High-throughput Top-down FAIMS Data Analysis with ProSightPD Nodes in Proteome Discoverer Software

Mick Greer[1], Vincent Gerbasi[2], Joseph Greer[3], Ryan Fellers[3], David Horn[4], Romain Huguet[4], Susan Abbatiello[5], Michael Belford[4], Philip Compton[2], Kenneth Durbin[3], Neil Kelleher[2], Richard LeDuc[3], Scott Peterman[4], Paul Thomas[2]
[1]Thermo Fisher Scientific, Austin, TX, [2]Northwestern University, Evanston, IL, [3]Proteinaceous Inc., Evanston, IL, [4]Thermo Fisher Scientific, San Jose, CA, [5]Northeastern University, Boston, MA

ABSTRACT

Purpose: Evaluate Thermo Scientific™ ProSightPD™ nodes for analysis of top-down FAIMS data.

Methods: Samples were fractionated offline prior to LC separation coupled to a FAIMS enabled Orbitrap mass spectrometer. RAW files were imported directly into Thermo Scientific™ Proteome Discoverer™ 2.3 for analysis using ProSightPD™ 2.0 nodes.

Results: ProSightPD software supports analysis of FAIMS separated top-down data collected at multiple CVs. Inclusion of FAIMS separation improved several top-down experimental metrics.

INTRODUCTION

Efficient chromatographic separation of intact proteoforms remains a significant barrier to the adoption of Top-Down proteomics due to structural similarity and the wide dynamic range of biologically expressed proteoforms. Several technologies have recently been adapted for intact protein separation, including capillary electrophoresis and Gel Elution Liquid Fraction Entrapment Electrophoresis (GELFrEE,).[1] One such novel approach, High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS), shows promise for proteomics applications. FAIMS separates ions based on their gas phase mobilities under varying electrical fields.[2] Here we demonstrate the feasibility of FAIMS separations for complex biological top-down workflows supported with data analysis by ProSightPD integrated in the Proteome Discoverer platform.

MATERIALS AND METHODS

Sample Preparation

Proteins from primary human neutrophils were extracted and separated using the GELFrEE system or other fractionation techniques. Low molecular weight GELFrEE fractions (0-30kDa) were separated over RP4H monolithic chromatography columns installed on an Easy-nLC chromatography pump. CD3+ T cells were fractionated using two different methods: 1. Total primary human CD3+ T cell proteins were extracted and separated using the GELFrEE system or 2. primary human CD3+ T cells were lysed and separated into cytosol and nuclear protein fractions that were precipitated, subjected to PLRP-S chromatography on an Ultimate 3000 chromatography pump, and analyzed on a hybrid iontrap Orbitrap™ instrument by top-down.

Experimental

Chromatography-separated proteins were subjected to electrospray using the Thermo Scientific™ EASY-Spray™ LC column or a custom nanospray source in line with the Thermo Scientific™ FAIMS Pro™ interface. Proteins subjected to different FAIMS compensation voltages (CV) were analyzed on a Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (**Figure 1.**) using high resolution MS1 and high resolution MS2 modes (High/High). Resulting raw data were imported into Proteome Discoverer and analyzed using the ProSightPD nodes.

Data Analysis

Resulting raw data were imported into Proteome Discoverer and analyzed using the ProSightPD nodes. The 3 Tier Hi Hi processing template and Hi Hi Consensus workflows were used with the default parameters. The data was searched against the January 2019 version of the homo sapiens proteome warehouse downloaded from proteinaceous.net .

Figure 1. Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with FAIMS Pro interface, EASY-Spray LC source and Thermo Scientific™ EASY-nLC™ 1200

RESULTS

FAIMS improves detection of low abundance proteoforms

FAIMS separates ions based on their unique gas phase mobilities, under various electric fields. An asymmetric waveform is applied to the inner cylindrical electrode while the outer is held at ground. The user controls a compensation voltage (CV) which filter ions for transmission into the MS; ions which are not transmitted into the MS are discharged on the electrodes.

In conventional sources ions are transmitted in a nonselective way such that both abundant and low abundance precursor ions (which may not be detected in the presence of more abundant precursors) enter the MS. Changing the CV in a stepwise fashion allows users to optimize transmission of low and high abundance precursors without having to isolate a narrow m/z range (**Figures 2 and 3.**)

Figure 2. Top-down LC-MS traces (TIC) of the cytosolic fraction of CD3 cells collected without FAIMS separation (top) and with varying CVs (blue, purple, orange, and green background) using a FAIMS Pro interface. MS1 from retention time: 27.5-28.0 min highlighted in red across all traces is shown in Figure 2b highlighting the proteoforms revealed by applying multiple CVs with the FAIMS Pro interface.

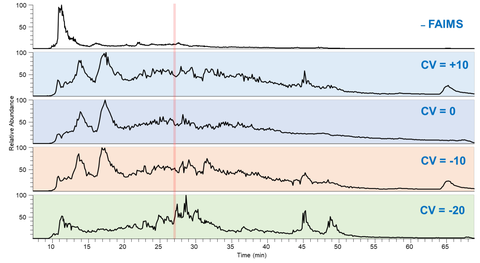
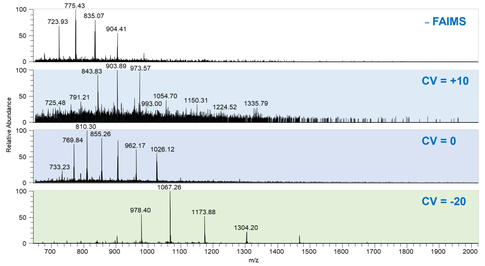


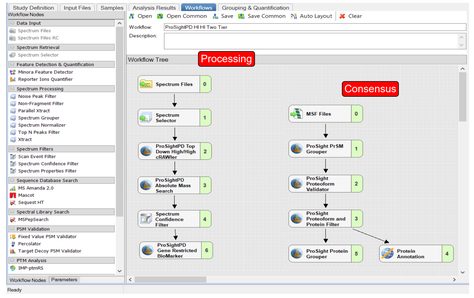
Figure 3. MS1 spectra without FAIMS (top) and with FAIMS (blue, purple and green background) from retention times highlighted in red from Fig.2a. Scanning across CV's (colorful backgrounds) reveals several low abundance proteoforms not readily detected above the noise in the top spectra (without FAIMS.)



Top Down FAIMS Pro data fully supported in Proteome Discoverer with ProSightPD nodes.

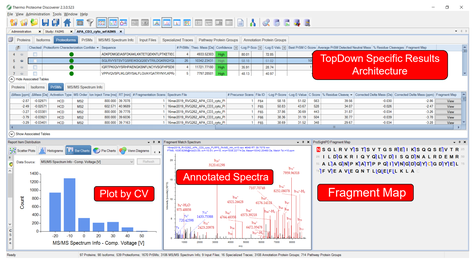
The cRAWler nodes (**Figure 4.**) included in ProSightPD 3.0 and later have been optimized for analysis of FAIMS Pro data collected using CV switching or multiple CVs in a single RAW file. The RAW files are deconstructed by CV and correct MS1 and MS2 scans are correlated and sent as targets for subsequent database searching.

Figure 4. Example ProSightPD workflows for top down data analysis shown in the Proteome Discoverer user interface.



Results from ProSight analysis of top down FAIMS data are now displayed in tables specific for top down (**Figure 5.**) New tables for proteoforms and proteoform spectral matches (PrSMs) have been added. Additionally results visualizations for FAIMS have been added allowing results to be plotted by CV.

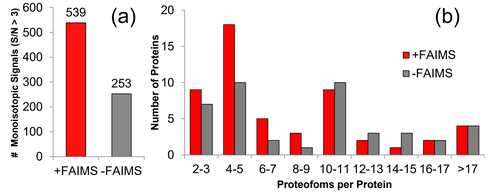
Figure 5. ProSightPD results view in Proteome Discoverer. The results view has been reconfigured to display results in a proteoform-centric design, including new tables for PrSMs, and proteoforms.



High thoughput Top down FAIMS workflow improves proteome coverage

Comparing results of a high throughput top down analysis of CD3 cytosolic proteins with and without the use of a FAIMS Pro interface on the front end show changes in several key metrics. Collecting high-data at several CVs increases the number of detected monoisotopic precursors 113% (**Figure 6a**) suggesting that many species go undetected in conventional LC-MS experiments.

Figure 6. a) Number of monoisotopic species detected with S/N > 3 in data with and without FAIMS. b) Number of identified proteins binned by the number of proteoforms per protein. Only proteins with greater than 1 proteoform per protein were plotted.



Furthermore conventional top-down LC-MS workflows often result in identifying many proteoforms of very few abundant proteins, due to a large amount of the MS2 scans devoted to these few species. This can be avoided by sample depletion or stringent dynamic exclusion. These approaches are non-ideal as they add complexity to the experiment (depletion) and often do not directly address the issue of abundance (dynamic exclusion.) An ideal solution would separate out the low abundance species while retaining the high abundance species without added sample preparation. **Figure 6b** suggests that FAIMS is able to increase the overall number of proteins with multiple proteoforms identified (larger increases in the 2-9 proteoforms per protein bin) while the results without FAIMS show a few proteins with a large number of proteoforms identified (<5 proteins in the 12 to >17 proteoforms per protein bins.)

FAIMS Separation of Proteoforms by Mass

To investigated FAIMS as a separation technique the low molecular weight GELFrEE fraction of human primary neutrophils was analyzed under CVs ranging from 0 to -80. The number of identified PrSMs was plotted by CV and binned by size in kilo-Daltons (kDa.) **Figure 7** shows a bimodal distribution of PrSM sizes where the lower masses are concentrated in CVs -80 to -40 and larger masses are found in CVs -30 to -10 .

Figure 7. Histogram of PrSM masses across 9 FAIMS CVs

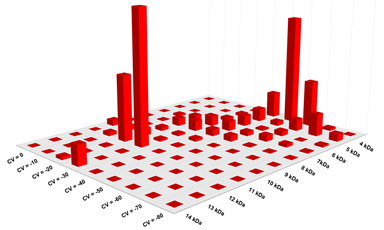


Table 1 summarizes the proteomics metrics generated by ProSightPD from data collected on a sample of cytocolic CD3 proteins with and without FAIMS. Results from FAIMS data shows significant increases not only in proteoform identifications but alsoimportantly, proteoform characterization. Interestingly most of the metrics increase greater than 10% however the number PrSM's increases only slightly. This result suggests that, for this sample, the PrSMs for data with and without FAIMS the PrSMs are distributed differently amongst the identified proteoforms (ie fewer PrSMs per proteoform resulting in more proteoform identifications) which is supported by **Figure 6b**.

	- FAIMS	+ FAIMS	
PrSM's	1607	1670	+4%
Identified Proteoforms	459	539	+17%
Protein Accessions	68	97	+43%
Characterized Proteoforms (C-Score > 40)	42	57	+35%

Table 1. Improvement in top-down proteomics metrics with use of FAIMS Pro interface, including: number of PrSMs, proteoform identifications, protein identifications and characterized proteins.

CONCLUSIONS

Addition of the FAIMS Pro interface to conventional top-down workflows shows potential to increase breadth and depth of proteome coverage. Analysis of FAIMS data in ProSightPD provides a convenient and comprehensive means to generate top-down FAIMS results. Continued efforts to optimized the applied CVs, and switching times will solidify the gains observed.

- ProSightPD is a fully integrated top-down analysis environment supporting FAIMS data analysis.
- The FAIMS Pro interface achieved separation of intact proteins by size.
- Use of the FAIMS Pro interface on a cytosolic fraction of CD3 cells improved identifications and characterization

REFERENCES

1. Lee JE, Kellie JF, Tran JC, et al. A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *J Am Soc Mass Spectrom.* 2009;20(12):2183–2191. doi:10.1016/j.jasms.2009.08.001

2. Hebert AS, Prasad S, Belford MW, et al. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem.* 2018;90(15):9529–9537. doi:10.1021/acs.analchem.8b02233

For Research Use Only. Not for use in diagnostic procedures.

TRADEMARKS/LICENSING

PO65509-EN0519S

Thermo Fisher SCIENTIFIC