New and Automated MSⁿ Approaches for Top-Down Identification of Modified Proteins

Vlad Zabrouskov and Michael W. Senko

Thermo Electron Corporation, San Jose, California, USA

Yi Du, Richard D. Leduc, and Neil L. Kelleher

Department of Chemistry, University of Illinois, Urbana, Illinois, USA

An automated top-down approach including data-dependent MS³ experiment for protein identification/characterization is described. A mixture of wild-type yeast proteins has been separated on-line using reverse-phase liquid chromatography and introduced into a hybrid linear ion trap (LTQ) Fourier transform ion cyclotron resonance (FTICR) mass spectrometer, where the most abundant molecular ions were automatically isolated and fragmented. The MS² spectra were interpreted by an automated algorithm and the resulting fragment mass values were uploaded to the ProSight PTM search engine to identify three yeast proteins, two of which were found to be modified. Subsequent MS³ analyses pinpointed the location of these modifications. In addition, data-dependent MS³ experiments were performed on standard proteins and wild-type yeast proteins using the stand alone linear trap mass spectrometer. Initially, the most abundant molecular ions underwent collisionally activated dissociation, followed by data-dependent dissociation of only those MS² fragment ions for which a charge state could be automatically determined. The resulting spectra were processed to identify amino acid sequence tags in a robust fashion. New hybrid search modes utilized the MS³ sequence tag and the absolute mass values of the MS² fragment ions to collectively provide unambiguous identification of the standard and wild-type yeast proteins from custom databases harboring a large number of post-translational modifications populated in a combinatorial fashion. (J Am Soc Mass Spectrom 2005, 16, 2027–2038) © 2005 American Society for Mass Spectrometry

Protein identification via mass spectrometry (MS) mainly relies on two general strategies. With the *bottom-up* approach, proteins, purified or in complex mixtures, are proteolytically or chemically digested, followed by analysis using MS and tandem MS (MS/MS) of the resulting peptides, with identification provided by a database search of the product ion MS/MS spectra [1, 2]. Alternatively, with the *top-down* approach, the intact protein ions, individually or in mixtures, are mass analyzed and then fragmented inside the mass spectrometer without prior digestion [3, 4]. The advantage of the latter method is the ability to measure the intact protein molecular weight, thus preserving both the protein sequence and the integrity of most post-translational modifications [5, 6]. This allows one to proceed from protein identification to primary sequence characterization in the same experimental dataset.

With a few exceptions [7–10] to date, the top-down approach has been restricted to FTICR instruments be-

cause of the need for high resolving power and mass accuracy for protein identification and characterization via accurate mass analysis of the intact protein molecular ions and their fragment ions. Intact protein and fragment molecular weights can be searched against a corresponding database in a manner similar to that of the bottom-up approach to provide protein identification [11–13]. At the moment, ProSight PTM is the only available database search engine for top-down MS [14, 15]. The probability of a correct protein identification improves dramatically if the MS² fragment masses are known accurately (e.g., <25 ppm), the correct protein form is in the database [16], or a sequence of amino acids, known as a sequence tag, can be identified through interpretation of MSⁿ data. Within ProSight PTM, a search using a combination of MS² fragments and sequence tags is known as hybrid search [14, 15].

A hybrid search involves first compiling a list of sequence tags consistent with a list of fragment ion mass values. Next, the database is searched for unmodified sequences that contain some or all of the sequence tags. The gene ID of each unmodified protein returned from the sequence tag search is then used to search a database populated with many possible protein forms via shotgun

Published online October 25, 2005

Address reprint requests to Dr. N. L. Kelleher, Department of Chemistry, University of Illinois, 53 RAL, 600 S. Mathew Ave., Urbana, IL 61801, USA.
E-mail: kelleher@scs.uiuc.edu

annotation [16]. This allows the intact ion mass and the fragment ion mass list to be compared with all modified or unmodified protein forms in the database that contain any of the possible sequence tags. The hybrid search is well suited for identifying proteins harboring several post-translational modifications not annotated in the database being queried. Even if a few of the MS^n fragments should end on a modified amino acid, the remaining fragment ions are usually sufficient to generate sequence tags for protein identification. All fragment ions, including those with modified amino acids, can be used for matching the MS^2 fragments in the database search. However, extensive sequential amino acid loss by a technique such as collisionally activated dissociation (CAD), while common for peptides, occurs less often when working with larger molecular ions (>5 kDa) [3, 4]. This is primarily due to the ergodic nature of this fragmentation, where weakest amide bonds break first [11]. Recently introduced, electron capture dissociation (ECD) [17–19] is far more efficient in producing sequential fragments from intact protein molecular ions; however its utility was limited to FTICR instruments until 2004 [20–23].

Reported efforts to automate the fragmentation of ions >10 kDa have so far been few. In a recent paper published by Karger and coworkers [24], semi-automated targeted LC/ MS^2 top-down analysis of human growth hormone was demonstrated. Successful coupling of top-down and bottom-up approaches to speed up protein identification/characterization has also been described for bacterial 70S ribosome [25]. In the Kelleher laboratory, a considerable degree of automation was achieved by coupling 2D acid labile electrophoresis/capillary RP HPLC to a home built 8.5 T Q-FTMS, either directly to measure intact molecular weights or via an off-line but automated nanospray to perform selective ion accumulation followed by isolation and fragmentation with an infrared (IR) laser [13, 26]. Here we expand the capability of automated LC/ MS^n identification and characterization of intact proteins [24]. Three yeast proteins were identified in a single LC/ MS^2 run. Two of the identified proteins were post-translationally modified; the nature and the location of the modification sites were determined in an off-line, targeted MS^3 experiment. Additionally, proteins $<3\%$ abundant were identified readily in off-line MS^2 experiments. To extend the top-down approach to a wider range of instrumental platforms, an automated CAD MS^3 stage was introduced into the standard top-down experiment to reliably generate sequence tags from protein molecular ions using a stand alone linear ion trap, in a manner similar to the bottom-up MS^3 experiments described by Olsen and Mann [27] for improved peptide identification. Initially, the intact molecular ion was dissociated and isotopically resolved fragments were automatically identified in the MS^2 spectra. These fragments were isolated and dissociated again (MS^3) on-the-fly. As a result, a number of sequential MS^3 product ions were formed, allowing robust protein identification based on sequence tag searching. The MS^2 fragment(s) and a sequence tag(s) identified from MS^3 spectra were used to search a corresponding

database via the hybrid search mode [14, 15]. The entire experimental sequence is performed on a chromatographic time scale and extends top-down capability to a wider range of instrumental platforms. It also significantly improves the confidence of the protein identification and the degree of characterization of protein primary structures.

Methods

Protein Samples

Bovine ubiquitin, bovine cytochrome *c*, and horse heart myoglobin were from Sigma (St. Louis, MO). Human apolipoprotein A1 was from Calbiochem (La Jolla, CA). *S. cerevisiae* cells (strain S288C) grown under aerobic condition were harvested right before they reached the stationary phase. The yeast cells (3 g, wet mass) were lysed by a French press (15,000 psi) and all the soluble proteins were fractionated using a combination of preparative electrophoresis and RPLC as previously described [13]. The protein fractions were further separated on-line with a Surveyor LC (Thermo Electron, San Jose, CA) using a 100×0.15 mm C_{18} column (Microtech Scientific, Orange, CA) at a flow rate of 1 μ L/min using a 30 min 10–80% acetonitrile/water gradient. Both solvents contained 0.1% formic acid. For direct infusion, protein mixtures were dissolved in water/acetonitrile/formic acid (50:50:0.1), and loaded into an externally-coated nanospray emitter with a 2 μ m i.d. (New Objective Inc., Woburn, MA) using a spray voltage of 1.0–1.4 kV versus the inlet of the mass spectrometer, resulting in a flow of 20–50 nL/min.

Mass Spectrometry

Proteins were analyzed using a linear trap/FTICR (LTQ FT) hybrid mass spectrometer (Thermo Electron Corp., Bremen, Germany). Ion transmission into the linear trap and further to the FTICR cell was automatically optimized for maximum ion signal. The number of accumulated ions for the full scan linear trap (LT), FTICR cell (FT), MS^n linear trap, and MS^n FTICR cell were 3×10^4 , 10^6 , 10^4 , and 5×10^5 , respectively. The resolving power of the FTICR mass analyzer was set at 50,000. The flexibility of LTQ FT platform allows the use of the FTICR and linear ion trap mass analyzers independently or simultaneously, depending on experimental requirements. Individual charge states of the protein molecular ions were automatically selected for isolation and collisional activation in the linear ion trap. The product ions were measured by either the FTICR or linear trap analyzer. All FTICR spectra were processed using Xtract (Thermo Electron Corp., San Jose, CA) to produce monoisotopic mass lists. For clarity, the mass difference (in units of 1.00235 Da) between the most abundant isotopic peak and the monoisotopic peak is denoted in italics after each M_r value. In data-dependent LC/MS experiments, Dynamic Exclusion was used with a single repeat count and 7 min duration. Full scan spectra on the FTICR were acquired using a single microscan

Figure 1. The automated LC/FT MS² top-down experiment to identify yeast proteins present in the complex mixture. **Top inset:** LC/MS base peak trace of the yeast protein mixture separated on-line by RPLC. **Top:** Mass spectra averaged across the corresponding LC peaks. **Middle:** Data-dependent MS² spectra of the parent ions (insets) marked with asterisk. **Bottom:** Protein sequences retrieved with ProSight PTM when corresponding MS² spectra were searched against yeast database. The identified b and y fragments are shown.

lasting ~500 ms. For MS/MS, precursors were activated using 25% normalized collision energy at the default activation q of 0.25. FTMS² data were the average of 5–10 microscans while LTMS² data were the average of 2 microscans. Multiply charged short MS² fragments were further isolated, dissociated, and analyzed in the linear trap. Here, a linear trap/FTICR hybrid is used; however any mass spectrometer capable of MS³ can be employed. The benefits of the FTICR for detection in the MS and MS² stages is high-resolution and mass accuracy, which allow separation of the isotope peaks and thus direct assignment of precursor and fragment charge. Sufficient resolution (up to 1.5×10^4) can also be achieved with a stand alone ion trap through the use of slower scan speeds to reliably determine the charge states of MS³ precursors up to 5 kDa. In addition, other methods are available for charge state determination if isotopic resolution is not possible. The benefit of using an ion trap for MS³ is its high sensitivity attributable to the use of electron multipliers in the detec-

tion circuit, but the experiment can also be performed entirely with an FTICR detector. The resulting MS³ spectra were processed using DenovoX software (Thermo Electron Corp., San Jose, CA) to identify 5–16 amino acids sequence tags. Software tools within ProSight PTM (<https://prosigthptm.scs.uiuc.edu>) were adapted to support the combined MS² and MS³ experiments described below.

The experimental method included a single full scan followed by data-dependent FTMS² conducted on the most abundant parent ion. The resulting spectra were processed with Xtract and searched using ProSight PTM. Alternatively, the same fraction was further purified off-line by RPLC to yield several subfractions; these were examined using nanospray to identify low abundance components and pinpoint sites of post-translational modifications by MS³.

Additionally, the mixture of standard proteins and the yeast protein subfraction were analyzed by LC/MS using

ID	Description	Sequence	Length	Mass	Mass Difference	PPM Difference	B- ions	Y- ions	Total Ions	McLukkey Score	Probability Score (0.01 = 1%)
14469	>NR_SC:SW-HS12_YEAST SW:HS12_YEAST p2294.3 saccharomyces cerevisiae (baker's yeast). 12 kDa heat shock protein (glucose and lipid-regulated protein). 5/2000; PIR: HHEB12 heat shock protein 12 - yeast (S acchar)	(51)SDAGRGKGFGEKASEALKPDSQKSYAE QGKEYITDKADKVAGKVQPEDNKGVPFGVH DSAEKGGKDAEGQGESLADQARDYMGAAKS KLNDAVEYVSGRVHGEEDPTTK	108	11596.6	-0.116845	-10.0759	8	9	17	269.231	1.0502e-28
14468	>NR_SC:SW-HS12_YEAST SW:HS12_YEAST p2294.3 saccharomyces cerevisiae (baker's yeast). 12 kDa heat shock protein (glucose and lipid-regulated protein). 5/2000; PIR: HHEB12 heat shock protein 12 - yeast (S acchar)	SDAGRGKGFGEKASEALKPDSQKSYAEQGKE YITDKADKVAGKVQPEDNKGVPFGVHDSAE KGGKDAEGQGESLADQARDYMGAAKSLND AVEYVSGRVHGEEDPTTK	108	11554.6	41.8938	3612.62	0	9	9	107.692	7.44443e-12
14467	>NR_SC:SW-HS12_YEAST SW:HS12_YEAST p2294.3 saccharomyces cerevisiae (baker's yeast). 12 kDa heat shock protein (glucose and lipid-regulated protein). 5/2000; PIR: HHEB12 heat shock protein 12 - yeast (S acchar)	MSDAGRGKGFGEKASEALKPDSQKSYAEQGGK EYITDKADKVAGKVQPEDNKGVPFGVHDSAE KGGKDAEGQGESLADQARDYMGAAKSLND DAVEYVSGRVHGEEDPTTK	109	11685.7	-89.1467	-7687.37	0	9	9	107.692	7.44443e-12

Figure 2. The output of ProSight PTM search listing possible matching proteins based on precursor and fragment molecular weights, and associated probability scores.

only the ion trap as a detector. In the first LC run the most abundant charge states were identified, followed by the second run where they were fragmented. The MS² spectra were acquired using slower scan speeds to achieve the increase in resolution of the fragments. This was followed by data-dependent MS³ stage where only the MS² fragments with resolved isotopes were automatically selected and dissociated. The resulting spectra were analyzed with DenovoX sequencing program and the identified sequence tags were searched against modified human database using the hybrid search mode. Similarly, an off-line nanospray automated MS³ experiment was conducted on human lipoprotein A1 (28 kDa) to identify its multiple isoforms.

Results and Discussion

LC/MS² of Intact Proteins

The yeast cell lysate was fractionated using a combination of preparative gel electrophoresis and RPLC. Fraction number 20 from RPLC of ALS-PAGE fraction number 9 was used for on-line top-down LC/MS analysis. Proteins eluted from the LC column as two broad, partially separated peaks that were dominated by three species with masses of 11,602.7–6, 11,934.8–6, and 9929.1–5 Da (Figure 1, top). The most abundant charge states, 15–19+ of

11,934.8–6, 16+ and 17+ of 11,602.7–6, and 10–12+ of 9929.1–5, were automatically selected and fragmented (Figure 1, middle). These MS/MS spectra (average of 10 microscans) required 5 s each to acquire. The MS/MS spectra were converted to monoisotopic mass lists using Xtract. The data were entered into ProSight PTM and searched against the yeast database in absolute mass mode. The input parameters for the database search are precursor and fragment masses, mass tolerances (e.g., ± 10 ppm), fragment ion types (b/y), organism (yeast), and all potential protein modifications (e.g., methylation, formylation etc). The output of the database search is a list of possible matching protein sequences and associated probability scores. The search results for the MS² spectra obtained from the molecular ion at m/z 726.17 are presented in Figure 2. They indicate that the most probable identification is a 12 kDa heat shock protein, matching 8 b-type ions and 9 y-type ions. The precursor molecular weight is accurate to 0.1 Da at 11.6 kDa (10 ppm). The probability that this is a random match is 10^{-28} [6]. The second and third best matches are N-terminal variants of the heat shock protein, which do not agree with experimental intact mass; their scores are 15 orders of magnitude lower, because of the absence of matching fragments from the N-terminus. The best scoring protein with an unrelated primary sequence had a 71% probability of being a

Figure 3. Off-line MS³ analysis of the subfraction containing heat shock protein to localize the acetylation site. **Top:** Full scan high-resolution spectra. **Middle:** FTMS² spectra of the 13+ precursor. **Bottom:** LTMS³ spectra of the 1645.8 Da MS² fragment. The sequential loss of three N-terminal amino acids is shown.

random match. In addition to the heat shock protein, S25 ribosomal protein (gi|13230211, 11,935 Da) and endozepine (gi|13230211, 9930 Da) were identified from the same LC run with P scores of 10^{-17} and 10^{-8} , respectively. The graphic fragmentation maps are presented in Figure 1, bottom. All three proteins lacked N-terminal Met, a common PTM.

The heat shock and S25 proteins were, respectively, 42 Da and 28 Da higher than predicted, indicating that they are post-translationally modified. There were 17 fragment ions which matched the amino acid sequence of the heat shock protein with an RMS error of 1.7 ppm. All *b* ions were on average 42.010 Da heavier than predicted. To localize this mass discrepancy, the protein mixture was further separated by off-line RPLC with fraction collection. The corresponding subfraction was nanosprayed (Figure 3, top) and 13+ molecular ion of heat shock protein was fragmented (Figure 3, middle), followed by further isolation and fragmentation of 1645.8 Da doubly charged *b*₁₆ ion, with data acquisition using the linear trap analyzer (Figure 3, bottom). The consecutive losses of 129,

115, and 71 suggest an N-terminal sequence of *Ser-Ala-Asp*, with a 42 Da modification on the N-terminus. Accurate mass measurements on *b* ions from MS² spectra indicate that the modification is most likely acetylation (42.011 Da, RMS = 1.39 ppm). The accurate mass analysis rules out the possibility for trimethylation (42.046 Da). Alternatively, N-terminus can end with *Glu*, isomeric with acetylated *Ser*, but this is highly improbable. Additionally, the RPLC subfraction containing the heat shock protein was examined for low abundance proteins, with two minor components detected. These were the 3% abundant isotopic cluster at *m/z* 883 and 1% abundant isotopic cluster at *m/z* 858. Xtract processing indicated monoisotopic molecular weights of 11,468.70 Da and 11,142.49 Da, respectively (Figure 4a). The MS² spectra of these components were averaged for 1 min to obtain high quality data for a database search (Figure 4b and c). Surprisingly, both minor components matched the original heat shock protein. The molecular weights were lower than the expected molecular weight by 128 Da and 454 Da, respectively. This in turn corresponds to the removal of *Lys* and *Pro-Tyr-Lys*

Lys from the C-terminus (Figure 4d). These minor components are not CAD fragments produced in the source, but are in fact ragged C-termini as they are 18 Da heavier than the corresponding source CAD fragments.

The b_{11} ions of the S25 ribosomal protein were 28.029 Da higher than predicted, suggesting two methylations (28.031 Da) located between N-terminus and Lys11 (Figure 1, bottom right). Formylation (27.9944 Da) is unlikely because of the large mass discrepancy (31 ppm). As with the heat shock protein, the corresponding RPLC subfraction was nano-sprayed off-line, the 14+ molecular ion (m/z 853) of S25 protein was dissociated, and MS³ was performed on a 3043.8 Da product ion (b_{29} , 4+) to localize the 28 Da modification (Figure 5). The MS³ spectra contained sequential fragments Ala-Ala-Gln/Lys-Ala-Ala-Gln/Lys (Figure 5, middle). The mass accuracy of the linear ion trap is not sufficient to distinguish between glutamine (128.0586 Da) and lysine (128.09496). Neither of the Gln/Lys residues appeared to be modified, suggesting that the 28 Da modification is confined between the N-terminus and Ser7, where only N-terminus, Lys3, and Ser7 can be methylated. This agrees with previously obtained results [13] where N-terminus of this protein was found to be doubly methylated. Fragmentation (MS³ of y_{17} identified Gln/Lys-His-Ser-Gln/Lys-Gln/Lys-Ala-Leu/Ile-Tyr-Thr-Arg-Ala-Thr-Ala-Ser-Glu sequence tag (Figure 5, bottom). This sixteen residue tag is of comparable length to those obtained in bottom-up peptide identification experiments demonstrating that it is entirely possible to use an additional data-dependent MS³ stage for top-down protein identification.

Calculation of the Probability Score for MS³ Data

Considering that MS² and MS³ fragments are obtained independently, the probability associated with matching an MS² fragment(s) simply by chance to a protein sequence in the database (P-score, [6]) is independent from the probability of matching an MS³-based sequence tag to the same sequence simply by chance. Hence, the intersection of these probabilities would be the probability score of an MS³ hybrid search (eq 1):

$$P(\text{hybrid}) = P(\text{absolute search}) \times P(\text{sequence tag}),$$

$$P(\text{seqtag}) = (\text{protein length} / \sum n_j) \cup_i P_i \quad (1)$$

where P_i = the frequency that the i 'th amino acid occurs in proteins, and n_j is the number of amino acids in each of the j sequence tags found in the protein [28].

In the last example, the resulting hybrid P-score for S25 based on the length of this protein, accurate mass of y_{17} , and a sixteen a.a sequence tag is 6.3×10^{-22} indicating that the confidence of the protein identification in the MS³ experiment based only on intact molecular weight, single MS/MS fragment, and the sequence tag within that fragment is sufficient to uniquely retrieve the protein from the

Figure 4. (a) Off-line FTMS analysis of the subfraction dominated by heat shock protein and its proteolytic fragments at low abundance. **Inset:** 11 468.7 Da and 11 142.5 Da components at m/z 883 and 858, respectively. (b), (c) FTMS² spectra for low abundant components. Each spectrum was acquired for a minute to provide high quality data to maximize identification confidence and degree of PTM localization during the database search. (d) The ProSight PTM output for the low abundant components. C-terminal truncations via proteolysis are indicated.

corresponding database. In fact the P -scores of 4.4×10^{-20} were sufficiently good even when mass accuracy tolerance for an MS/MS fragment was increased to 250 ppm, indicating that the entire experiment is now possible on the stand alone ion trap, provided that the charge state of the MS² fragment ion can be determined.

Identification of Low Abundant Isoforms of Human Apolipoprotein A1

Human apolipoprotein A1 (28,078 kDa, 250 fmol/ul) was ionized by ESI off-line at 10–20 nL/min on a stand alone

Figure 5. Off-line MS³ analysis of the subfraction containing S25 protein to localize the methylation sites. **Top:** FTMS² spectra of the 14+ precursor (inset). **Middle:** LTMS³ spectra of the 3043.839 Da MS² fragment. The sequential loss of seven N-terminal amino acids is shown. **Bottom:** LTMS³ spectra of the 1987.05 Da MS² fragment. The sequential loss of sixteen N-terminal amino acids is shown.

linear ion trap MS. There were 28,087 Da major and 27,952, 28,244, and 28,460 Da minor protein forms (<6%) detected in the full scan (Figure 6 top); these protein forms were chosen for an automated data dependant MS³ experiment. All three protein forms (29–32+ precursors) produced a 4+ ion at m/z 864.2 (Figure 6, middle), which automatically triggered MS³. The resulting spectra contained [Gln/Lys+Thr]-Asn-Leu/Ile-Gln/Lys-Gln/Lys-Thr-Tyr-Glu-Glu-[Ala+Leu/Ile]-Ser-Leu/Ile sequence tag or its portions (Figure 6, bottom) which identified all three forms of apolipoprotein A1 to be modified with highly significant P-score of 4×10^{-17} . The exact nature of these modifications is yet to be determined.

LC/MS³ of Standard Proteins

We further tested the possibility of using an ion trap instrument alone to provide high quality protein iden-

tifications in an automated LC/MS³ experiment. Five hundred fmol of an equimolar mixture of bovine ubiquitin, cytochrome *c*, and horse heart myoglobin was loaded on a C₁₈ column. The two LC runs were performed consecutively. This is necessary because software tools available on LTQ FT do not allow automatic selection of charge state dependent MS³ events without a corresponding charge state dependant MS² event. Hence, if the charge state cannot be determined for an MS² precursor (as in the case of intact protein molecular ions acquired on a stand alone ion trap), then MS² and consequently MS³ do not occur. Therefore, in the first run, the size of the proteins and the most abundant charge states were identified (Figure 7, top). During the second run, the molecular ions at 779, 816, and 739 m/z were fragmented (Figure 7, middle) followed by charge state-dependent MS³ fragmentation (Figure 7, bottom). The Tyr-Asn-Leu/Ile-Gln/Lys-Gln/Lys-Glu-Ser, Glu-Asn-

Figure 6. The automated MS^3 top-down experiment performed on a stand alone linear trap to identify isoforms of human apolipoprotein A1 present in the mixture. **Top:** Averaged spectrum of apolipoprotein A1 molecular ions containing 27,952 Da, 28,087 Da, 28,244 Da, and 28,460 Da proteins. **Middle:** MS^2 spectra of the m/z 1003.91 molecular ion corresponding to 28,087 Da protein. **Bottom:** Data dependant MS^3 spectra of the MS^2 fragment which charge state was determined automatically. The sequential loss of amino acid residues is shown. The MS/MS spectra of m/z 965, 976, and 982 corresponding to 27,952 Da, 28,244 Da, and 28,460 Da proteins also contained 4+ 3456.9 ions with similar MS^3 fragmentation pattern (data not shown), indicating that the above species are modified forms of the same protein.

Figure 7. The automated LC/MS³ top-down experiment to identify standard proteins present in the mixture performed on a stand alone linear trap. **Top inset:** LC/MS base peak trace of the standard protein mix separated on-line by RPLC. **Top:** Mass spectra averaged across the corresponding LC peaks. **Middle** (second LC run): MS² spectra of the parent ions marked with asterisk. **Bottom:** Data dependant MS³ spectra of the MS² fragments, the charge state of which was determined automatically. The sequential loss of amino acid residues is shown.

Thr-Ala-Gln/Lys-Gln/Lys-Leu/Ile-Tyr and a combination of *Ala-Gly-Met-Thr-Gln/Lys-Ala*, *Glu-Leu/Ile-Gly-Phe-[Gln/Lys+Gly]*, and a *Gln/Lys-Ala-Ala-Leu/Ile* sequence tags were identified in MS³ spectra of molecular ions at 779, 816, and 739 *m/z*, respectively. These sequence tags were sufficient to uniquely identify ubiquitin, cytochrome *c*, and myoglobin, using a hybrid search when their sequences had been added to human database with corresponding low *P*-scores of 6×10^{-9} , 4×10^{-11} and 2×10^{-8} .

Identification of a Wild-Type Yeast Protein by LC/MS³

Approximately fifty femtomoles of yeast protein fraction obtained by preparative electrophoresis [13] was loaded on C₁₈ column, and two consecutive runs were performed in the same fashion as for the standard proteins. In the first LC run, an 8555 Da protein was identified in the full scan spectra (Figure 8 top). In the second run, molecular ions at 715 *m/z* were dissociated (Figure 8, middle), followed by the data-dependant

fragmentation of all isotopically resolved MS² ions (Figure 8, bottom). The 2726.1 Da 5+ ion produced an easily identifiable *Glu-Gln/Lys-Gln/Lys-Leu/Ile-Asn-Tyr-Asp* tag which, together with the MS³ precursor, were hybrid searched against the yeast database. With *P*-score of 5×10^{-9} , yeast ubiquitin (8556 Da) was the only protein matching both size of the molecular ion, MS² fragment, and MS³ sequence tag. In contrast, when only the intact mass and resolved MS/MS fragments were searched against the yeast database in absolute mass mode, the resulting retrieval came with the statistically insignificant *P*-score of 0.24.

Thus, this MS³ top-down approach can be used reliably to generate sequence tags sufficient for not only greatly improved intact protein identification confidence but also protein characterization. By improving data acquisition software to allow yet smarter decisions to be made independently in each MSⁿ stage, it will be possible to perform the experiments described above in a single LC run with both MS² and MS³ precursor ions being data-dependently selected. Further developments in this area would include merging such techniques as

Figure 8. The automated LC/MS³ top-down experiment performed on a stand alone linear trap to identify unknown yeast proteins present in the mixture. **Top:** Mass spectra averaged across the corresponding LC peak of yeast protein mix separated on-line by RPLC. **Middle** (second LC run): MS² spectra of the parent 715 *m/z* molecular ions. **Bottom:** Data dependant MS³ spectra of the MS² fragment, the charge state of which was determined automatically. The sequential loss of amino acid residues is shown.

ECD [29] or ETD [23] as MS³ fragmentation stages with CAD (MS² stage) in an on-line LC/MS experiment. It is worth reiterating that CAD/CAD, CAD/ETD, and potentially CAD/ECD MS³ [7, 30] experiments make it entirely possible to perform top-down protein identification in the absence of FTICR instrument on any stand alone ion trap. Double stage mass analyzers (i.e., QTOF or triple quadrupole) are capable of CAD/CAD pseudo MS³ runs where one has to measure the intact protein molecular ions followed by their dissociation in the

source (nozzle/skimmer fragmentation) [8, 9]. The rest of the experiment is identical to that with an ion trap.

Conclusions

Automation of a top-down experiment allows straightforward identification and characterization of intact proteins. Here, three wild-type yeast proteins were identified and characterized using a combination of automated on-line LC/MS² and off-line LC/MS³ data-

dependent experiments. Further improvements in separation would allow both experiments in the same run. The data-dependent MS³ fragmentation produces an extended sequence tag from an MS² fragment; this sequence tag alone with the mass of intact protein molecular ion and the mass of MS² precursor fragment was used to unambiguously identify standard proteins and wild-type yeast proteins in the highly annotated database. This approach can be applied to any protein, provided that its multiple charge states are resolved and MS² fragmentation forms isotopically resolved precursor for MS³ across many instrumental platforms capable of MS³/pseudo MS³ experiments and, in addition to protein identification, allows mapping the modification sites near or at the termini.

Acknowledgments

The authors would like to thank Ian Jardine, Iain Mylchreest, George Stafford, and Stevan Horning of Thermo Electron Corp. for their assistance. The acid-labile analogue of SDS was a generous gift from Edward Bouvier of the Waters Corp. The laboratory of NLK received support from the National Institutes of Health (GM 067193), the Research Corporation (Cottrell Scholars Program), and the Sloan Foundation. The support from the Center of Neuroproteomics in the University of Illinois funded through PHS 1 P30 DA 018310 is also gratefully acknowledged.

References

1. Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011–5015.
2. Yates, J. R., III; Carmack, E.; Hays, L.; Link, A. J.; Eng, J. K. Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry. *Methods Mol. Biol.* **1999**, *112*, 553–569.
3. Reid, G. E.; McLuckey, S. A. Top down protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **2002**, *37*, 663–675.
4. Kelleher, N. L. Top down proteomics. *Anal. Chem.* **2004**, *76*, 197A–203A.
5. Reid, G. E.; Stephenson, J. L.; McLuckey, S. A. Tandem mass spectrometry of ribonuclease A and B: N-linked glycosylation site analysis of whole protein ions. *Anal. Chem.* **2002**, *74*, 577–583.
6. Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat. Biotech.* **2001**, *19*, 952–957.
7. Baba, T.; Hashimoto, Y.; Hasegawa, H.; Hirabayashi, A.; Waki, I. Electron capture dissociation in a radio frequency ion trap. *Anal. Chem.* **2004**, *76*, 4263–4266.
8. Ginter, J. M.; Zhou, F.; Johnston, M. V. Generating protein sequence tags by combining cone and conventional collision induced dissociation in a quadrupole time-of-flight mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1478–1486.
9. Nemeth-Cawley, J. F.; Tangarone, B. S.; Rouse, J. C. "Top Down" characterization is a complementary technique to peptide sequencing for identifying protein species in complex mixtures. *J. Proteome Res.* **2003**, *2*, 495–505.
10. Amunugama, R.; Hogan, J. M.; Newton, K. A.; McLuckey, S. A. Whole protein dissociation in a quadrupole ion trap: Identification of an a priori unknown modified protein. *Anal. Chem.* **2004**, *76*, 720–727.
11. Senko, M. W.; Speir, J. P.; McLafferty, F. W. Collisional activation of large multiply charged ions using Fourier transform mass spectrometry. *Anal. Chem.* **1994**, *66*, 2801–2808.
12. Mortz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence data bases. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8264–8267.
13. Meng, F.; Du, Y.; Miller, L. M.; Patrie, S. M.; Robinson, D. E.; Kelleher, N. L. Molecular-level description of proteins from *Saccharomyces cerevisiae* using quadrupole FT hybrid mass spectrometry for top down proteomics. *Anal. Chem.* **2004**, *76*, 2852–2858.
14. Taylor, G. K.; Kim, Y. B.; Forbes, A. J.; Meng, F.; McCarthy, R.; Kelleher, N. L. Web and database software for identification of intact proteins using "top down" mass spectrometry. *Anal. Chem.* **2003**, *75*, 4081–4086.
15. LeDuc, R. D.; Taylor, G. K.; Kim, Y. B.; Janusz, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res.* **2004**, *32*, W340–W345.
16. Pesavento, J. J.; Kim, Y. B.; Taylor, G. K.; Kelleher, N. L. Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry. *J. Am. Chem. Soc.* **2004**, *126*, 4081–4086.
17. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
18. Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity. *J. Am. Chem. Soc.* **1999**, *121*, 2857–2862.
19. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **2000**, *72*, 563–573.
20. Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Plasma electron capture dissociation for the characterization of large proteins by top down mass spectrometry. *Anal. Chem.* **2003**, *75*, 1599–1603.
21. Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1774–1779.
22. Ge, Y.; El-Naggar, M.; Sze, S. K.; Oh, H. B.; Begley, T. P.; McLafferty, F. W.; Boshoff, H.; Barry, C. E. Top down characterization of secreted proteins from *Mycobacterium tuberculosis* by electron capture dissociation mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 253–261.
23. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
24. Wu, S. L.; Jardine, I.; Hancock, W. S.; Karger, B. L. A new and sensitive on-line liquid chromatography/mass spectrometric approach for top-down protein analysis: The comprehensive analysis of human growth hormone in an *E. coli* lysate using a hybrid linear ion trap/Fourier transform ion cyclotron resonance mass spectrometer. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2201–2207.

25. Strader, M. B.; Verberkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B. Characterization of the 70S Ribosome from *Rhodopseudomonas palustris* using an integrated “top-down” and “bottom-up” mass spectrometric approach. *J. Proteome Res.* **2004**, *3*, 965–978.
26. Du, Y.; Meng, F.; Patrie, S. M.; Miller, L. M.; Kelleher, N. L. Improved molecular weight-based processing of intact proteins for interrogation by quadrupole-enhanced FT MS/MS. *J. Proteome Res.* **2004**, *3*, 801–806.
27. Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13417–13422.
28. LeDuc, R. D.; Roth, M. J.; Boyne, M. T., II; Kim, Y.; Forbes, A. J.; Kelleher, N. L. The bioinformatics of human top down proteomics. *Proceedings of the 53rd Annual Meeting of the American Society for Mass Spectrometry*; San Antonio, TX, June 2005.
29. Patrie, S. M.; Charlebois, J. P.; Whipple, D.; Kelleher, N. L.; Hendrickson, C. L.; Quinn, J. P.; Marshall, A. G.; Mukhopadhyay, B. Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1099–1108.
30. Coon, J. J.; Ueberheide, B.; Syka, J. E.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 9463–9468.