Application Note: 484

N- and C-terminal Sequencing of Proteins Using Top-down Electron Transfer Dissociation Mass Spectrometry

Zhiqi Hao, Lihua Jiang, Iman Mohtashemi, Andreas FR Hühmer, Thermo Fisher Scientific, San Jose, CA, USA
Yuehe Ding, Mengqiu Dong, National Institute of Biological Sciences, Beijing, China

Key Words

- Orbitrap
- C-terminal
- Electron Transfer Dissociation (ETD)
- N-terminal
- Protein Sequencing
- Top-down

Introduction

N-terminal sequencing is widely accepted as a reliable tool for protein characterization throughout all stages of drug discovery and biopharmaceutical manufacturing. The two major direct methods of protein N-terminal sequencing are Edman degradation and mass spectrometry. Although the traditional Edman technique is very robust and provides *de novo* capabilities, the method suffers from several limitations, namely throughput, sensitivity, cost, and the need for specific data interpretation expertise. Mass spectrometry-based methods use the bottom-up approach, which is not only labor intensive but also has limitations in terms of digestion efficiency and efficient capture of the N-terminal peptides. In contrast, emerging top-down approaches involve direct analysis of intact proteins. Direct analysis can preserve the post-translationally modified forms of proteins, resolve protein-level variations, and determine expression ratios of intact protein forms.

Mass spectrometry using electron transfer dissociation (ETD) is particularly advantageous for top-down sequencing applications because ETD is relatively unaffected by protein size, amino acid composition, and post-translational modifications. ETD randomly cleaves protein backbone bonds while preserving post-translational modifications.[1] ETD technology has been implemented in hybrid linear ion trap – Thermo Scientific Orbitrap mass spectrometers whose high mass resolution and mass accuracy facilitate top-down analysis of intact proteins.[2] The ability of a hybrid Orbitrap™ mass spectrometer to perform multiple stages of MS/MS with a variety of activation types extends the rapid N-, and C-terminal sequencing capability to complete sequencing of the terminus – even for proteins with modifications. In this study, the utility of ETD mass spectrometry for N- and C-terminal sequencing of intact proteins is presented. The high-resolution, high-mass-accuracy capabilities of Orbitrap ETD hybrid mass spectrometry coupled with multiple fragmentation techniques enabled complete characterization of N- and C-termini of a truncated protein.

Goal

- To apply electron transfer dissociation (ETD) mass spectrometry to top-down protein sequencing
- To apply and optimize such an approach for protein N- and C-terminal sequencing

Experimental

Sample Preparation

Standard proteins were purchased from Sigma. Partially purified DAF-16 protein fragment was a gift from NIBS, Beijing, China. Desalted intact protein was diluted in acetonitrile/water/formic acid (50:50:0.1 vol%) to a final concentration of 2 to 5 pmol/L. The sample was directly infused using static nanospray with a 4-micron tip (PicoTip®, New Objective).

Mass Spectrometry

The energy used for source fragmentation was 20 V for full MS1 scans. ETD was performed using a Thermo Scientific LTQ Orbitrap XL ETD mass spectrometer (Figure 1) with HCD gas turned off. Both MS^n automatic gain control (AGC) and reagent anion AGC target values were set to 5e5. ETD reaction time was adjusted from 2 milliseconds to 50 milliseconds, as specified. Spectra were averaged for 5 minutes (200-300 microscans) using a resolving power of 60,000 (FWHM at *m/z* 400).

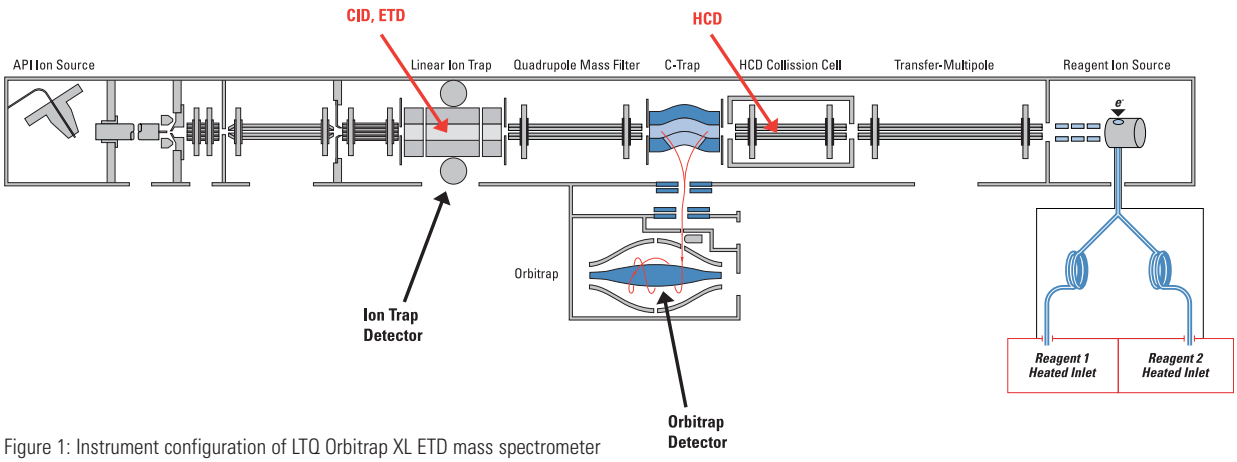


Figure 1: Instrument configuration of LTQ Orbitrap XL ETD mass spectrometer

Data Analysis

High-resolution MS² spectra were deconvoluted using the Xtract functionality of Qual Browser in Thermo Scientific Xcalibur instrument control software. Data were analyzed using Thermo Scientific ProSightPC 2.0 software in single-protein mode for deconvoluted, high-resolution, top-down ETD or CID spectra with 5-ppm fragment ion tolerance. Database searching for protein truncation, and therefore the identification of the C-terminus of truncated DAF-16, was performed using ProSightPC™ delta M mode under the BioMarker search category. When the observed molecular mass was from ion trap detection, a large precursor mass tolerance of up to 40 Da was used in combination with a 5-ppm tolerance for fragment ions. ETD MS³ spectra were analyzed using the SEQUEST® algorithm.

Results and Discussion

Top-down Protein Sequencing Using ETD with High Resolving Power and Accurate Mass

Intact protein carbonic anhydrase was used as a standard protein to optimize experimental conditions for top-down protein terminal sequencing. Precursor ions carrying 36 charges were isolated for the ETD experiment and the product ions were detected in the Orbitrap analyzer. As shown in Figure 2, ETD of carbonic anhydrase generated a very informative, complex spectrum with the N-terminal acetylation preserved. The charge states of the product ions were well resolved using Orbitrap MS detection. From the deconvoluted, accurate-mass ETD spectrum, extensive sequence coverage was obtained through ProSightPC search.

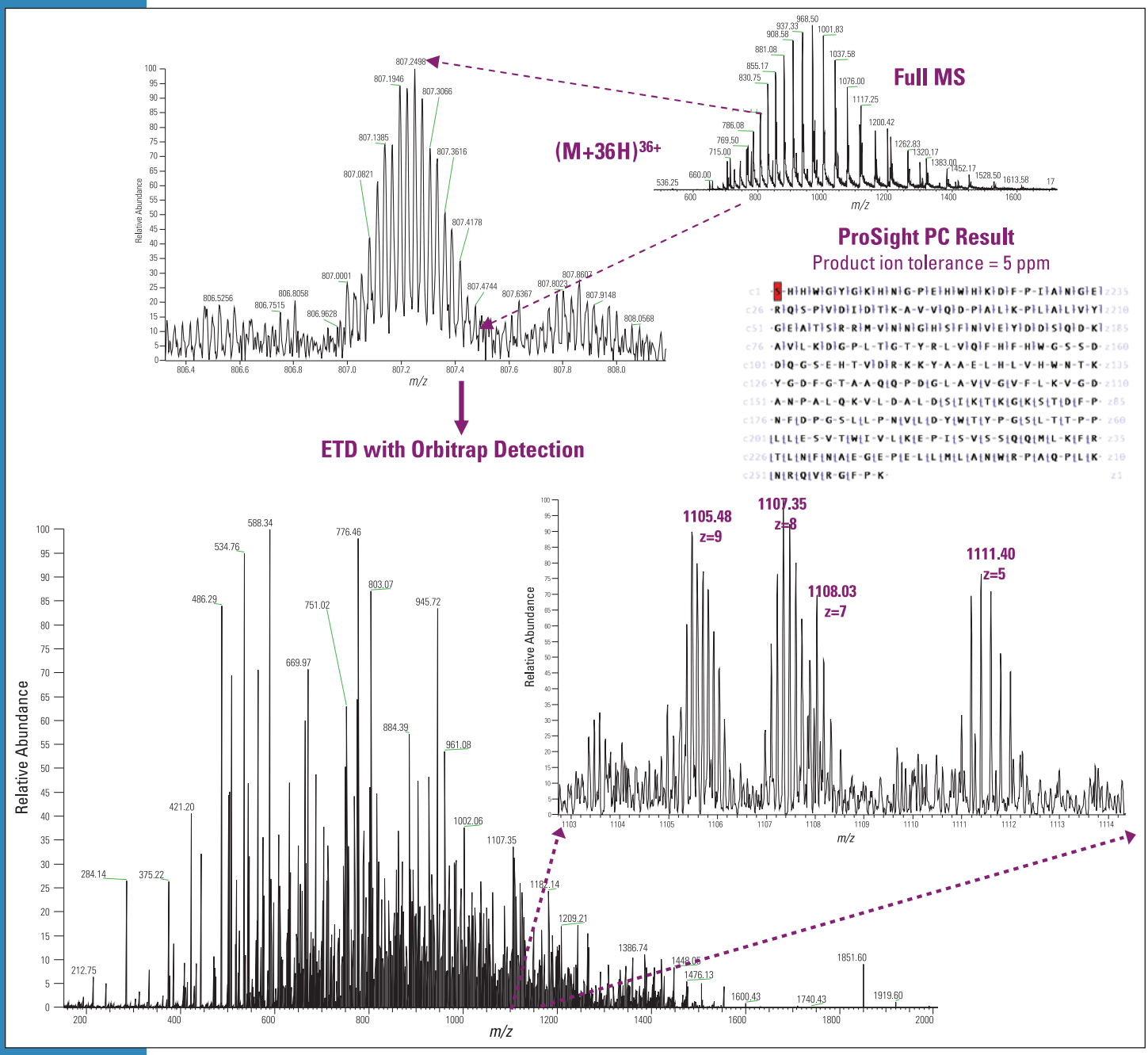


Figure 2: Electron transfer dissociation of intact carbonic anhydrase (MW= 29 kDa) with Orbitrap detection

Optimization of ETD Reaction Time for Terminal Sequence Coverage

The ETD reaction time required for an intact protein is much shorter than that for a peptide. This is consistent with the fact that the reaction rate of ETD is proportional to the square of the number of charges carried by the precursor ion. As listed in Table 1 under the experimental conditions, a 4- to 8-millisecond ETD activation time produced the best sequence coverage for this protein. With an increase of ETD reaction time, loss of bigger product ions was observed with increased appearance of shorter terminal-sequence ions (Table 1 and Figure 3). The best N- and C-terminal sequence coverage was obtained using 50 milliseconds of ETD activation time. Under these conditions, 26 (out of 27) c ions for N-terminal 30 amino acids (with 3 Pro), and 25 (out of 26) z· ions for C-terminal 30 amino acids (with 4 Pro) were identified. Results obtained using other intact proteins, such as myoglobin (16.9 kDa) and enolase (47 kDa), were similar (data not shown). Shorter ETD reaction times generated larger product ions and better overall sequence coverage, while longer ETD reaction times produced mostly shorter, terminal-sequence ions, which is an ideal result for protein N- and C-terminal sequencing. A simple ETD MS^2 with Orbitrap detection produced enough information for both N- and C-terminal sequencing.

Reaction time (ms)	Total number of c ions (out of 239)	Total number of z· ions (out of 239)	Number of c ions for N-terminal 30 amino acid (out of 27)	Number of z· ions for C-terminal 30 amino acid (out of 26)
2	65	51	21	11
4	84	61	25	16
8	74	67	25	22
20	66	64	25	23
50	40	43	26	25
100	38	35	25	22

Table 1: Effect of ETD reaction time on total sequence and terminal sequence coverage

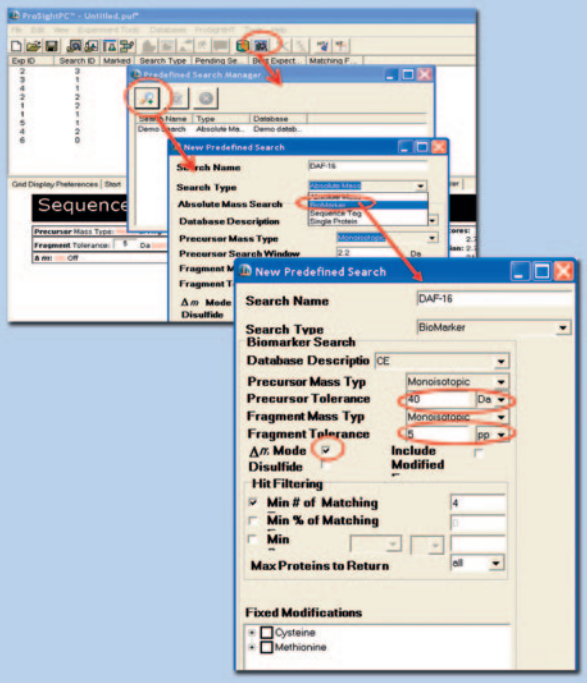
Figure 3: Effect of ETD reaction time on sequence coverage

Top-down N- and C-terminal Sequencing of DAF-16 from *C. elegans*

SDS PAGE of His-tagged DAF-16 protein from *C. elegans* suggested that it was a truncated fragment of about 40 kDa. The top-down ETD MS^2 spectrum of this truncated DAF-16 protein was deconvoluted and analyzed by ProSightPC 2.0 software against full length DAF-16 sequence. Near complete sequence ion series for the N-terminus were identified, while almost no sequence ions from the C-terminus of intact DAF-16 were found (data not shown), indicating that truncation occurred on the C-terminus. Two different approaches were used to localize the C-terminus of this truncated protein as described in the following sections. Confident results were obtained independently using either of the approaches. The complete C-terminal sequence was obtained using the top-down, combined ETD and collision-induced dissociation (CID) approach.

Localization of C-terminus Using ProSightPC BioMarker Search Category

When a high-resolution MS^2 spectrum and the molecular mass of a truncated protein are available, ProSightPC can be used to locate truncation by searching against the intact protein sequence under the delta M mode of the BioMarker search category (Figure 4). Using a stringent product ion tolerance, ProSightPC can produce extremely confident results even if the calculated molecular mass contains a relatively large error. In this case, a wide precursor tolerance was needed to allow the localization of the actual terminus by the software. A 40-Da precursor tolerance was used since the molecular mass was calculated from ion trap MS (data not shown). Using a 5-ppm product ion tolerance in ProSightPC, unambiguous localization of the C-terminus was achieved together with extensive sequence coverage on the truncated protein. With both the N- and C-terminus of this truncated DAF-16 protein identified, the theoretical molecular mass was calculated to be 36695.4 Da. Top-down ETD MS^2 identified 31 c and 52 z· ions with more-complete sequence coverage observed for the N-terminus than for C-terminus. A reported P score of 2.4 E-59 by ProSightPC indicates a high-level confidence for this identification (Figure 5).



- The Delta M mode of BioMarker search type allows identification of protein terminal truncation while also providing sequence coverage.
- When accurate molecular mass is not available, wide precursor ion tolerance can be used to accommodate experimental error in molecular mass.
- Stringent product ion tolerance is necessary to ensure confident identification.

Figure 4: ProSightPC BioMarker search for identification of protein terminal truncation

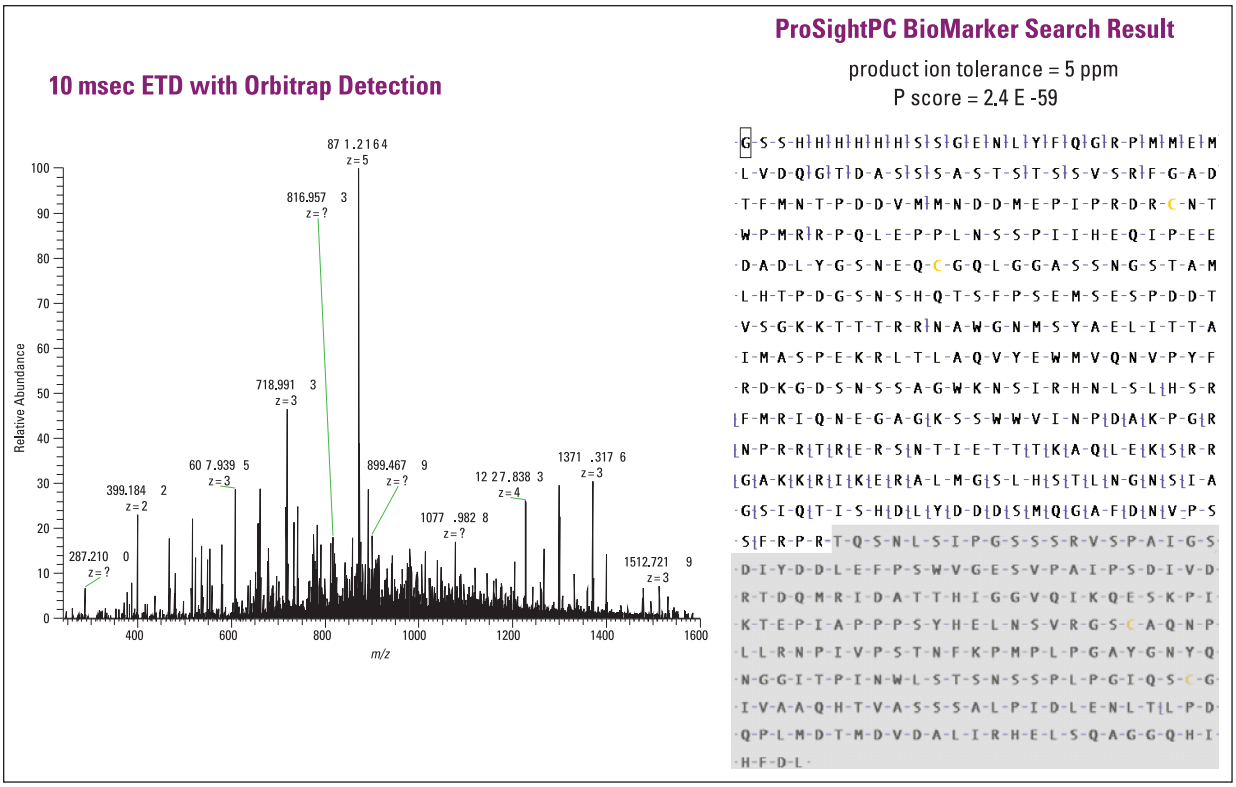
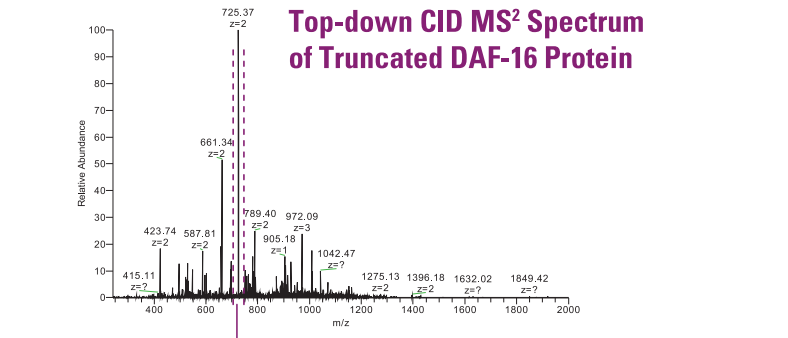
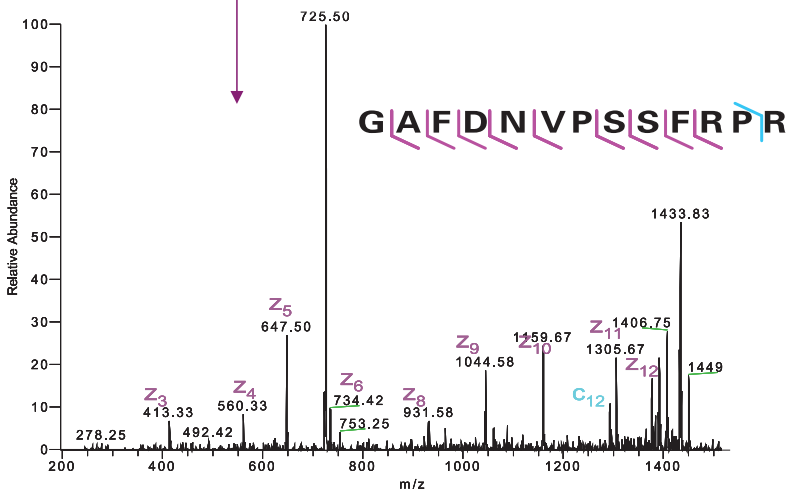


Figure 5: Identification of C-terminal truncation and top-down sequencing of DAF-16





GSSHHHHHHSSGENLYFQGRPMMEM
LVDQGTDASSSASTSTSSVSRFGAD
TFMNTPDDVMMNDDMEPIPRDRCNT
WPMRRPQLEPPLNSSPIIHEQIPEE
DADLYGSNEQCGQLGGASSNGSTAM
LHTPDGSNSHQTSFPSEMSESPDDT
VSGKKTTTRRNAWGNMSYAELITTA
IMASPEKRLTLAQVYEWMVQNVPYF
RDKGDSNSSAGWKNSIRHNLSLHSR
FMRIQNEGAGKSSWWVINPDAKPGR
NPRRTRERSNTIETTTKAQLEKSRR
GAKKRIKERALMGSLHSTLNGNSIA
GSIQTISHDLYDDDSMQGAFDNVPS
SFRPRTQSNLSIPGSSSRVSPAIGS
DIYDDLEFPSWVGESVPAIPSDIVD
RTDQMRIDATTHIGGVQIKQESKPI
KTEPIAPPPSYHELNSVRGSCAQNP
LLRNPIVPSTNFKPMPLPGAYGNYQ
NGGITPINWLSTSNSSPLPGIQSCG
IVAAQHTVASSSALPIDLENLTLPD
QPLMDTMDVDALIRHELSQAGGQHI
HFDL

Figure 6: Identification and complete sequencing of C-terminus using ETD MS³ of y13 ions

C-terminal Identification and Complete Sequencing Using a Combined Top-down ETD and CID Experimental Approach

The DAF-16 truncated protein was subjected to top-down CID followed by ETD MS³ of product ions from CID MS². A SEQUEST search of the ETD MS³ spectrum of the y13 ion against the full length DAF-16 sequence identified the C-terminus of this truncated protein and returned the complete C-terminal sequence information as well (Figure 6). In this experiment, information from the ETD MS³ spectrum covered every possible amino acid in the C-terminal peptide, including the first several amino acids that were missed when using only top down ETD MS². With the C-terminus identified, the ProSightPC search could be carried out under the Single Protein category against the truncated sequence. The same sequence coverage was obtained when using the BioMarker search (data not shown). The combined sequence coverage for the N- and C-termini from ETD MS² and ETD MS³ is shown in Figure 7. Confident characterization and complete sequence of both N- and C-termini of the truncated DAF-16 were achieved using a top-down approach combining the high-resolution, accurate-mass, multiple activation methods, and the MS^n capabilities of the LTQ Orbitrap XL ETD mass spectrometer and the ProSightPC software.

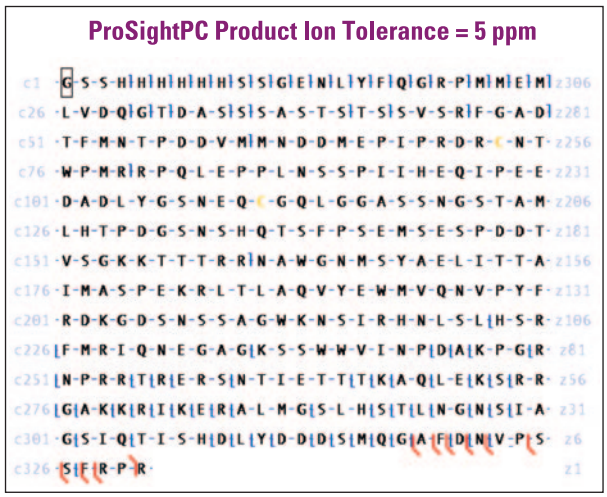


Figure 7: Top-down N- and C-terminal sequence coverage of DAF -16 fragment of 36695.4 Da

Conclusion

- Confident and complete characterization of the N-terminus and C-terminus of a truncated protein can be achieved using a top-down approach on an LTQ Orbitrap XL ETD hybrid mass spectrometer with ProSightPC software.
- In many cases, a simple top-down ETD MS^2 approach is sufficient to obtain intact protein N- and C-terminal sequence.
- Top-down CID followed by ETD MS^3 of product ions from CID MS^2 can be used to locate truncation as well as to sequence the terminus of the protein. This approach can provide complete terminal sequence coverage and is therefore particularly useful when the first several amino acids from the terminus were missed using top-down ETD MS^2.
- For a truncated protein with unknown terminus, ProSightPC is the software of choice for terminus identification as well as for N- and C-terminal sequencing. Extremely confident identifications can be obtained by analyzing high-resolution ETD data using the delta M mode of ProSightPC search under the BioMarker category.

References

1. Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. Chi A, Bai DL, Geer LY, Shabanowitz J, Hunt DF. *Int J Mass Spectrom*. 2007 Jan 1;259(1-3):197-203.
2. A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, Stafford G, Swaney DL, Syka JE, Zabrouskov V, Coon JJ. *J Proteome Res*. 2008 Aug;7(8):3127-36.

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

Africa-Other
+27 11 570 1840
Australia
+61 2 8844 9500
Austria
+43 1 333 50 34 0
Belgium
+32 53 73 42 41
Canada
+1 800 530 8447
China
+86 10 8419 3588
Denmark
+45 70 23 62 60
Europe-Other
+43 1 333 50 34 0
Finland/Norway/Sweden
+46 8 556 468 00
France
+33 1 60 92 48 00
Germany
+49 6103 408 1014
India
+91 22 6742 9434
Italy
+39 02 950 591
Japan
+81 45 453 9100
Latin America
+1 608 276 5659
Middle East
+43 1 333 50 34 0
Netherlands
+31 76 579 55 55
South Africa
+27 11 570 1840
Spain
+34 914 845 965
Switzerland
+41 61 716 77 00
UK
+44 1442 233555
USA
+1 800 532 4752

www.thermo.com

AN63237_E 08/16M