Poster Note 64418

Optimization of LC/MS Intact /Top-Down Protein Analysis on an Orbitrap Fusion Mass Spectrometer

Rosa Viner, Seema Sharma, Jesse D. Canterbury, David M.Horn, Vlad Zabrouskov
Thermo Fisher Scientific, San Jose, CA

Overview

Purpose: Development of a general RP-LC-MS method for intact/ top-down analysis on a Thermo Scientific™ Orbitrap Fusion™ mass spectrometer.

Methods: LC/MS was performed using an Thermo Scientific™ UltiMate™ 3000 RSLCnano system in microflow mode and Orbitrap Fusion mass spectrometer in protein mode. Mixture of standard proteins were separated using a Thermo Scientific™ ProSwift™ RP-4H or 5H monolithic capillary columns (200 um x 25 cm). For top down MS/MS experiments, ETD, EThcD or CID/HCD MS2 fragmentation were used at a resolution of 120K@m/z 200.
Results: Using protein mode at standard conditions (2-3 mtorr gas pressure in the IRM), all proteins in the mixture including contaminants (enolase 2, SOD, and others) and their truncated forms were efficiently separated and identified within 15 minutes gradient. The optimized LC/MS method was successfully applied for analysis of *E.coli* ribosome sample.

Introduction

Intact/top-down protein analysis provides the unique capabilities to identify and quantify proteoforms unlike the bottom up MS approach. However, LC-MS analysis of intact proteins on a proteomics scale is challenging and requires significant method optimization of front–end separation, instrument parameters and data analysis. In this study, we developed a general RP-LC-MS method for intact/ top-down analysis on an Orbitrap Fusion mass spectrometer using standard protein mixture and monolithic capillary columns and the ProSightPD™ nodes for Thermo Scientific™ Proteome Discoverer™ 2.0 software. The optimized method was applied for characterization of a mixture of E.coli ribosomal proteins.

Methods

Sample Preparation

Protein standards (cytochrome c, myoglobin, trypsin inhibitor, bovine serum albumin, enolase, carbonic anhydrase, and RNAseA) were obtained from Sigma-Aldrich (MO) and mixed as shown in Table 1. *E.coli* ribosomes were purchased from New England Biolabs (MA) and ribosomal proteins were prepared as described in reference 1. 3.8 µg of reduced and alkylated sample was used per injection.

Liquid Chromatography

The seven protein mixture and ribosomes were separated using an UltiMate 3000 RSLCnano system in microflow mode and an Orbitrap Fusion mass spectrometer in protein mode. Proteins were separated using a ProSwift RP-4H or 5H monolithic capillary column (200 um x 25 cm), 1 ul/injection. For the seven protein mixture, gradient elution was performed from 10–25% over 6 min, from 25–40% over 5 min and from 40-70% over 3 min with ACN in 0.1% formic acid at a flow rate of 10-12 uL/min. For ribosomal proteins, gradient elution was performed from 5–40% over 38 min and from 40–70% over 5 min with ACN in 0.1% formic acid at flow rate of 10 uL/min.

TABLE 1. Seven Protein Mixture. The theoretical masses include known sequence variants, post-translational modifications, and disulfide bonds.

Elution Order	Protein Name	Uniprot Accession	Amount (pmol/inj)	Theo. Average Mass (Da)	Theo. Mono Mass (Da)	RP-4H, RT	RP-5H, RT
1a 1b	SoyBean Trypsin Inhibitor	P01070	20	19977.489 20090.649	19964.955 20078.039	3.7 4.3	1.7 2.35
2	Bovin RNaseA	P61823	10	13682.23	13674.24	5.6	3.5
3	Horse Cythochrome C	P00004	2	12358.76	12351.32	6.6	5.2
4	Yeast SOD	P00445	unknown, contaminant	15721.43	15711.81	6.9	5.4
5	Horse Myoglobin	P68082	2	16951.48	16940.96	9.6	8
6	BSA	P02769, var. 214 A-T	10	66428.69	66390.83	9.9	8.5
7	Bovine Carbonic Anhydrase II	P00921	2	29024.63	29006.74	10.6	9.4
8	Yeast Enolase I	P00924, var. 242 I-V	2	46670.91	46642.2	10.8	9.6
9	Yeast Enolase II	P00925	unknown, contaminant	46782.98	46753.97	11.1	9.8

Mass Spectrometry

Samples were analyzed on an Orbitrap Fusion mass spectrometer in intact protein mode using 2-3 mTorr ion-routing multipole (IRM) pressure. MS/MS spectra were acquired using Orbitrap HCD, CID, ETD, EThcD, and ETciD MS2 fragmentation modes with Top 3-5 DDA methods. OT MS1 data was acquired at resolution settings of 15–120K at *m/z* 200 and OTMS2 at a resolution of 120K at *m/z* 200. Precursor ion isolation was performed with the mass selecting quadrupole and the isolation window set to 3 m/z. The AGC target value was set to 5e5 for both MS1 and MS2; maximum injection times of 100 msec x 5 uscans for MS1 and 200-250 msec x 5 uscans for MS2 were used.

Data Analysis

Intact protein spectra were deconvoluted with ReSpect™ (for 15k resolution) or Xtract (for 120K resolution) using the sliding window deconvolution algorithm in Thermo Scientific™ Protein Deconvolution™ 4.0 software. The top down data were analyzed with Thermo Scientific™ ProSightPC™ 3.0 and Thermo Scientific™ Proteome Discoverer™ 2.0 (utilizing the ProSightPD™ node) software packages (Figure 1. All searches were performed against databases of 105 candidate ribosomal sequences and 20 candidate sequences for standard mixture analysis. Final results were filtered using E value cutoff of 1 $x10^{-5}$ and search engine rank 1.

Thermo SCIENTIFIC

FIGURE 1. The processing and consensus ProSight PD workflows in Proteome Discoverer 2.0 software including final annotation in ProSight Lite.

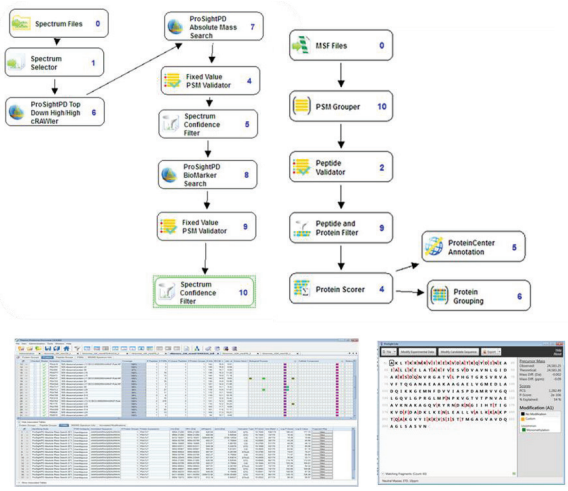


FIGURE 2. Base Peak Chromatograms of the Seven Protein Mixture Analyzed by LC-MS using ProSwift C4-RP-5H (A) or ProSwift RP-4H (B) columns.

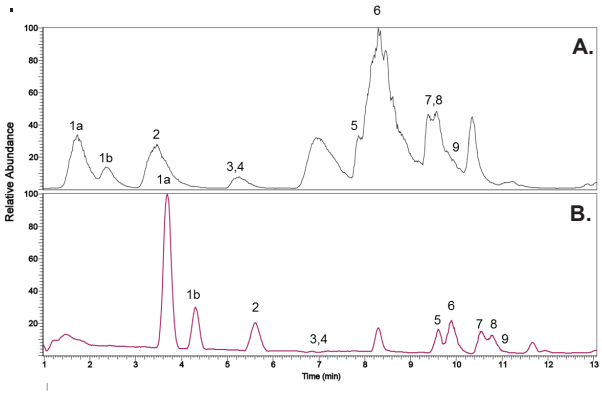


FIGURE 3. LC/MS Intact Mass Measurement Results for the Seven Protein Mixture using the sliding window deconvolution method with ReSpect (A) or Xtract (B) algorithms. The upper plot results were from the ProSwift C4-RP-5H separation while the lower plot are from ProSwift RP-4H column.

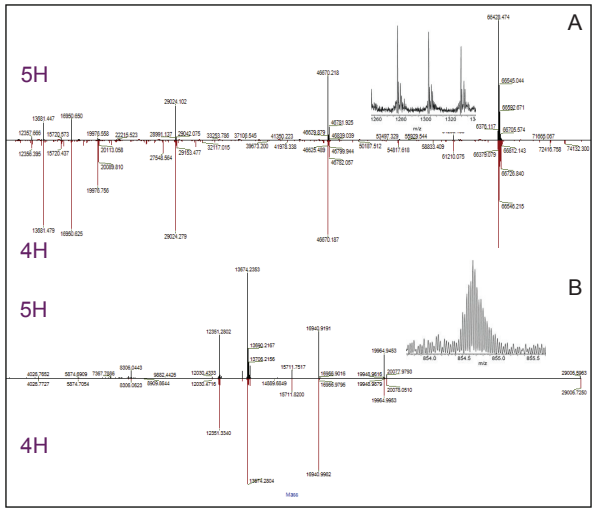


TABLE 2. Summary of the total number of assigned fragment ions from the different fragmentation methods for the Seven Protein Mixture. Data were acquired using the same LC gradient as in figure 2, 15K MS1/120K MS2 resolution@m/z200 and a Top 3 DDA method.

Protein Name	CID, CE 35	HCD, NCE 20	ETD, 6msec	EThcD, 6@10NCE	ETciD, 6@15CE
Soybean Trypsin Inhibitor	35	0	0	8	0
Bovine RNaseA	28	29	38	52	40
Horse Cythochrome C	30	21	38	32	36
Yeast SOD	36	30	43	60	51
Horse Myoglobin	49	38	78	85	71
BSA	0	19	20	18	18
Bovine Carbonic Anhydrase II	51	58	40	42	40
Yeast Enolase I	29	27	27	30	26
Yeast Enolase II	23	23	27	15	26

Results

Optimizing LC/MS conditions for intact protein analysis.

The quality of intact protein and top-down analysis using LC-MS generally depends on MS1/MS2 resolution, fragmentation type, precursor selection width, acquisition speed, and separation method. To optimize each of these parameters, we used seven standard proteins mixed in different concentrations (2-20 pmol/uL range, Table 1). These proteins were selected based on several criteria to mimic a typical top-down proteomics experiment: hydrophobicity range, molecular weight range, and the existence of multiple proteoforms. For LC separation optimization, two newly developed monolithic columns for intact protein analysis were compared (Figure 2). Using protein mode at 2 mtorr IRM pressure, all proteins in the mixture including contaminants (enolase 2, SOD, and others) and their truncated forms were efficiently separated on both columns within a 15 minute gradient. As expected for small and medium size proteins, the 4H column provided slightly better separation than the 5H[2]. All but enolase and BSA were isotopically resolved at 120K (Figure 3, A vs B) and mass accuracy was <10 ppm for isotopically resolved species (Figure 3, B) and <20 ppm for average masses (Figure 3, A).

Top down protein characterization on an LC-MS time scale is very challenging and requires informative product spectra to be obtained for each scan without excessive signal averaging. Using an MS^1 medium (15K)/MS^2 high (120k) resolution approach and optimized LC conditions (4H column, Figure 2. B) we achieved a good balance of speed and spectral quality. Five fragmentation techniques available on the Orbitrap Fusion instrument were evaluated, one per run in a total 9 different combinations. All proteins in the seven protein mixture were confidently identified via the top-down approach in a data dependent experiment with MS2 EThcD[3] fragmentation in a single LC run (Table 2) . In general, EThcD was the most efficient fragmentation technique as almost all proteins in the mixture contain disulfide bonds and no reduction/alkylation had been performed. However, as expected, the optimal fragmentation method or its conditions were protein dependent. For example as shown in Figure 4 and Table 2 , CID was the best fragmentation method for trypsin inhibitor, while ETD, HCD and ETciD failed to produce any positive identification for this protein. Results presented in Table 2 and Figure 4 were collected using protein mode at 3 mtorr IRM pressure. We could still identify all proteins in the mixture at 2 mtorr pressure, but sequence coverage and P score were reduced, especially for HCD fragmentation.

FIGURE 4. Top down Analysis of Trypsin Inhibitor by CID(A) and EThcD (B).

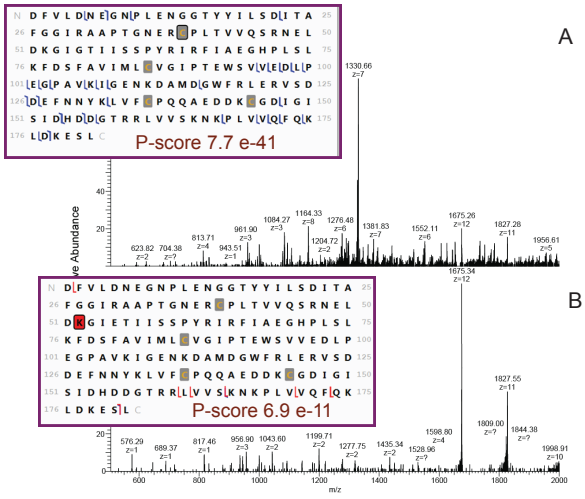
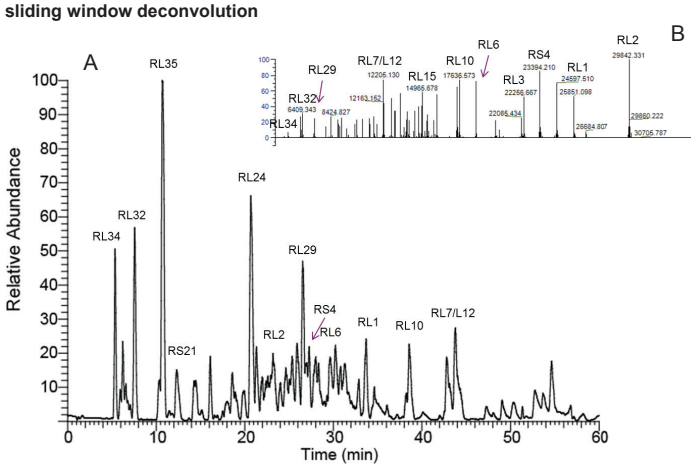


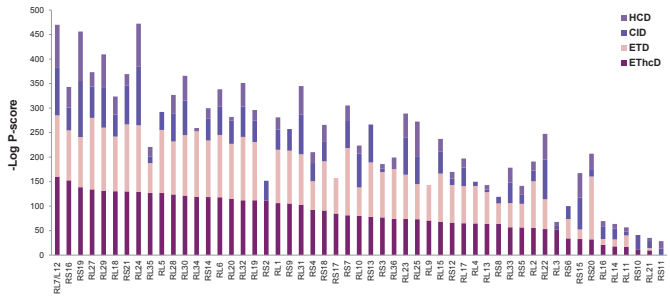
FIGURE 5. LC-MS Analysis of *E.coli* ribosomal proteins using ProSwift RP-4H column: A. Total ion chromatogram; B. Deconvolution results using ReSpect sliding window deconvolution



Analysis of E.coli ribosomal proteins by LC-MS/MS.

The ribosomal proteins are a medium complexity sample (105 unique sequences) and require better separation than a simple protein mixture. We were able to achieve highly reproducible and efficient separation using a 4H column and a 60 min gradient (see “Methods’) at a flow rate of 10 ul/min (Figure 5). 91 proteoforms were identified with at least 5% abundance (Figure 5, B). For the ribosomal protein analysis we used 2 mtorr IRM pressure and medium/high (MS^1 15k/MS^2 60 K) or high/high (MS^1 120K/MS^2 120K) top down workflows. For the top down experiments, we used ETD (6 msec), EThcD (4 msec, 10% NCE), HCD (25% NCE), and CID (35%) fragmentation methods. EThcD outperformed all other methods in terms of both identification and characterization for the majority of proteins observed (Figure 6). From the single analysis, EThcD identified 52 unique proteins while CID produced the second most with 50. ETD identified the most proteoforms (68). N-terminal acetylation and lysine methylation were the most commonly identified PTMs. Overall, using all 4 fragmentation methods combined, we identified 46 proteins and 77 proteoforms via the medium/high workflow vs. 53 proteins and 118 proteoforms via high/high. High/high workflow produced more identifications than medium/high as average mol. weight of ribosomal proteins is around 15kDa. Our results obtained using microflow LC-MS/MS and the high/high method are on par with previously reported data for nanoflow based experiments.[1]

FIGURE 6. *E.coli* Ribosomal Proteins Identified by LC-MS/MS using the High/High approach.



Conclusions

- LC at microliter flow rates in combination with ProSwift columns, 200 um id, is well suited for MS intact/top down proteomics analysis in the pmol of sample range.
- The sliding window method in Protein Deconvolution 4.0 and ProsightPD node in Proteome Discoverer 2.0 software provide complete data analysis for LC-MS and LC-MS/MS intact protein characterization.
- Comparison of different fragmentation methods for top down analysis on an Orbitrap Fusion demonstrated their complementarity and strong dependence upon protein sequence with hybrid techniques such as EThcD being the most versatile.
- An optimized microflow LC-MS/MS intact/top down workflow using multiple fragmentation methods on Orbitrap Fusion MS resulted in confident identification of 53 unique sequences from prepared *E.coli* ribosomes.

References

1. Cannon *et al. Anal.Chemistry.* **2014,** *86(4):* 2185-2192
2. Thermo Fisher Scientific Product Specifications #20842
3. Brunner *et al. Anal.Chemistry.* **2015,** PMID25803405

Acknowledgements

We would like to thank Kelly Flook, Thermo Fisher Scientific for providing prototypes of the ProSwift C4-RP-5H and ProSwift RP-4H columns and Lani Cardasis, Thermo Fisher Scientific for seven protein mixture sample.

www.thermoscientific.com

Thermo Fisher Scientific, San Jose, CA USA is ISO 13485 Certified.

ISO 13485

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China 800 810 5118 (free call domestic) 400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 45 453 9100
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand

PN64418-EN 0615S