Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Mass Spectrometry 234 (2004) 175–184
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Strategies for automating top-down protein analysis with Q-FTICR MS

Steven M. Patrie, Dana E. Robinson, Fanyu Meng, Yi Du, Neil L. Kelleher*

Department of Chemistry, University of Illinois, 53 RAL, 600 S. Mathews, Urbana, IL 61801, USA

Received 13 November 2003; accepted 20 February 2004

Abstract

Thirty years after its invention, the Fourier transform ion cyclotron resonance (FTICR) mass spectrometer continues its evolution into an enabling technology to deepen our understanding of biological systems. For contemporary protein analysis, a quadrupole FTICR hybrid with electrospray ionization is forwarded here as an engine for high resolution tandem mass spectrometry in a high-throughput setting. Three basic strategies for MS/MS of ions >10 kDa are illustrated by identification and characterization of proteins from stationary-phase yeast cells. From samples containing 1–13 proteins, introduced by a nanospray robot, 2–6 can be isolated automatically and fragmented in 15–45 min. Features used commonly for peptide analyses (e.g., multidimensional separations, data dependent acquisition, and probability-based protein identification) are now available in an off-line top-down platform. In one set of 9 samples, 20 proteins (6–17 kDa) were processed through the platform yielding a mean $P_{score} = 0.002$ (99.8% identification confidence) upon database retrieval with characterization of N-terminal post-translational modifications. On occasion of his 60th birthday, we offer this work in celebration of the steadily advancing technology that Alan Marshall helped to invent (now in over 400+ labs worldwide).

Keywords: FTICR; FTMS; Top-down; Proteomics

1. Introduction

With the staggering complexity of biological systems, measurement approaches with high resolution, dynamic range and throughput are required. Particularly, "functional-proteomics" methodologies are needed for studying the evolution of cellular phenotype from the genome through the proteome and complex networks that develop within a cell [1]. One aspect of functional proteomics emphasizes the identification of cellular proteins, in addition to the characterization of their post-translational state (i.e., protein expression ratios [2–6], modifications [7,8] and protein interactions [9–11]). The prevailing methodology for identification and characterization relies upon separation of protein mixtures (often with 2D gel electrophoresis [12–14]), enzymatic digestion, followed by chromatographic fractionation and analysis with mass spectrometry [15–17]. The peptide mass and fragmentation data, generated by tandem MS (MS/MS), are used to search the organism's protein sequence database [16,18–22] to identify the original proteins present. Although reliable for separation of thousands of species [23], protein identification is limited by dynamic range, high sample consumption, reproducibility, and the ability to separate proteins of extreme acidity, basicity, or mass [16,24,25].

"Shotgun-proteomics" eliminates the gel separation by digesting whole cell lystates without prior fractionation [15,26,27]. Several studies combining multi-dimensional separations with mass spectrometry have extended the dynamic range for protein identification [28,29]. Conrads, et al. created a version of the "shotgun" approach, where protein identification is based upon accurate mass analysis of proteolytic digests (<1 ppm mass error) [30,31]. This "accurate mass tag" (AMT) approach was used to obtain 6997 AMTs identifying >61% of the predicted proteins in the genome of *Deinococcus radiodurans* [32].

Even though high-throughput analysis of the proteome is possible with bottom-up-based approaches, validity of protein identification is often limited by low peptide match redundancy, a high false positive rate and the complexity of data analysis as the proteome size increases [29,33,34]. A

* Corresponding author. Tel.: +1-217-244-3927; fax: +1-217-244-8068.
E-mail address: kelleher@scs.uiuc.edu (N.L. Kelleher).

1387-3806/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2004.02.020

recent study by Peng et al. [29] estimated that 13% of all yeast proteins identified were false positives before manual interpretation of the dataset. Of the remaining identified proteins, 34% were identified with only one peptide match. This lack of sequence coverage leads to incomplete characterization of proteins identified.

A complementary technique to peptide-based proteomics, called the "top-down" approach, is based on MS/MS analysis of intact protein ions without prior digestion [35–38]. Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) [39–41] is one of a few techniques capable of top-down protein analysis from mixtures (ionized by electrospray, ESI) due to its high resolution (and mass accuracy [42]) in MS/MS mode [43]. This approach is advantageous because all fragment ions generated are formed within the instrument itself, via MS/MS, and correlate to the specific protein being analyzed. Additionally, with the intact molecular weight information, 100% sequence coverage is obtained, improving detection of post-translational processing events [44]. The ability to identify and characterize intact proteins with ESI/FTMS makes top-down proteomics a desirable target for technology development as the evolution of large molecule MS continues to unfold.

Classical and shotgun-proteomics both utilize multiple separation techniques to enhance dynamic range for MS/MS of peptides [45]. Incorporating this kind of front end methodology has the same effect for the top-down approach [46]. In the results reported here, most samples created by a relatively new two-dimensional approach for proteome fractionation [46] contain 1–13 proteins >10 kDa. Even with the increased dynamic range of FTMS at high magnetic field, limitations associated with sample complexity, ESI signal suppression, and chemical noise create a significant challenge above 10 kDa. We outline here our current progress toward automated acquisition of MS/MS using quadrupole-FTICR hybrid technology [47–49]. Three different data acquisition modes are demonstrated, with one run of nine samples yielding 34 distinct species detected, 24 targeted with MS/MS, and 20 of these unambiguously identified from a custom database of yeast proteins. With this technology we wish to ameliorate a major limitation of top-down mass spectrometry which is lack of dedicated data acquisition software.

2. Experimental methods

Protein samples from *Saccharomyces cerevisiae* (yeast) were produced by two-dimensional fractionation of cell lysates using continuous-elution gel electrophoresis with an acid-labile surfactant (ALS, Waters Corporation, Milford, MA) to facilitate subsequent reversed-phase liquid chromatography (RPLC). SDS–PAGE was used to determine molecular weight ranges of the fractions from the first dimension of separation. These ALS–PAGE/RPLC fractions, generated as previously described [46,50], were then lyophilized and resuspended in electrospray solution (78:20:2 = CH_3CN:H_2O:CH_3COOH). Solvents and other reagents were obtained from Sigma Chemical Co. (St. Louis, MO, USA).

The quadrupole FTICR mass spectrometer (Q-FTMS) is quite similar to the 9.4 T passively-shielded system maintained at the NHMFL by Marshall and co-workers [48,51]. A separate report currently submitted will describe the system in detail [49]. The instrument is controlled by the MIDAS data station [52] which utilizes an embedded tool command language (Tcl) interpreter to facilitate automated control of the instrument hardware [53]. A library of Tcl modules were created in-house to control various functions used in automation, such as file naming, fragmentation method, isolation mode (quadrupole and/or SWIFT [54]), spectrum processing (deconvolution [55] or THRASH [56]), and transition from sample to sample, Fig. 1A. The modular design of the library facilitates the rapid generation of different experimental event sequences used in automation. Fig. 1B outlines the platform used for the automatic detection and fragmentation of intact protein ions. From the modular event sequence (Fig. 1A, right) the user defines specific parameters unique to each Tcl module (Fig. 1B, left). This format for automation development releases the responsibility of the user to understand the details of the underlying Tcl code/hardware interface, while creating a dynamic environment for an automation platform.

For experiments targeting proteins <25 kDa, accurate protein M_r values by direct charge state assignment is achieved by comparison of resolved isotopic envelopes with theoretical envelopes [56]. This method can be unreliable for determination of masses >25 kDa. Therefore, for ALS–PAGE/RPLC fractions with proteins >25 kDa, a deconvolution method, based on charge state determination from low resolution data [50], is used to determine the average M_r values of the proteins present.

Current methodologies used in the top-down platform have general formats which include steps for molecular ion determination, isolation of targeted species, and finally fragmentation. With current capabilities each of these steps may be linked under a fully-automated platform or separated by some manual intervention. Outlined in Fig. 2 are three examples of modular event sequences implemented to date. A fully-automated deconvolution-based method [50] is used (Fig. 2A) to automatically acquire and process the intact spectrum (25 scans) into protein M_r values with known charge states [55]. The five most abundant charge states for an individual protein are SWIFT isolated on-the-fly and subsequently fragmented. In a multiplexing strategy, (Fig. 2B) several protein candidates are isolated simultaneously using the quadrupole filter and fragmented in parallel. In this fully-automated script, after the intact spectrum (25 scans) is acquired MS/MS of intact proteins is done by arbitrarily fragmenting all protein ions in predefined 20–60 m/z sections (1–5 consecutive sections, 25–50 scans each) [49]. With this strategy proteins are identified by iterative

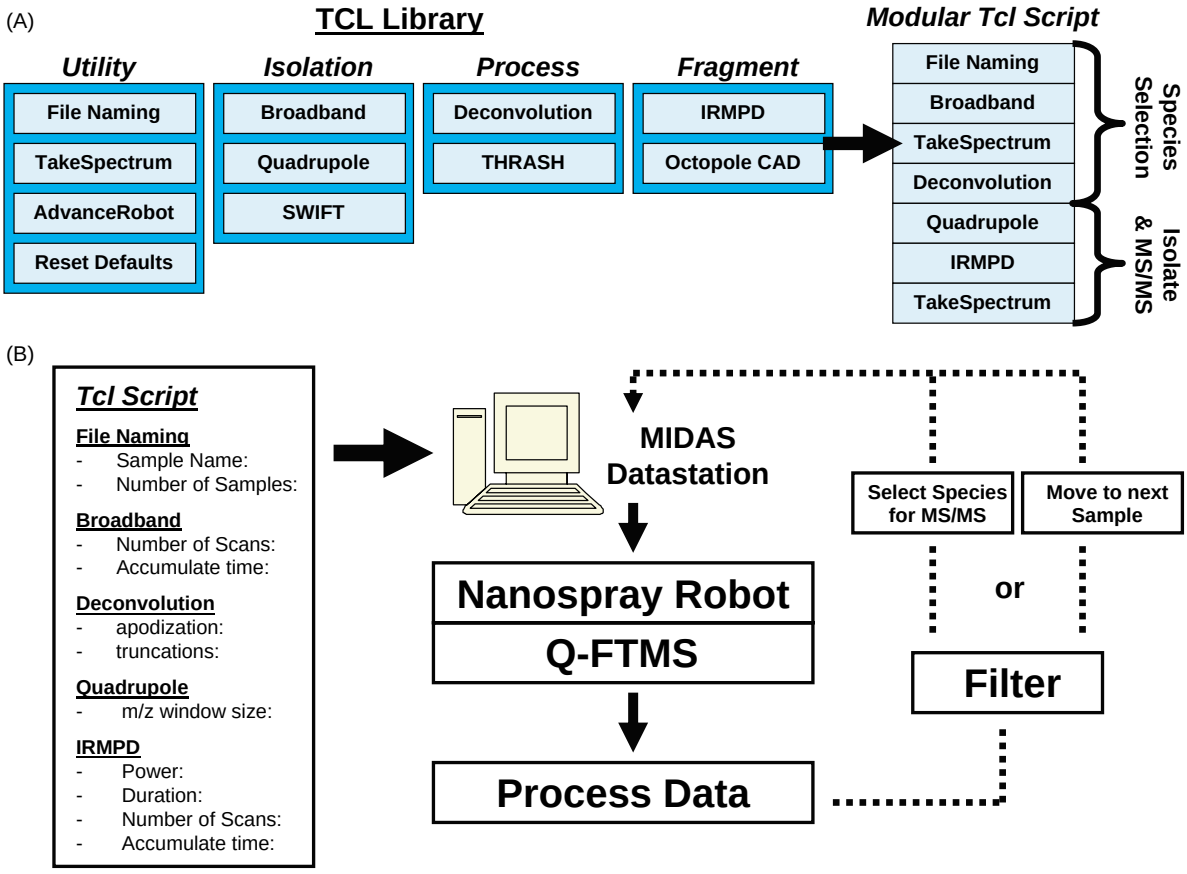

Fig. 1. (A) Modular Tcl library used for automatic sample processing on a custom quadrupole-FTMS with a nanospray robot and the MIDAS datastation. The modular format of the library facilitates generation of diverse event sequences (at right). (B) General platform of the highly automated top-down platform starting with selection of user-definable properties for the modular experimental sequence.

database searching, starting with the most abundant precursor ions.

Lastly, a method based on isotopically-resolved peaks, automatic processing, filtering, and a combination of quadrupole and/or SWIFT isolation of precursor ions is used (Fig. 2C). In this format the quadrupole mass filter serves a dual function. First, the intact spectrum is sampled with enhanced sensitivity by selective accumulation of ~60 *m/z*

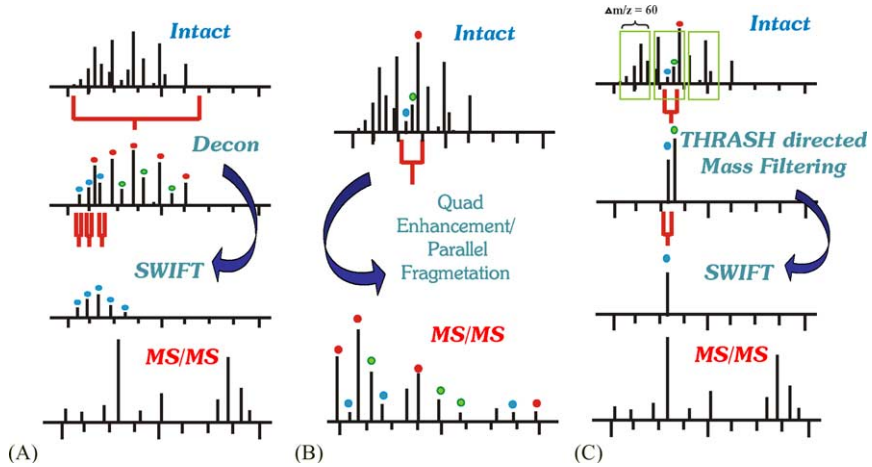

Fig. 2. (A) Deconvolution-directed automation where the intact spectrum is deconvoluted, processed and either Quadrupole or SWIFT used for subsequent isolation of intact ions prior to fragmentation. (B) In a multiplexing experiment, multiple species are isolated simultaneously for fragmentation in parallel. (C) A quadrupole targeting experiment, including sampling of the intact spectrum in 60 *m/z* sections with THRASH-directed detection of intact ions for fragmentation.

windows, defined as a "quad-march." The data generated is automatically sent to a remote computer (2.8 GHz Pentium IV, 512 MB DDR RAM) for reduction by a modified version of THRASH. Generated lists of pseudo-molecular ions are combined and filtered (off-line) to obtain protein charge state distributions in a custom data filter. The filter enables features such as dynamic exclusion of previously identified proteins and exclusion of adduct peaks (i.e., phosphate, sodium, and potassium). The filter also integrates user definable searches of molecular ion spectra to facilitate detection of common post-translational modifications of known mass shift. After manual inspection of the four to six species to be targeted, new Tcl scripts are automatically generated to isolate precursor ions with the quadrupole mass filter and finally fragmented. All masses reported correspond to monoisotopic values unless otherwise specified.

Samples were presented to the instrument with the NanoMate 100 nanospray robot (Advion BioSciences, Ithaca, NY). Ten microliters of each sample was loaded into a 96-well plate and covered with an aluminum seal. Larger sample loads (15–20 μl) reliably gave over 1 h of running time with ~200 nl/min flow rate at 0.2 psi back pressure and chip voltage of 1500 V.

In most MS/MS experiments infrared multiphoton dissociation (IRMPD) [57] or collisional dissociation [58,59] in the accumulation octopole was used. IRMPD is performed using a 75 W CO_2 laser (10.6 μm) with a user-determined irradiation period (~0.25 ms) and power level (37–75 W). Collisional dissociation was induced in the front octopole by lowering the axial offset of the octopole (at static pressure) to −35 V. To optimize fragmentation, a secondary manually-controlled power supply was coupled to the octopole DC offset supplied by MIDAS to allow on-the-fly adjustments of the fragmentation power.

Fragment mass values, generated by the THRASH algorithm, were analyzed for *b*, *y* ions and sequence tag [60,61] information using custom software for database retrieval and protein characterization through the ProSight PTM website, available at https://prosightptm.scs.uiuc.edu/ [62]. The probability scores reported are with 50 ppm fragment ion tolerance and a ±1000 Da search window (unless stated otherwise) around the candidate protein to accommodate mass shifts associated with post-translational modifications. All data was externally calibrated on a bovine ubiquitin spectrum. For identification of proteolytic products, the 0–60 kDa region of the database was searched in 5 kDa windows at 50 ppm fragment mass tolerance to localize protein candidates based on number of fragment ion matches. In all cases presented, only one protein is identified at one time with >99% confidence unless multiple species were fragmented in parallel or the identified protein is part of a gene family with nearly identical sequences. If proteins were identified as products of duplicate non-identical genes and the fragmentation data could not discern between the two, they were still considered unique identifications only if each intact protein form was observed.

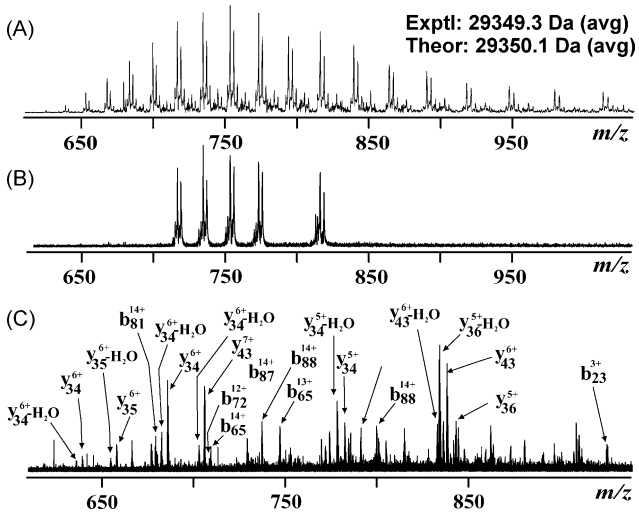

Fig. 3. Implementation of the deconvolution-based method for identification of a 29.3 kDa yeast protein. The method included automatic reduction of the intact spectrum (A), with on-the-fly SWIFT isolation (B), and fragmentation (C). The contaminating species in (B) is a +98 Da phosphate adduct.

3. Results and discussion

3.1. Decon-directed automation

Automatic deconvolution of the spectrum obtained from intact yeast proteins allowed the detection (Fig. 3A), isolation (Fig. 3B), and fragmentation (Fig. 3C) of a species observed in one 25–30 kDa ALS–PAGE/RPLC fraction. The five most abundant charge states for this 29.3 kDa protein (and its phosphate adduct, +98 Da) were SWIFT isolated and fragmented by IRMPD. Ten fragment ions (six *b* and four *y*) matched the *40s ribosomal protein s4* with <25 ppm mass accuracy and a probability score of 5×10^{-8}. This protein was found to lack the N-terminal Met but otherwise harbors no other PTMs. There is no commercial instrument that executes data acquisition in this manner. The future of this approach lies in processing mixtures of ever-larger proteins, while those of smaller M_r values can be measured with isotopic resolution even on lower B_o FTICR systems.

3.2. Multiplexed fragmentation and the informatic challenge

The quadrupole enhancement to FTMS applied to one ALS–PAGE/RPLC fraction containing 10–15 kDa proteins yielded the Fig. 4B versus Fig. 4A improvement over standard FTMS. The broadband spectrum revealed three proteins in the 860–870 *m/z* region (Fig. 4A) whereas targeted accumulation yielded nine proteins with a ~5× increase in S/N using about ~45% less sample (Fig. 4B). Collisional fragmentation of all proteins in the 860–870 *m/z* region as they exited the quadrupole yielded 75 fragment ions from 1 to 13 kDa (Fig. 4C). Fig. 5 contains a screenshot

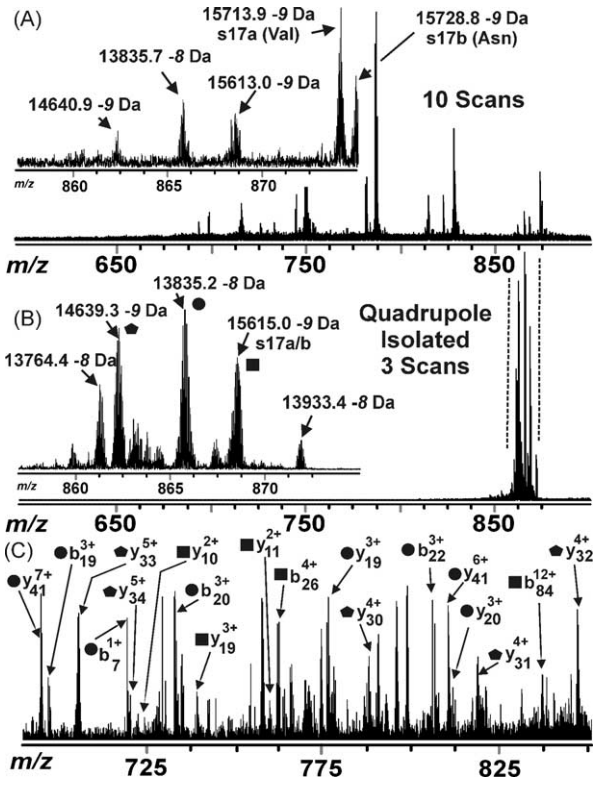

Fig. 4. (A) Broadband ESI/FT mass spectrum (no mass selection) for a yeast ALS–PAGE/RPLC fraction (4.1 s experiment length, 10 scans). (B) Selective accumulation (with mass selection) of all ions in the 860–870 *m/z* region (7.8 s experiment length, 3 scans), compare inset from (A) and (B). (C) Fragment ions identified in the 690–850 *m/z* region the spectrum after parallel fragmentation of all species in (B).

of the Prosight PTM output for a search of the 12–16 kDa region of the yeast database (50 ppm mass tolerance), showing the six most probable candidates for identification. Two proteins were identified with >99% confidence (P_{score} = 0.0002 and 0.01) resulting from 13 to 11 *b*/*y* fragment matches, respectively. Both proteins were observed with start methionine removed and are products of the duplicate *40s ribosomal protein s17* genes which differ by one amino acid on the C-terminal end (Val versus Asn). Neither of the identified proteins were present in the quadrupole isolated region targeted for fragmentation (Fig. 4B), however, both of these values matched molecular ions in the broadband spectrum outside the quadrupole window (Fig. 4A, inset) (15713.6-*9* Da theoretical versus 15713.9-*9* Da experimental, and 15728.7-*9* Da theoretical versus 15728.8-*9* Da experimental). Upon further inspection of the intact masses within the quad-window (Fig. 4B), one of the dominant species, 15615.0-*9* Da, is consistent within 20 ppm with the C-terminally truncated version of both *s17 ribosomal proteins* (proteolytic loss of either a C-terminal valine or asparagine). This explains why both proteins were primarily identified from N-terminal fragment ions, with the two C-terminal *y* ions due to random fragment matches. This type of proteolysis could either be artifactual or biologically-relevant, but many such cases have been observed to date with yeast (vide infra) [46,50], an organism notorious for its proteolytic capacity.

As the data from Figs. 4 and 5 illustrates, >11 of 75 fragment masses must match to obtain 99% confidence in identification (P_{score} = 0.01), when searching ~2500 protein forms in the 4000 Da window [63]. Fig. 5 also shows four other candidate proteins were returned in the search. The intact mass for two of them corresponds to species observed in Fig. 4B (13835.0-*8* Da theoretical versus 13835.2-*8* Da experimental and 14639.2-*9* Da theoretical versus 14639.3-*9* Da experimental), with 10 and 6 *b*/*y* fragment ions matching, respectively. However, with P_{score} > 0.01 these were not considered *unambiguously* identified. Exclusion of fragment masses (and their ammonia and water losses) related to the already identified 15.6 kDa protein lowered the P_{score} of the 13.8 kDa protein to 0.0008. A similar exclusion of these fragment products from the peak list and searching again resulted in a P_{score} = 0.01 for the 14.6 kDa protein present. In both cases no fragment ions for the identified proteins were lost during the filtering event. Improved confidence in identification may also be achieved by internal calibration of the spectrum with fragment ions identified during the "first pass." Thus, lowering the fragment mass search tolerance in Prosight PTM from ±50 to ±10 ppm makes reported P_{scores} drop by ~4 orders of magnitude.

The above data conveys that as the fragmentation spectra increase in complexity (due to a combination of secondary fragmentation, water and ammonia loss, as well as fragmenting multiple proteins) the probability of spurious matches also increases. This is compounded by the increased chance for a false hit when searching the whole database or even a large portion of it. On the other hand, lowering the tolerance and restricting searches by the knowledge of observed intact mass values (i.e., top-down) all serve to improve overall identification confidence. However, the latter presumes the database contains the correct protein form (or nearly so). As protein PTM complexity increases (e.g., in higher eukaryotes), alternative identification techniques such as the sequence tag approach will be needed. Application of ECD [64] to the quad-enhanced multiplexing approach is one avenue to explore, with the MS/MS spectral complexity anticipated to be exceptionally high.

3.3. THRASH directed automation

Two-dimensional fractionation of yeast lysates followed by ESI/Q-FTMS of the resulting samples typically yields 1–13 molecular ion masses per fraction, visible after 25 scans (no prior mass selection). However, spectral quality of samples with low protein concentration is reduced further by incomplete desolation, chemical noise, and ESI signal suppression. This is typified by the results in Fig. 6A where two yeast proteins were discernable from the high level of background noise in the broadband spectrum. Selective accumulation of 60 *m/z* sections enhanced the dynamic range

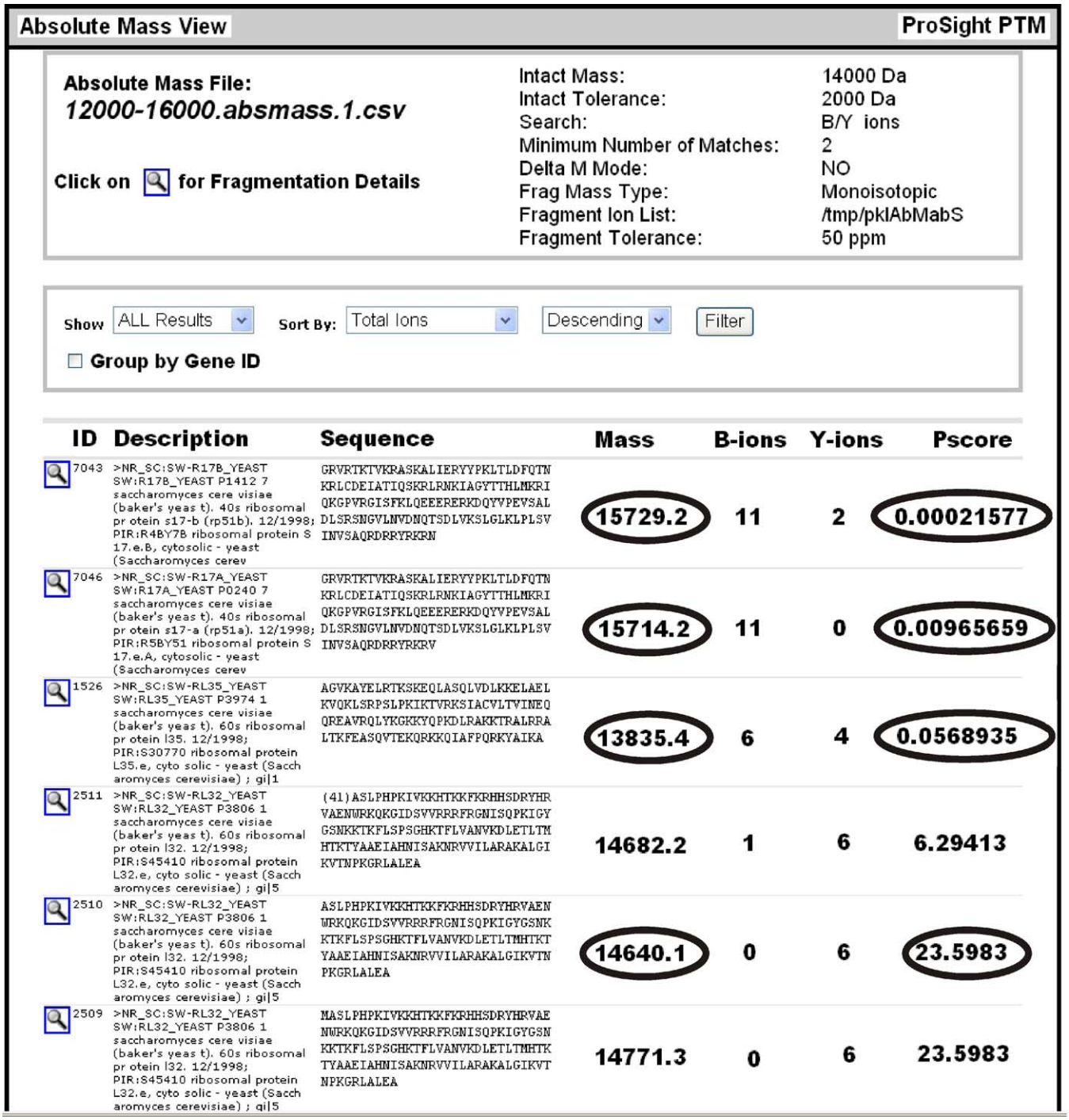

Fig. 5. A representation of the Prosight PTM output for the fragmentation data shown in Fig. 4C. Circled numbers indicate either the intact theoretical average mass values for the protein candidates (middle column) or there respective P_{score} (far right column). $P_{score} < 0.01$ indicates >99% confidence in the identification. Search was performed on the 14,000 ± 2000 Da region of the yeast database with a 50 ppm fragment mass tolerance on all monoisotopic fragment mass values. Two uncircled mass values correspond two different protein forms of identified *60s ribosomal protein l32*.

of the localized areas ~20-fold (Fig. 6B). Using this procedure across the 900–1100 *m/z* region increased the number of observed charge state distributions from 2 to 13 with observed masses from 11 to 16 kDa.

After filtering, five of the observed intact ions were subsequently targeted for fragmentation by IRMPD. To maximize transmission of targeted ions the mass selection window is kept at ~30 *m/z* and a SWIFT isolation is used to remove residual molecular ions that contaminate the spectrum. Fig. 6 also contains the isolation (Fig. 6C and D) and MS/MS spectra (Fig. 6E and F) for two of the five proteins targeted from the "quad-march" spectra (Fig. 6B). Four of

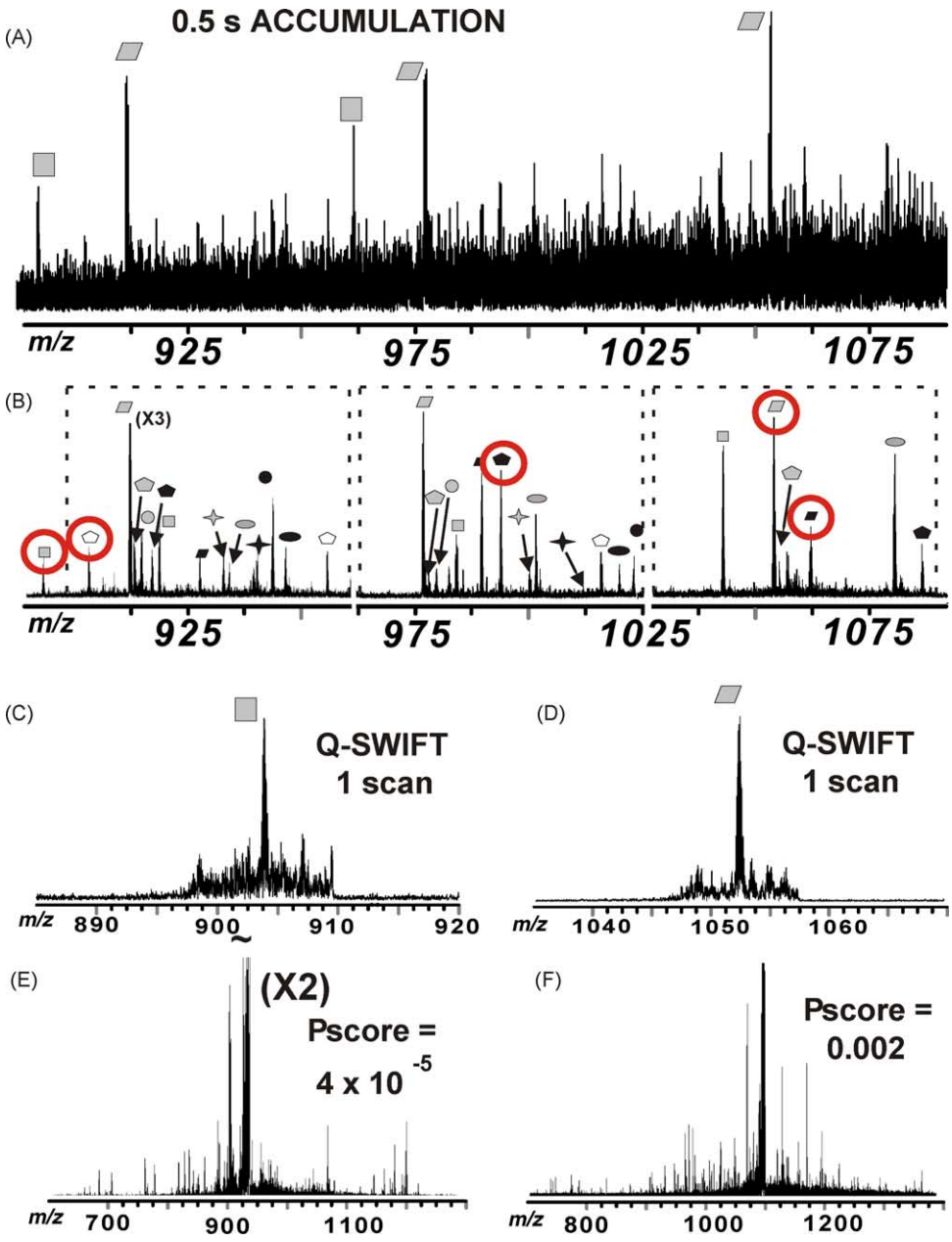

Fig. 6. (A) The 900–1100 m/z region of a broadband ESI/FT mass spectrum for a yeast ALS–PAGE/RPLC fraction (10 scans). (B) Sampling of the same region in (A) in $m/z = 60$ sections (10 scans). (C and D) Selective accumulation followed by SWIFT isolation (one scan) of two (of five targeted) proteins with more than two charge states in (B). (E and F) IRMPD fragmentation data (25 scans) for isolated ions shown in C and D, respectively. Proteins identified were *hypothetical 12.0 kDa protein*, *glyceraldehyde 3-phosphate dehydrogenase 3*, *acidic ribosomal protein P0.e*, and *2-phosphoglycerate dehydratase* with the later three were observed as 16.2, 13.7, and 12.5 kDa proteolysis products.

five proteins targeted for fragmentation were identified with an average of nine fragment ions yielding $P_{\text{scores}} < 0.002$. To obtain yet more backbone cleavages, each of the proteins were manually targeted by collisional induced fragmentation in the accumulation octopole yielding at least a four amino acid sequence tags for each of the proteins (data not shown).

For complex mixtures, such as in Fig. 6, increased experimental duration due to sampling the spectrum in 60 m/z sections is offset by the >20-fold increase in S/N for the observed species. This is explained by improved accumulation efficiency for the mass selected species per unit time in the second accumulation octopole as well as reduced dephasing of ion packets in the ICR cell at high ion populations [49]. Also, the size dependent fractionation leads to mixtures of proteins within a ~5–7 kDa window. Therefore, with limited mass ranges at least two charge states for all proteins present are usually detectable within three adjacent 60 m/z sections eliminating the need for quad-based sampling of the entire spectral region.

Table 1
Proteins identified from nine ALS–PAGE/RPLC yeast fractions

Observed mass (Da)	Theoretical mass (Da)	Delta M	Name	*b* ions	*y* ions	P_{score}	Notes
6454.5	22540.4	x	60s ribosomal protein 113-a or -b	0	9	$3 \times 10(-5)$	G, PP
8319.4	8319.3	0.1	Cytochrome c oxidase copper chaperone. 7	7	0	0.003	M
11596.5	11596.6	-0.1	12 kDa heat shock protein	28	18	$4 \times 10(-38)$	M, Ac
11612.6	11596.6	16.0	12 kDa heat shock protein	11	6	$2 \times 10(-10)$	M, Ac, Ox
13535.6	13535.7	−0.1	40s ribosomal protein s26-b	12	1	$9 \times 10(-6)$	M
13593.6	13593.6	0	40s ribosomal protein s26-a	12	1	$9 \times 10(-6)$	M
13728.5	13727.9	0.6	Ribosomal protein L34.e.A	9	2	0.005	M
13728.5	13729.9	−1.4	Ribosomal protein L34.e.B	8	2	0.005	M
13827.2	13827.0	−0.2	60s ribosomal protein l35	6	4	0.0008	M
14094.9	14094.0	0.9	60s ribosomal protein l26-a	9	27	$3 \times 10(-10)$	M
14094.9	14095.0	−0.1	60s ribosomal protein l26-b	7	28	$1 \times 10(-9)$	M
14631.3	14631.2	0.1	Ribosomal protein L32.e	0	6	0.01	M
15230.4	15230.4	0	Ribosomal protein S24.e	5	14	$2 \times 10(-11)$	M, Ac
15605.9	15605.6	x	40s ribosomal protein s17-a or -b	11	0	0.01	M, C-TT
15704.7	15704.6	−0.1	40s ribosomal protein s17-a	15	0	$7 \times 10(-7)$	M
15719.8	15719.6	−0.2	40s ribosomal protein s17-b	15	0	$7 \times 10(-7)$	M
16303.9	16304.8	−0.3	Hypothetical protein YIL051c	5	3	0.0002	M, Ac
16579.6	16580.0	−0.4	Ribosomal protein L27a.e	15	0	$2 \times 10\ (-7)$	M
17081.2	22765.5	x	DNA polymerase epsilon subunit c	7	0	0.003	PP
17822.0	35707.5	x	Glyceraldehyde-3-phosphate dehydrogenase	8	1	$1 \times 10(-4)$	M, PP

Identified proteins from highly-automated Q-FTMS/MS analysis of nine ALS–PAGE/RPLC fractions processed with a THRASH-directed methodology. x or PP: proteolysis product; M: start methoinine loss; Ac: N-terminal acetylation; C-TT: C-terminal truncation; Ox: Oxidation; G: could not differentiate duplicate genes; ( ): P_{score} exponent.

3.4. Throughput for current top-down strategies

Nine ALS–PAGE/RPLC fractions in the 10–20 kDa molecular weight range were processed with the "quad-march"/THRASH-directed approach (Fig. 2C), with intact proteins detected in three 60 *m/z* Q-FTMS spectra. About 690 individual isotopic clusters were detected yielding 34 charge state distributions (after filtering), 10 of which were believed to be oxidation or phosphate adducts. The remaining 24 were targeted with the quadrupole (no SWIFT clean up) using collisional dissociation in the accumulation octopole for fragmentation. Table 1 contains the 20 proteins identified with an average of ∼16 *b*/*y* fragment ions matching with all $P_{scores} < 0.003$. These were predominantly ribosomal proteins with 18 of 20 detected without their start met and four of these also N-terminally acetylated. An oxidized form (+16 Da) of a 12 kDa heat shock protein was observed in one ALS–PAGE/RPLC fraction. The modification was localized to a 47 amino acid stretch on the backbone (Pro45–Ala91). Three other proteins detected were truncated products due to proteolysis reactions that can occur during cell lysis or perhaps within the yeast cells.

With our current method of operation we can typically perform the broadband (10 scans), three quad-march spectra (10 scans each), process the data, and fragment six proteins (25–50 scans depending on initial abundance) in 45 min. Identification of the proteins is performed manually, facilitated by ProSight PTM, with general processing times of <5 min per protein except were multiple proteins have been isolated and fragmented in parallel. Batch file processing and more sophisticated data handling is expected to make top-down database retrievals as streamlined as those for bottom-up in the very near future. Manual validation of $P_{scores} < 0.005$ will not be required.

4. Conclusion

Three general strategies are shown for top-down analysis of proteins in an automated setting. A general format for intact protein detection followed by selective accumulation with a quadrupole mass filter and MS/MS is achieved for up to six proteins in 45 min. The quadrupole mass filter is also used to sample broadband spectra in 60 *m/z* sections, significantly increasing the dynamic range for both protein detection and MS/MS. Data analysis facilitated by ProSight PTM requires ∼5 min per protein unless ions are multiplexed, in which case data filtering to remove fragment ions from already identified proteins is required to give >99% confidence in identifications. Overall throughput is expected to rise significantly as the sophistication of the automation platform increases and more of the decision making process is taken from the user and put under software control. Future emphasis will be placed on automatic control of ion and scan number, as well as the dynamic control of fragmentation conditions for improved MS/MS spectral quality in shorter acquisition times. Also, extension of the sequence tag approach by incorporation of electron capture dissociation (ECD) into the automation platform will ensure both

protein identification and characterization even in complex proteomes.

Acknowledgements

We thank many members of the Kelleher Group for technical assistance and helpful discussions. The authors acknowledge the National Science Foundation (CHE-0134953), the National Institutes of Health (GM 067193), and the Packard and Sloan Foundations. Finally, we thank Alan Marshall, Chris Hendrickson, John Quinn, and Greg Blakney for their gracious support of our FTICR MS construction project, made possible by the NSF National High-Field FTICR Mass Spectrometry Facility (CHE 99-09502), Florida State University, and the National High Magnetic Field Laboratory in Tallahassee, Florida.

References

[1] J. Godovac-Zimmermann, L.R. Brown, Mass Spectrom. Rev. 20 (2001) 1.
[2] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.
[3] R. Tonge, J. Shaw, B. Middleton, R. Rowlinson, S. Rayner, J. Young, F. Pognan, E. Hawkins, I. Currie, M. Davison, Proteomics 1 (2001) 377.
[4] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[5] S.P. Gygi, B. Rist, T.J. Griffin, J. Eng, R. Aebersold, J. Proteome Res. 1 (2002) 47.
[6] B. B. Haab, M. J. Dunham, P. O. Brown, Genome Biol. 2 (2001).
[7] M. Mann, O.N. Jensen, Nat. Biotechnol. 21 (2003) 255.
[8] S. Meri, M. Baumann, Biomol. Eng. 18 (2001) 213.
[9] M.F. Templin, D. Stoll, J.M. Schwenk, O. Potz, S. Kramer, T.O. Joos, Proteomics 3 (2003) 2155.
[10] D. Figeys, L.D. McBroom, M.F. Moran, Methods 24 (2001) 230.
[11] S. Fields, O. Song, Nature 340 (1989) 245.
[12] A. Gorg, C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, W. Weiss, Electrophoresis 21 (2000) 1037.
[13] S.J. Fey, P.M. Larsen, Curr. Opin. Chem. Biol. 5 (2001) 26.
[14] T. Rabilloud, Proteomics 2 (2002) 3.
[15] R. Aebersold, M. Mann, Nature 422 (2003) 198.
[16] D. Lin, D.L. Tabb, J.R. Yates, Biochim. Biophys. Acta. Proteins Proteom. 1646 (2003) 1.
[17] W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 5011.
[18] J.K. Eng, A.L. Mccormack, J.R. Yates, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[19] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.
[20] K.R. Clauser, P. Baker, A.L. Burlingame, Anal. Chem. 71 (1999) 2871.
[21] H.I. Field, D. Fenyo, R.C. Beavis, Proteomics 2 (2002) 36.
[22] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, Anal. Chem. 75 (2003) 4646.
[23] J. Klose, U. Kobalz, Electrophoresis 16 (1995) 1034.
[24] S.E. Ong, A. Pandey, Biomol. Eng. 18 (2001) 195.
[25] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 9390.
[26] C.C. Wu, M.J. MacCoss, Curr. Opin. Mol. Ther. 4 (2002) 242.
[27] W.H. McDonald, J.R. Yates, Dis. Markers 18 (2002) 99.
[28] D.A. Wolters, M.P. Washburn, J.R. Yates, Anal. Chem. 73 (2001) 5683.
[29] J. Peng, J.E. Elias, C.C. Thoreen, L.J. Licklider, S.P. Gygi, J. Proteome. Res. 2 (2003) 43.
[30] T.P. Conrads, G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic, R.D. Smith, Anal. Chem. 72 (2000) 3349.
[31] R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra, H.R. Udseth, Proteomics 2 (2002) 513.
[32] M.S. Lipton, L. Pasa-Tolic, G.A. Anderson, D.J. Anderson, D.L. Auberry, J.R. Battista, M.J. Daly, J. Fredrickson, K.K. Hixson, H. Kostandarithes, C. Masselon, L.M. Markillie, R.J. Moore, M.F. Romine, Y. Shen, E. Stritmatter, N. Tolic, H.R. Udseth, A. Venkateswaran, K.K. Wong, R. Zhao, R.D. Smith, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 11049.
[33] A. Keller, S. Purvine, A.I. Nesvizhskii, S. Stolyar, D.R. Goodlett, E. Kolker, Omics 6 (2002) 207.
[34] R.D. Smith, Comp. Funct. Genome. 3 (2002) 143.
[35] N.L. Kelleher, H.Y. Lin, G.A. Valaskovic, D.J. Aaserud, E.K. Fridriksson, F.W. McLafferty, J. Am. Chem. Soc. 121 (1999) 806.
[36] N.L. Kelleher, Chem. Biol. 7 (2000) R37.
[37] F.W. McLafferty, Int. J. Mass Spectrom. 212 (2001) 81.
[38] N.C. VerBerkmoes, J.L. Bundy, L. Hauser, K.G. Asano, J. Razumovskaya, F. Larimer, R.L. Hettich, J.L. Stephenson, J. Proteome. Res. 1 (2002) 239.
[39] M.B. Camisarow, A.G. Marshall, Chem. Phys. Lett. 25 (1974) 282.
[40] A.G. Marshall, C.L. Hendrickson, G.S. Jackson, Mass Spectrom. Rev. 17 (1998) 1.
[41] A.G. Marshall, Int. J. Mass Spectrom. 200 (2000) 331.
[42] N.L. Kelleher, M.W. Senko, M.M. Siegel, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 8 (1997) 380.
[43] N.L. Kelleher, S.V. Taylor, D. Grannis, C. Kinsland, H.J. Chiu, T.P. Begley, F.W. McLafferty, Protein Sci. 7 (1998) 1796.
[44] A.J. Forbes, M.T. Mazur, H.M. Patel, C.T. Walsh, N.L. Kelleher, Proteomics 1 (2001) 927.
[45] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates, Nat. Biotechnol. 17 (1999) 676.
[46] F. Meng, B.J. Cargile, S.M. Patrie, J.R. Johnson, S.M. McLoughlin, N.L. Kelleher, Anal. Chem. 74 (2002) 2923.
[47] M.E. Belov, E.N. Nikolaev, G.A. Anderson, K.J. Auberry, R. Harkewicz, R.D. Smith, J. Am. Soc. Mass Spectrom. 12 (2001) 38.
[48] C.L.Q. Hendrickson, J.P. Quinn, M.R. Emmett, A.G. Marshall, in: Proceedings of the 48th American Society for Mass Spectrometry, 2000, p. 834.
[49] S.M. Patrie, J.P. Charlebois, J.P. Quinn, D. Whipple, A.G. Marshall, C.L. Hendrickson, N.L. Kelleher, 2004 (submitted).
[50] J.R. Johnson, F.Y. Meng, A.J. Forbes, B.J. Cargile, N.L. Kelleher, Electrophoresis 23 (2002) 3217.
[51] M.W. Senko, C.L. Hendrickson, M.R. Emmett, S.D.H. Shi, A.G. Marshall, J. Am. Soc. Mass Spectrom. 8 (1997) 970.
[52] M.W. Senko, J.D. Canterbury, S.H. Guan, A.G. Marshall, Rapid Commun. Mass Spectrom. 10 (1996) 1839.
[53] M.A. Freitas, E. King, S.D.H. Shi, Rapid Commun. Mass Spectrom. 17 (2003) 363.
[54] S.H. Guan, A.G. Marshall, Int. J. Mass Spectrom. Ion Process. 158 (1996) 5.
[55] Z.Q. Zhang, A.G. Marshall, J. Am. Soc. Mass Spectrom. 9 (1998) 225.
[56] D.M. Horn, R.A. Zubarev, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 11 (2000) 320.
[57] D.P. Little, J.P. Speir, M.W. Senko, P.B. Oconnor, F.W. Mclafferty, Anal. Chem. 66 (1994) 2809.
[58] R.M. Danell, M.L. Easterling, S. Van Orden, C.B. Berg, J. Anderson, J. Meier, P. Speir, in: Proceedings of the 51st ASMS Conference on Mass Spectrometry, Allied Topics, 2003.

[59] M.A. McFarland, C.L. Hendrickson, A.G. Marshall, in: Proceedings of the 51st ASMS Conference on Mass Spectrometry, Allied Topics, 2003.
[60] E. Mortz, P.B. OConnor, P. Roepstorff, N.L. Kelleher, T.D. Wood, F.W. McLafferty, M. Mann, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 8264.
[61] M. Mann, M. Wilm, Anal. Chem. 66 (1994) 4390.
[62] G.K. Taylor, Y.B. Kim, A.J. Forbes, F.Y. Meng, R. McCarthy, N.L. Kelleher, Anal. Chem. 75 (2003) 4081.
[63] F. Meng, B.J. Cargile, L.M. Miller, A.J. Forbes, J.R. Johnson, N.L. Kelleher, Nat. Biotechnol. 19 (2001) 952.
[64] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, J. Am. Chem. Soc. 120 (1998) 3265.