method was combined with that from MALDI-TOF MS to yield >85% coverage in most cases relative to the known sequence for the protein. Also, a comparison of sequence coverage to previous CE-MS work is shown for several proteins. The combined MALDI and ESI-MS procedure was performed for ten proteins from a malignant breast cancer cell line in the pH 7.5 fraction. Also, tandem mass spectrometric analysis confirmed the presence of a unique peptide in different isoforms of lamin.

2 Materials and methods

The experimental overview is shown in Fig. 1. The proteins from lysed cells are subjected to 2-D liquid-phase separations according to their pI and hydrophobicity. The

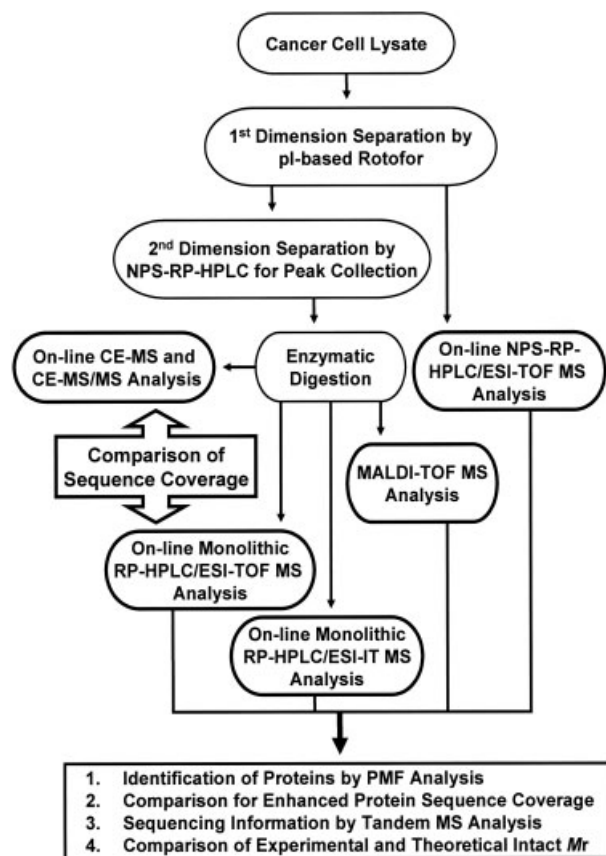


Figure 1. Experimental scheme of the liquid-phase 2-D separation techniques followed by PMF analysis and sequencing from different MS methods for identification of proteins in human breast cancer cell lines with high overall sequence coverage.

isolated proteins are then enzymatically digested and analyzed by MALDI-TOF MS and monolith-based LC-MS for protein identification by PMF. The sequence coverage results from CE-MS analysis from previous work are also compared for several proteins. The intact M_r of the proteins is obtained using nonporous (NPS)-RP-HPLC separations interfaced on-line with ESI-TOF MS. The results from PMF analysis using the two different MS methods are combined to yield high sequence coverage and a comparison between the experimental and theoretical intact M_r is used to provide highly reliable protein identification and to identify the presence of isoforms. The sequencing information of the selected tryptic peptides is provided by the monolith-based LC directly interfaced with linear IT ESI-MS to confirm their presence.

2.1 Cell lines

The cells used in this work are CA1a.cl1 and CA1d.cl1, cloned and fully malignant lines derived from the MCF10A human breast cancer cell line, which are maintained and prepared by the Barbara Ann Karmanos Cancer Institute (Wayne State University, Detroit, MI) as previously described [23].

2.2 Preparative liquid-phase IEF

The first-dimensional separation based on pI was performed using the Mini-Rotofor (BioRad, Hercules, CA) to separate the cell extracts as previously described [9, 15]. Briefly, cell extracts were mixed with IEF running buffer composed of 8 M urea, 2 M thiourea, and 2% Biolyte of pH 3–10 (BioRad). The Rotofor chamber was loaded with 18 mL of the mixture, and the separation constantly controlled at 12 W for 3.5 h. The separated pI fractions were harvested into 20 tubes for pH measurements using an Orion pH meter (model 250A, Allometrics, Baton Rouge, LA) and Accumet combination electrodes (Fischer, Pittsburgh, PA). Each fraction was subjected to a colorimetric-based Lowry assay (RC Protein Assay, BioRad) for relative quantitation and stored at -80°C until use.

2.3 NPS-RP-HPLC separation

The second-dimensional separation was performed using NPS-silica columns. The NPS-RP-HPLC column (33 mm length \times 4.6 mm id) packed with 1.5 μm C18 NPS ODSIII silica beads (Eprogen, Darien, IL) was used at a flow rate of 0.5 mL/min. Approximately 250 μg of protein obtained from each Rotofor fraction was loaded for separation with the HPLC System Gold equipped with UV detector (Beckman Coulter, Fullerton, CA). The solvent system comprised of

solvent A: helium-degassed deionized (DI) water (Millipore, Billerica, MA) with 0.1% TFA (TFA, 99.5%; Sigma, St. Louis, MO) and solvent B: ACN (ACN, 99.93% HPLC grade; Sigma) with 0.1% TFA. The column was maintained at 65°C (model 7971 column heater, Jones Chromatography, Resolution Systems, Holland, MI) to improve resolution and increase the separation speed by using a gradient elution profile as follows: 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 for 2 min; and 100–5% B in 1 min. The proteins separated by NPS-RP-HPLC were monitored at 214 nm and collected into 1.5 mL Eppendorf tubes using a Beckman SC-100 fraction collector controlled by semiautomated acquisition program. Protein collection was performed off-line according to the peaks detected from the HPLC separation where 40–50 fractions, each containing a volume in the range of 100–500 μL , were obtained.

2.4 NPS-RP-HPLC/ESI-TOF MS

Intact M_r analysis was performed by analyzing eluent from NPS-RP HPLC for on-line ESI-TOF MS (LCT, Waters-Micromass, Manchester, UK). The separation was performed under the same experimental conditions as in the previous section except that TFA was substituted with 0.3% formic acid (Sigma) in both mobile phases to improve ESI efficiency. A splitter system was used so that 40% of eluent from the HPLC was delivered to the LCT. The capillary voltage for electrospray was set at 3200 V, sample cone at 40 V, extraction cone at 3 V, and reflection lens at 750 V. Desolvation was enhanced by controlling the desolvation temperature at 300°C and source temperature at 120°C . The nitrogen gas flow was maintained at approximately 650 L/h. One mass spectrum was acquired *per* second. The intact M_r value was obtained by deconvoluting the combined ESI spectra of the protein utilizing the MaxEnt1 feature of MassLynx software version 4.0 (Waters-Micromass).

2.5 Protein digestion

The proteins collected off-line by NPS-RP-HPLC separation were completely dried down using a SpeedVac (Labconco, Kansas City, MO). Fifty microliters of 50 mM ammonium bicarbonate at pH 7.8, filtered through 0.22 μm and 0.5 μg of L-1-tosylamido-2-phenylethyl chloromethyl ketone modified sequencing-grade trypsin (Promega, Madison, WI), was added for digestion at 37°C for 18 h and stopped by adding 1 μL of 10% v/v TFA. The digestion mixture was divided into 10 μL for MALDI-TOF MS analysis and the rest dried completely with a SpeedVac. Prior to monolith-based LC-MS and LC-MS/MS analysis, peptides were reconstituted in 5 μL of deionized water.

2.6 MALDI-TOF MS

Each digested protein was analyzed by MALDI-TOF MS. Prior to spotting, each of the digested samples were desalted and concentrated by using C18 ZipTip (Millipore, Bedford, MA) and eluted into 0.5% v/v TFA/60% v/v ACN. The MALDI-matrix solution was prepared by diluting saturated α -cyanohydroxycinnamic acid (α -CHCA; Sigma) with 0.5% v/v TFA/60% v/v ACN at 1:4 ratio v/v. The internal standards included angiotensin I, adrenocorticotrophic hormone (ACTH) fragments 1–17 and 18–39 (all from Sigma). Internal standard peptides were added so that a final concentration of 50 fmol each was reached in each spot of the MALDI plate.

Peptide masses were measured by the TofSpec2E (Waters-Micromass) with delayed extraction in reflectron mode with positive polarity using a nitrogen laser (337 nm). The operating voltage was 20 kV and reflectron voltage of 24.5 kV. The pulse voltage used for delayed extraction was set at 2300 V at a 520 ns delay time. The sampling rate was 2 GHz. Peptide mass spectra were internally calibrated resulting in a mass accuracy of 50 ppm or less. The calibrated spectra were processed using MassLynx version 4.0 to obtain monoisotopic experimental masses for submission to MS-Fit available from <http://prospector.ucsf.edu> to search the Swiss-Prot and NCBI databases for protein identification under the species of *Homo sapiens* with no restrictions on M_r and pI range. A maximum of two missed cleavages was allowed, and cysteine was unmodified. The possible modifications including oxidation of methionine, N-terminal acetylation, and phosphorylation at S, Y, and, T were allowed.

2.7 Monolith-based HPLC separation and on-line interfacing with ESI-TOF MS

The Ultra-Plus II MD Capillary Pump module (Micro-Tech Scientific, Vista, CA) with a homebuilt column heater utilizing a variable autotransformer (Staco Energy Product, Dayton, OH) was used for all chromatographic experiments with the monolithic-capillary column. Monolithic-capillary columns of dimension of 360 μm od \times 200 μm id \times 60 mm L were prepared by copolymerizing styrene and divinylbenzene according to the protocol described elsewhere [19]. The capillary column was directly mounted to a microinjector (model C4-1004-.5, Valco Instruments, Houston, TX) with a 500 nL internal sample loop and a microtight union with 5 nL swept volume (Upchurch Scientific, Oak Harbor, WA) was used to connect a capillary tubing to the LCT. The flow rate of the solvent delivery pump was set at 0.5 mL/min, which was split precolumn to produce a flow rate of 2.5 $\mu\text{L}/\text{min}$ at 60°C through the

monolithic-capillary column. A mobile-phase system of two solvents was used, where solvents A and B are composed of 0.05% formic acid (Sigma) in DI water and ACN, respectively. A linear gradient of 0–100% B in 18 min was applied, immediately followed by column equilibration.

Each digested protein was analyzed using monolith-based HPLC interfaced on-line with the LCT. The desolvation temperature was maintained at 120°C and source temperature at 80°C. The nitrogen gas flow was controlled at 90 L/h. One mass spectrum was acquired *per* second. The monoisotopic peptide M_r values were obtained by utilizing MaxEnt3 feature of MassLynx version 4.0 software by allowing maximum charge state of up to +3 in the mass range of 500–4000 Da. All processed mass spectra were subjected to PMF analysis by database searching in the manner described in Section 2.6.

2.8 Monolith-based HPLC-MS/MS

In order to sequence the tryptic peptide sequence of interest, HPLC-MS/MS experiments were performed using linear IT MS (LTQ, Thermo Finnigan, San Jose, CA). An identical platform for HPLC separation, described in the previous section, was utilized. The capillary transfer tube was set at 175°C, and electrospray needle was held at +3.5 kV. A sheath gas flow of 15 arbitrary units was used. The ion activation was achieved by utilizing helium at a normalized collision energy of 35%. All MS/MS data obtained were analyzed by using the TurboSequest feature of Bioworks 3.1 SR1 (Thermo Finnigan). By allowing the maximum missed cleavage of two, peptide ions are automatically assigned with the X_{corr} values to consider >3.5 for +3 ions, >2.5 for +2 ions, and >1.5 for +1 ions.

3 Results and discussion

The monolithic column was used in this work to analyze proteins separated from breast cancer cell lines. All proteins from the human breast cancer cells in this experiment were analyzed with the same monolithic-capillary column. The column exhibits excellent ruggedness, where the stability of this type of column has been described elsewhere [24]. The salts and other species that might otherwise clog the column in CE-MS were cleaned up by running a 100% aqueous solvent prior to RP separation to wash off the impurities. In addition, only a few femtomoles of injection was possible with CE-MS in previous work [15], making it difficult to analyze proteins of low abundance, whereas monolith-based LC offers much higher loadability to improve detection.

3.1 High protein sequence coverage with monolithic LC-MS

A CA1a cell lysate was separated by 2-D liquid separations where NPS-RP-HPLC separation of proteins pre-fractionated at the pH of 7.55 is shown in Fig. 2. A total of ten proteins contained in this fraction were analyzed for identification through PMF analysis. The protein digests were analyzed by capillary monolithic LC-MS where a typical high-resolution separation profile, completed in less than 10 min, is shown in Fig. 3 for one of the proteins (peak no. 6 in Fig. 2). Typical full peak widths at half height are in the range of 3–5 s. The speed of separation is important since in analysis of complex cell lysates, there may be large numbers of proteins that need to be analyzed. Considering that recovery of proteins separated by NPS silica column is approximately 80% [25], it is estimated by performing peak quantitation with manual

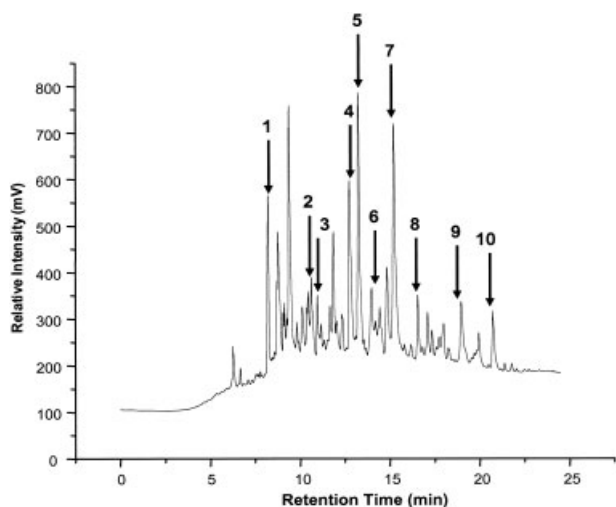


Figure 2. NPS-RP-HPLC chromatogram of proteins in pH fraction of 7.55 from CA1a cell line with peaks annotated for protein identification results shown in Table 1.

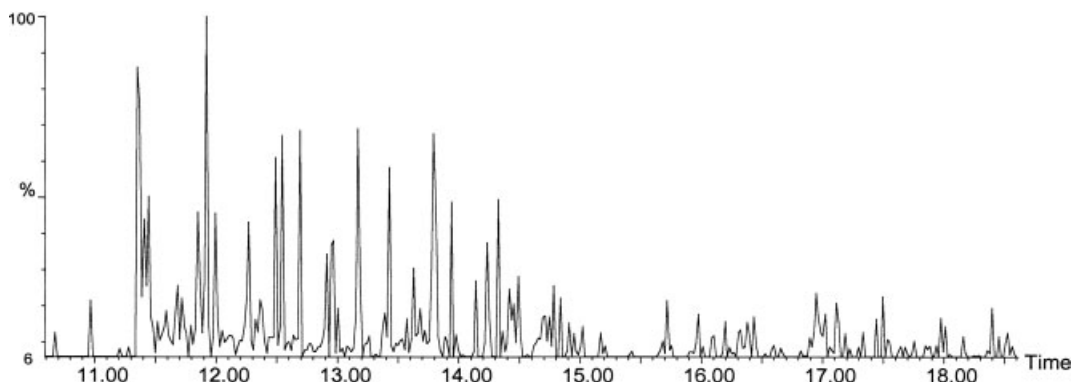


Figure 3. Monolith-based RP-HPLC-MS chromatogram of tryptic digest of fructose biphosphate aldolase A.

baseline by Origin software (version 6.0, Microcal Software, Northampton, MA, USA) that approximately 40 ng of protein digest was injected into the monolithic-capillary column. The mass analysis and database search identify this protein as fructose biphosphate aldolase A, a 40 kDa protein so that approximately 1 pmol was consumed for analysis.

Table 1 summarizes the comparison between theoretical *versus* experimental intact M_r , pI , and the sequence coverage by monolith-based LC/MS and MALDI-TOF MS for each protein denoted in Fig. 2. Also shown is the overall sequence coverage obtained by combining peptide masses from each method into database searching. In many cases, high sequence coverage was readily obtained solely by the monolithic-LC-MS method. Figure 4A illustrates the protein coverage map by each MS method for annexin II, which is found to be highly expressed in the malignant cell of human breast cancer and suggested as a potential biomarker based on the previous differential expression study [9]. The sequence coverage of 90% for this protein of ca. 40 kDa was analyzed by monolithic LC-MS alone. As shown in the coverage map, monolithic LC-MS missed only three peptides, including 1–10, 79–88, and 207–212, excluding three tryptic peptides composed of one or two amino acids that cannot be detected in the mass range set for 500–4000 Da for both of the mass spectrometric methods. This implies that the protein coverage by monolithic LC-MS was almost complete. A similar observation can be made with superoxide dismutase, for which the coverage map is shown in Fig. 4B. For this protein, a single tryptic peptide sequence composed of 38 amino acids (157–194; 4236.16 Da) without any cleavage site and a single amino acid residue (222) were not detected, of which detection is limited by the mass range examined in this experiment.

Although significantly higher protein sequence coverage was consistently obtained with monolith-based LC/MS as compared to MALDI-TOF MS, it is observed that the



Figure 4. Illustration of protein sequence coverage maps (red bar – by monolith-based HPLC/ESI-TOF MS; blue bar – by MALDI-TOF MS) of (A) annexin II and (B) superoxide dismutase.

Table 1. Comparison between theoretical and experimental intact M_r and pI and sequence coverage between different MS methods for proteins in pH fraction of 7.55 of CA1a cell line (see Fig. 2 for peak no.)

Peak number	Protein name	Accession number	Theoretical		Experimental		Protein sequence coverage (number of amino acids)		
			M_r	pI	M_r	pI	MALDI-TOF	Monolith LC/MS	Overall
1	Ubiquitin	P68198 M	8 565	6.6	8 565	7.55	61% (46/76)	95% (72/76) 100% (76/76) ^{b)}	95% (72/76)
2	Peptidyl-prolyl <i>cis-trans</i> isomerase A	P62937	18 013	7.7	18 052	7.55	46% (76/165)	78% (128/165)	90% (149/165)
3	Nucleophosmin	P06748	32 575	4.6	33 104	7.55	59% (173/294)	69% (203/294)	84% (248/294)
4	Lamin A	P02545 ^{a)}	74 140	6.6	74 601	7.55	44% (290/664)	82% (545/664)	86% (573/664)
	Lamin C		65 135	6.4	66 056	7.55	55% (316/572)	69% (457/664) ^{b)} 91% (520/572) 77% (438/572) ^{b)}	96% (548/572)
	Lamin A Δ 10		70 660	8.6	70 114	7.55	44% (272/634)	83% (527/634) 71% (451/634) ^{b)}	88% (555/634)
5	Superoxide dismutase	P04179	24 722	8.3	24 606	7.55	60% (134/222)	78% (174/222)	82% (183/222)
6	Fructose biphosphate aldolase A	P04075	39 420	8.3	39 282	7.55	40% (147/364)	71% (257/364)	80% (293/364)
7	α -Enolase	P06733	47 169	7.0	47 073	7.55	48% (210/434)	91% (397/434)	91% (397/434)
8	Annexin II	P07355	38 604	7.6	38 614	7.55	47% (158/339)	90% (304/339)	95% (321/339)
9	ATP synthase α chain	P25705	59 751	9.2	59 750	7.55	47% (259/553)	76% (423/553)	80% (445/553)
10	Phosphoglycerate kinase 1	P00558	44 615	8.3	44 528	7.55	55% (231/417)	88% (369/417)	88% (369/417)

a) Isoforms found. Database searching performed with both Swiss-Prot and NCBI databases.

b) These percentage coverages are the previous results from CE-MS, adapted from [15] with permission.

addition of peptide mass data from MALDI-TOF MS contributed to an increase of >10% of the overall protein sequence coverage for many of the proteins analyzed. It shows that this unique approach of combining different MS methods typically covers from 80% to over 90% of the overall sequence of the proteins ranging in size from a few thousands to over 70 kDa. Table 2 shows that proteins contained in other pH fractions of CA1a and a different cell line, CA1d, were also successfully analyzed.

Occasionally, the tryptic peptide sequences that were detected by neither monolith-based LC-MS or MALDI-TOF MS were detected and successfully sequenced by linear IT MS. Two of the tryptic peptides of heterogeneous nuclear ribonucleoproteins A2/B1, LFVGGIK, shown in Fig. 5 with the MS/MS spectrum, and EESGKPGAHVTVK (110–112) were not detected by either method for PMF analysis, yielding the overall coverage of only 74%. However, the tandem mass spectrometric analysis identified these peptides, therefore increasing the overall sequence coverage to 80%.

3.2 More enhanced peptide detection by monolithic LC-MS

In previous work, lamin, which helps maintain and establish the shape and strength of the interphase nucleus [26, 27], was analyzed with CE-MS with sequence coverage ranging from ca. 68–76% for a few isoforms (Table 1) [15]. The same protein was analyzed by monolithic LC-MS to reveal sequence coverage of higher than 80–90%, which increased up to 96% with the addition of the results from MALDI-TOF MS. Figure 6A shows the monolithic LC-MS

spectrum scanned over the actual separation time of a digest of lamin identifying five peptide sequences, mostly within mass accuracy of less than 100 ppm, that were not detected by the CE-MS method. Given that both methods utilized the same ionization method, it is strongly believed that the separation characteristics of monolithic LC, including high recovery and loadability, may have contributed to revealing peptides that went undetected by CE. This is an important observation because peptides bearing certain PTMs are difficult to detect due to their presence in relatively low concentrations and the monolithic LC may be a suitable means to overcome this problem. Two additional tryptic peptide sequences in lamin, key to identifying sequence variations involving isoforms, were detected by monolithic LC-MS analysis and are discussed in the following section. Figure 6B shows the direct comparison of sequence coverage for lamin isoforms obtained by monolithic LC-MS and CE-MS, where it is clearly visualized that enhanced sequence coverage was observed by monolithic LC-MS.

3.3 Analysis of isoforms and PTMs using monolithic LC-MS and NPS-RP-LC-MS

In addition to very high protein sequence coverage by PMF analysis, an excellent match between experimental M_r obtained by NPS-RP-HPLC/ESI-TOF MS and the theoretical M_r of many of the proteins analyzed in this experiment serves to further suggest that the protein identification procedures are highly reliable, as several proteins in Tables 1, 2 exhibited a very close agreement. Ubiquitin, for example, contains no suggested modifications in its sequence, and therefore experimental and theoretical M_r 's are identical.

Table 2. Comparison between theoretical and experimental intact M_r and pI and sequence coverage between different MS methods for proteins in other pH fractions of CA1a and CA1d cell lines

Cell line	Protein name	Accession number	Theoretical		Experimental		Protein sequence coverage (number of amino acids)		
			M_r	pI	M_r	pI	MALDI-TOF	Monolith LC-MS	Overall
CA1a	Splicing factor, arginine/serine-rich 1	Q07955	27 745	10.4	28 163	7.00	58% (145/248)	73% (182/248)	88% (218/248)
CA1a	Actin ^{a)}	P63260	41 793	5.3	41 714	6.21	53% (198/375)	81% (305/375)	88% (331/375)
	(γ -actin)	P60710	41 737	5.3	41 670	6.21	48% (181/375)	77% (287/375)	83% (313/375)
	(β -actin)	Q04695	48 106	5.0	48 077	5.71	56% (241/432)	72% (309/432)	84% (365/432)
CA1d	Keratin, type I cytoskeletal 17 (CK17)	P22626	37 430	9.0	37 411	6.91	43% (153/353)	74% (261/353)	80% (281/353)
CA1d	Heterogeneous nuclear ribonucleoproteins A2/B1 ^{b)}								

a) Isoforms found.

b) Overall protein sequence coverage includes two tryptic peptides (110–112 and 114–120) analyzed by monolithic LC/MS/MS.

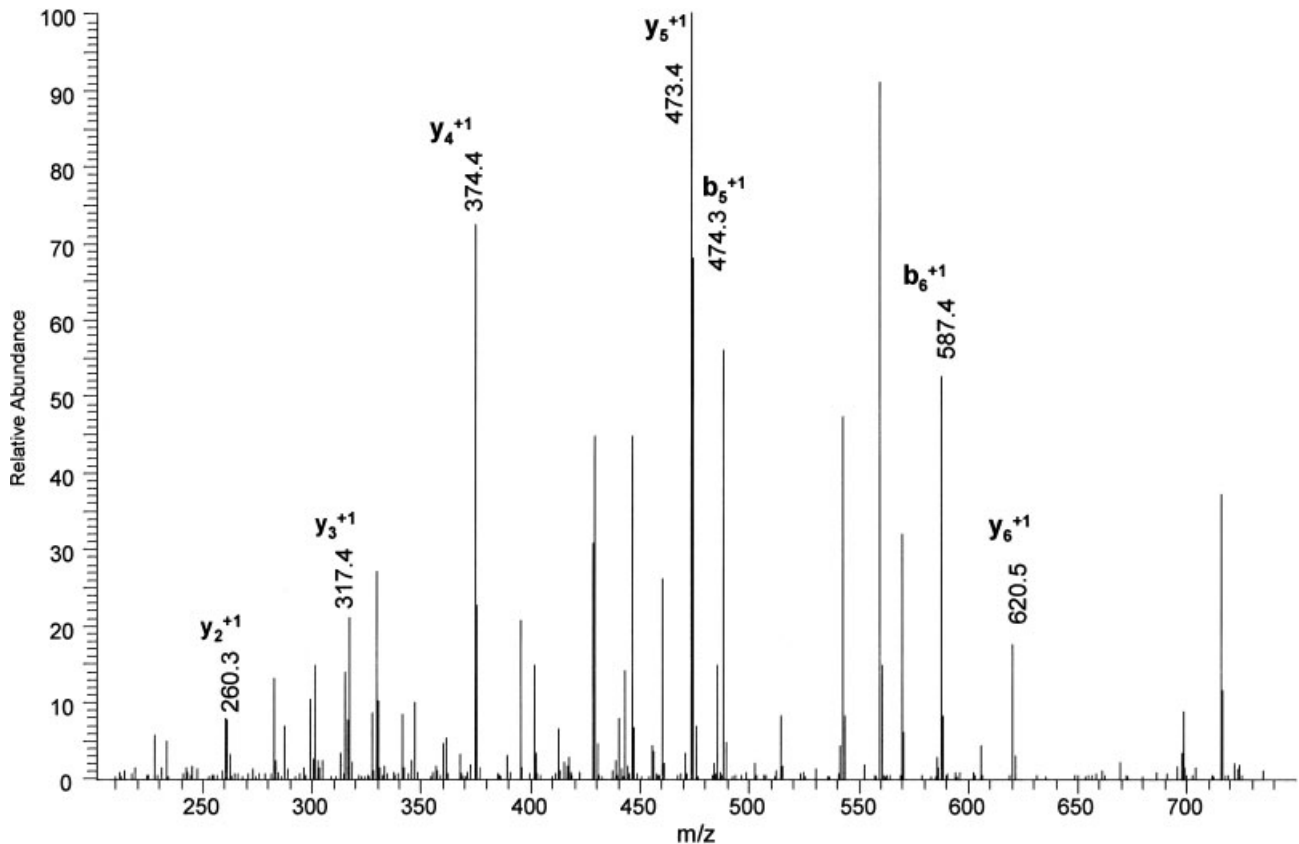


Figure 5. Monolith-based LC-MS/MS spectrum of the tryptic peptide, LFVGGIK (114–120), of heterogeneous nuclear ribonucleoprotein A2/B1.

In our previous study [15], the usefulness of intact M_r in determining sequence variations of proteins was demonstrated where unique peptides were detected by MALDI-TOF MS. The use of M_r measurements allowed the identification of the presence of isoforms of lamin. A single gene of lamin A/C encodes four isoforms, the most widely studied being lamins A and C, as well as ADelta10, resulting from alternative splicing [26, 28]. The lamin C is identical to that of lamin A for the sequence (1–566), while a slight difference is observed in (567–572) where GSHCSS in lamin A is replaced by VSGSRR in lamin C. The rest of the sequence in lamin A (573–664) is absent in lamin C. In lamin ADelta10, a sequence (537–566) is missing from lamin A, the rest being identical [27]. In the present study, we successfully identified unique peptides with no missed cleavages contained in two isoforms, lamins A and C, with both monolith LC-MS and MALDI-TOF MS, as summarized in Table 3. Figures 7A and B show the detection of the peptide sequences unique to lamins A and C, respectively, by monolithic LC-MS. Figure 8 shows that one of these unique peptides, TALINST-GEEVAMR (528–541), was successfully sequenced, further confirming its detection by utilizing monolith-based

LC separation. The unique peptide of lamin ADelta10 was not detected by the present study. It is interesting to note that none of these unique peptides were observed by CE-MS analysis in our previous study, while their presence was observed using monolithic LC-MS, further suggesting its usefulness in PMF analysis.

Another interesting observation of isoforms was made with cytoplasmic actin. There are three main groups of actin isoforms in vertebrates, including α , β , and γ forms [29, 30]. β -Actin and γ -actin are known to coexist in most cell types as components of the cytoskeleton and as mediators of internal cell motility. This protein has been studied earlier based on intact M_r from NPS-RP-HPLC [31], although unique peptides were not found by MALDI-TOF MS analysis due to their very high degree of homology where 98.9% (371/375) of the sequence is identical. In this work, both β -actin and γ -actin were identified with overall coverage of 83 and 88%, respectively, with closely matching intact M_r values, when peptide maps from different MS methods were combined (Table 2). The unique peptides of these two isoforms are found in four of the first ten amino acids in their sequences, MEEEIAALVI for

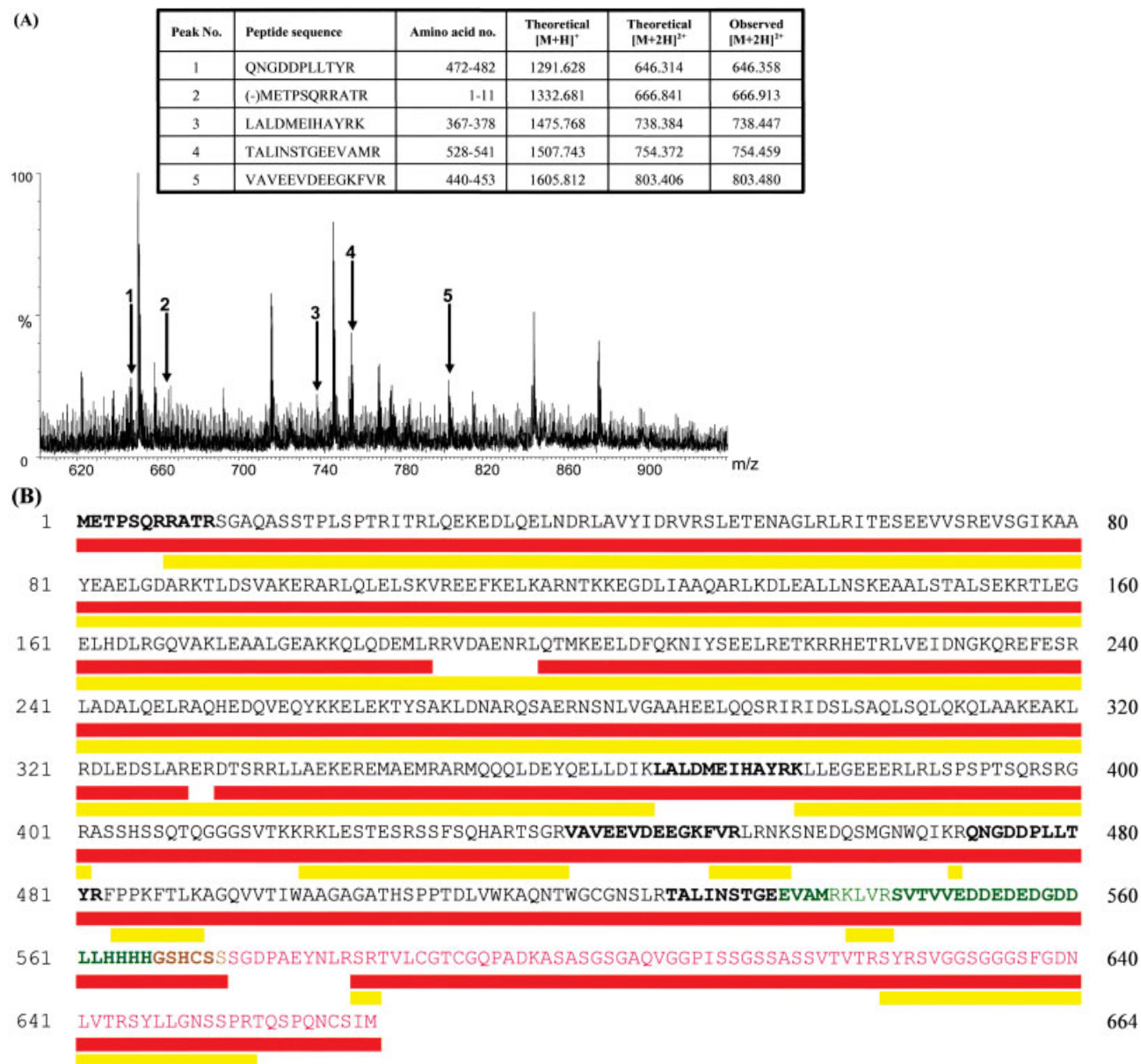


Figure 6. (A) Illustration of several tryptic peptides (sequence information in table inset) from lamin detected by monolith-based HPLC/ESI-TOF MS that were not detected by CE-ESI-TOF MS in previous study. (B) Comparison of protein sequence coverage for lamin A/C (red bar – monolith-based LC-MS; yellow bar – CE-MS, adapted from [15] with permission). All tryptic peptides only detected by monolithic LC-MS are in bold. Note: Sequence in green (537–566) is missing in lamin Adelta10. Sequence in pink (573–664) is missing in lamin C. Sequence in brown (567–572) is replaced with VSGSRR in lamin C.

γ -actin and MDDDI AALV V for β -actin, while the remainder of the 371 amino acids in the sequence of these proteins are identical, making it a difficult task to distinguish them.

In their mature forms, the first amino acid residue of the sequence, methionine, of both forms of actins is missing and the second amino acid in each of the sequence, glutamic acid for γ -actin and aspartic acid for β -actin, are acetylated at the N-terminus [32–34]. Many studies report

that the histidine residue at position 73 is methylated in actins [35, 36]. Considering these modifications and alterations in sequence, theoretically, γ -actin should have MW of 41 718 Da, which closely matches our experimentally determined intact MW of 41 714 Da, or less than 100 ppm of mass accuracy. The same consideration can be made with β -actin which should have MW of 41 662 Da, in comparison to the MW of 41 670 Da determined experimentally. In Fig. 9, the observation of the

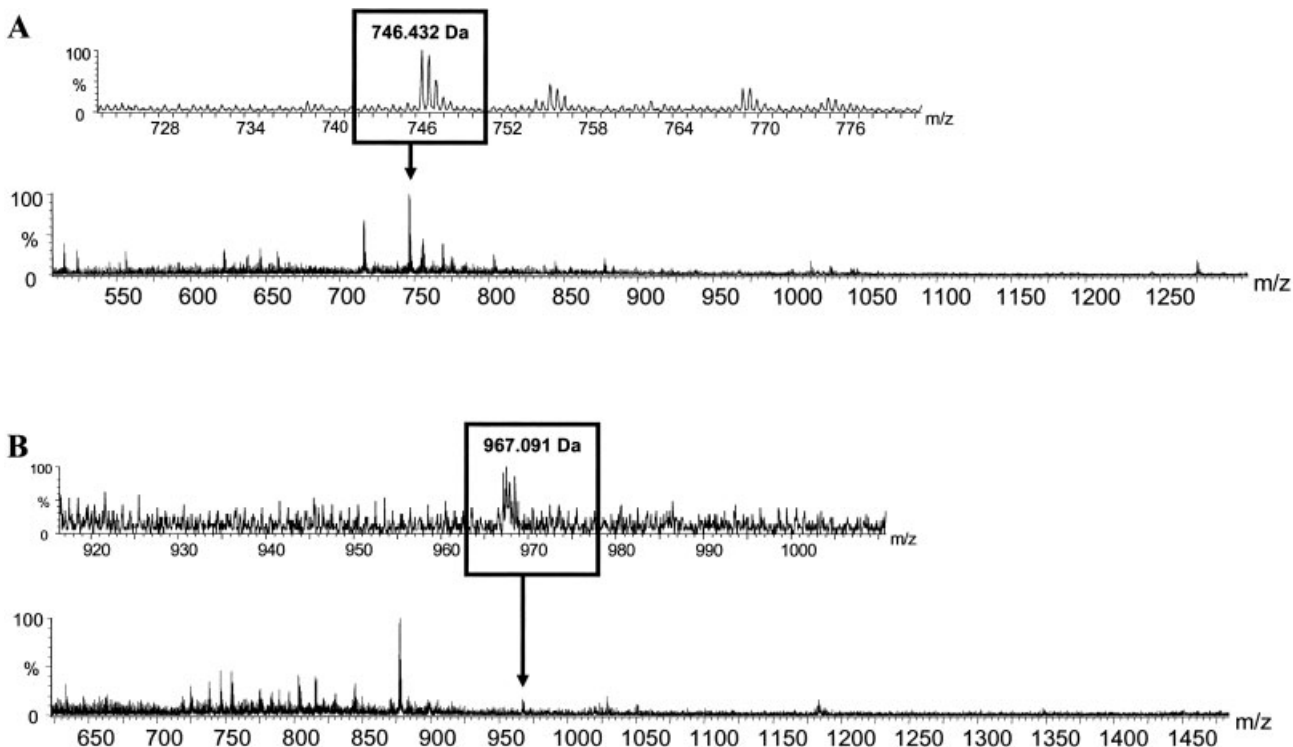


Figure 7. Identification of unique peptides (multiply charged; see Table 3 for peptide sequence information) in the isoforms of lamin, (A) lamins A and C and (B) lamin C, detected by monolith-based HPLC/ESI-TOF MS with zoomed-in view.

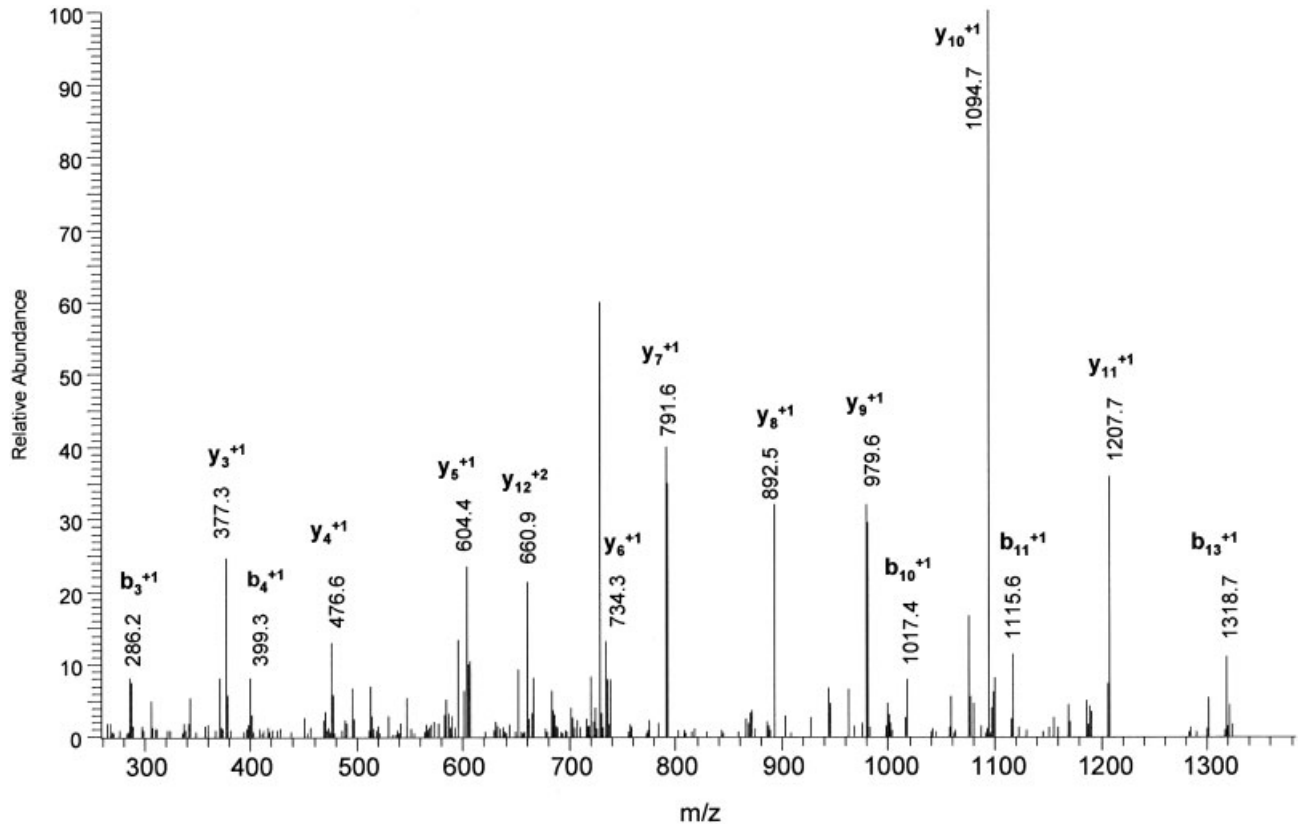


Figure 8. Monolith-based LC-MS/MS spectrum of the tryptic peptide, TALINSTGEEVAMR (528–541), of lamins A and C.

Table 3. Unique tryptic peptides detected by different MS methods to distinguish isoforms

Protein name	Isoform(s)	Unique tryptic peptide sequence (amino acids)	[M + H] ⁺	MS method to detect		
				MALDI-TOF MS	Monolith LC-MS	CE-MS
Lamin A/C	Lamins A and C	TALINSTGEEVAMR (528–541)	1491.747	Yes	Yes	No
	Lamin C	SVTVVEDEDEGGDLLHHHVSGSR (546–571)	2899.272	Yes	Yes	No
Actin ^{a)}	β-Actin	DDDIAALWDNGSGMCK (2–18)	1764.778	No	No	NA
	γ-Actin	EEEIAALVIDNGSGMCK (2–18)	1820.841	Yes	Yes	NA

a) [M + H]⁺ values include *N*-terminal acetylation for each of the peptide.

presence of the unique peptide sequence of γ-actin with *M_r* including acetylation at a glutamic acid residue in its mature form by both monolithic LC-MS and MALDI-TOF MS is illustrated. Neither MS method revealed the unique peptide sequence contained in β-actin.

A closer examination of the mass spectrum obtained for actin allowed for the detection of the presence of another PTM. Shown in Fig. 10 is a doubly and triply charged peptide sequence containing methylated histidine, YPIEHGIVTNWDDMEK (69–84, 1960.911 Da), by monolithic LC-MS. This modified sequence was also found by MALDI-TOF MS analysis (data not shown). The modification at this particular residue of this peptide sequence, common to both β- and γ-actins, was suggested by a recent study where it appears to play a role in polymerization of actin and ATP hydrolysis [37]. Although detailed structural study by MS/MS is necessary for characterization, the results presented here strongly indicate that monolith-based LC-MS followed by PMF analysis is capable of identifying the presence of PTMs.

4 Concluding remarks

A method for obtaining consistently high sequence coverage of proteins separated from lysates of human breast cancer cell lines has been successfully demonstrated by utilizing a polymer-based monolithic-capillary column for LC-MS to obtain a rapid and high-resolution separation. Very high sequence coverage of proteins readily obtained by combining PMF results from monolithic LC/MS and MALDI-TOF MS enhances the reliability of the protein identification procedures. Although CE-MS could also be used to analyze these samples, an important advantage of the monolith HPLC-ESI-TOF MS is the ruggedness of the technique. In CE-MS using a sheathless interface the electrical connection has been found to be the limiting factor where the connection needs to be redone after a couple of runs. Although CE can provide much improved resolution [38] compared to the monolithic columns in the 10 min separation interval, with the use of MS analysis

sufficient resolution is obtained for distinguishing each peak. In addition, the monolithic columns provide much improved loadability compared to CE and much shorter separation times compared to packed capillary chromatography.

The monolithic column separations coupled to MS was also used to elucidate the presence of sequence variations, such as isoforms, and PTMs of proteins, aided by intact *M_r* information and MS/MS using linear IT MS. It is important to emphasize the uniqueness of this study due to the nature of samples analyzed, which originate from highly complex biological mixtures, suggesting the versatility of this approach for many other applications. In addition, this method requires no sample preparation or purification upon completion of enzymatic digestion of proteins prior to MS analysis.

Several proteins analyzed in this work, including annexin II, fructose bisphosphate aldolase A, phosphoglycerate kinase 1, and peptidyl-prolyl *cis-trans* isomerase A, have been reported to be highly expressed in tumor cells and described as potential biomarkers based on differential expression studies [9, 39]. Although further work will be necessary to fully characterize these proteins, it is clearly demonstrated in this study that the method of monolith-based LC on-line hyphenated with MS has great potential to become a high-throughput methodology. It may be used to characterize large numbers of potential biomarkers in various types of tumor and other diseased cells and to study the presence of a wide variety of modifications and structural changes in protein expression.

This work was supported in part by the National Institutes of Health under grant No. R01 GM 49500(DML) and the National Cancer Institute under grant No. R01CA90503(FRM, DML) and the National Science Foundation under NSF SBIR Award No. 0321763 to Syagen Technology, and a subaward to the University of Michigan. The MALDI-TOF MS instrument used in this work was funded by the National Science Foundation under grant No. DBI 99874. We thank Frank J. Yang of

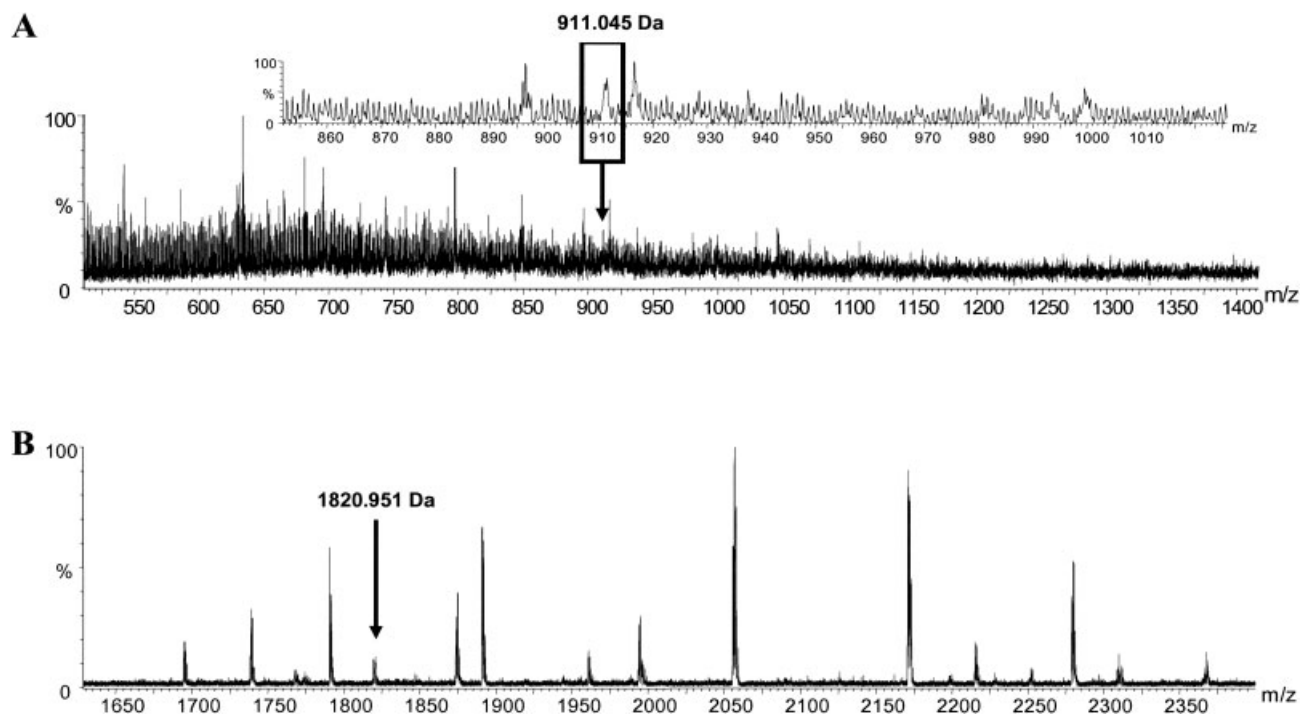


Figure 9. Identification of unique peptide (see Table 3 for peptide sequence information) in γ -actin by (A) monolith-based HPLC/ESI-TOF MS with zoomed-in view and (B) MALDI-TOF MS.

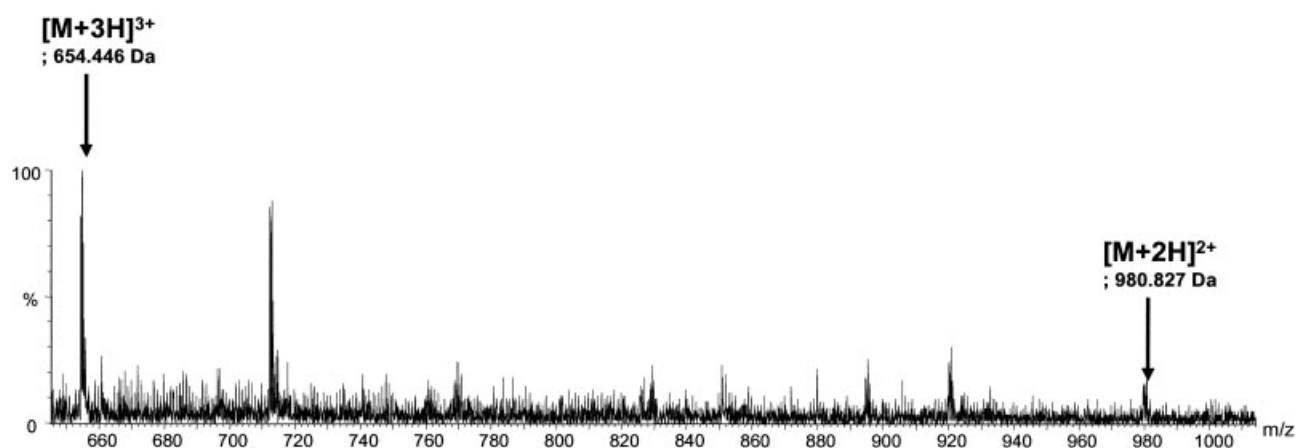


Figure 10. Identification of multiply charged tryptic peptide containing methylated histidine residue in actin by monolith-based HPLC/ESI-TOF MS.

Micro-Tech Scientific for the use of the Ultra-Plus II MD Capillary Pump and Reiner Wintringer for preparing the monolithic separation columns used in this study.

5 References

- [1] Vlahou, A., Fountoulakis, M., *J. Chromatogr. B* 2005, 814, 11–19.
- [2] O'Donovan, C., Apweiler, R., Bairoch, A., *Trends Biotechnol.* 2001, 19, 178–181.
- [3] Mann, M., Jensen, O. N., *Nat. Biotechnol.* 2003, 21, 255–261.
- [4] Han, K. K., Martinage, A., *Int. J. Biochem.* 1992, 24, 19–28.
- [5] Wu, W., Hu, W., Kavanagh, J. J., *Int. J. Gynecol. Cancer* 2002, 12, 409–423.
- [6] O'Farrell, P. H., *J. Biol. Chem.* 1975, 250, 4007–4021.
- [7] Bernova-Giorginni, S., *Trends Anal. Chem.* 2003, 22, 273–281.
- [8] Goodlett, D. R., Yi, E. C., *Funct. Integr. Genomics* 2002, 2, 138–153.

- [9] Hamler, R. L., Zhu, K., Miller, F. R., Lubman, D. M. et al., *Proteomics* 2004, 4, 562–577.
- [10] Kreunin, P., Urquidi, V., Lubman, D. M., Goodison, S., *Proteomics* 2004, 4, 2754–2765.
- [11] Wang, H., Kachman, M. T., Schwartz, D. R., Cho, K. R., Lubman, D. M., *Proteomics* 2004, 4, 2476–2495.
- [12] Padliya, N. D., Wood, T. D., *Proteomics* 2004, 4, 466–473.
- [13] Krause, E., Wenschuh, H., Jungblut, P. R., *Anal. Chem.* 1999, 71, 4160–4165.
- [14] Hale, J. E., Butler, J. P., Knierman, M. D., Becker, G. W., *Anal. Biochem.* 2000, 287, 110–117.
- [15] Zhu, K., Kim, J., Yoo, C., Miller, F. R., Lubman, D. M., *Anal. Chem.* 2003, 75, 6209–6217.
- [16] Moore, R. E., Licklider, L., Schumann, D., Lee, T. D., *Anal. Chem.* 1998, 70, 4879–4884.
- [17] Walcher, W., Oberacher, H., Troiani, S., Huber, C. G. et al., *J. Chromatogr. B* 2002, 782, 111–125.
- [18] Holzl, G., Oberacher, H., Pitsch, S., Stutz, A., Huber, C. G., *Anal. Chem.* 2005, 77, 673–680.
- [19] Premstaller, A., Oberacher, H., Huber, C. G., *Anal. Chem.* 2000, 72, 4386–4393.
- [20] Premstaller, A., Oberacher, H., Walcher, W., Timperio, A. M. et al., *Anal. Chem.* 2001, 73, 2390–2396.
- [21] Lee, D., Svec, F., Frechet, J. M. J., *J. Chromatogr. A* 2004, 1051, 53–60.
- [22] Kimura, H., Tanigawa, T., Morisaka, H., Ikegami, T. et al., *J. Sep. Sci.* 2004, 27, 897–904.
- [23] Santner, S. J., Dawson, P. J., Tait, L., Soule, H. D. et al., *Breast Cancer Res. Treat.* 2001, 65, 101–110.
- [24] Toll, H., Oberacher, H., Swart, R., Huber, C. G., *J. Chromatogr. A* 2005, 1079, 274–286.
- [25] Wall, D. B., Lubman, D. M., Flynn, S. J., *Anal. Chem.* 1999, 71, 3894–3900.
- [26] Broers, J. L., Hutchinson, C. J., Ramaekers, F. C., *J. Pathol.* 2004, 204, 478–488.
- [27] Moir, R. D., Spann, T. P., *Cell. Mol. Life Sci.* 2001, 58, 1748–1757.
- [28] Wilson, K. L., Zastrow, M. S., Lee, K. K., *Cell* 2001, 104, 647–650.
- [29] Vandekerckhove, J., Weber, K., *J. Mol. Biol.* 1978, 126, 783–802.
- [30] Garrels, J. I., Gibson, W., *Cell* 1976, 9, 793–805.
- [31] Wall, D. B., Kachman, M. T., Gong, S. S., Lubman, D. M. et al., *Rapid Commun. Mass Spectrom.* 2001, 15, 1649–1661.
- [32] Rubenstein, P. A., Martin, D. J., *J. Biol. Chem.* 1983, 258, 3961–3966.
- [33] Vandekerckhove, J., Weber, K., *Eur. J. Biochem.* 1978, 90, 451–462.
- [34] Abe, A., Saeki, K., Yasunaga, T., Wakabayashi, T., *Biochem. Biophys. Res. Commun.* 2000, 268, 14–19.
- [35] Asatoor, A. M., Armstrong, M. D., *Biochem. Biophys. Res. Commun.* 1967, 26, 168–174.
- [36] Raghavan, M., Smith, C. K., Schutt, C. E., *Anal. Biochem.* 1989, 178, 194–197.
- [37] Nyman, T., Schuler, H., Korenbaum, E., Lindberg, U. et al., *J. Mol. Biol.* 2002, 317, 577–589.
- [38] Pelzing, M., Neusüss, C., *Electrophoresis* 2005, 26, 2717–2728.
- [39] Bini, L., Magi, B., Marzocci, B., Tosi, P. et al., *Electrophoresis* 1997, 18, 2832–2841.