Application Note: 30080

Top-Down LC/MS^n Identification and Characterization of Unknown Yeast Proteins

Vlad Zabrouskov and Mike Senko, Thermo Electron Corporation, San Jose, CA

Key Words

- Finnigan™ LTQ FT™
- Finnigan Surveyor™ HPLC
- MS^n
- Proteomics
- Top-Down

Introduction

Protein identification via mass spectrometry (MS) currently 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 of the resulting peptides, with identification provided by a protein database search of the product ion spectra.[1,2] Alternatively, with the *top-down* approach, the intact proteins, individually or in mixtures, are mass analyzed and then fragmented inside the mass spectrometer without prior enzymatic digestion.[3] The advantages of the latter method are the ability to measure the intact protein molecular weight, thus preserving both the protein sequence and the integrity of post-translational modifications. Currently, the top-down approach has been almost exclusively restricted to FTICR instruments due to the need for high resolving power and mass accuracy. Intact protein and fragment molecular weights can be searched against a corresponding database in a manner similar to that of the bottom-up approach in order to provide protein identification.[4-7] The most widely used top-down database search engine is ProSight™ PTM developed by the Kelleher research group at the University of Illinois.[8] Here we use a Finnigan LTQ FT to perform top-down analysis of unknown yeast proteins. Five proteins were identified and their primary sequences characterized. Two proteins were found to be post-translationally modified and the locations of these determined by MS^3. The robustness of the LTQ FT, combined with unprecedented ease of use, makes it the ideal instrument for top-down proteomics.

Methods

Protein samples

Yeast proteins were fractionated using a combination of preparative electrophoresis and RP-HPLC as previously described.[7] The protein fractions were further separated on-line with a Finnigan Surveyor HPLC using a 100×0.15 mm C_{18} column (Microtech Scientific, Orange, CA) at a flow rate of 1 µL/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 µm i.d. (New Objective Inc., Woburn, MA) using a spray voltage of 1.2–1.4 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 hybrid mass spectrometer. Ion transmission into the linear trap and further to the FTICR cell was automatically optimized for maximum ion signal. The targets for the full scan linear trap, FTICR cell, 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 the Finnigan 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™ to produce monoisotopic mass lists. 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, requiring approximately 500 milliseconds of total time. For MS/MS, precursors were activated using 25% normalized collision energy at the default activation q of 0.25. FT MS/MS data was the average of 5–10 microscans while linear trap MS/MS data was the average of 2 microscans.

Thermo
ELECTRON CORPORATION

Results and Discussion

The yeast cell lysate was fractionated using a combination of preparative gel electrophoresis and RP-HPLC. Fraction #20 from RP-HPLC was used for on-line top-down LC/MS analysis. The experimental method included a single full scan, followed by Data Dependent FT MS/MS conducted on the most abundant parent ion. The resulting MS/MS spectra were processed with Xtract and searched using Prosight PTM. Alternatively, the same fraction was further purified off-line by RP-HPLC to yield several subfractions; these were examined using nanospray to identify low abundance components and pinpoint sites of post-translational modifications by MS^3.

LC/MS/MS of intact proteins

Proteins eluted from the LC column as two broad, partially separated peaks which were dominated by three species with masses of 11602.7-*6**, 11934.8-*6* and 9929.1-*5* (Figure 1, top). The most abundant charge states, 15-19+ of 11934.8-*6*, 16+ and 17+ of 11602.7-*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 processed results for the MS/MS spectra of the *m/z* 726.17 parent are presented in Figure 2. The data was entered into ProSight PTM and searched against the yeast database (Figure 3). The input parameters for the database search are precursor and fragment masses, mass tolerances, fragment ion types, organism,

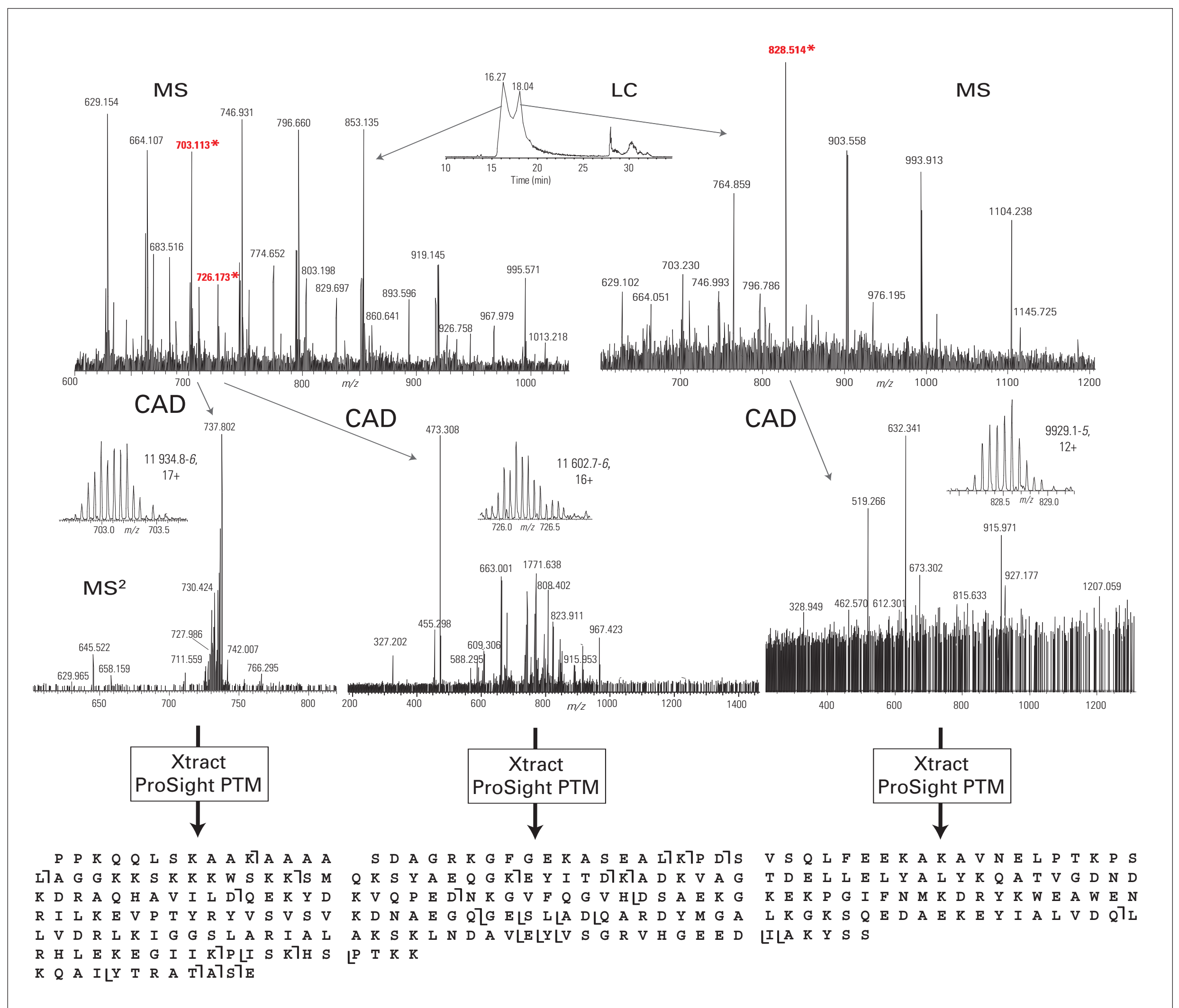

Figure 1: The automated LC/MS/MS top-down experiment to identify yeast proteins present in the complex mixture. Top inset: LC/MS base peak trace of the yeast protein mix separated on-line by RP-HPLC. Top: Mass spectra averaged across the corresponding LC peaks. Middle: Data Dependent MS/MS spectra of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

*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.

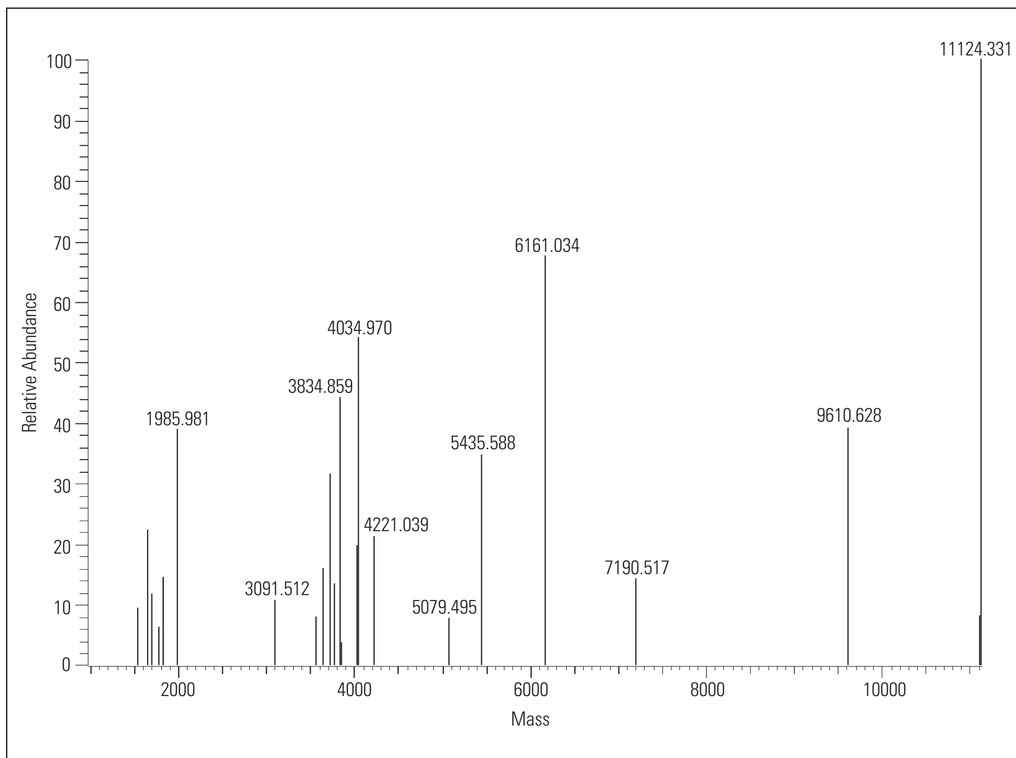

Figure 2: Deconvolution of the MS/MS spectrum from the *m/z* 726.17 parent ion (Figure 1, middle) to the corresponding zero charge monoisotopic masses (Xtract output).

Figure 3: The interface of the Prosight PTM search engine. The database searches are conducted by specifying a precursor and fragment masses, mass tolerances, fragment types, organism, and possible protein modifications.

and potential protein modifications. The output of the database search is a list of possible matching proteins sequences and associated probability scores. The search results for the MS/MS spectra obtained from *m/z* 726.17 are presented in Figure 4. It indicates 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} (*see reference* 6). The second and third best matches are N-terminal variants of the heat shock protein, with scores that are 13 orders of magnitude lower, due to the absence of matching N-terminal fragments. The best scoring protein with an unrelated primary sequence had a 71% probability of being a random match. In addition to the heat shock protein, S25 ribosomal protein (gi|13230211|, 11935 Da) and endozepine (gi|13230211|, 9930 Da) were identified from the same LC run. The protein fragmentation maps are presented in Figure 1 (bottom). All three proteins lacked N-terminal Met; this post-translational modification is common for eukaryotes. The heat shock and S25 proteins were respectively 42 Da and 28 Da

File Edit View Favorites T» Address C:\Documents and Settings\vzabrousk Go

ID	Description	Sequence	Length	Mass	Mass Difference	PPM Difference	B- ions	Y- ions	Total Ions	McLuckey Score	Probability Score (0.01 == 1%)
14469	>NR_SC:SW-HS12_YEAST SW:HS12_YEAST P2294 3 saccharomyces cere visiae (baker's yeas t). 12 kda heat shoc k protein (glucose a nd lipid-regulated p rotein). 5/2000; PIR :HHBY12 heat shock p rotein 12 - yeast (S acchar	(51)SDAGRKGFGEKASEALKPDSQKSYAEQGKEYITDKADKVAGKVQPEDNKGVFQGVHDSAEKGKDNAEGQGESLADQARDYMGAAKSKLNDAVEYVSGRVHGEEDPTKK	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 cere visiae (baker's yeas t). 12 kda heat shoc k protein (glucose a nd lipid-regulated p rotein). 5/2000; PIR :HHBY12 heat shock p rotein 12 - yeast (S acchar	SDAGRKGFGEKASEALKPDSQKSYAEQGKEYITDKADKVAGKVQPEDNKGVFQGVHDSAEKGKDNAEGQGESLADQARDYMGAAKSKLNDAVEYVSGRVHGEEDPTKK	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 cere visiae (baker's yeas t). 12 kda heat shoc k protein (glucose a nd lipid-regulated p rotein). 5/2000; PIR :HHBY12 heat shock p rotein 12 - yeast (S acchar	MSDAGRKGFGEKASEALKPDSQKSYAEQGKEYITDKADKVAGKVQPEDNKGVFQGVHDSAEKGKDNAEGQGESLADQARDYMGAAKSKLNDAVEYVSGRVHGEEDPTKK	109	11685.7	-89.1467	-7687.37	0	9	9	107.692	7.44443e-12

Figure 4: The output of Prosight PTM search listing possible matching proteins based on precursor and fragment molecular weights, and associated probability scores.

higher than predicted, indicating that they are post-translationally modified. There were seventeen fragments which matched the 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 RP-HPLC with fraction collection. The corresponding subfraction was nanosprayed (Figure 5, top) and 13+ molecular ion of heat shock protein was fragmented (Figure 5, middle), followed by further isolation and fragmentation of 1645.8 Da doubly charged $\boldsymbol{b}_{16}$ ion, with data acquisition using the linear trap analyzer (Figure 5, 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 MS/MS spectra indicate that the modification is most likely acetylation (42.010 Da, RMS = 1.39 ppm). The accurate mass analysis rules out the possibility for triple methylation (42.046 Da).

The $\boldsymbol{b}_{11}$ ion of the S25 ribosomal protein was 28.029 Da higher than predicted suggesting two methylations (28.031 Da), located between N-terminus and *Lys*11

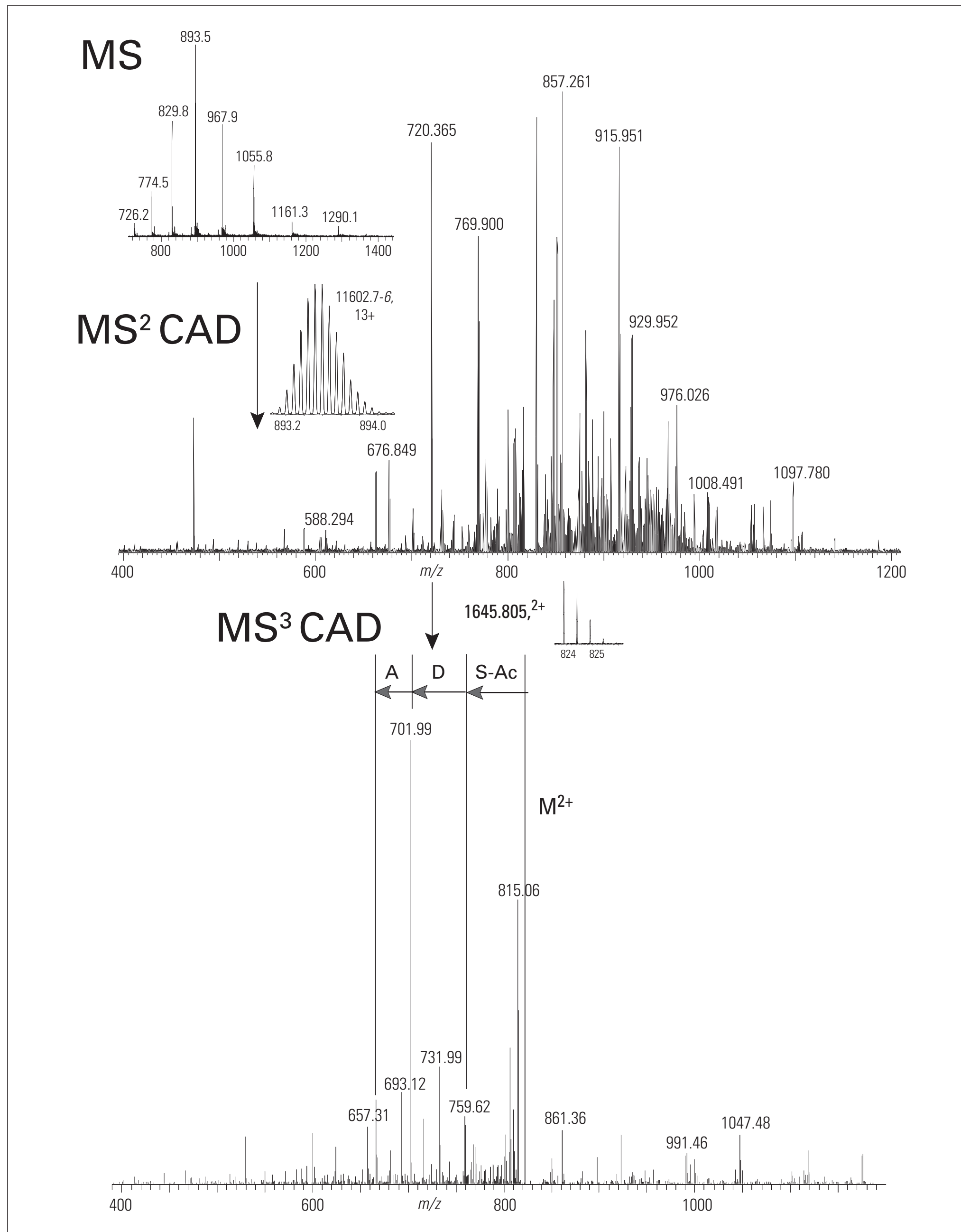

Figure 5: Off-line MS³ analysis of the subfraction containing heat shock protein to localize the acetylation site. Top: Full scan high resolution spectra. Middle: MS/MS spectra of the 13+ precursor ion. Bottom: MS³ spectra of the 1645.8 Da MS/MS fragment. The sequential loss of three N-terminal amino acids is shown.

(Figure 1, bottom). Formylation (27.9944 Da) is unlikely due to the large mass discrepancy. As with the heat shock protein, the corresponding subfraction was nanosprayed off-line, the 14+ molecular ion (*m/z* 853) of S25 protein was dissociated and MS^3 was performed on 3043.8 Da product ion (b_{29}, 4+) to localize the 28 Da modification (Figure 6). The MS^3 spectra contained sequential fragments *Ala-Ala-Gln/Lys-Ala-Ala-Gln/Lys* (Figure 6, bottom). 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 *Ser*7 residues, where only N-terminus, *Lys3* and *Ser*7 can be methylated. Additionally, the RP-HPLC 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

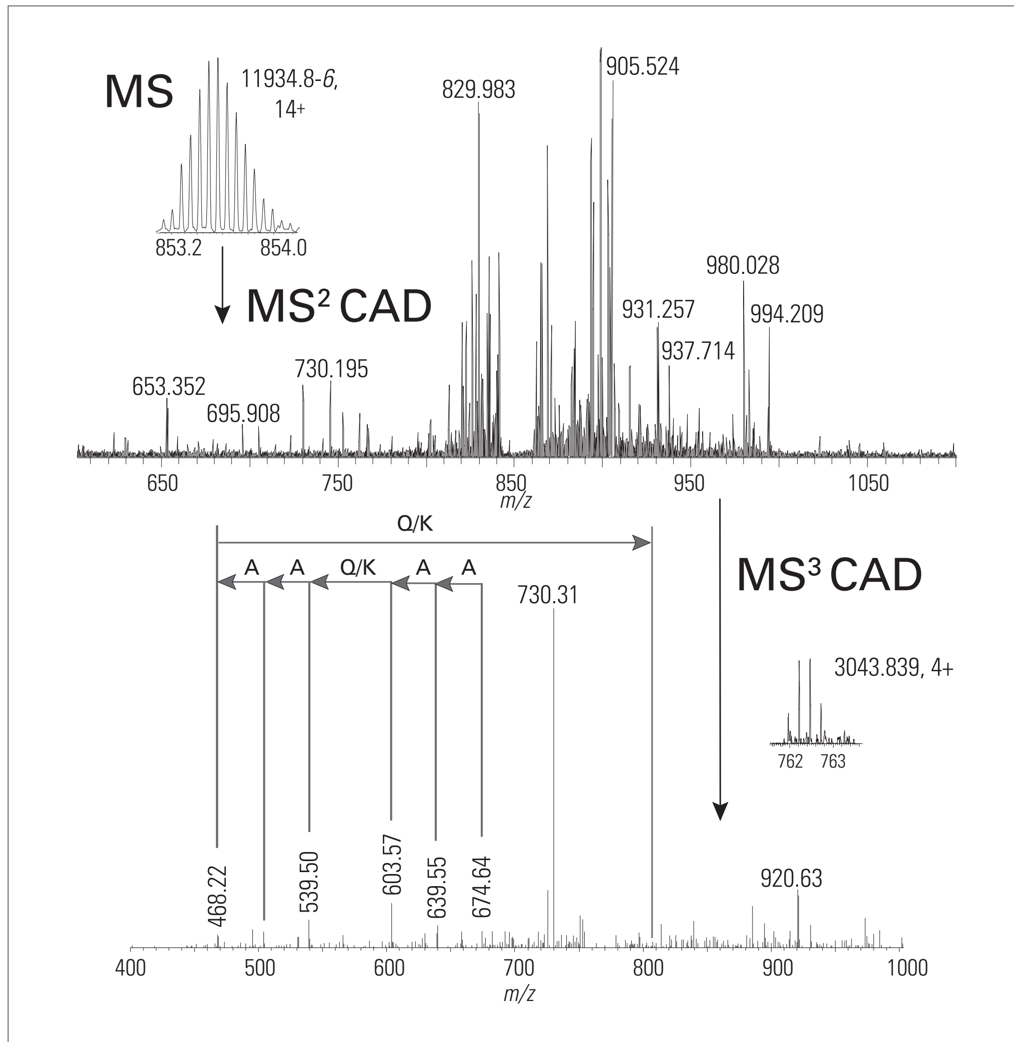

Figure 6: Off-line MS^3 analysis of the subfraction containing S25 protein to localize the methylation sites. Top: The precursor from full scan high resolution spectra. Middle: MS/MS spectra of the 14+ precursor. Bottom: MS^3 spectra of the 3043.839 Da MS/MS fragment. The sequential loss of seven N-terminal amino acids is shown.

and 11,142.49 Da respectively (Figure 7A). The MS/MS spectra of these components are shown in Figure 7B, C. Spectra were averaged for 1 min to obtain high quality data for a database search. Surprisingly, both minor components matched to 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 7D). These minor components are not source CAD fragments, but are in fact ragged C-terminii. It is easy to distinguish the difference between these two possibilities because the ragged-ended proteins are 17 Da heavier than the corresponding source CAD fragments due to the OH group on the C-terminus. This example demonstrates the ability of ProSight PTM to match proteins which differ in sequence from that listed in the data base.

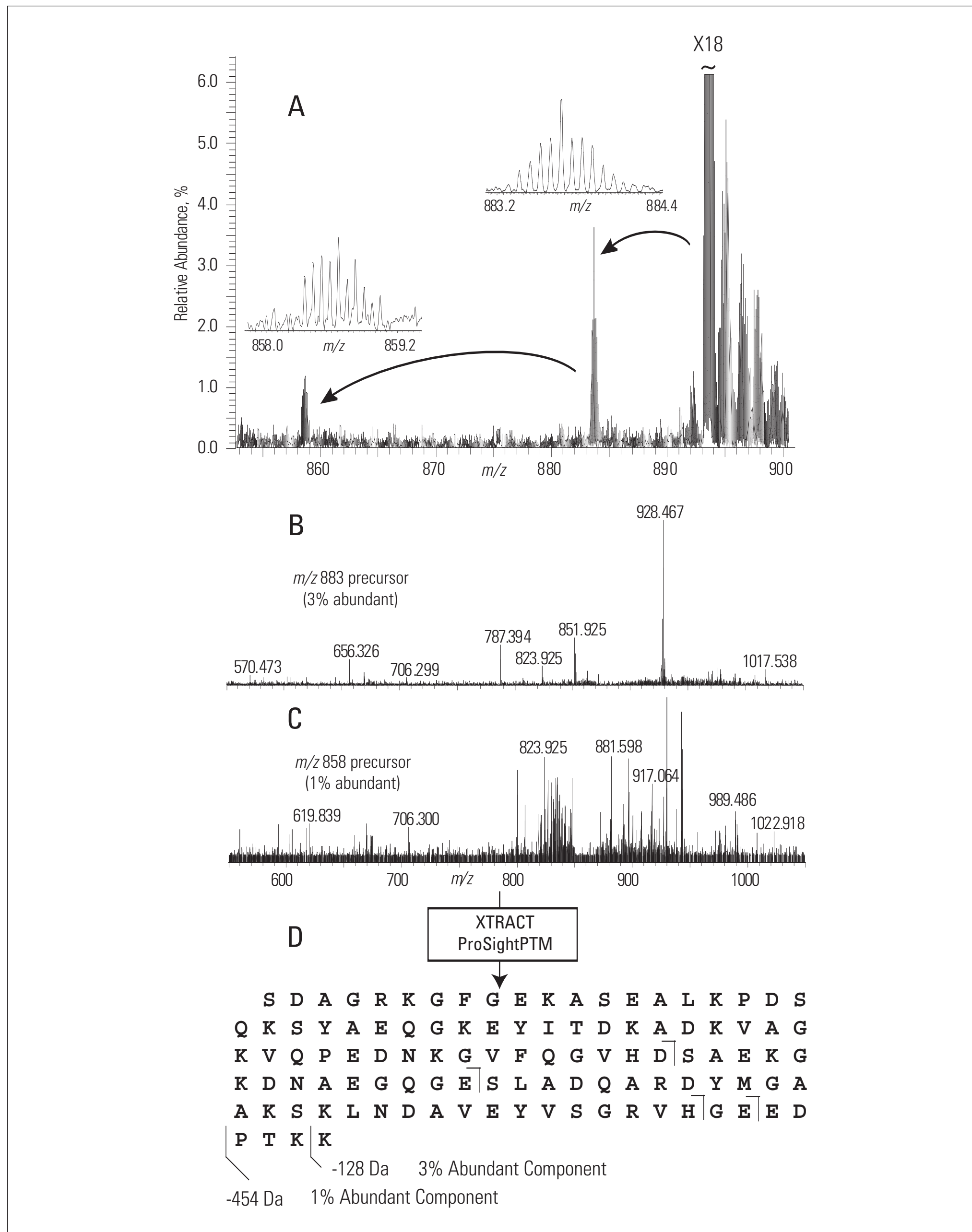

Figure 7:

A. Off-line FTMS analysis of the subfraction dominated by heat shock protein and its proteolytic fragments at low abundance. Insets: 11 468.7 Da and 11 142.5 Da components at *m/z* 883 and 858 respectively.

B, C. FT MS^2 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.

Conclusions

Automation of top-down protein analysis allows identification and characterization of intact proteins. Here, three unknown yeast proteins were identified and characterized using a combination of on-line LC/MS/MS and off-line LC/MS^3 Data Dependent experiments. Off-line analysis allowed identification of two minor components which were present at abundances of less than 5% of the major components

Acknowledgments: We would like to thank Professor Neil Kelleher and Yi Du from the University of Illinois for supplying the prepurified samples.

References

1. Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. *Proc Natl Acad Sci U S A* 1993; **90**: 5011.
2. Yates JR 3rd, Carmack E, Hays L, Link AJ, Eng JK. *Methods Mol Biol.* 1999; **112**: 553.
3. Reid GE, McLuckey SA. *J Mass Spectrom.* 2002; **37**: 663.
4. Senko MW, Beu SC, McLafferty FW. *Anal Chem.* 1994; **66**: 415.
5. Mortz E, O'Connor PB, Roepstorff P, Kelleher NL, Wood TD, McLafferty FW, Mann M. *Proc Natl Acad Sci U S A* 1996; **93**: 8264.
6. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. *Nature Biotech.* 2001; **19**: 952.
7. Meng F, Cargile BJ, Patrie SM, Johnson JR, McLoughlin SM, Kelleher NL. *Anal Chem.* 2002; **74**: 2923.
8. Taylor GK, Kim YB, Forbes AJ, Meng F, McCarthy R, Kelleher NL. *Anal Chem.* 2003; **75**: 4081.

In addition to these offices, Thermo Electron Corporation maintains a network of representative organizations throughout the world.

Australia
+61 2 8844 9500
Austria
+43 1 333 50340
Belgium
+32 2 482 30 30
Canada
+1 800 532 4752
China
+86 10 5850 3588
France
+33 1 60 92 48 00
Germany
+49 6103 4080
India
+91 22 2778 1101
Italy
+39 02 950 591
Japan
+81 45 453 9100
Latin America
+1 512 251 1503
Netherlands
+31 76 587 98 88
Nordic
+46 8 556 468 00
South Africa
+27 11 570 1840
Spain
+34 91 657 4930
Switzerland
+41 61 48784 00
UK
+44 1442 233555
USA
+1 800 532 4752

www.thermo.com

©2005 Thermo Electron Corporation. All rights reserved. ProSight PTM is a trademark of the University of Illinois at Urbana-Champaign. All other trademarks are the property of Thermo Electron Corporation and its subsidiaries.

Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details.

Thermo Electron (Bremen) GmbH is certified DIN EN ISO 9001:2000

AN61813_E 04/05S

Thermo
ELECTRON CORPORATION