Optimization of wide window acquisition methods for improved proteome coverage

Tabiwang N. Arrey[1], Bernard Delanghe[1], David Horn[2], Daniel Hermanson[2], Daniel Zolg[3], and Martin Frejno[3]
[1]Thermo Fisher Scientific GmbH, Bremen, Germany; [2]Thermo Fisher Scientific, San Jose, CA, USA; [3]MSAID GmbH, Garching b Munchen, Germany

ABSTRACT

Purpose: The novel CHIMERYS™ intelligent search algorithm provides a means to deconvolute chimeric spectra that have long been a challenge for bottom-up proteomics data analysis. To investigate the impact of data processing with CHIMERYS on data acquisition methods we varied multiple parameters to optimize the coverage and throughput of single-shot proteomics data.

Methods: We performed comparative analyses of standard runs using various gradient lengths, sample loads, and data-dependent acquisition with variable MS^2 precursor isolation widths. Data was acquired using a Thermo Scientific™ Orbitrap Exploris™ 480 mass spectrometer with or without a Thermo Scientific™ FAIMS Pro Duo™ interface. Data was processed using Thermo Scientific™ Proteome Discoverer™ 3.0 software.

Results: Increasing the precursor isolation width to values between 2-4 Th results in a marked increase in identified peptide spectrum matches (PSMs), unique peptides, and proteins across multiple gradient lengths and protein loads. The combination of wide window acquisition and data processing with the CHIMERYS intelligent search algorithm provides similar numbers of identified proteins at substantially shorter run times and improved coverage at the same run time when compared to traditional narrow isolation window methods and standard data processing with Sequest™ HT.

INTRODUCTION

The identification of peptides in bottom-up proteomics relies on matching tandem mass spectra to peptide fragmentation patterns. Tandem mass spectra containing only 1 peptide present a straightforward matching process, but typical data-dependent acquisition proteomics data sets contain chimeric spectra with more than 1 peptide in an estimated over 40% of spectra [1]. The CHIMERYS intelligent search algorithm rethinks the analysis of tandem mass spectra from the ground up by using artificial intelligence, including the accurate prediction of fragment ion intensities, to enable the deconvolution of complex chimeric spectra. Here, we evaluate the impact of using CHIMERYS for data processing on data acquisition for single-shot proteomics using wide isolation window data-dependent acquisition.

MATERIALS AND METHODS

Sample Preparation

Thermo Scientific™ Pierce™ HeLa Digest Standard (20 µg/vial) was reconstituted by adding 200 µL of 5% acetonitrile (ACN) in 0.1% formic acid in water. Sample was aspirated and pipetted approximately 10 times and then transferred to an autosampler vial. 2 µl of this solution was injected on to the trapping column. Yeast protein digest (Promega, 100 µg) was reconstituted in 200 µl of 5% ACN in 0.1% formic acid in water. It was further diluted with 0.1 % formic acid to obtain a concentration of 500 ng/µl. Sample was aspirated and pipetted approximately 10 times and then transferred to an autosampler vial. 1 µl of this solution was injected on to the trapping column.

Liquid Chromatography Parameters

Samples were run using either Trap and Elute mode or Direct Injection mode. For Trap and elute mode a Thermo Scientific™ Easy-Spray™ PepMap™ Neo 2 µm C18 75 µm X 150 mm column was used for separation along with a Thermo Scientific™ PepMap™ Neo 5 µm C18 300 µm X 5 mm Trap Cartridge. Flow rate was varied between 1.3 to 0.4 µl/min to achieve gradients of 5.5 minutes (180 samples per day, SPD), 11 min (100 SPD), 20.1 min (60 SPD), 44.4 min (30 SPD), and 56.4 min (24 SPD). For Direct Injection mode a Thermo Scientific™ Easy-Spray™ PepMap™ Neo 2 µm C18 75 µm X 500 mm column was used with a flow rate of 250 nl/min along with 60- and 90-minute gradients.

Mass Spectrometry Methods

Data was acquired using an Orbitrap Exploris 480 mass spectrometer with a scan range of 350-1,200 *m/z*. Full scan resolution was set to 45,000 for 180, 100, and 60 SPD; 60,000 for 30 and 24 SPD; and 120,000 for 60- and 90-minute direct injection gradients. For MS^2 acquisition 180, 100, and 60 SPD were run with 7,500 resolution and 30 and 24 SPD and 60- and 90-minute gradients were acquired with 15,000 resolution.

Data Analysis

Data analysis was performed using Proteome Discoverer 3.0 software with the default Sequest HT_Percolator, INFERYS_Rescoring_SequestHT_Percolator, and CHIMERYS_Percolator workflows paired with a standard consensus workflow.

RESULTS

Comparison of different search strategies for various gradient lengths

To determine the benefits of the CHIMERYS intelligent search algorithm, we performed data-dependent acquisition using 2 Th isolation for HeLa and Yeast digest using different gradient length. The average number of proteins and unique peptides identified from the same .raw data files were compared to those obtained when searched with Sequest HT and Sequest HT with INFERYS™ Rescoring. To eliminate the difference in performance as a result of sample load, we also varied the amount loaded on column for the yeast sample. CHIMERYS produces increases in unique peptides and protein groups across all gradient length and protein loads, demonstrating its utility across a wide variety of throughput conditions.

Figure 1. Average and standard deviation for identified unique peptides and proteins from 5 replicates of 250 ng of Yeast lysate with different gradient lengths and search strategies. CHIMERYS doubles protein identifications versus Sequest HT for an 8-minute gradient and identifies between 20-60% more proteins and unique peptides for longer gradients.

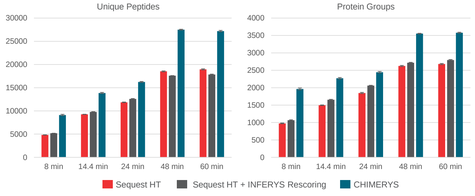


Figure 2. Average and standard deviation for identified unique peptides and proteins from 5 replicates of 500 ng of Yeast lysate with different gradient lengths and search strategies. CHIMERYS doubles protein identifications versus Sequest HT for an 8-minute gradient and identifies between 30-60% more proteins and unique peptides for longer gradients.

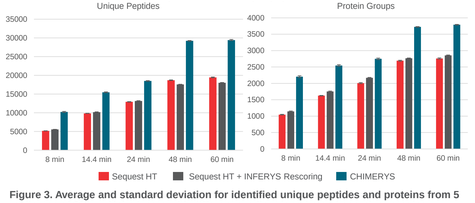
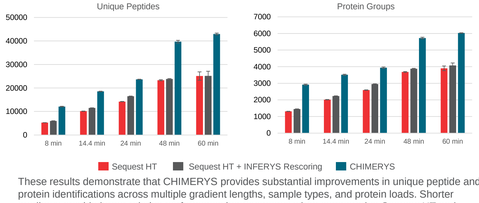


Figure 3. Average and standard deviation for identified unique peptides and proteins from 5 replicates of 200 ng HeLa lysate with different gradient lengths and search strategies. CHIMERYS provides an increase in identified proteins and unique peptides, with a more than doubling of unique peptides and proteins for an 8-minute gradient and 30-70% improvements in protein and unique peptide identifications over Sequest HT for longer gradient lengths.



These results demonstrate that CHIMERYS provides substantial improvements in unique peptide and protein identifications across multiple gradient lengths, sample types, and protein loads. Shorter gradients provide larger relative performance improvements when compared to Sequest HT and Sequest HT with INFERYS Rescoring, and a higher load provides slightly improved identifications.

Comparison of different MS^2 isolation widths for trap-elute short gradients

Given the ability of CHIMERYS to deconvolute spectra containing multiple PSMs, we investigated the impact of wider MS^2 isolation width on the number of identified proteins and unique peptides using identical gradients. These results demonstrate that the CHIMERYS intelligent search algorithm in the Proteome Discoverer software framework can be paired with wide window acquisition strategies for synergistic improvements in performance across any gradient length.

Figure 4. Average unique peptides and proteins identified for 4 replicates of 200 ng of HeLa lysate run with an 8-minute gradient with varied isolation windows processed with Proteome Discoverer software and CHIMERYS. Increasing the isolation width from 0.4 Th to 4 Th provides a 68% increase in proteins and a 120% increase in unique peptides identified.

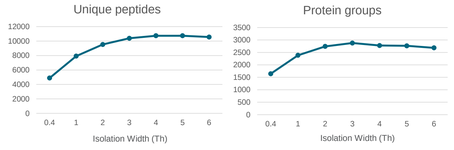


Figure 5. Average unique peptides and proteins identified for 4 replicates of 200 ng of HeLa lysate run with a 14.4-minute and 24-minute gradient with varied isolation windows processed with Proteome Discoverer software and CHIMERYS. Increasing the isolation width to between 2 and 4 Th provides similar performance for proteins but increased unique peptide identifications.

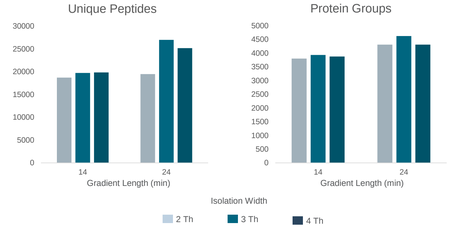
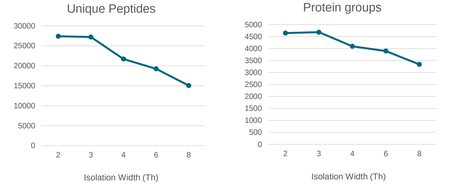


Figure 6. Average unique peptides and proteins for 3 replicates of 200 ng of HeLa lysate run with a 46-minute gradient with varied isolation windows processed with Proteome Discoverer software and CHIMERYS. Increasing the isolation width to between 2 and 3 Th provides improved performance, while wider isolation windows decreased performance.



Impact of CHIMERYS on longer runs with direct injection

To investigate the impact of CHIMERYS and wide window acquisition on longer gradient direct injection methods we also analyzed HeLa digest using 60- and 90-minute gradients. Data files were processed using Proteome Discoverer software with Sequest HT, Sequest HT with INFERYS Rescoring, and the CHIMERYS intelligent search algorithm. These results demonstrate that CHIMERYS provides increased unique peptide and protein identifications for both short and long gradient lengths and with both trap-elute and direct injection methods.

Figure 7. Average identified unique peptides and proteins for 4 replicates of 200 ng of HeLa lysate run with a 60-minute gradient with direct injection and a 2 Th isolation window. CHIMERYS provides a 37% increase in unique peptides and a 27% increase in proteins over Sequest HT with INFERYS Rescoring.

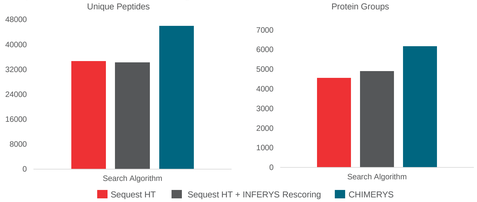
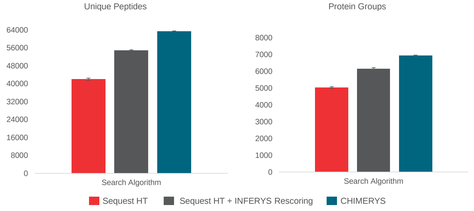


Figure 8. Average unique peptides and proteins for 4 replicates of 200 ng of HeLa lysate run with a 90-minute gradient with direct injection and a 2 Th isolation window. CHIMERYS provides a 16% increase in unique peptides and a 13% increase in proteins over Sequest HT with INFERYS Rescoring.



Comparison of different MS^2 isolation widths for direct injection long gradients

Results from HeLa digest using longer gradients with varied isolation widths revealed a similar pattern with improved performance between 2-4 Th isolation windows, suggesting that wide window acquisition is generalizable across gradient lengths and both trap-elute and direct injection methods. These results are achieved via an increase in PSMs per tandem mass spectra by collecting and deconvoluting chimeric spectra with CHIMERYS to achieve a higher instrument utilization.

Figure 9. Average unique peptides and proteins for 2 replicates of 200 ng of HeLa lysate run with a 60-minute gradient with varied isolation windows and a Thermo Scientific™ µPac™ Neo column. Increasing the isolation width to between 2 and 4 Th provides improved performance, while wider isolation windows decreased performance.

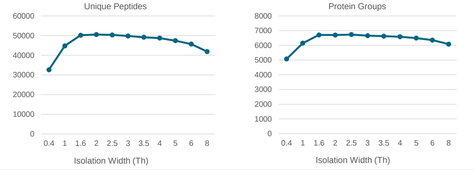


Figure 10. Average unique peptides and proteins for 3 replicates of 250 ng of HeLa lysate run with a 90-minute gradient with varied isolation windows. Increasing the isolation width to between 2 and 4 Th improved performance, while wider windows decreased performance.

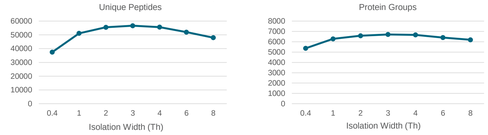


Figure 11. Average unique peptides and proteins for 3 replicates of 500 ng of HeLa lysate run with a 90-minute gradient with varied isolation windows. Increasing the isolation width to between 2 and 4 Th improved performance, while wider windows decreased performance.

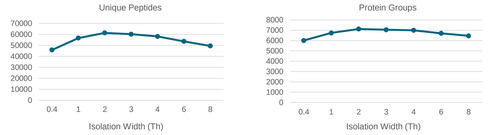
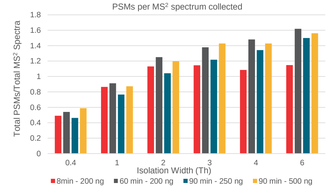


Figure 12. Average PSMs per MS^2 spectra collected for data processed with Proteome Discoverer software with CHIMERYS for varied isolation windows, gradient lengths, and protein loads. As isolation window is increased more PSMs per spectrum are identified, enabling an increased instrument utilization with more identifications per scan event.



CONCLUSIONS

- Proteome Discoverer software and the CHIMERYS intelligent search algorithm provide improvements across gradient lengths, with particularly large improvements for short run times
- Data-dependent acquisition methods can be improved by using wider isolation windows between 2-4 Th for synergistic benefits with processing with Proteome Discoverer software and the CHIMERYS intelligent search algorithm across multiple gradient lengths. These benefits arise from an increased instrument utilization without compromising identifications as shown through an increase in the number of PSMs per tandem mass spectrum.

REFERENCES

1. Dorfer V; Maltsev S; Winkler S; and Metchler K. CharmeRT: Boosting Peptide Identifications by Chimeric Spectra Identification and Retention Time Prediction. *Journal of Proteome Research* 2018 17 (8), 2581-2589. https://doi.org/10.1021/acs.jproteome.7b00836

TRADEMARKS/LICENSING