

Intact Protein N-terminal Sequencing by Mass Spectrometry; Alternatives to Edman Degradation

Iman Mohtashemi, Zhiqi Hao and Andreas F. Hühmer

• Thermo Fisher Scientific San Jose, CA, USA

Abstract

Purpose: N-terminal sequencing of intact proteins by electron transfer dissociation (ETD)

Methods: Intact proteins were directly infused into an Orbitrap mass spectrometer equipped with ETD technology. The ETD reaction time control feature allowed over-reaction of fragments to generate c-type ion series which can be read from the N-terminus. ProSightPC software was used to generate sequence specific information.

Results: Both ABRF study sample #1 and study sample #2 N-termini were identified using this approach.

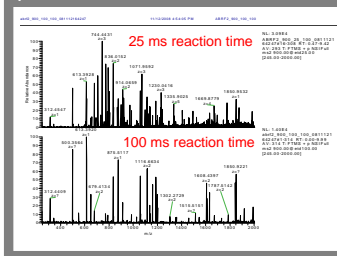
Introduction

N-terminal sequencing remains a robust and reliable tool to identify proteins, determine expression dynamics in recombinant systems, and implement quality control for signal peptide processing and protein degradation. Despite the fact that the traditional Edman technique is very robust and provides *de novo* capabilities, the technique suffers from several limitations: throughput, sensitivity, cost and specific data interpretation expertise. Mass spectrometry has been utilized as an alternative to Edman sequencing and results have thus far been variable. Most mass spec-based methods utilize the bottom-up approach in which the proteins are reduced/alkylated and digested into peptides prior to analysis. These labor-intensive techniques also have limitations in terms of digestion efficiency and efficient capture of the N-terminal peptides using various isolation methods such as immunoaffinity chromatography. In this work, we present data on N-terminal sequencing using top-down approaches in which the protein is directly infused into the mass spectrometer and fragmented via electron transfer dissociation (Figure 1). Using the appropriate ion/ion reaction conditions, charge state-reduced species are prominently evident, however a consecutive series of c-type ions are obtained by ETD that provide unambiguous sequence determination of the protein N-terminus with minimal sample manipulation. Results from the ABRF ESRG sample analysis are presented.

Methods

ABRF samples #1 and #2 were re-suspended in 0.1% formic acid. 50 picomoles of each sample was loaded onto a Michrom polymeric guard column and washed with 10 volumes with 0.1% formic acid. Proteins were then eluted with 0.1% formic/ACN 50/50 and directly infused using static nanospray with a 4 micron tip (Picotip™, New Objective). Samples were analyzed on a Thermo Scientific LTQ Orbitrap XL ETD mass spectrometer. Intact full-MS spectra (data not shown) were collected to determine precursor selection for ETD. Scans were averaged for as little as 5 minutes. Data was sent to Thermo Scientific ProSightPC 1.0 software for sequence determination as well as manually interpreted. Both samples were also sequenced using Edman techniques for reference and digested for bottom-up analysis.

FIGURE 1. ETD fragmentation of the ABRF Sample #2. Increased ETD reaction time results in a simplified spectrum which contains mostly charge-reduced fragment ion species from protein terminus



Results

N-terminal sequencing of sample #1 yielded 16 consecutive N-terminal c-type ion fragments that identified the protein as a potential His-tag fusion protein (Figure 2).

Further analysis of sample #1 using a bottom-up approach identified the sample as alcohol dehydrogenase-1 fusion protein (Figure 3).

FIGURE 2. C-type ion series of the His-tagged protein was easily identified by ETD fragmentation and accurate mass detection.

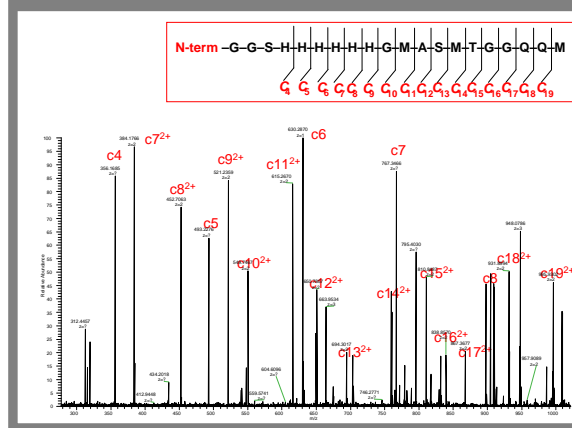
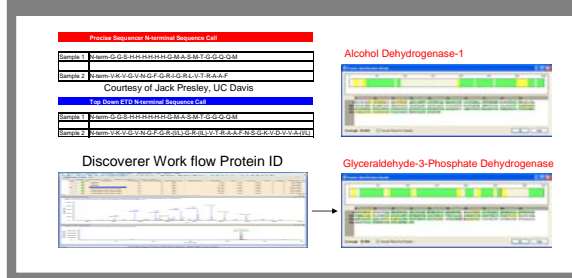


FIGURE 3. Sequence coverage and protein identification (top-down and bottom-up)



- Additional z-type ion fragments from the C-terminus of the sample #1 protein further supported these results (Figure 4).
- High mass accuracy increased confidence in fragment identifications (Figure 5).

FIGURE 4. N- and C-terminal fragmentation spectrum of ABRF sample #1

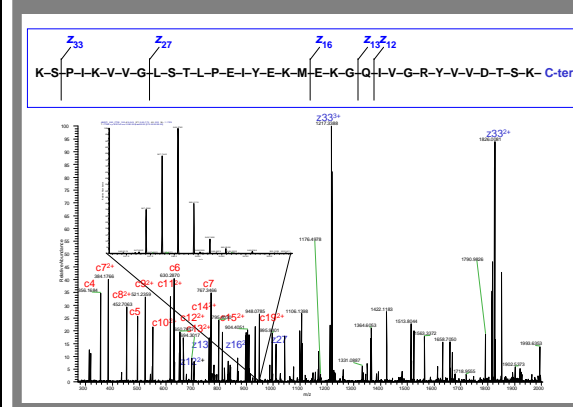
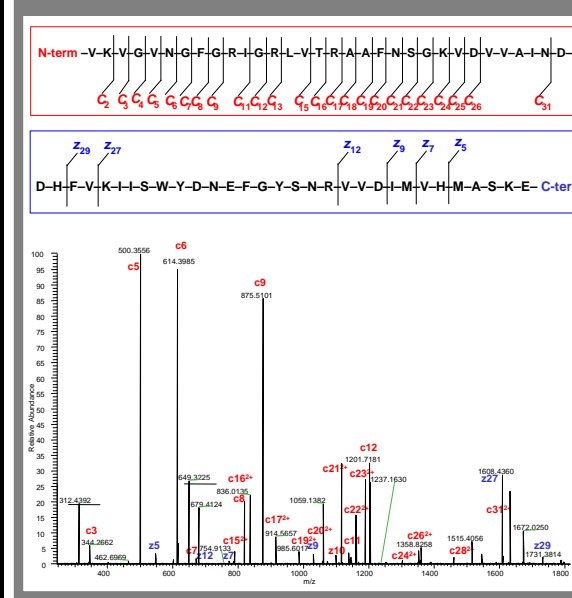


FIGURE 5. N-terminal c-type ion series with corresponding mass accuracy error for ABRF Sample #1 and #2. A combined mass accuracy average of 1.13 ppm was obtained using the LTQ Orbitrap XL with ETD.

Sample #1				Sample #2			
C Ion	Theoretical	Experimental	Error (PPM)	C Ion	Theoretical	Experimental	Error (PPM)
c4	356.1677	356.1695	-2.25	c2	245.1972	245.1974	-0.82
c5	493.2286	493.2276	2.83	c3	344.2656	344.2662	-1.74
c6	630.2855	630.287	-2.38	c4	401.2871	401.2880	-2.24
c7 2+	384.1759	384.1766	-1.82	c5	500.3555	500.3556	-0.20
c8 2+	452.7053	452.7053	-2.51	c6	614.3984	614.3985	-0.16
c9 2+	521.2348	521.2359	-2.11	c7	671.4199	671.4199	0.00
c10 2+	549.7455	549.7467	-2.18	c8	818.4883	818.4889	-0.73
c11 2+	615.2657	615.267	-2.11	c9	875.5101	875.5101	-0.34
c12 2+	650.7843	650.7857	-2.15	c11	1144.6949	1144.6968	-1.66
c13 2+	814.063	814.063	0.00	c12	1201.7164	1201.7181	-1.41
c14 2+	759.8217	759.8206	1.45	c13 2+	679.4124	679.4124	0.00
c15 2+	810.3444	810.3463	-2.34	c15 2+	785.4886	785.4888	-2.74
c16 2+	838.8551	838.857	-2.26	c16 2+	836.0125	836.0135	-1.20
c17 2+	867.3659	867.3677	-2.08	c17 2+	914.0630	914.0630	0.00
c18 2+	931.3851	931.3973	-2.26	c18 2+	949.5816	949.5845	-3.05
c19 2+	995.4244	995.426	-1.61	c20 2+	985.1001	985.0974	2.74
				c20 2+	1056.6343	1056.6364	-3.87
				c21 2+	1115.6558	1115.6584	-2.33
				c22 2+	1159.1718	1159.1710	0.69
				c23 2+	1187.6926	1187.6929	-0.17
				c24 2+	1251.7300	1251.7314	-1.12
				c25 2+	1301.2642	1301.2692	-3.84
				c26 2+	1358.7777	1358.7719	4.27
				c31 2+	1606.9282	1606.9354	-4.48

- Sample #2 was identified as glyceraldehyde-3-phosphate dehydrogenase using 24 nearly consecutive N-terminal c-type ion fragments (Figure 6). The identification was further supported by a number of C-terminal z-type fragment ions that are available for protein identification using this ETD-based sequencing approach. The bottom-up approach confirmed the findings.
- N-terminal amino acid sequences using traditional Edman degradation confirmed N-terminal sequencing results obtained by the ETD-based top-down approach.

FIGURE 6. C- and z-type ion series evident with sub 3 ppm mass accuracy using the LTQ Orbitrap XL with ETD providing amino acid sequence composition for up to 31 N-terminal residues.

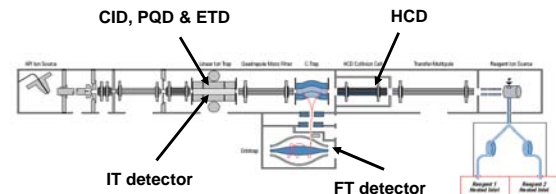


Discussion

Edman chemistry, while robust, requires long cycle times (~1 hour/amino acid) preventing high-throughput applications with high sample loads. Furthermore, C-terminal sequencing is not a trivial task. Reaction time control is critical to ETD fragmentation and longer fragmentation times result in smaller terminal, charge-reduced fragments. C- and z-type ions can be easily read from the terminus *de novo* or searched against a known database using ProSightPC™ software. Most "bottom-up" approaches result in the loss of terminal sequences necessitating alternative techniques to applications such as QC recombinant protein expression systems. Instrument scan time of 5 minutes and robust terminal fragmentation allows for adaptation for high-throughput analysis with large sample loads increasing the versatility of the LTQ Orbitrap XL ETD™ for protein characterization. Due to the random nature of ETD, both N- and C-terminal sequences information is obtained in a single experiment with excellent mass accuracy.

ETD capability added to LTQ Orbitrap XL Software Overview

- Data-dependent decision tree
- ETD data mining using Proteome Discoverer software
- Middle-down data mining with ProSightPC software



Conclusions

- Both N-termini of ESRG samples were sequenced by electron transfer dissociation (ETD) using the accurate mass detection capabilities of the LTQ Orbitrap XL ETD.
- The ETD-based sequencing approach provided the same amino acid composition information in a fraction of the time of the traditional Edman degradation technique.
- Additional sequence information covering the internal sequences of the protein is easily obtained by changing the ETD reaction conditions (see accompanying poster Hao et al.)
- The versatility of the LTQ Orbitrap XL has been increased with the addition of ETD to provide high confidence N-terminal sequencing capabilities.

Picotip is a trademark of New Objective, Inc. ProSightPC is a trademark of the University of Illinois at Urbana-Champaign. All other trademarks are the property of Thermo Fisher Scientific, Inc. and its subsidiaries.