Deep Metaproteome Analysis using a Vanquish Neo UHPLC System Coupled to an Orbitrap Eclipse Tribrid Mass Spectrometer with FAIMS Pro Interface

Amirmansoor Hakimi[1], Kevin L. Schauer[1], Runsheng Zheng[2], Alexander Boychenko[2], David Horn[1], Steven L. Reeber[1] and Daniel Lopez Ferrer[1]
[1]Thermo Fisher Scientific, San Jose, California, USA, [2]Thermo Fisher Scientific, Germering, Germany

ABSTRACT

Purpose:

LC-MS-based proteomics is a powerful tool for deep profiling of peptides and proteins in complex biological samples. Analyzing the proteome of microbial communities represents a challenge for current proteomics workflows due to the wide dynamic range of metaproteomes. Extensive fractionation is required to address this challenge. Here we use a newly developed high performance nLC column that provides extensive peak capacity and is hyphenated to a High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) interface coupled to a Thermo Scientific™ Orbitrap Eclipse™ Tribrid™ mass spectrometer to maximize the proteome coverage of metaproteome samples with very high protein diversity.

Methods:

1µg of peptides were separated using a Thermofisher Scientific™ EASY-Spray™ PepMap™ Neo column on a Vanquish Neo UHPLC system coupled to an Orbitrap Eclipse Tribrid MS with and without FAIMS. For the CV evaluation, eight CVs from -20 V to -90 V were tested with a 60-minute gradient. Three of the CVs that showed the best proteome coverage and the least overlap in the peptides were selected. The final FAIMS MS method was set to switch between different CVs with a top-speed method in a 3 second total cycle time over a 140 minutes gradient.

Results:

The effect of gas-phase fractionation using the FAIMS was evaluated in a bottom-up proteomics setup. A Vanquish Neo UHPLC system utilizing a 75µm x 75cm EASY-Spray PepMap Neo column was used to separate 1µg load of peptides from different microbiome standards. The raw files were searched against concatenated databases downloaded from UniPort using Thermofisher Scientific™ Proteome Discoverer™ 3.0 software utilizing, SEQUEST HT, and INFERYS re-scoring algorithm. Preliminary results showed over 10,000 proteins and 70,000 peptides in the ZymoBIOMICS Microbial Community standard for the No-FAIMS dataset. The addition of FAIMS improved protein identifications by about 19%. Similar improvement was observed in the ZymoBIOMIC Gut Microbiome standard dataset with identification of over 11,000 proteins and 80,000 peptides in the No-FAIMS experiment with a 17.5% improvement in protein identification when FAIMS is used.

INTRODUCTION

Metaproteomics aims to characterize microbial functions in environmental samples such as soil[1]. The first metaproteomics study performed on human samples was in 2009 by Verberkmoes et al.[2] but the complexity and wide dynamic range of the microbiome samples have always been a major challenge for researchers. In most cases, extensive offline fractionation is required to reduce the dynamic range and complexity of the samples. Recent advancement in liquid chromatography coupled to mass spectrometry has shown a great capacity to overcome the complexity of the biological samples. High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS™) has shown several advantages when used for gas-phase fractionation compared to traditional off-line fractionation. Here we use FAIMS coupled to a Thermo Scientific™ Orbitrap Eclipse™ Tribrid™ mass spectrometer to characterize two metaproteome standards and asses the impact of FAIMS to reduce sample complexity and maximize the proteome coverage.

MATERIALS AND METHODS

The two metaproteome standards used in this study were purchased from Zymo Research Corporation called ZymoBIOMICS Microbial Community Standard (Catalog No. D6300) and ZymoBIOMICS Gut Microbiome Standard (Catalog No. D6331). Samples were removed from -80˚C freezer and defrosted before adding lysis buffer. After a quick vortex, samples were transferred to bead beater compatible tubes and approximately 350ul worth (each) of 0.1mm glass beads and 0.5mm zirconium oxide beads were added to each sample. The bead beater was set at the speed of 10 for 5 minutes.

Species	Theoretical Composition (%)				
	Genomic DNA	16S Only	16S & 18S	Genome Copy	Cell Number
Faecalibacterium prausnitzii	14	17.63	15.96	14.77	14.82
Veillonella rogosae	14	15.87	14.37	19.94	20.01
Roseburia hominis	14	9.89	8.95	12.43	12.47
Bacteroides fragilis	14	9.94	9.00	8.33	8.36
Prevotella corporis	6	4.98	4.51	6.26	6.28
Bifidobacterium adolescentis	6	8.78	7.95	8.83	8.86
Fusobacterium nucleatum	6	7.49	6.79	7.53	7.56
Lactobacillus fermentum	6	9.63	8.72	9.68	9.71
Clostridioides difficile	1.5	2.62	2.37	1.10	1.10
Akkermansia muciniphila	1.5	0.97	0.87	1.62	1.62
Methanobrevibacter smithii	0.1	0.066	0.060	0.17	0.17
Salmonella enterica	0.01	0.009	0.008	0.007	0.0065
Enterococcus faecalis	0.001	0.0009	0.0008	0.0011	0.0011
Clostridium perfringens	0.0001	0.0002	0.0002	0.00009	0.00009
Escherichia coli (JM109)	2.8	2.53	2.29	1.82	1.83
Escherichia coli (B-3008)	2.8	2.53	2.29	1.82	1.82
Escherichia coli (B-2207)	2.8	2.29	2.07	1.64	1.65
Escherichia coli (B-766)	2.8	2.31	2.09	1.66	1.66
Escherichia coli (B-1109)	2.8	2.46	2.23	1.77	1.77
Candida albicans	1.5	N/A	3.11	0.31	0.16
Saccharomyces cerevisiae	1.4	N/A	6.35	0.32	0.16
Pseudomonas aeruginosa	12	4.2	3.6	6.1	6.1
Escherichia coli	12	10.1	8.9	8.5	8.5
Salmonella enterica	12	10.4	9.1	8.7	8.8
Lactobacillus fermentum	12	18.4	16.1	21.6	21.9
Enterococcus faecalis	12	9.9	8.7	14.6	14.6
Staphylococcus aureus	12	15.5	13.6	15.2	15.3
Listeria monocytogenes	12	14.1	12.4	13.9	13.9
Bacillus subtilis	12	17.4	15.3	10.3	10.3
Saccharomyces cerevisiae	2	NA	9.3	0.57	0.29
Cryptococcus neoformans	2	NA	3.3	0.37	0.18

Sample Preparation

Samples were then centrifuged, and the supernatant was transferred to a new tube, cleaned using 0.5 mL Pierce™ Detergent Removal Spin Column (product number: 87777) and dried down in SpeedVac and resuspended in 50 µl of lysis buffer in preparation for reduction, alkylation, digestion and clean-up step. Samples were reduced and alkylated followed by digestion with Lys-C and trypsin. Peptides were later cleaned up using SPE columns. Peptide concentration was measured by UV and the digest was then dried down in a SpeedVac. Dried samples were resuspended in 0.1% FA in water to have a final concentration of 1µg/µl.

Figure 1. Deep metaproteomic workflow including a Vanquish Neo UHPLC system and an Eclipse Tribrid mass spectrometer coupled with FAIMS Pro Interface.

LC Method

- Analytical Column: EASY-Spray™ PepMap™ Neo, 2 µm C18, 75 µm X 75 cm
- Mobile phases: [A] 0.1% FA in H_2O, [B] 0.1% FA in 80% ACN
- Flow rate: 250 nL/min
- Workflow: Direct injection
- Column temperature: 50˚C
- 1 µL injections (1µg on the column)
- Gradient: 5% to 28% B in 105 minutes and 28% to 40% B in 15 minutes.

MS Method

For the No-FAIMS experiment, a data dependent acquisition (DDA) method was used by having 240K resolution, AGC Target at 4.0e5, and Maximum IT set to Auto for the full scan. Intensity threshold was set to 5.0e3 and Dynamic Exclusion was set to 60 seconds. Data Dependent MS2 scan properties were set as shown in figure 2:

For the FAIMS dataset, identical experiments with the same parameters as the No-FAIMS experiment were made with different FAIMS compensation voltage. After evaluation of different CVs, the 3 CVs of -35V, -50V and -65V with the least overlap in peptide identification were selected for the final run.

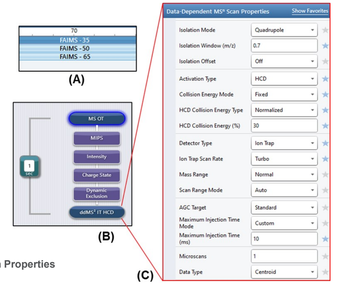

Figure 2. A) Three identical experiments with different CVs. B) The instrument method used for data acquisition.

C) Data-Dependent MS2 Scan Properties

Data Processing

Proteome Discoverer 3.0 software was used for data processing. In the Processing Workflow, Spectrum Files RC node was used for recalibration of the runs. Precursor Detector with S/N 1.5 was used to handle chimeric spectra by identifying additional precursors within the isolation window of the precursor spectrum in the results. SEQUEST HT was used to search the data with custom fasts files loaded for each standard and standard variable/static modifications were selected. Trypsin was the enzyme of choice with maximum number of 2 missed cleavage sites per peptide.

Data Processing

The Minora Feature Detector was used for label-free quantitation. The INFERYS Rescoring node was attached to the Sequest HT used to improve confidence using a deep learning-based method as described in reference 3.

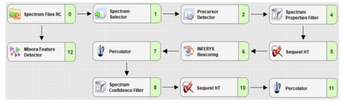

Figure 3. Processing Workflow on Proteome Discoverer software.

RESULTS

Identification

Using Proteome Discoverer software approximately 10,800 proteins and over 74,000 peptides were identified from the Microbiome Community Standard and over 11,000 proteins and close to 84,000 peptides from the Gut Microbiome Standard (Figure 4). The addition of the FAIMS Pro Interface to the workflow improved the protein identification by about 19.8% and 17.5% in Microbiome Community Standard (12,938 proteins) and Gut Microbiome Standard (13,332 proteins), respectively. This improvement clearly shows the advantage of using different compensation voltages (CV) with FAIMS Pro Interface for gas-phase fractionation of the complex samples (Figure 5). The Venn diagram shows a minimum overlap between the selected CVs and low number of peptides identified in all 3 CVs for both Microbiome Community Standard and Gut Microbiome Standard.

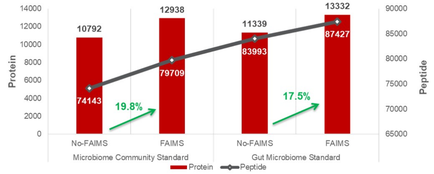

Figure 4. Protein and peptide identification with and without FAIMS Pro Interface shows the advantage of using FAIMS for gas-phase fractionation.

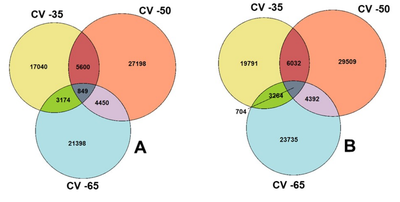

Figure 5. Minimum overlap was observed between the peptides identified using FAIMS CV of -35 V, -50 V and -65 V. A) Peptides identified in Microbiome Community standard **B)** peptides identified in Gut Microbiome Standard.

Quantitation

Label-free quantitation of those identified proteins and peptides using "Minora Feature Detector", "Feature Mapper" and "Precursor Ion Quantifier" nodes was performed in Proteome Discoverer Software. More than 95% (on average) of the identified proteins and peptides were quantified in both samples regardless of whether the FAIMS Pro Interface (Figure 6) was used with very high degree of quantitative reproducibility for triplicate injection (Figure 7).

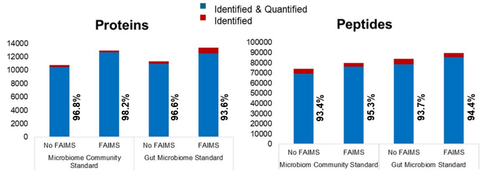

Figure 6. Percentage of quantified proteins and peptides vs only identified ones. Great quantification rate (on average 95%) was observed across the study for both proteins and peptides.

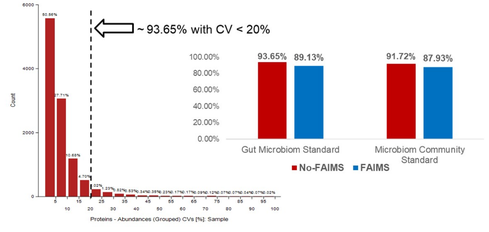

Figure 7. Quantitative reproducibility for triplicate injection. More than 93% of the proteins quantified with %CV of less than 20% in triplicates injections of Gut Microbiome Standard. Similar observation was made for other sample with and without FAIMS Pro Interface.

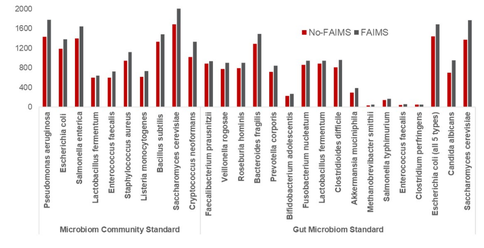

Figure 8. Proteins identified from different species in each sample. FAIMS Pro Interface improved protein identification in each species by about 20% (on average).

Figure 9 is a histogram showing the comparison of different species abundance obtained in this study by label-free quantitation to the equivalent value obtained by NGS as described in the certificate of analysis (briefly, 75 µl of the standard was used for DNA isolation using the ZymoBIOMICS™ DNA Miniprep [Cat. No. D4300]. 100 ng of the isolated DNA was used for shotgun library preparation. The final library was sequenced on HiSeq™ or MiSeq™ [Illumina]). In most of the species (~70%) , an agreement between the abundances was observed except for the lower abundances where NGS is over-estimating the abundances.

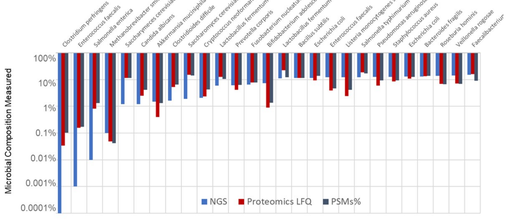

Figure 9. Predefined microbial composition measured by NGS and Proteomics for different species in Microbiome Community Standard and Gut Microbiome Standard.

CONCLUSIONS

- Analysis of metaproteome standards using a Vanquish Neo UHPLC system with an EASY-Spray PepMap Neo column coupled to an Orbitrap Eclipse Tribrid enables great proteome coverage. By identifying over 10,000 and 11,300 proteins in Microbiome Community Standard and Gut Microbiome Standard, respectively.
- Addition of the FAIMS Pro Interface allows digging even deeper into the proteome by increasing the number of identified proteins by about 20% in Microbiome Community Standard and 17.5% in Gut Microbiome Standard.
- NGS abundances for each spices in the same range as proteomics label-free quantitation except for the ones at lower abundance where NGS is overestimating the abundance.

REFERENCES

1. Wilmes, P. & Bond, P. L. Metaproteomics: studying functional gene expression in microbial ecosystems. Trends Microbiol. 14,92–97 (2006).
2. Verberkmoes, N. C. et al. Shotgun metaproteomics of the human distal gut microbiota. ISME J. 3, 179–189 (2009).
3. Zolg, D. P. et al. INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results. Rapid Communications in Mass Spectrometry, Special Issue, May 2021.

ACKNOWLEDGEMENTS

We would like to thank Corey Broeckling and Kitty Brown from Colorado State University for their help in this project.

TRADEMARKS/LICENSING

PO66125 EN0921S