RAPID COMMUNICATIONS IN MASS SPECTROMETRY
Rapid Commun. Mass Spectrom. 2001; **15**: 1649–1661

Isoelectric focusing nonporous silica reversed-phase high-performance liquid chromatography/electrospray ionization time-of-flight mass spectrometry: a three-dimensional liquid-phase protein separation method as applied to the human erythroleukemia cell-line

Daniel B. Wall[1], Maureen T. Kachman[1], Siyuan S. Gong[1], Stephen J. Parus[1], Michael W. Long[2] and David M. Lubman[1]*

[1]Department of Chemistry, The University of Michigan, Ann Arbor, MI 48109-1055, USA
[2]Department of Pediatrics, School of Medicine, The University of Michigan, Ann Arbor, MI 48109, USA

Received 13 July 2001; Accepted 13 July 2001

SPONSOR REFEREE: Prof. Neil Kelleher, 53 Roger Adams Laboratory Box 47-5, MC-712600 S. Matthews Avenue, Urbana, IL 61801, USA

A liquid-phase three-dimensional protein separation method has been developed that is used to separate the cytosolic fraction of a HEL cell lysate via isoelectric focusing (IEF), nonporous silica (NPS) reversed-phase high-performance liquid chromatography (RP-HPLC) and electrospray ionization time-of-flight mass spectrometry (ESI-TOFMS), respectively. Several hundred unique protein molecular weights were observed in a pI range from 4.8 to 8.5 and a mass range from 5 to 85 kDa. Proteins were positively identified by analysis of the pI (±0.5 pI units), an intact protein molecular weight (±150 ppm), and peptide mass mapping results. Using the molecular weight (MW) and peptide mapping results of identified proteins it was possible to characterize their posttranslational (PTMs) and/or sequence modifications. PTMs were detected on both forms of cytosolic actin, heat shock 90 β, HINT and α-enolase. Sequence modifications or conflicts were observed for β-and γ-actin, ATP β-synthase and heat shock 90 β. IEF-NPS-RP-HPLC/ESI-TOFMS was used to determine experimental pI, MW and relative hydrophobicity values for each protein detected. This data was used to generate a 2-D pI-MS protein map, where proteins are displayed according to their pI and molecular weight. Protein molecular weight peaks are represented as bands in the 2-D pI-MS image where the gray scale of each band is proportional to the intensity of the protein molecular weight peak. In addition, a third hydrophobicity dimension (%B) was added as the % acetonitrile elution to generate a 3-D pI-MS-%B plot where each protein can be tagged according to three parameters.

An important area of research in modern biology involves the study of the proteome. With the initial publication of the human genome map, the need for methods to profile the proteome to monitor changes in protein expression or structure due to cell transformation and cancer progression or the response of cells to chemotherapeutic treatment will become increasingly important. The method generally used to separate and image large numbers of proteins from cells is two-dimensional polyacrylamide gel electrophoresis (2-D PAGE).[1] This technique is capable of resolving thousands of proteins[2] based upon pI in one dimension and protein size in the second dimension.

A major limitation of the gel method is that the proteins are separated according to size and not MW. This mechanism is prone to reproducibility problems and can only generate a rough estimate of the protein's MW (±10% mass accuracy). Method reproducibility becomes important when comparing the results of different experiments. The mass accuracy and resolution of the method become important when one is studying posttranslational or sequence modifications to the protein as well as proteins that have similar MW values that are difficult to resolve. While a gel-separated protein can be analyzed by mass spectrometry based peptide mass mapping and MS/MS methods, to determine the protein's posttranslational and sequence modifications, this method generally provides for limited sequence coverage.

**Correspondence to*: D. M. Lubman, Department of Chemistry, The University of Michigan, Ann Arbor, MI 48109-1055, USA.
E-mail: dmlubman@umich.edu
Contract/grant sponsor: National Institutes of Health; Contract/grant number: 2-R01GM49500-5.
Contract/grant sponsor: National Science Foundation; Contract/grant number: DBI-9987220.

DOI:10.1002/rcm.421

Alternatively, an accurate intact protein MW, determined by mass spectrometry, would reflect the full sequence and all modifications present.[3]

There have been several strategies developed in previous work to determine, by mass spectrometry based methods, intact MW values of proteins separated in gels.[4] The major difficulty is finding a means to analyze intact proteins which are embedded in the gel. MW analysis of gel-separated intact proteins has been achieved by extracting proteins from the gel using electro-elution[5] and solvent extraction[6] methods. Other strategies have involved electro-blotting the proteins to membranes and either performing matrix-assisted laser desorption/ionization (MALDI) directly from the PVDF membrane[7–9] or using procedures to dissolve a nitrocellulose membrane so that it can be directly applied to the MALDI-MS sample probe.[10,11] In other work, the MW values of intact proteins have been obtained directly from 1-D IEF gels in vacuum using MALDI-MS.[12] Although reasonable mass accuracy of ± 5000 ppm[8] has been achieved, the resolution and sensitivity of MALDI-MS becomes limited at high mass, which in turn limits the accuracy of the method.

More recent work has used liquid-based methods to separate proteins prior to mass spectrometric analysis. One such method is on-line CIEF-FTICRMS, which has been used to obtain masses of hundreds of putative proteins with high resolution.[13,14] RP-HPLC/ESI/MS has also been performed using a triple quadrupole instrument to detect the masses of a simple mixture of intrinsic thylakoid membrane proteins though with poor mass accuracy.[15] Other groups have used multi-dimensional liquid-phase separations to prepare proteins for analysis by MALDI and/or ESI mass spectrometry.[16] This kind of work has included 2-D liquid chromatography,[17,18] 1-D liquid-phase IEF,[16,19,20] 1-D continuous elution PAGE,[21] 2-D preparative LPE,[19,20,22] 2-D semi-preparative electrophoresis,[16,20,23] and IEF-RP-HPLC.[24]

In this work, a novel gel-free 3-D separation method for the determination of accurate protein MW values, protein mapping and protein identification is described. The sample analyzed is the cytosolic fraction of a whole cell lysate of the human erythroleukemia (HEL) cell-line. A liquid-phase IEF method is initially used to fractionate proteins from the HEL cell lysate according to pI. The protein pI fractions are then analyzed using nonporous silica (NPS) RP-HPLC with on-line protein detection by ESI-TOFMS. ESI-TOFMS provides rapid mass analysis of specific protein pH fractions and yields high mass resolution and high mass accuracy of intact protein molecular weights. The proteins are identified by the use of MW, pI, and tryptic digest mass mapping results. The presence of sequence modifications and posttranslational modifications (PTMs) is determined from the intact protein MW values and identified from the peptide mass map analysis and Swiss-Prot database information. Detected proteins are imaged in a number of different protein mass map formats. One format images the proteins according to pI and MW in a 2-D pI-MS protein map. The map provides information analogous to 2-D gels in that each protein can be tagged based upon pI and MW. In an alternative 3-D format, proteins are mapped according to their pI, MW and percent acetonitrile at time of elution (%B). The addition of the third parameter allows for greater separation of closely overlapping proteins in mixtures and may have value as an additional parameter with which to tag a protein.

EXPERIMENTAL

HEL Con5d cytosolic protein sample preparation

The HEL cell-line samples were prepared in the laboratory of Dr. Michael Long at the University of Michigan in the Department of Pediatrics. The cells were grown in RPMI-1640 medium,[25] washed three times in PBS and then pelletized and frozen at $-80\,^{\circ}$C. The proteins were extracted from the cells by thawing to room temperature and adding three volumes of lysis buffer consisting of 6 M urea (Biorad), 2 M thiourea (Biorad), 1% n-octyl-β-D-galactopyranoside (Sigma), 6% ampholytes (Biorad, 3/10), 10 mM TCEP (Pierce), 10 mM DTT (Biorad) and 10 mM PMSF (Biorad). The mixture was vortexed for 5 s and allowed to stand for 30 min. The resulting solution was then ultra-centrifuged at 150 000 RCFG for 1.5 h in a Beckman L-70 ultracentrifuge to produce a supernatant that contained primarily cytosolic proteins. The supernatant was then harvested and stored at $-80\,^{\circ}$C until loading to the IEF separation.

IEF with the mini-Rotofor

The HEL Con5d cytosolic protein fraction was loaded to the mini-Rotofor along with a separation buffer containing all the same ingredients as the lysis buffer with a total final separation volume of 15 mL. The mini-Rotofor was then run as in previous work[24] and the pH protein fractions were harvested. The pH of the fractions was determined by use of a mini-pH electrode (PH/C 900, Amersham Pharmacia Biotech.). The concentration of protein in each fraction was determined using the Bradford-based assay (Biorad). Protein pH fractions were stored at $-80\,^{\circ}$C until loading to the nonporous RP-HPLC column.

NPS-RP-HPLC protein separations

The separations were performed as in previous work.[24] Identical separation conditions were used for HPLC fraction collection as well as HPLC on-line mass spectrometry. However, in this work, the 4.6×33 mm columns were replaced with 3×33 mm columns packed with the same 1.5 μm diameter nonporous silica ODS I particles (Eichrom Technologies Inc.). This allowed for lower flow rates (0.2 mL/min.) that were more compatible with the ESI-MS method. The binary gradient went from 0 to 20% B in 1 min, then 20 to 30% B in 2 min, then 30 to 54 %B in 8 min, then 54 to 65% B in 1 min and finally up to 100% B in 1 min. Accounting for a 1-min dwell-time the actual gradient reached 54% in 12 min.

Fraction collection and tryptic digestion

Protein fractions were collected to polypropylene 1.5-mL microtubes using a semi-automated fraction collection system built in-house. Accounting for the delay time between the detection of the proteins and their elution to the tubes, it was possible to accurately collect specific peaks from the HPLC eluent and record the %B at which these fractions eluted. The liquid fraction was speed-vacuumed at 65 °C for 12 min to remove most of the TFA and acetonitrile.

Then 1 M NH_4HCO_3 was added to the tubes to reach a final concentration of 100 mM NH_4HCO_3. 2 μL of a solution of 0.05 μg/uL porcine trypsin (Promega) were then added. The tubes were then vortexed and stored in warm room at 37 °C for 24 to 48 h to allow digestion to occur. Digests were prepared for MALDI-MS by either solid-phase extraction (SPE) using ZipTips (Millipore) or direct application to the MALDI probe on top of a thin layer of pure nitrocellulose (NC, BA83, 0.2 μm pore, Schleicher and Schuell). The NC method provided more complete peptide mass maps and significantly higher sensitivity. In both cases, the MALDI matrix consisted of a freshly prepared solution of α-cyanohydroxycinnamic acid (α-CHCA, Sigma) in 1:1 water/acetonitrile (0.1% TFA) diluted 4-fold from a saturated solution. The digest and matrix solutions were added to the probe surface in a 1:1 (v/v) ratio and allowed to air-dry. Digests for capillary RP-HPLC/ESI-TOFMS were frozen at −20 °C until analysis.

MALDI-MS

MALDI-MS was performed using a delayed extraction reflectron equipped MALDI-TOF instrument (STR, Perseptive). The repeller voltage was set at +25 kV, the grid voltage was 72% of the repeller voltage, the delay time was 100 ns and the reflectron was set at a ratio of 1.12. Each mass spectrum was generated from an average of 75 individual mass spectra. Peptides were calibrated with close spot external calibration using a standard peptide calibration mixture (Std 2, Sequazyme, Perseptive) of angiotensin I, ACTH clips and bovine insulin (Sigma).

ESI-MS

Electrospray ionization mass spectrometry was performed using an ESI TOF instrument (LCT, Micromass). Ions were generated from a z-spray source with the nitrogen desolvation gas at 400 °C and a flow rate of 600 L/h. The source block was held at 150 °C and the nebulizer gas was held at a relatively high flow rate. The capillary voltage was + 2500 V, the sample cone was at + 45 V, the extraction cone was at +3 V, the hexapole RF was + 1000 V with a DC offset of +7 V. The second hexapole was biased to −2 V and the detector was held at −3000 V. The maximum TOF of each spectrum was 75 μs and the mass spectra were combined every 0.95 s. This yields a mass spectral acquisition rate of 0.95/0.000075 or 12666 Hz. Data was processed using Masslynx v3.4 software and protein multi-charged umbrellas were deconvoluted using Maxent 1 software (Micromass).[26–28] The mass precision was optimized by loading 0.1 μg of bovine insulin with each NPS-RP-HPLC sample and then using the 1912.197 peak as a lock mass to fine tune the external calibration for each mass spectrum. External calibration was performed prior to each separation using direct infusion of a standard NaI CsI solution. Typical flow rates to the mass spectrometer were 0.1 to 0.2 mL/min with a binary gradient of water (0.1% TFA, 0.3% formic acid) and acetonitrile (0.1% TFA, 0.3% formic acid) ranging from 0 to 100% organic.

Table 1. pH values for Rotofor fractions and protein loads for NPS-RP-HPLC separations

pH	Rotofor Data Conc. μg/uL	NPS-RP-HPLC Data V (μL) load HPLC	Mass load (μg)
4.8	0.282	200.00	56.4
5.1	0.462	200.00	92.4
5.3	0.570	200.00	114.0
5.6	0.650	200.00	130.0
6.0	0.595	200.00	119.0
6.3	0.463	200.00	92.6
6.5	0.399	200.00	79.8
6.8	0.401	200.00	80.2
7.0	0.330	200.00	66.0
7.3	0.272	200.00	54.4
7.6	0.300	200.00	60.0
8.0	0.307	200.00	61.4
8.5	0.302	200.00	60.4

RESULTS AND DISCUSSION

The sample studied in this work is the cytosolic protein fraction of a human erythroleukemia cell-line (HEL), which was grown from the bone marrow tissue of a patient with leukemia. The HEL cell-line constitutively expresses an erythroid phenotype but is tri-phenotypic and can develop into each of the three myeloid lineages: granulocyte, erythrocyte or megakaryocyte.[25]

NPS-RP-HPLC/ESI TOFMS and protein MW

The mini-Rotofor was used to prepare pI focused protein fractions in the liquid phase from the cytosolic whole cell lysate fraction. Each fraction was then analyzed by NPS-RP-HPLC with on-line detection using ESI-TOFMS. The protein concentration and fraction pH values are listed in Table 1. For the purpose of illustrating the general method the fraction at pH 5.1 will be used as an example.

NPS-RP-HPLC/ESI-TOFMS generates a raw data file total ion chromatogram (TIC) (Fig. 1(A)). From this TIC, it is possible to detect proteins as an array of mass/charge peaks representing a series of differing charge states for a particular protein. This ion series is termed a protein umbrella (Fig. 1(B)) and can be deconvoluted using MaxEnt software to yield a protein molecular weight peak (Fig. 1(C)). This protein molecular weight peak retains the high 5000 mass resolution of the smaller multiply charged peaks as well as the quantitative value of the protein umbrella signal.[26–28]

The two examples given in Fig. 1 eluted at 49.3 %B and 53.3 %B with the 49.3 %B peak containing a protein of MW 50148 Da. The 53.3 %B peak contains a protein of MW 51843 Da. From this data alone it is not possible to identify these proteins but in the next section these MW values will be associated with protein digest identification data by association of the MW and the peptide mass map via a particular percent acetonitrile at time of elution.

NPS-RP-HPLC/MALDI-MS peptide mass mapping

Given that the protein elutes at a particular point in time, it is possible to determine the percent acetonitrile at time of

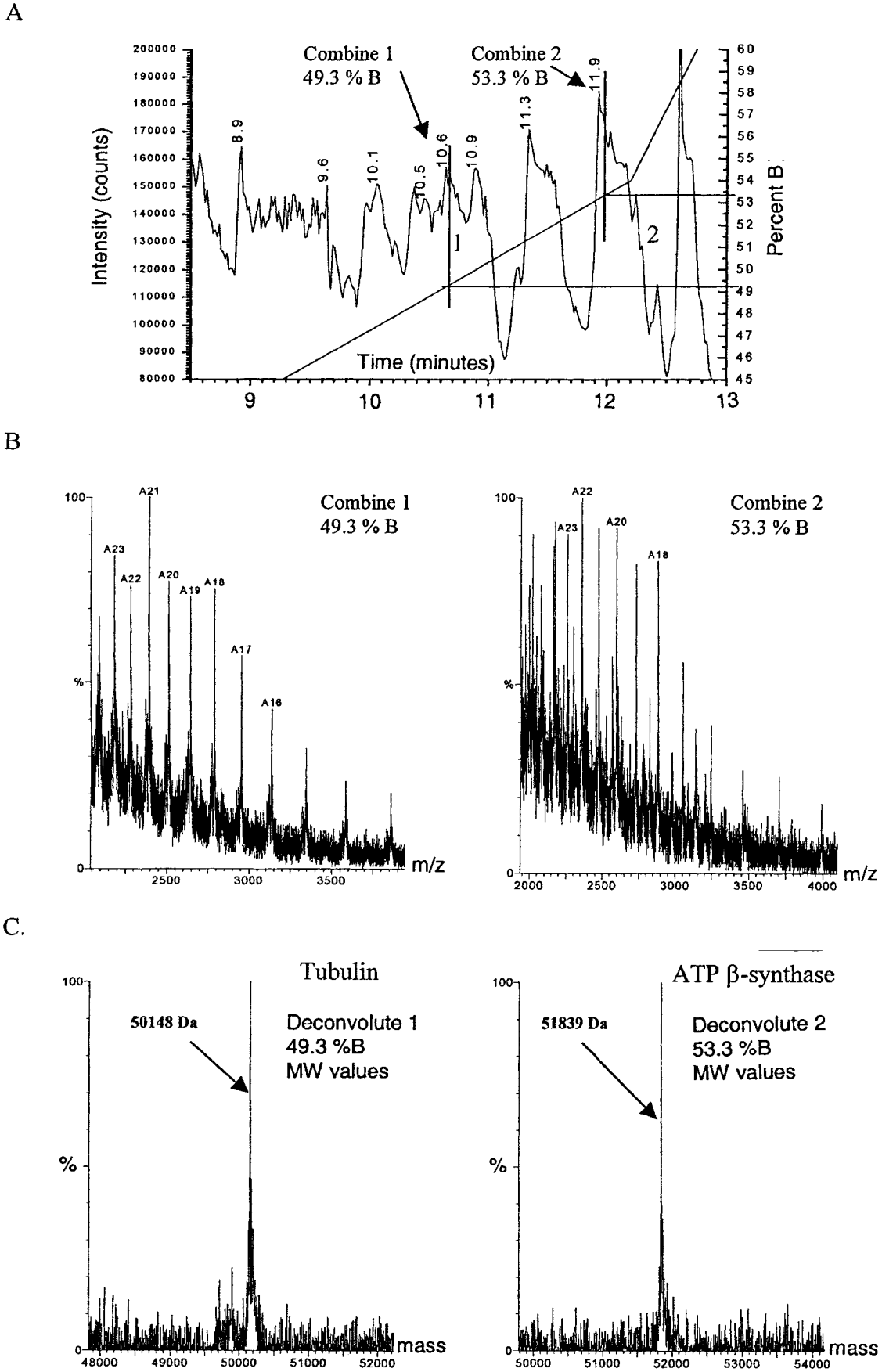

Figure 1. Generation of protein molecular weight mass values from the total ion chromatogram of a NPS RP HPLC-ESI TOF/MS analysis where (A) TIC of the NPS RP HPLC separation of Rotofor fraction with pH 5.1 obtained using detection by ESI-TOFMS, (B) multiply charged peaks obtained from ESI-TOFMS detection of intact proteins at 49.3 and 53.3 %B elution from the NPS-RP-HPLC separations, and (C) deconvoluted protein MW mass spectra of the protein umbrellas from (B).

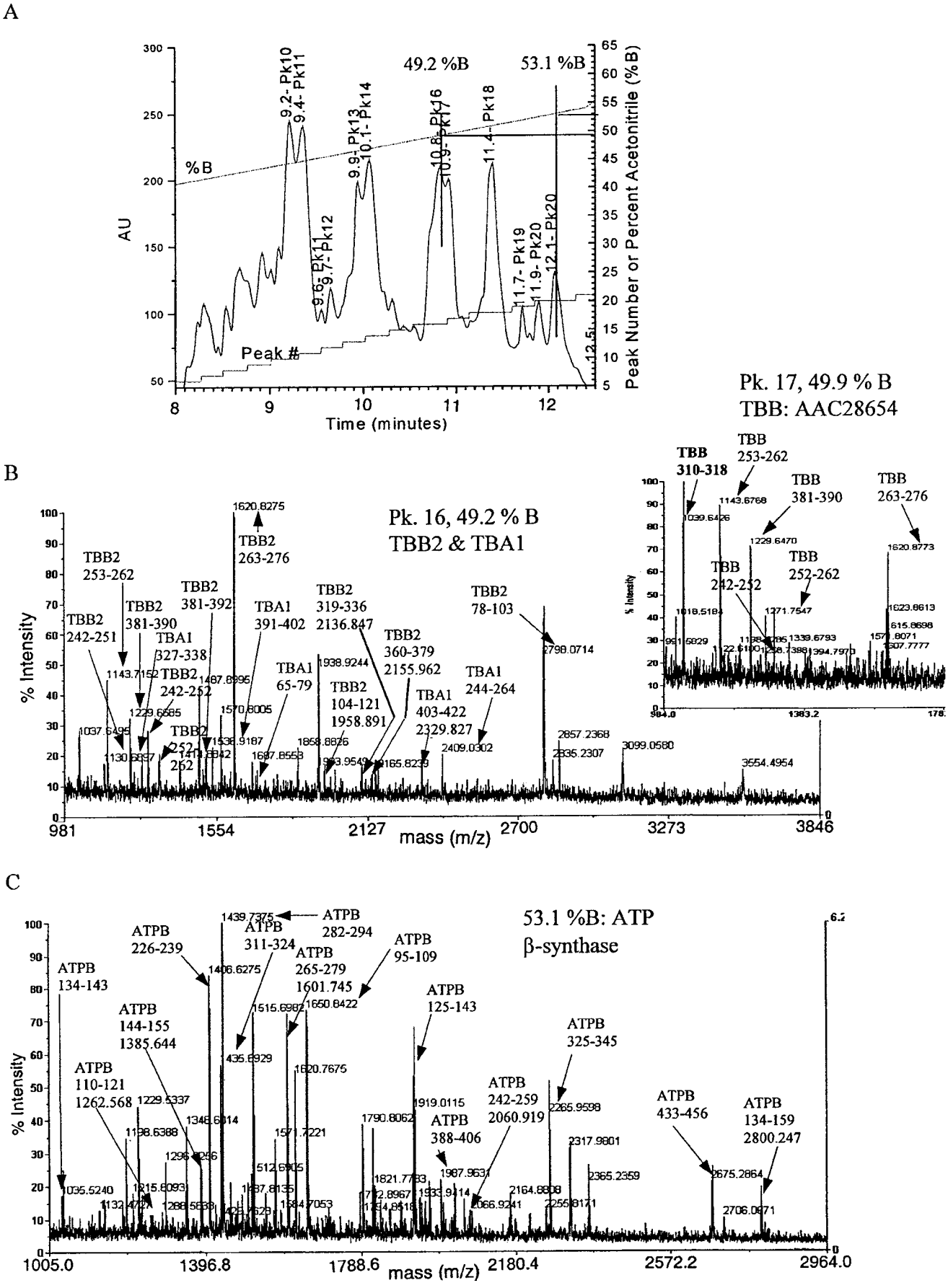

Figure 2. (A) NPS-RP-HPLC separation of the protein content of Rotofor fraction with pH 5.1 by UV detection, (B,C) the tryptic digest peptide mass maps of the protein eluent of the separation at 49.2 and 53.1 %B with MALDI-MS detection.

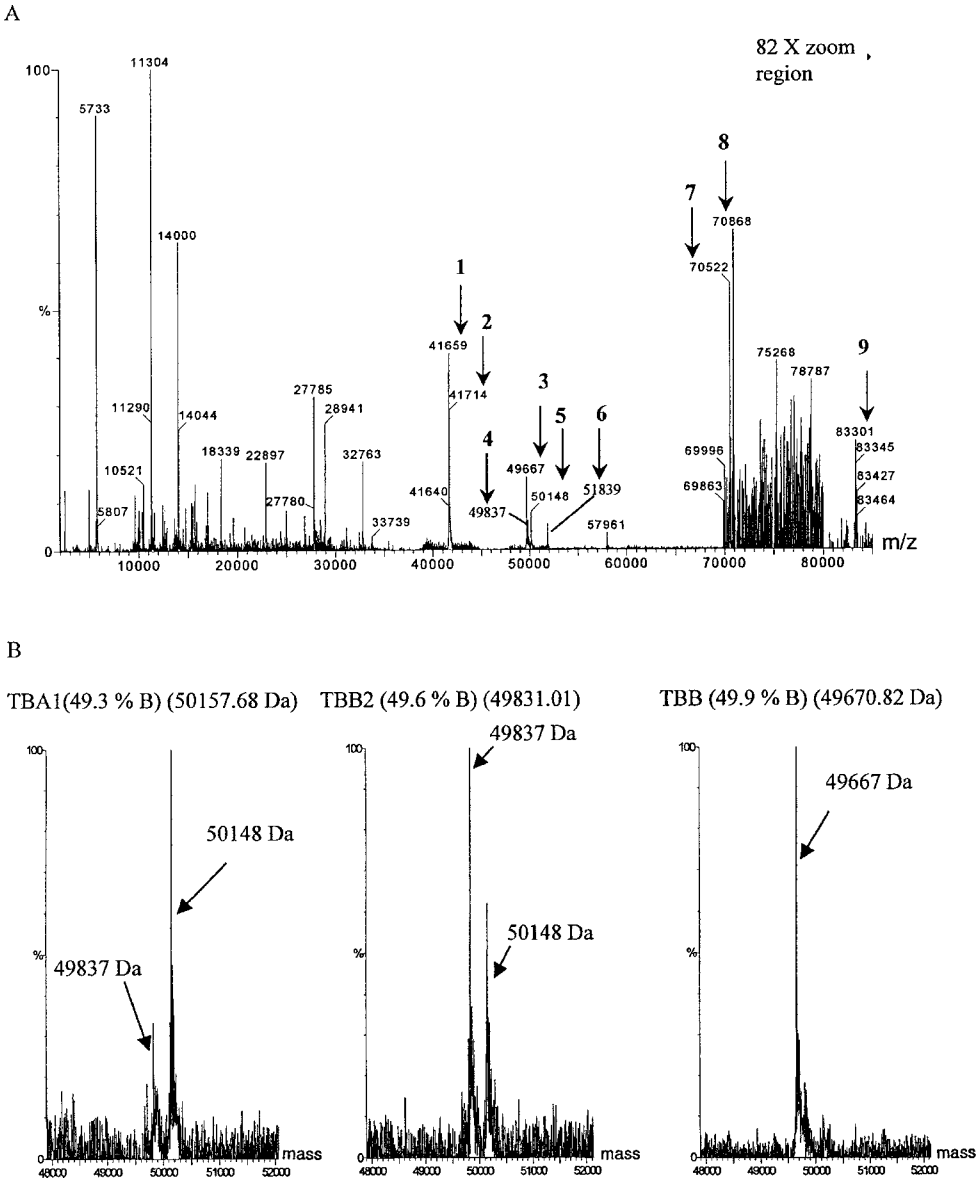

Figure 3. (A) Summed molecular weight mass spectrum of the protein content of the Rotofor pH 5.1 fraction and (B) mass spectra of the tubulin isoforms as detected as a function of %B.

elution (%B) after accounting for the system dwell time and flow rate. This value is reproducible from separation to separation (± 0.2 %) and can be used to predict the time of elution of a protein because it is based on a fundamental and stable property of the protein, its hydrophobicity. The protein molecular weight and percent acetonitrile can thus be linked. When subsequent separations are performed (Fig. 2(A)) and protein fractions collected for peptide mass mapping, the protein identification obtained from the mass analysis of the digest of a particular percent acetonitrile fraction can thus be associated with the full intact protein molecular weight that eluted in the same percent acetonitrile. Although the resolution of the RP-HPLC separation in Fig. 2(A) is limited, due to the rapid (11 min) gradient and the high sample loading to the 3 mm i.d. NPS column, the use of ESI-MS in this work readily allows resolution of proteins that are not separated by the 2-D liquid separation.

To illustrate how intact protein ESI-MS and protein digest MALDI-MS data can be effectively associated with one another, the MW (Fig. 1(C)) and peptide mass maps (Fig. 2(C)) for tubulin and ATP β-synthase will be examined. In the case of tubulin, the protein detected at 49.3 %B has a MW of 50148 Da (Fig. 1(C)) while the peptide mapping results from the fraction at 49.2 %B identified the tubulin α1 (50157.68 Da, 5.02 pI) (Fig. 2(B)) yielding a mass accuracy of 192 ppm. For the case of ATPB, the protein eluting at 53.3 %B has a MW of 51839 Da (Fig. 1(C)) while the peptide mass map analysis of the fraction at 53.1 %B identifies the mature form of ATPB (51830.3 Da, 5.00 pI) yielding a mass accuracy of 168 ppm. These two examples show that this method can associate a protein digest with a protein MW via the %B at time of elution. The validity of this method in which protein MW and digest data can be associated was confirmed by direct infusion of small aliquots of selected protein fractions for analysis by ESI-TOFMS prior to their tryptic digestion (data not shown).

Table 2. Protein identifications from pH 5.1 digest and MW information (Fig. **3)**

MS Pk number	Protein Name	Swiss-Prot & PTM pI	Swiss-Prot & PTM MW*	PTM	pI Exp	MW Exp	Mass Acc. Ppm**	%B MW ESI-MS	%B digest
1	ACTB_HUMAN_1	5.29	41661.61	Acet, Meth	5.4	41659	63	51.3	50.8
2	ACTG_HUMAN_1	5.31	41717.72	Acet, Meth	5.4	41714	89	51.3	50.8
3	TBB_AAC28654	4.78	49670.82	None	5.1	49667	77	49.9	49.8
4	TBB2_HUMAN	4.79	49831.01	None	5.2	49837	120	49.6	49.4
5	TBA1_HUMAN	5.02	50157.68	None	5.5	50148	192	49.3	49.2
6	ATPB_HUMAN_1	5.00	51830.30	Seq. change	5.1	51839	168	53.3	53.1
7	GR78_HUMAN_1	4.98	70261.26	Unknown	5.1	70522	NA	45.2	44.9
8	HS76_HUMAN	5.67	70853.88	None	5.6	70868	199	48.3	48.2
9	HS9B_HUMAN	4.97	83292.90	SeqCh & 2PO4	5.1	83301	97	53.0	52.8

* Protein MW values reflect the sequence and posttranslational modifications.
** Mass accuracy between modified protein MW values and experimental MW values.

Summation of individual protein MW mass spectra into a complete protein mass map for the IEF fraction at pH 5.1 with sequence and PTM analysis

After deconvoluting individual mass spectra from a TIC (Fig. 1) it is possible to sum all MW mass spectra from one TIC into a single mass spectrum (Fig. 3(A)). In this figure one can see over 50 distinct protein MW peaks. Table 2 shows a list of nine of these proteins that were selected for identification and detailed modification analysis.

Protein posttranslational and sequence modification analysis

The tubulins (TBB, TBB2, TBA1)

The tubulins are cytoplasmic proteins that function as the primary components of the microtubules, which in turn play a major role in many cellular processes including mitosis and cellular structure. The protein tubulin is made up of an alpha and a beta chain.

Using the protein MW values it was possible to observe two isoforms of the beta chain of tubulin and one form of the alpha chain. The beta tubulins observed were TBB2 (49831.01 Da, 4.79 pI) and TBB (49670.82 Da, 4.78). The experimental MW values for TBB2 and TBB are 49837 and 49667 Da, respectively, yielding a mass accuracy of 120 and 77 ppm. From Fig. 3(B) it is clear that TBB2 (peak at 49.3–49.6 %B) elutes earlier than TBB (peak at 49.9 %B) demonstrating how the RP separation is capable of separating two forms of the tubulin β-chain. In addition, the digest peptide mass maps (Fig. 2(B)) identify TBB2 in the 49.2 %B fraction while TBB is only identified in the 49.9 %B fraction. The TBB form (AAC28654) sequence (MW 49670.82, pI 4.78) is almost identical to that of TBB1 (MW 47758.9, pI 4.75) as determined by a BLAST search of the TBB sequence that gave TBB1 as the closest match (data not shown). Accordingly, the digest data fit to both TBB1 and TBB, but, due to the highly accurate MW of 49667, it is possible to determine that this protein is indeed the TBB form and not the TBB1 form. The alpha chain of tubulin, TBA1 (50157.58 Da, 5.02 pI), was observed in the 49.3 %B fraction and had a mass of 50148 Da for a mass accuracy of 192 ppm. The peak collected at 49.2 %B produced a peptide mass map (Fig. 2(B)) that identified the α1 form of tubulin as well as β2. This data clearly indicates the detection of two of the tubulin beta chain forms and one of the alpha chain forms and also shows how the RP separation can be effective in separating closely related proteins (Fig. 3(B)).

β- and γ-Actin (ACTB and ACTG) MW and digest analysis

The beta and gamma actins are cytosolic proteins that play a major role in the cytoskeleton and cellular motility.

The experimental and database masses listed in Table 2 illustrate the typical 150 ppm mass accuracy of this method. A detailed analysis of β-actin and γ-actin, as detected from the pH 5.6 Rotofor fraction, is shown in Fig. 4 where the protein molecular weight mass spectrum and peptide mass map can be seen. Table 3 shows the Peptident Swiss-Prot search results for the tryptic digest of this protein as well as the typical search parameters used.

The most obvious feature of the intact protein mass spectrum is that there are two mass peaks at 41656 and 41712 Da that differ by 56 Da. These two peaks correspond to the two mature and posttranslationally modified forms of actin, these being β-actin (5.29, 41661.61) and γ-actin (5.31, 41717.72) with the isoform mass difference being 56.11 Da. The mass accuracy for the beta and gamma actins is therefore 134 and 136 ppm, respectively. Both forms of actin elute at 51.2 %B, and the peptide mass map search from the same elution region identifies ACTB and/or ACTG (Fig. 4(B)). The peptide mass map identification of ACTB and/or ACTG does not conclusively prove that both forms are present due to the high degree of sequence homology, only that at least one form is present. However the protein MW information combined with the peptide mass map results conclusively show the presence of both ACTB and ACTG.

The initial search using MSFit showed the masses of the immature unmodified forms of these proteins, but the Peptident search, which takes into account database protein sequence modifications, reported the mature form protein MW values reflecting the cleavage of the first amino acid. Clearly, the Peptident search engine, which accounts for known database protein modifications, is an improved program for matching protein digest search results to high mass accuracy protein MW data. The database information from Swiss-Prot indicated that both forms of actin are known to have one methylation and one acetylation; and, with these modifications, the experimental protein MW falls within the 150 ppm mass accuracy expected. Peptident could further

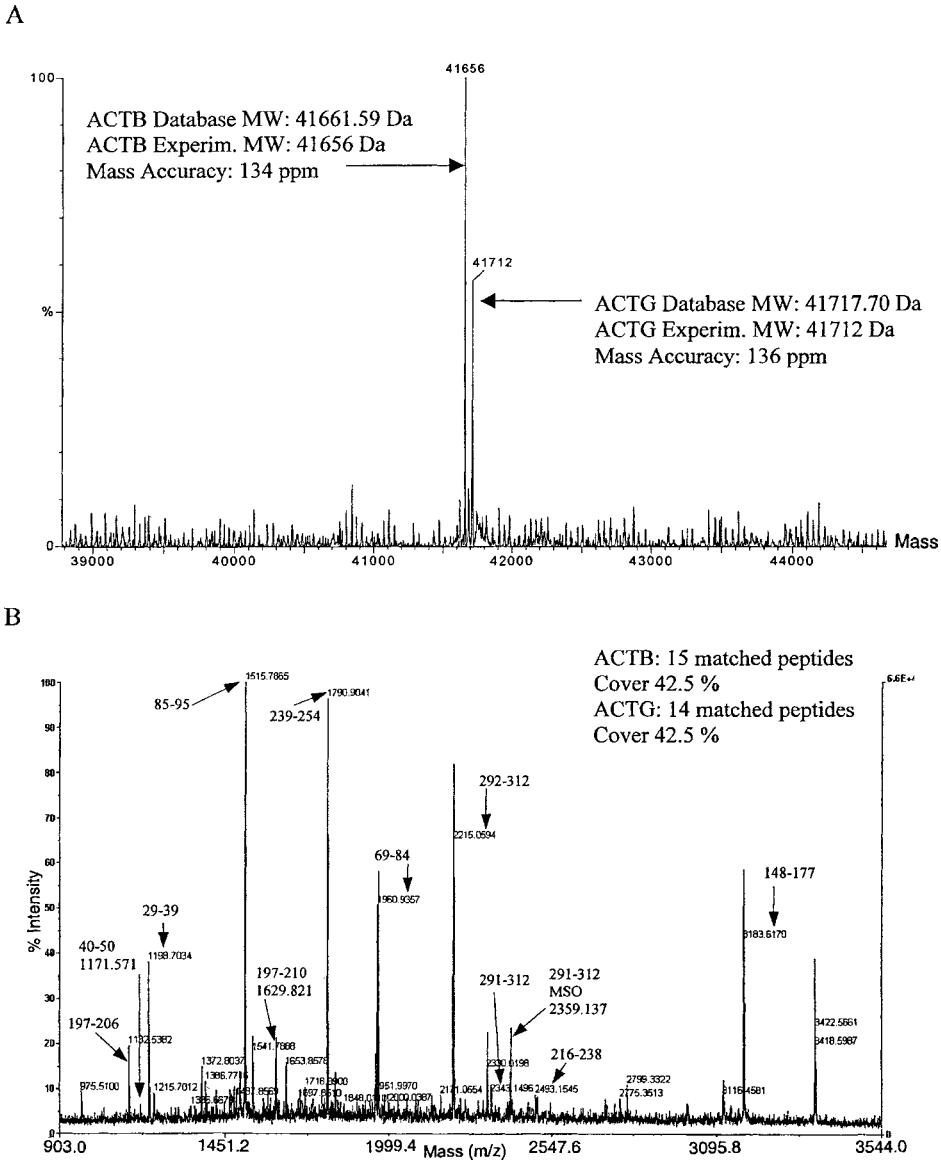

Figure 4. (A) Detail of β- and γ-actin protein molecular weight MS and (B) MALDI-TOFMS of tryptic digest peptide mass map of actin isoforms with labeled peptides.

improve its recognition of modified protein forms by including PTM mass shifts in the reported protein MW as well as the peptide mass mapping results.

The digest data (Fig. 4(B), Table 3) support what the MW data has already suggested regarding posttranslational modification of the actins. This is evidenced by the identification of a methylated peptide (69–84: 1960.911 Da). Not only does this methylation fit the PTM profile of actin, but also its location on the 69–84 peptide matches the information given in the Swiss-Prot database. In addition, it is clear from the MW data that both forms of actin are present and both forms are modified in the same way.

Glyceraldehyde 3PO4 dehydrogenase (G3P2) and α-enolase (ENOA)

G3P2 is a cytoplasmic homotetramer protein that performs the first step in the second phase of glycolysis and has been linked to many types of cancer due to up-regulation of glycolysis during the rapid neoplasia characteristic of carcinogenesis. Similarly α-enolase has been linked to many cancers due to its role in glycolysis in the catalysis of 2-phospho-D-glycerate to phosphoenolpyruvate and water.

In the case of G3P2, the Swiss-Prot database indicates that there are no PTMs or sequence changes. The experimental mass for this protein (pH 8.0 fraction) is 35927 Da and the database mass is 35922.02 giving a mass accuracy of 139 ppm (Table 4). This suggests that the protein is not modified in any way, and this idea is supported by the digest data that do not reveal any PTMs.

From the pH 7.1 fraction α-enolase was determined to have an experimental mass of 47077 Da (Table 4). The Peptident results for the digest of this protein identified α-enolase (47037.77 Da, 7.0 pI) yielding a mass accuracy of 914 ppm. This poor mass accuracy indicates that some modification to the protein is not being taken into account. A detailed LC/MS/MS analysis of α-enolase from the human breast carcinoma cell-line (MDA-MB231)[29] determined that this protein has an N-terminal acetylation, which

Table 3. MALDI-MS peptide mass map results for β- and γ-actin. Peptident search parameters: pI: 4–6, MW 32000–48000, 200 ppm mass error, 2 MC, allow for oxidation of methionines

Score: 0.11, 15 matching peptides: P02570 (ACTB_HUMAN) pI: 5.29, Mw: 41605.54
CHAIN 1: ACTIN, CYTOPLASMIC 1. – Homo sapiens (Human)

User mass	Matching mass	Δmass (ppm)	#MC	Modification	Position	Peptide
1132.538	1132.527	−9.83	0		197–206	GYSFTTTAER
1171.571	1171.571	0.06	0		40–50	HQGVMVGMGQK
1198.703	1198.705	2.06	0		29–39	AVFPSIVGRPR
1515.791	1515.749	−27.6	0		85–95	IWHHTFYNELR
1629.821	1629.823	1.27	1		197–210	GYSFTTTAEREIVR
1744.933	1744.886	−26.73	1		192–206	ILTERGYSFTTTAER
1790.904	1790.892	−6.77	0		239–254	SYELPDGQVITIGNE R
1960.934	1960.911	−11.74	0	1xMETH	69–84	YPIEHGIVTNWDDME K
2215.059	2215.07	4.91	0		292–312	DLYANTVLSGGTTMY PGIADR
2343.15	2343.165	6.35	1		291–312	KDLYANTVLSGGTTM YPGIADR
2359.137	2359.16	9.66	1	MSO: 305	291–312	KDLYANTVLSGGTTM YPGIADR
2493.155	2493.152	−1.05	0		216–238	LCYVALDFEQEMATA ASSSSLEK
3183.617	3183.614	−0.85	0		148–177	TTGIVMDSGDGVTHT VPIYEGYALPHAILR

42.5% of sequence covered:

```
     1          11         21         31         41         51
     |          |          |          |          |          |
   1  dddiaalvv dngsgmckag fagddaprAV FPSIVGRPRH QGVMVGMGQK dsyvgdeaqs    60
  61 krgiltlkYP IEHGIVTNWD DMEKIWHHTF YNELRvapee hpvllteapl npkanrekmt   120
 121 qimfetfntp amyvaiqavl slyasgrTTG IVMDSGDGVT HTVPIYEGYA LPHAILRldl   180
 181 agrdltdylm kILTERGYSF TTTAEREIVR dikekLCYVA LDFEQEMATA ASSSSLEKSY   240
 241 ELPDGQVITI GNERfrcpea lfqpsflgme scgihettfn simkcdvdir KDLYANTVLS   300
 301 GGTTMYPGIA DRmqkeital apstmkikii apperkysvw iggsilasls tfqqmwiskq   360
 361 eydesgpsiv hrkcf
```

Table 4. Summary of protein identifications from a selection of pH fractions and HPLC peaks

Fraction pH	%B	pI database	MW with modification Database	Accession #	ID and *modifications*	pI experim.	MW (Da) experiment	Mass Accuracy ppm
pH 5.3 fraction	51.0	5.3	41661.59	P02570	ACTB_HUMAN_1 *acet, meth*	5.4	41661	−14
	51.0	5.3	41717.70	P02571	ACTG_HUMAN_1 *acet, meth*	5.4	41717	−17
pH 5.6 fraction	48.9	5.4	70853.88	P17066	HS76_HUMAN	5.6	70857	44
	51.2	5.3	41661.59	P02570	ACTB_HUMAN_1 *acet, meth*	5.4	41656	−134
	51.2	5.3	41717.70	P02571	ACTG_HUMAN_1 *acet, meth*	5.4	41712	−137
	50.6	5.7	57962.86	P10809	CH60_HUMAN	5.4	57961	−32
pH 6.5 fraction	35.2	6.4	13712.72	P49773	HINT_HUMAN *acet*	6.5	13712	−53
	41.2	6.5	26538.30	P00938	TPIS_HUMAN	6.5	26534	−162
pH 7.1 fraction	42.1	7.0	47079.81	P06733	ENOA_HUMAN acet	6.7	47077	−60
pH 7.2 fraction	41.1	7.6	57746.62	P14618	KPY1_HUMAN	8.0	57743	−63
	42.2	7.0	47079.81	P06733	ENOA_HUMAN Acet	6.7	47079	−17
pH 7.6 fraction	38.5	8.4	39288.83	P04075	ALFA_HUMAN	8.1	39288	−21
	39.1	8.6	35922.02	P04406	G3P2_HUMAN	8.2	35921	−28
	41.2	7.6	57746.62	P14618	KPY1_HUMAN	8.0	57743	−63
	41.1	8.0	57782.66	P14786	KPY2_HUMAN	8.0	57791	144
	42.0	7.0	47079.81	P06733	ENOA_HUMAN Acet	6.7	47082	47
pH 7.8 fraction	36.4	7.8	17881.30	P05092	CYPH_HUMAN	7.4	17883	95
	41.3	8.4	39288.83	P04075	ALFA_HUMAN	8.1	39289	4
	39.3	8.6	35922.02	P04406	G3P2_HUMAN	8.2	35924	55
	41.3	7.6	57746.62	P14618	KPY1_HUMAN	8.0	57741	−97
	42.4	7.0	47079.81	P06733	ENOA_HUMAN Acet	6.7	47086	131
	39.9	8.0	57782.66	P14786	KPY2_HUMAN	8.0	57778	−81
pH 8.0 fraction	39.0	8.6	35922.02	P04406	G3P2_HUMAN	8.2	35927	139
	38.7	8.4	39288.83	P04075	ALFA_HUMAN	8.1	39294	132

would increase its mass by 42.037 Da. The protein MW should therefore be 47079.81 Da, which would yield a mass accuracy of 60 ppm.

β-Heat shock 90 (HS9B) and GR78

HS9B is a cytoplasmic protein that functions as a molecular chaperone and is in the heat shock 90 family. GR78 is a protein found in the endoplasmic reticulum and is classified with the heat shock 70 protein family.

Heat shock 90 has an experimental mass of 83301 Da while the database mass is 83163.09 Da. Clearly, this mass error of 1659 ppm is very high, indicating some modified form of the protein. The Swiss-Prot database indicates that this protein may have two sequence changes (146: R to T and 176: M to R) and two phosphorylations (PO4) (225 and 254). The mass of *β*-heat shock 90 with these changes applied is 83292.9 Da giving a mass accuracy of 97 ppm indicating that this protein is in the posttranslationally modified form. The protein could also have no sequence modifications and 2 PO4. In this case, the protein's mass would be 83323.09 Da giving a mass error of 276 ppm. This error is high but not too high to eliminate this possibility. The Peptident analysis of the peptide digest data indicates the presence of a phosphorylated peptide (223–241: 2317.908 Da) and a sequence modified peptide (168–185: 1916.042 Da), which supports the presence of the sequence and phosphorylation modified form of the protein. The conclusion may be drawn therefore that this HS9B protein has both the sequence modifications and the PO4s indicated in the Swiss-Prot entry.

The GR78 protein exhibits an experimental mass that is 261 Da above that predicted from the database. At this time there is no explanation for this mass error.

ATP β-synthase (ATPB) and HINT (HINT)

ATPB is a mitochondrial protein that functions in the conversion of ADP to ATP in the presence of a proton gradient across the mitochondrial membrane. The beta form is one of five subunits that make up the catalytic core of the ATP synthase complex.

ATPB is reported to be a 56560 Da protein in the Swiss-Prot database, but there are many possible modifications to the protein. The protein contains a transit peptide from 1–47 that is cleaved off in the mature form of the protein; thus, the mature form of the protein has a mass of 51793 Da. The experimental MW for ATPB with no sequence conflicts incorporated is 51837 Da. Clearly, this protein MW value is not possible due to the high mass error. After taking into account the sequence conflicts for this protein, the alternative database protein MW value is 51830.30 (all Swiss-Prot sequence changes were factored into the protein MW value, see Swiss-Prot entry for ATPB_HUMAN). These mass values agree with 129 ppm mass accuracy. The data suggests that the mature form of the protein matches well with the sequence changes suggested from the Swiss-Prot entry for ATPB.

HINT is also linked to the energy production machinery of the cell in that it catalyzes the cleavage of ADP to AMP and inorganic phosphate. The HINT protein is known to have an N-terminal acetylation by similarity (Swiss-Prot database, P49773) and the MW of this protein with the acetylation accounted for is 13712.72 Da while the experimental MW from the pH 6.5 fraction is 13712 Da yielding a mass accuracy of 53 Da.

2-D mapping of the IEF-NPS-RP-HPLC/ESI-TOFMS detected proteins

Having demonstrated how deconvoluted protein mass spectra can be summed into one mass spectrum with the pH 5.1 fraction (Fig. 3(A)), it is possible to look at a series of summed mass spectra from consecutive pH fractions and generate a 2-D pI-MS protein mass map (Fig. 5). The protein MW peaks are displayed as bands in the pI-MS image with the band's gray scale intensity representing the intensity of the protein MS peak. The pI-MS map is shown with bands labeled with protein identifications. In each case, proteins were identified by association of the peptide mass mapping results with the experimentally determined protein MW values via the %B at time of elution. The pI-MS map can be used to identify patterns of proteins and can be used for comparisons between different samples. In addition, as will be explained below, the image can be used to estimate the pI of a particular protein that may occur in multiple lanes due to the longitudinal diffusion and cathodic drift inherent to the carrier ampholyte liquid-phase IEF method.

pI estimation

The Rotofor IEF separation suffers from diffusion in the liquid phase and cathodic drift in the basic region. These two factors lead to proteins being spread out over a number of different pH fractions. Clearly, the basic proteins are more spread out than the acidic ones as a result of cathodic drift (Fig. 5). As a result of this experimental inefficiency, it is important to realize that a protein may appear in a pH fraction that significantly deviates from its actual pI. There must therefore be an experimentally determined pI as distinguished from the fraction pH of the protein.

Since the proteins will tend to focus to the correct pH fraction but diffuse out to neighboring fractions, it is logical to assume that the protein will be most abundant in the pH fraction that is closest to its pI value and diminish in abundance the further from this fraction the protein is found. The plots in Fig. 6 show the abundance distribution of α-enolase and glyceraldehyde 3PO4 dehydrogenase, as determined by the heights of the MaxEnt deconvoluted peaks, over the pH range that the protein is detected. With gaussian fitting methods, it is possible to determine an experimental pI estimate for a given protein. For α-enolase (pI 6.99), the experimental pI is 6.7, and for glyceraldehyde 3PO4 dehydrogenase (pI 8.58), the experimental pI is 8.2. These values are well within the quoted pI accuracy of 0.5 pI units. The experimental pI values of all proteins identified in this work are listed in Tables 2 and 4.

3-D pI-MS-protein hydrophobicity map

Since the 3-D separation method generates quantitative data on three fundamental properties of the protein, the pI, MW and hydrophobicity, it is possible to generate a 3-D protein map from the experimental data. The ability of the protein mass map to resolve a large number of proteins from a cell

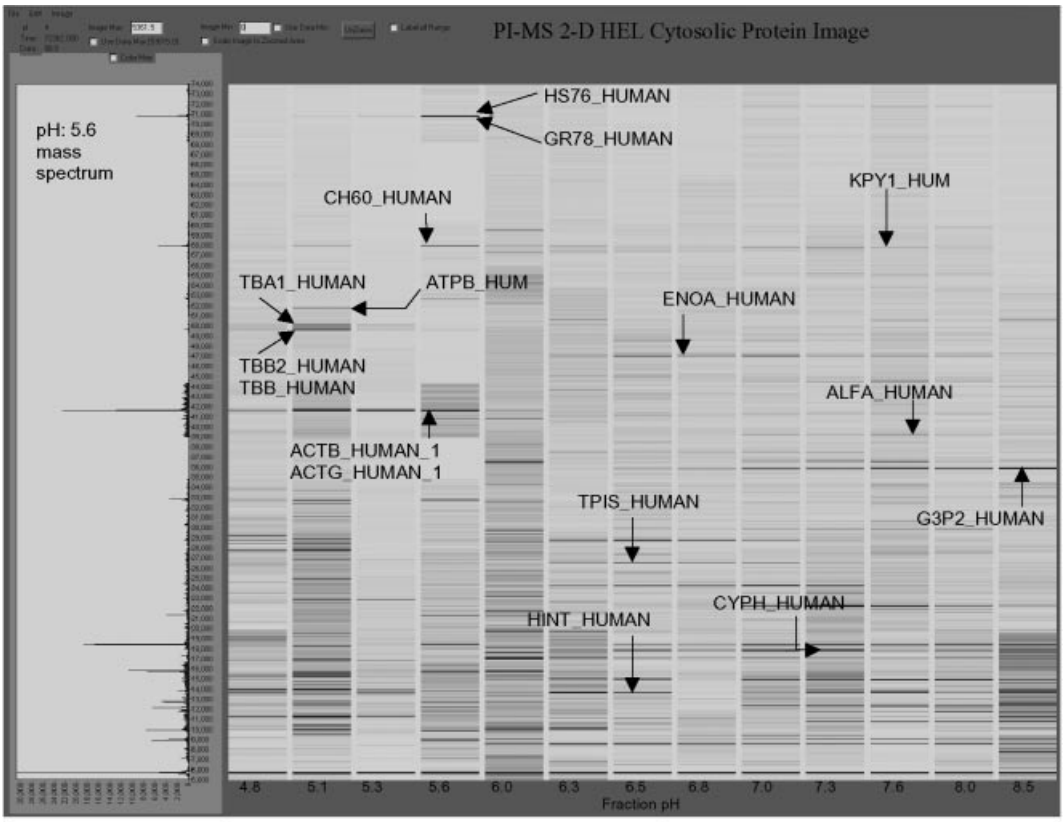

Figure 5. pI-MS 2-D image with labeled protein bands where pH of the Rotofor fractions are plotted on the x-axis and the MW of the intact proteins are plotted on the y-axis.

should clearly be enhanced by use of three versus two dimensions.

The addition of a third hydrophobicity dimension to the protein map relies upon a linear relationship that exists between the % acetonitrile at time of elution and the hydrophobicity of the protein. In order to characterize the nature of this relationship an initial plot of %B vs. the hydrophobicity factor F1 (F1 = log of the protein MW times the ratio of the nonpolar to the polar amino acids (NP/P)) was plotted (data not shown). To control for protein pI effects on solubility, the first plot was done using only data from the pH 5.1 fraction. The data shows an excellent linear

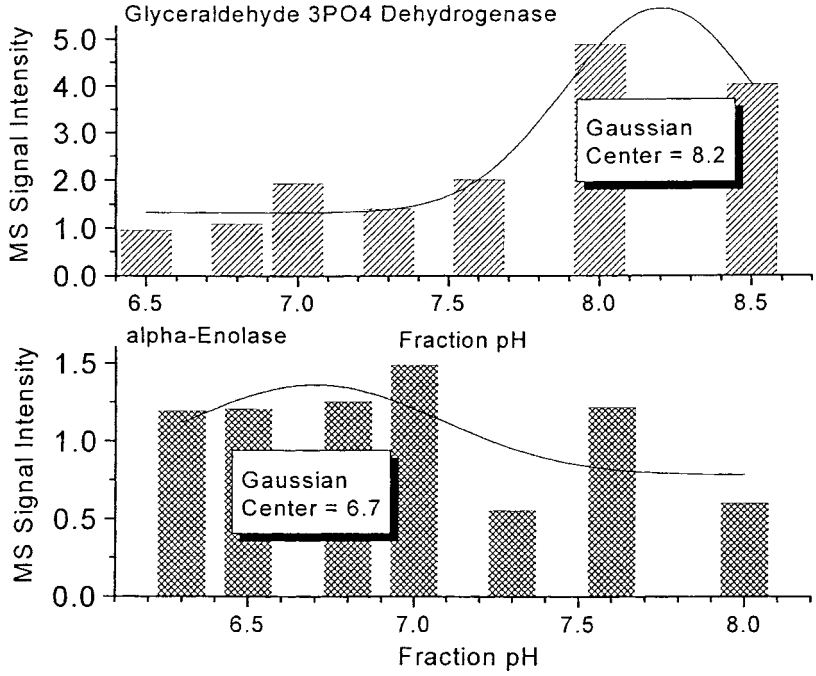

Figure 6. pI estimation plots for α-enolase and glyceraldehyde 3PO4 dehydrogenase.

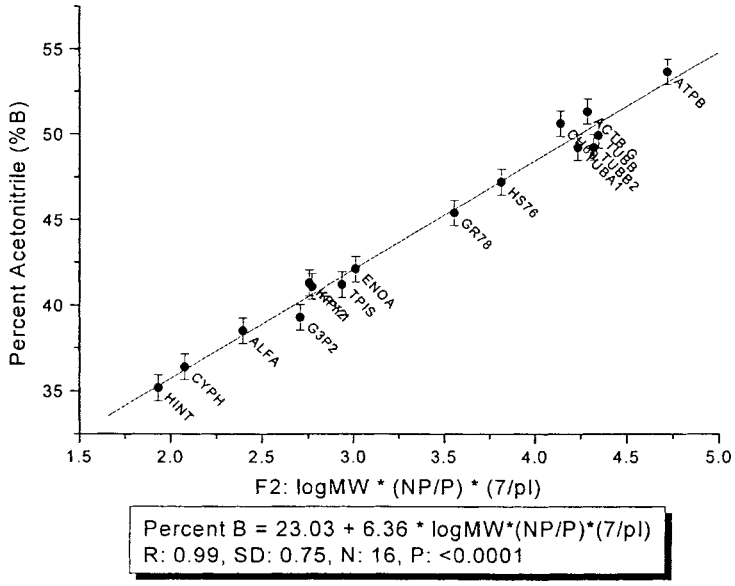

Figure 7. Linear correlation between percent acetonitrile at time of elution and the logMW * (ratio nonpolar amino acids/polar amino acids) * (7/pH).

fit for the pH 5.1 fraction. Addition of the basic proteins to the plot destroyed the linear relationship as all the basic proteins eluted earlier than was predicted by the pH 5.1 %B vs. F1 plot. This data suggests that basic proteins are more soluble in an acidic HPLC mobile phase than acidic proteins. This solubility effect was accounted for by modifying the hydrophobicity factor F1 to hydrophobicity factor F2 as follows: %B vs. logMW*(NP/P)*(7/pI). This plot is seen in Fig. 7, and the linear fit is quite good (R: 0.99, SD: 0.75, N: 16, P: <0.0001) with both basic and acidic proteins considered. Realizing this linear relationship (%B = 23.03 + 6.36 * (NP/P) * (7/pI)), it should be possible to calculate the ratio of nonpolar to polar amino acids, or absolute protein hydrophobicity, in a particular protein from the experimental pI, MW and %B data.

This correlation between %B and protein hydrophobicity (F2) is consistent with the protein identifications reported earlier. The order of elution of the tubulins shown in Fig. 3(B) agrees with the data in Fig. 7 in that TBB2 and TBA1 elute first due to their lower hydrophobicity and TBB elutes later

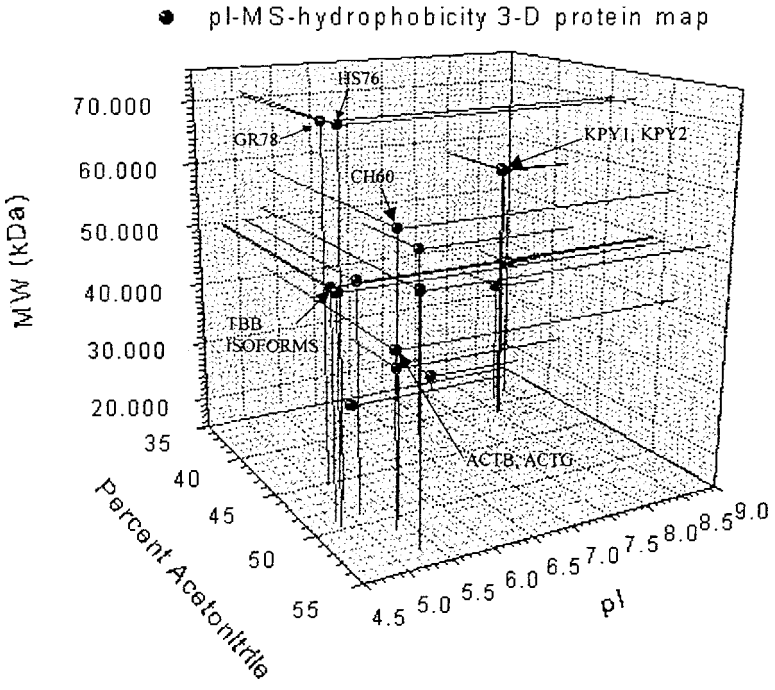

Figure 8. 3-D plot of the proteins listed in Tables 2 and 4 based upon the intact MW of the protein, the pI and the % acetonitrile on the RP-HPLC gradient.

due to its higher hydrophobicity. The large proteins GR78 and HS76 elute earlier than the smaller proteins ACT(B/G) and ATPB, and this is in strong agreement with their significantly lower hydrophobicities (Fig. 7). In addition, the hydrophobicities of these proteins are significantly different from one another and allow for the RP-HPLC separation of these two proteins that have similar MW values.

Based upon our experimental results, it is possible to create a 3-D protein plot by use of the protein MW, pI and %B (Fig. 8) for all proteins listed in Tables 2 and 4. This plot can be used to separate proteins in three dimensions and to tag a protein based upon a protein's pI (±0.5 pI units), MW (± 150 ppm) and %B (±0.2% points). It should also be possible to use the linear relationship described in Fig. 7 to calculate the ratio of the nonpolar to the polar amino acids in the protein. This calculation of the absolute hydrophobicity of the protein may be useful in characterizing the nature of the protein. It is yet to be determined how certain modifications such as prenylations or lipoylation may affect the protein hydrophobicity and the ability to use this relationship in its present form.

CONCLUSIONS

IEF-NPS-RP-HPLC/ESI-TOFMS has been shown to be a three-dimensional method for separating proteins based upon pI, hydrophobicity and MW. The method provides a means of separating and mapping hundreds of proteins in the liquid phase which can be displayed in a 2-D or 3-D image and where each protein can be tagged according to an exact MW, pI and %B elution or hydrophobicity. The ability of this method to provide an image of the protein content of a human cell sample using pI vs. MW may provide a complementary method to 2-D gel electrophoresis for the Human Proteome Project. In addition, the use of the intact MW along with the pI and peptide mapping results provides a unique method for identification of proteins and for the characterization of the protein sequence and posttranslational modifications. In this work, acetylated forms of β- and γ-actin, HINT and α-enolase were found as well as the phosphorylated form of heat shock 90 β. Sequence modifications were found including well-characterized cleavages of transit peptides and N-terminal methionines (β- and γ-actin, ATP β-synthase). Improvements in the databases will be needed to take full advantage of the high accuracy protein MW data in that the true modified masses of the proteins should be built into the search process and reported in the results. Peptident has already begun to address this problem by taking into account some of the well-known Swiss-Prot annotated protein modifications during the peptide mass map search. Peptident albeit only accounts for PTMs and their characteristic mass shifts when matching the experimental and database peptide mass values, and not for the experimental and database protein MW values. With the advent of high mass accuracy protein MW data it may now be possible to incorporate a protein MW value into the database search process thereby greatly facilitating the identification and characterization process.

Acknowledgements

We gratefully thank Dr. Tim Barder of Eichrom Technologies Inc. for helpful suggestions during the course of this work. We would like to thank Tim Riley and Matt Dilts of Micromass for loan of the LCT. The authors gratefully acknowledge partial support of this work by the National Institutes of Health under grant No. 2-R01GM49500-5 and the National Science Foundation grant No. DBI-9987220.

REFERENCES

1. O'Farrell PH. *J. Biol. Chem.* 1975, **250**, 4007.
2. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. *Proc. Natl. Acad. Sci.* 2000, **97**, 9390.
3. Kelleher NL. *Chem. Biol.* 2000, **7**, R37.
4. Vestling MM, Fenselau C. *Mass Spectrom. Rev.* 1995, **14**, 169.
5. Chang H, Yergey AL, Chrambach A. *Electrophoresis* 2001, **22**, 394.
6. Cohen SL, Chait BT. *Anal. Biochem.* 1997, **247**, 257.
7. Vestling MM, Fenselau C. *Anal. Chem.* 1994, **66**, 471.
8. Sutton CW, Wheeler CH, US, Corbett JM, Cottrell JS, Dunn MJ. *Electrophoresis* 1997, **18**, 424.
9. Schreiner M, Strupat K, Lottspeich F, Eckerscorn C. *Electrophoresis* 1996, **17**, 954.
10. Liang XL, Bai J, Liu YH, Lubman DM. *Anal. Chem.* 1996, **68**, 1012.
11. Dukan S, Turlin E, Biville F et al.. *Anal. Chem.* 1998, **70**, 4433.
12. Loo RO, Stevenson TI, Mitchell C, Loo JA, Andrews PC. *Anal. Chem.* 1996, **68**, 1910.
13. Yang L, Lee CS, Hofstadler SA, Pasa-Tolic L, Smith RD. *Anal. Chem.* 1998, **70**, 3235.
14. Jensen PK, Pasa-Tolic L, Peden KK, Martinovic S, Lipton MS, Anderson GA, Tolic N, Wong KK, Smith RD. *Electrophoresis* 2000, **21**, 1372.
15. Zolla L, Timperio AM. *Proteins: Struct. Funct. Genet.* 2000, **41**, 398.
16. Nilsson CL, Larsson T, Gustafsson E, Karlsson K, Davidsson P. *Anal. Chem.* 2000, **72**, 2148.
17. Opiteck GJ, Lewis KC, Jorgenson JW, Anderegg RJ. *Anal. Chem.* 1997, **69**, 1518.
18. Chong BE, Yan F, Lubman DM, Miller F. *Rapid Commun. Mass Spectrom.* 2001, **15**, 291.
19. Nilsson CL, Puchades M, Westman A, Blennow K, Davidsson P. *Electrophoresis* 1999, **20**, 860.
20. Davidsson P, Westman A, Puchades M, Nilsson CL, Blennow K. *Anal. Chem.* 1999, **71**, 642.
21. Puchades M, Westman A, Blennow K, Davidsson P. *Rapid Commun. Mass Spectrom.* 1999, **13**, 344.
22. Masouka J, Glee PM, Hazen KC. *Electrophoresis* 1998, **19**, 675.
23. Hesse C, Nilsson C, Blennow K, Davidsson P. *Electrophoresis* 2001, **22**(9), 1834–1837.
24. Wall DB, Kachman MT, Gong S, Hinderer R, Parus S, Misek DE, Hanash SM, Lubman David M. *Anal. Chem.* 2000, **72**, 1099.
25. Long MW, Heffner CH, Williams JL, Peters C, Prochownik EV. *J. Clin. Invest.* 1990, **85**, 1072.
26. Ferrige AG, Seddon MJ, Green BM, Jarvis SA, Skilling J. *Rapid Commun. Mass Spectrom.* 1992, **6**, 707.
27. Ferrige AG, Seddon MJ, Jarvis S. *Rapid Commun. Mass Spectrom.* 1991, **5**, 374.
28. Evershed RP, Robertson DHL, Beynon RJ, Green BN. *Rapid Commun. Mass Spectrom.* 1993, **7**, 882.
29. Reid GE, Rasmussen RK, Dorow DS, Simpson RJ. *Electrophoresis* 1998, **19**, 946.