Quantitative Proteomics

High Resolution DIA: A Workflow for Highly Accurate Relative Label-Free Quantification of Microbial Proteins in Complex Cell Lysates

Julia Kraegenbring[1], Tabiwang N. Arrey[1], Jeff op de Beeck[2], Natalie van Landuyt[2], Alexander Harder[1],
[1]Thermo Fisher Scientific, Bremen, Germany, [2]Thermo Fisher Scientific MSS, Gent, Belgium

Abstract

Purpose: Demonstrate the accuracy, precision and robustness of a high-resolution data-independent acquisition (HR-DIA) label-free quantification (LFQ) workflow.

Methods: A micropillar array-based column for highly reproducible separation of peptide digests coupled to a Thermo Scientific™ Orbitrap mass spectrometer operated in data-independent acquisition mode.

Results: Different samples were quantified accurately and with a high precision of peptide quantities while maintaining a solid throughput of approximately 35 samples per day. Presented data sets are part of a larger study extending over approximately two months with an estimated net acquisition time of 500 hours.

Introduction

The increasing focus on large-scale mass spectrometry-based proteomics studies necessitates robust high-throughput methods while at the same time maintaining accurate quantification capabilities to answer biological questions. Strategies for meeting the required throughput often employ shorter gradients. However, these gradients can generate high back pressures in conventional columns, with increased strain in the whole chromatographic system, and compress peptide signals into very narrow chromatographic peaks. Here we show that the use of a versatile Vanquish™ Neo LC system with a micropillar array-based column is a suitable chromatographic setup with high reproducibility and excellent robustness over a whole set of experiments. Coupled to a Thermo Scientific™ Orbitrap Exploris™ 240 mass spectrometer operated with an optimized DIA method, results in a workflow that is suitable for reproducible analysis of large sample sets yielding great quantification performance in an LFQ approach and long-term robustness.

Materials and methods

Sample Preparation

Peptide digests were purchased from the following vendors: *E.coli* MassPREP digest from Waters, Pierce™ HeLa Protein Digest from Thermo Fisher Scientific. Peptide digests were mixed in defined ratios and separated on a 30 min gradient (see Table 1) using the Vanquish Neo HPLC system on a 50 cm Thermo Scientific™ µPAC™ HPLC column. A total of 824 ng to 992 ng of two-proteome mixed digest sample and 500 ng of three-proteome mixed digest sample was loaded onto the column. The eluting peptides were analyzed on an Orbitrap Exploris 240 mass spectrometer operated with an optimized DIA method (see Table 1).

Table 1. Summary of Method Parameters.

LC gradient		LC parameters		DIA method	
Time [min]	% B	flow rate	350 nL/min	resolution MS1/MS2	60k/15k
0	4	solvent A	0.1% formic acid	AGC target MS1/MS2	300%/800%
22.5	30	solvent B	80% ACN, 0.1% formic acid	MS1 scan range	m/z 400-900
30	45	column oven	50° C	isolation width	12 Th
30.1 – 33.0	97.5	column	50cm µPAC Neo	MS scan range	145-1450
33.1 - 39	4			HCD NCE	30

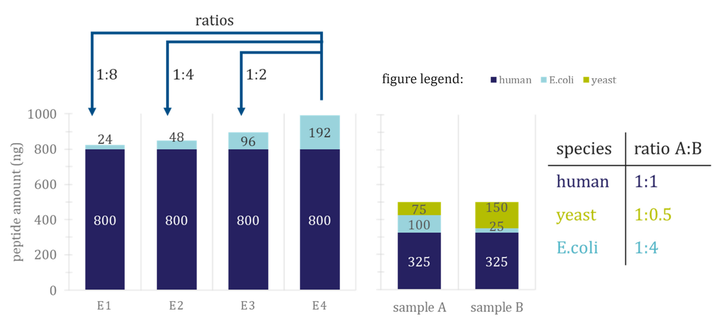


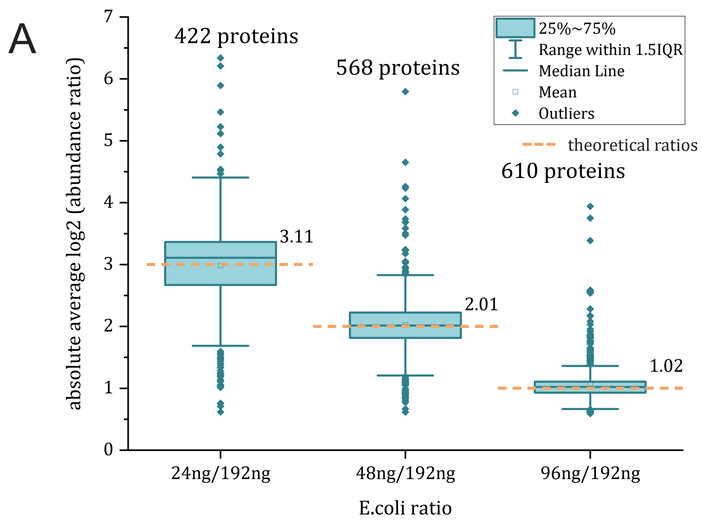
Figure 1. Sample Setup for HR-DIA LFQ Experiments

Two different sample mixtures with defined ratios were analyzed. A two-proteome mixture of high HeLa background spiked with low amounts of E.coli was analyzed at three different ratios of 1:2, 1:4, and 1:8. A three-proteome mixture (LFQ benchmark[1]) with human background and different amounts of yeast and E.coli at ratios ranging from 1:0.5 to 1:4 was analyzed as a second, more universal sample set.

Data Analysis

DIA raw data files were processed using Spectronaut™ 16 (v16.0.220606.53000). Quantification was based only on unique peptides with a false-discovery rate (FDR) of 1%. Protein groups were filtered 1% FDR on experiment level. Protein quantities are reported from such filtered protein groups exclusively.

DDA raw data files were processed with CHIMERYS® search and filtered for 1% FDR.



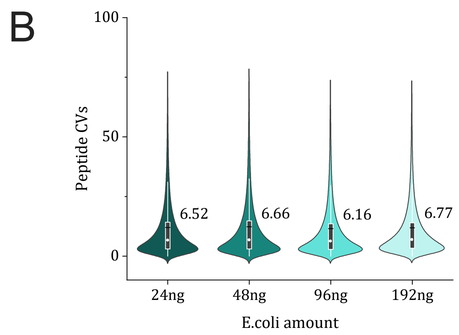


Figure 2. Accuracy and Precision of DIA Workflow for Label-Free Quantification of a Two-Proteome Mixture with a High Human Background.

(A) Whisker boxplots of logarithmic protein ratios for three different *E.coli* levels. Medians align well with the theoretical ratios (orange dotted line). The quantity ratios are narrowly distributed and a minimum of 50% (box) of all values are clearly distinguishable from the different ratios. (B) Violin plots in the four different conditions show median coefficients of variation of peptide quantities within technical replicates of well below 10%, demonstrating a high precision of the acquired data set.

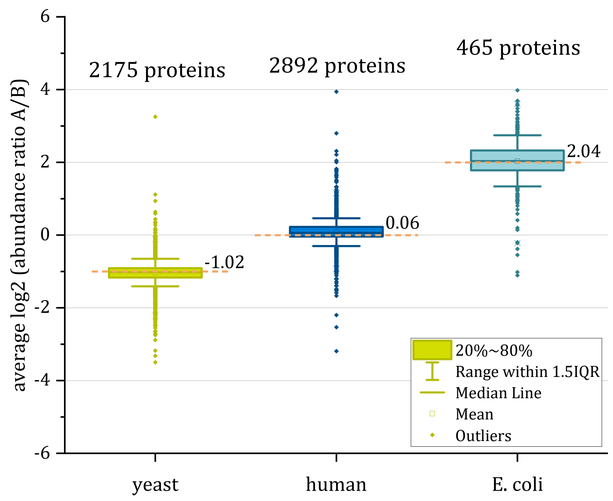


Figure 3. Accuracy and Precision of DIA Workflow for Label-Free Quantification of Two Microbial Species in a Three-Proteome Mixture.

Whisker boxplots of logarithmic protein abundance ratios of all three species. Median values align well with the theoretical ratios (orange dotted line). The quantity ratios are narrowly distributed and a minimum of 60% (box) of all values are clearly distinguishable interspecifically.

Results

Quantification Accuracy of Differential Abundances of Microbial Proteins

Samples of *E.coli* peptides at low levels in a mixture with high-abundant human peptides as background serves as a mimic in this study for differentially expressed proteins in biological samples from different conditions. Separation of these peptides using a high-performance µPAC columns giving sharp peaks with high intensities on a 30 min gradient resulted in great quantification accuracy of all three interrogated ratios, as well as high precision for the peptide quantities over three replicates (Figure 2).

Evaluating two samples with different spiked amounts of yeast and *E.coli* peptides representing both up- and down-regulation of protein expression levels and two species from different taxonomic kingdoms yields very similar quantification accuracy (Figure 3). The use of the 30 min method leads to high ID numbers for all three species while maintaining data quality and sample throughput.

Longevity of DIA Setup Assessed by DDA Performance

The experiments shown here are an excerpt from a much larger study that has been running for two months including idle time of both the MS and LC system, even with zero flow on the column. System performance has been verified by intermittent defined QC runs of 200 ng HeLa from two different sample batches in DDA mode. Identification numbers on protein group and peptide level could be maintained over the whole study, comprising 500 hours net instrument acquisition time, as well as a total estimated sample load of approximately 130 µg peptide digest (Figure 4). This outstanding robustness of column, LC, and MS is a valuable and important prerequisite for statistically meaningful, large-cohort studies.

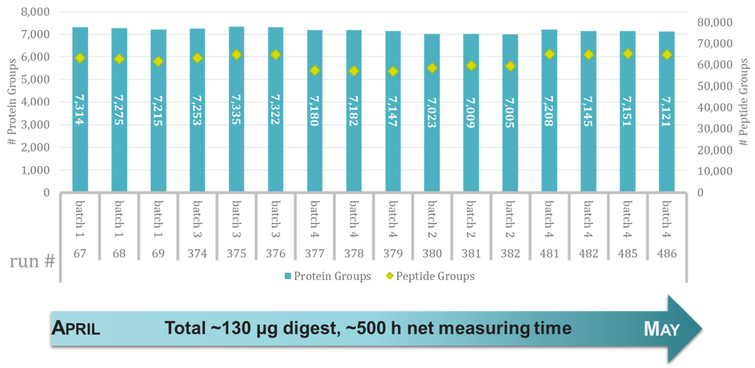


Figure 4. Protein and Peptide IDs of 67min QC DDA Runs with 200ng HeLa.

Defined QC experiments have been intermittently run throughout the whole experiment series. Protein and peptide IDs as obtained by a Chimerys search could be maintained over a time period of 2 months, including idle times, with a net measuring time of around 500 h and an estimated total of 130ug digest.

Conclusions

With the presentation of these two data sets from a robust high-resolution DIA workflow for label-free quantification we could show that:

- excellent quantification accuracy and precision can be obtained for complex sample mixtures with high background and low spiked amounts of bacterial or fungal proteomes
- sample throughput and quantitative quality of the data obtained by the shown method and setup does not compromise on proteome coverage
- the presented setup delivers long-term stability and reproducibility making it suitable for the application in large-cohort studies which often occur in clinical research

References

1. Navarro *et al.*, Nat Biotechnol. 2016;34(11):1130-1136. doi:10.1038/nbt.3685

Trademarks/licensing