DATA INDEPENDENT TOP-DOWN MASS SPECTROMETRY FACILITATED BY A NEW MSE PROCESSING TOOL

Lindsay J. Morrison and Barbara J. Sullivan
Waters Corporation, Beverly, USA

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

OVERVIEW

A software for processing of chromatographic top-down MSE data from the Synapt G2-Si is demonstrated using a degraded IdeS digested NIST mAb.

INTRODUCTION

• Full characterization of multiply post-translationally modified proteins is best approached by top-down methods

• Identification and characterization of truncation products is challenging by traditional bottom-up methods due to overlap with WT peptides

• We previously demonstrated improved sequence coverage and confidence scores for low abundance proteins using a top-down MSE-type approach in which the entire charge state envelope is fragmented compared to a more traditional targeted approach (Figure 1).[1] Recently, we have developed new column chemistries for intact proteins that facilitate separation of proteins on the basis of very small chemical differences; despite this, partial chromatographic resolution is often still the best that can be achieved

• Currently, options for processing chromatographic top-down data independent data are limited, often relying on a retention time window, approach were a series of scans are simply summed.

METHODS

MS peak processing was performed using the Sum and ScanByScan features in IMTBX+grppr from the University of Michigan.[2] A script for chromatographic peak processing was created in Python and an UI created using Tkinter. The software allows the user to select between two modes of apex processing, FastApex and ApexFit, the former of which relies simply upon the first and second derivative of the data and the second of which applies a GaussianExponential fitting algorithm to the detected chromatographic peaks. The high and low energy data are aligned following application of a clustering routine and the resulting deconvolved spectra visualized in the UI. Options for peak matching, truncation searching, and open searching are also included in the UI. At present, the open search is restricted to examination of truncations +/- 1000 Da from the target mass and having a single unspecified modification; however, these criteria will be expanded in future iterations. A flowchart that shows the processing workflow is shown in Figure 2.
Proof-of-concept top-down MSE experiments were performed on a degraded IdeS digest of NIST mAb using a Waters Synapt G2-Si. The protein mixture was separated using a 0.3x150 mm BioResolve column with a 15 min gradient in which the %ACN was ramped from 25 to 40%. Fragments were generated using a collision energy ramp of 20-30V in the trap region of the TriWave device prior to separation of the fragments by ion mobility and mass analysis in resolution mode.

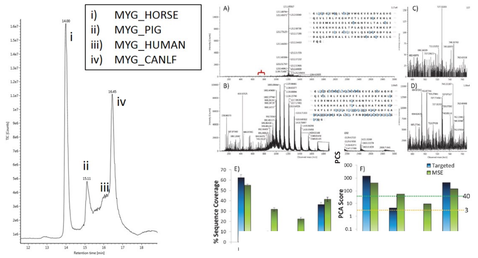

Figure 1. Top-down MS/MS (A,C) and MSE (B,D) of four myoglobin variants demonstrates the challenge of obtaining high quality fragments from low abundance proteins by targeted MS/MS. Because MSE utilizes the entire charge state envelope, better fragment S/N is obtained for the low abundance proteins (pig and human forms). This results in improved sequence coverage (E) and protein characterization scores (PCS) (F) for the low abundance variants.

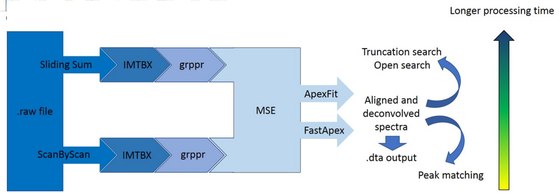

Figure 2. Flow chart for data processing options in the TD-MSE software. Two peak detection options are available for chromatographic data, ScanByScan and Sliding Sum. Sliding Sum requires significantly longer but is better suited for detecting ions in low signal peaks or fragments with very low relative abundance. Two chromatographic apexing options are available, FastApex and ApexFit, the latter of which is more appropriate for chromatographic peaks that elute with a higher degree of overlap.

RESULTS

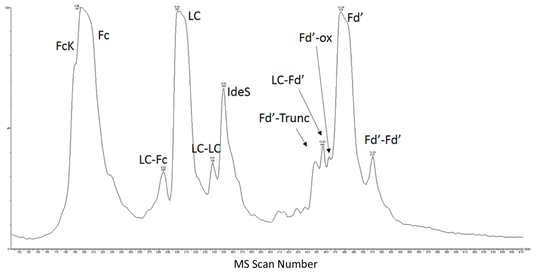

Figure 3.) Chromatographic separation of a degraded IdeS digested NIST mAb. Due to the chemical similarity of proteoforms, a single chromatographic separation is not sufficient for baseline resolution of many of the species. Lysine clips, truncations, and oxidized forms to elute with only partial chromatographic separation.

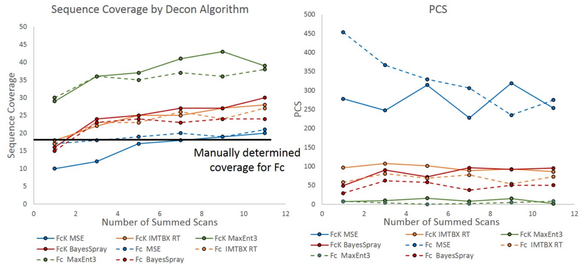

Figure 5.) Sequence coverage and PCS for the Fc and Fc-Lys subunit following sum of 1, 3, 5, 7, 9, and 11 scans and subsequent deconvolution and processing using BayesSpray, MaxEnt3, IMTBX+grppr, and IMTBX+grrpr with chromatographic apexing (MSE). A horizontal black line represents the sequence coverage of the Fc domain calculated by manual identification of b and y ions in a 5 scan average.

Due to the nature of intact protein separations, baseline resolution of proteoforms is often challenging. Using a modified HDMSE approach, charge state envelopes are subjected to CID prior to ion mobility in order to expand the peak capacity of fragment ion detection. The fragments from partially co-eluting proteoforms are then separated in post-processing according to their peak apexes. This strategy, shown in Figure 4, was applied to the separation of a degraded, IdeS digested NIST mAb, shown in Figure 3 and evaluated on the basis of sequence coverage and PCS. The manually determined sequence coverage for the Fc subunit was 18%, denoted by a horizontal black line in Figure 5. Although MaxEnt3, BayesSpray, and the retention window form of IMTBX+grppr provided higher sequence coverages than did the MSE approach, comparison of protein characterization scores shows that these deconvolutions include a large number of false positive fragment IDs. The MSE approach provides a sequence coverage consistent with manual fragment assignments and provides nearly double the PCS of the retention time window based methods. PCS was observed to decrease with increasing sliding sum window width; to better understand this phenomenon, manual assignment of internal ions was performed for the FcK subunit data using a the processed data from a 7 scan sliding sum and from the ScanByScan processed data, shown in Figure 6. A significantly larger number of internal ions were discovered in the sliding sum processing, presumably because these tend to be lower abundance. Manual removal of these ions resulted in a significant increase of the PCS, indicating many of these ions are being counted as false positives. Consequently, in the MS data viewer window in the TD-MSE UI, shown in Figure 7, high probability internal ions (those resulting from XP, DX, and EX cleavages) are colored in bronze. Primary ions are colored in blue (y type) and green (b type) to facilitate rapid identification of present or missing termini. Tools for searching for truncations, de novo sequence tags, and unknown modifications are also available.

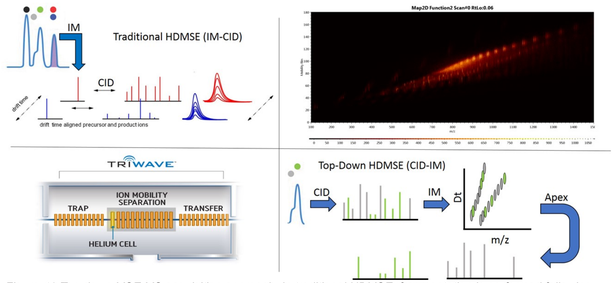

Figure 4.) Top-down MSE MS acquisition approach. In traditional HDMSE, fragmentation is performed following mobility to allow for alignment of precursor and product ions in drift space (top left). In top-down MSE, pecursor charge states are fragmented as an ensemble in the trap region of the TriWave device, allowing for improved peak capacity of the fragment ions due to the subsequence ion mobility cell, as shown in the top-right panel. Partially co-eluting protein precursors are separated in the MSE processing according to peak apex, as shown in the bottom right panel.

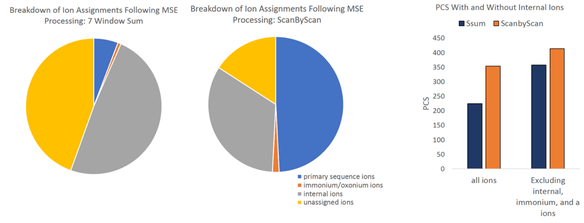

Figure 6. Observed percentage of high energy ion type using a) a sliding 7 scan window sum and b) ScanByScan mode for the FcK subunit. The sliding sum increases the number of primary ions as well as the number internal ions that are identified and unassigned species, ultimately resulting in an increase in coverage but overall reduction of the protein characterization score. In c), Protein characterization scores for the Fc subunit are shown using ScanbyScan and Sliding Sum modes for mass detection. The sliding sum approach improves the S/N of low abundance fragments, resulting in increased observance of internal and a ion species. The PCS does not takes these ions into account, resulting in a reduced PCS for the sliding sum approach, despite the higher sequence coverage. Manual removal of these ions prior to scoring with ProSight Lite improves this score.

RESULTS

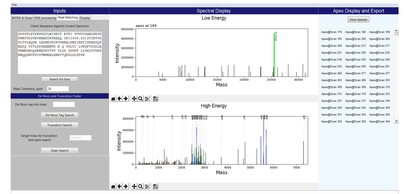

Figure 7. Screenshot of the TD-MSE user interface, demonstrating peak matching capabilities. Matched b and y type ions are highlighted in green and blue coloring to facilitate rapid identification of missing termini. Internal ions are colored in bronze. Tools for searching for N and C-terminal truncations, for unknown modifications, and de novo sequence tags are also available to facilitate identification of unknowns.

CONCLUSIONS

• Processing for data independent top-down acquisitions is demonstrated using a modified HDMSE approach for Synapt G2-Si data
• A sliding sum approach is demonstrated for improved sequence coverage of low abundance species
• Given that alternative fragmentation techniques often result in numerous but low abundance fragments, the sliding sum strategy may provide a reasonable strategy for processing of chromatographic top-down data generated using ECD or UVPD
• Using the ProSight protein characterization score as a metric for confidence, alignment by retention time apex is demonstrated to provide significantly higher confidence scores relative to retention time window based approaches
• Use of the sliding window sum feature is demonstrated to improve the number of internal ion hits, which may provide an avenue for improving sequence coverage by CID in the middle of proteins
• A copy of the TD-MSE tool may be obtained by contacting Lindsay Morrison at Lindsay_Morrison@waters.com

References

1. Morrison, L.; Williams, B.; Sullivan, B., Data Independent Top-Down Characterization of Proteins for Biotherapeautic Applications. ASMS poster presentation, **2016**.

2. Avtonomov, DM; Polasky, DA; Ruotolo, BT; Nesvizhskii, AI., IMTBX and Grppr: Software for Top-Down Proteomics Utilizing Ion Mobility-Mass Spectrometry. *Anal Chem*. **2018,** 90(3), 2369-2375.

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS