Unleashing the power of HT-DIA acquisition on Orbitrap Exploris 240 MS – Precise and accurate quantitation at 260 SPD

Dominic G. Hoch[1], Riccardo Stucchi[1], Jeff op de Beeck[2], Julia Krägenbring[3], Ece Aydin[4], Maciej Bromirski[5]

[1]Thermo Fisher Scientific, Reinach, Switzerland; [2]Thermo Fisher Scientific, Ghent, Belgium, [3]Thermo Fisher Scientific, Bremen Germany; [4]Thermo Fisher Scientific, Germering, Germany
[5]Thermo Fisher Scientific, Warsaw, Poland

Abstract

Assessing and demonstrating the qualitative and quantitative performance of fast data-independent acquisition (DIA) gradients (<270SPD) on a Thermo Scientific™ Orbitrap Exploris™ 240 mass spectrometer (MS), Thermo Scientific™ Vanquish™ Neo UHPLC system and micropillar array-based 5.5 cm Thermo Scientific™ µPAC™ Neo high-throughput HPLC column.

Introduction

The ability to gain global proteome-wide information is key in the understanding of biological processes. Biological systems operate in equilibrium states, so that meaningful insights can be drawn from quantitative information on identified, impacted proteins. Therefore, the mere identification of the proteins that are present in a given sample is rarely informative. In developing suitable workflows, the aim is always to achieve the highest possible number of high-quality, confident quantitative information on proteins of interest.

With increasing capabilities of todays' mass spectrometers (MS) in terms of throughput, depth of analysis, and consistency, MS-based quantitative proteomics has widely been adopted for clinical research, e.g., in the discovery of biomarkers or studying the mechanism-of-action of novel drugs. To gain statistically significant quantitative information, study cohorts can comprise several hundreds of biological samples and replicates. A suitable workflow must accommodate for the need to process, measure, and analyze a high number of samples with reproducible data and should maintain the same performance over the duration of the whole study.

Here, we present a rugged workflow for LC-MS based high-throughput label-free quantitation using a Vanquish Neo LC system equipped with a 5.5 cm µPAC Neo high throughput column, and an Orbitrap Exploris 240 mass spectrometer. We benchmark three different gradient lengths amounting to 100, 170, and 260 samples per day (SPD), showing high proteome coverage with reproducible quantitative results between replicates. We demonstrate the quantitative performance of these three methods by use of a three-species proteome mix, where we show accurate quantitation results for all three gradients. Additionally, we could show the robustness of the workflow with data from 1000 consecutive injections, as well as cross-site comparison of data collected at three different sites in Europe.

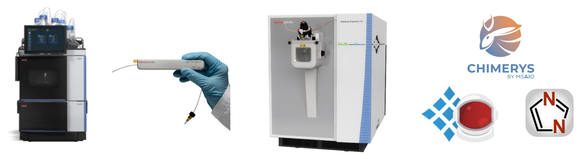
Materials and methods

Sample preparation

100 uL of 5% acetonitrile (ACN) in 0.01% TFA were added to 20ug Pierce HeLa protein digest standard. The vial was sonicated in a sonication bath for 5 minutes at room temperature, before transferring it to the autosampler vial.

Workflow components

	Description
Liquid chromatography	Thermo Scientific™ Vanquish™ Neo UHPLC system: Binary Pump N, Split Sampler NT, Solvent Rack, Vanquish System Controller, System Base with drawer, Vanquish Display (P/N 6036.1180), Vanquish Split Sampler Sample Loop, 25 µL (P/N 6252.1940), Vanquish Column Compartment N (P/N VN-C10-A-01)
Column	Thermo Scientific™ µPAC Neo high throughput HPLC column, 5.5cm (P/N COL-CAPHTNEOB)
Emitter	Stainless steel emitter, 30 µm (P/N ES542) Microtight Union
Source	Thermo Scientific™ Nanospray Flex™ ion source (P/N ES071)
Mass spectrometer	Orbitrap Exploris 240 mass spectrometer (TNG Tune v4.2 SP1)
Data analysis	Spectronaut™ 18 software (Biognosys) DIA-NN 1.8.1 Thermo Scientific™ Proteome Discoverer™ 3.1 software with CHIMERYS™ 2.0 intelligent search algorithm by MSAID



Methods

<u>260SPD</u>

LC: 3.5-minute active gradient at 3 µL/min, column temperature at 60˚C, Fast loading: PressureControl at 400 bar, loading volume: 1 µL

MS: Resolution MS1/DIA 30k/15k, scan range: 525-800, Rf lens: 70%, AGC target MS1/DIA: 300/800%, isolation width: 16 *m/z*, window overlap: 1 *m/z*, NCE: 28%

<u>170SPD</u>

LC: 5.5-minute active gradient at 1.25 µL/min, column temperature at 50˚C, Fast loading: PressureControl at 400 bar, loading volume: 1 µL

MS: Resolution MS1/DIA 30k/15k, scan range: 525-825, Rf lens: 70%, AGC target MS1/DIA: 300/800%, isolation width: 10 *m/z*, window overlap: 1 *m/z*, NCE: 28%

<u>100SPD</u>

LC: 11-minute active gradient at 1 µL/min, column temperature at 50˚C, Fast loading: PressureControl at 400 bar, loading volume: 1 µL

MS: Resolution MS1/DIA 30k/15k, scan range: 525-825, Rf lens: 70%, AGC target MS1/DIA: 300/800%, isolation width: 8 *m/z*, window overlap: 1 *m/z*, NCE: 28%

Data analysis

Data was processed using Spectronaut™ 18 software, using directDIA+™ (Deep), DIA-NN 1.8.1 or Thermo Scientific™ Proteome Discoverer ™ 3.1 software with CHIMERYS™ 2.0 intelligent search algorithm by MSAID. Settings have been set to default values, except that Cross-Run Normalization > Normalization Filter Type was set to "FASTA name filter" and the "FASTA name" was defined to be the human protein database. Peptide and protein identifications were filtered for 1% FDR, and a Q-value cutoff of 1% was used for the DIA analysis

Results

Workflow performance in high-throughput DIA (HT-DIA)

The performance of capillary flow chromatography was explored by evaluating three different LC gradient lengths (3.5, 5.5 and 10 min, resulting in approximately 260, 170, and 100 SPD and instrument productivities of 63, 65 and 70%, respectively). For each gradient length, the number of proteins and peptides identified in 200ng of HeLa digests were measured (Figure 1).

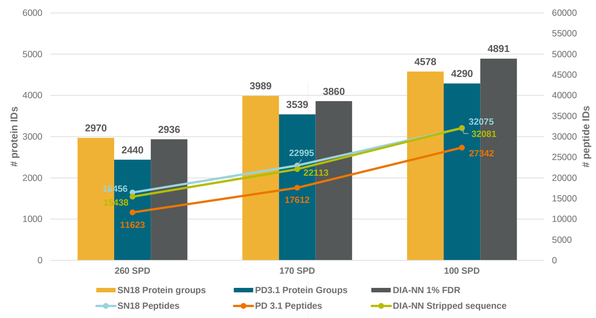


Figure 1: HT-DIA delivers exceptional proteome coverage depth combined with unprecedent productivity. Bar graph comparison of protein group and peptide numbers identified in 200ng of HeLa digest (three technical replicates per condition). Data analysis has been conducted with Spectronaut 18 software (DirectDIA), DIA-NN 1.8.1 (library-free) or Proteome Discoverer 3.1 software (CHIMERYS-DIA). All protein group and peptide results are filtered with 1% FDR.

Quantitative precision and accuracy for HT-DIA workflows

PRTC peptides were spiked in all samples and the XICs of the six most abundant transitions were assessed (Figure 2). Despite the reduced gradient length, data clearly demonstrate that a median of 5-6 points per peak could be measured for all PRTC peptides. This could be achieved by carefully adjusting the duty cycle in each of the three DIA methods tested.

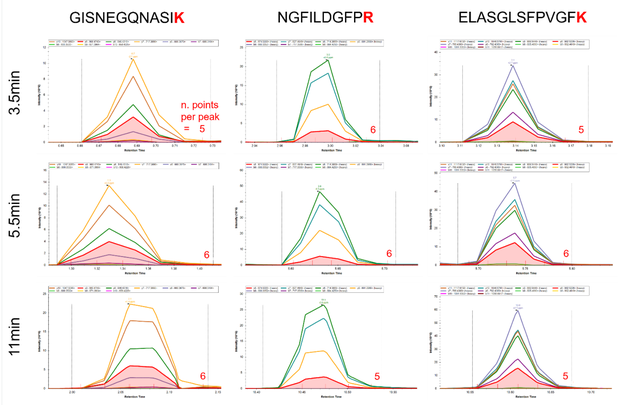


Figure 2: MS2 fragment profiling of PRTC peptides allows to evaluate quantitation performance of the HT-DIA methods. Extracted ion chromatograms (XICs) of PRTC peptides GISNEGQNASIK*, NGFILDGFPR* and ELASGLSFPVGFK*. Skyline™ output shows number of datapoints across the chromatographic peak. MS/MS spectra of selected PRTC peptides, with the six most abundant b and y fragments highlighted.

In LFQ experiments, the primary goal is the accurate and precise quantitation of proteins or peptides at differential abundance levels. To monitor the quantitative performance of the described methods, we analyzed three mixed-species proteome samples composed of tryptic digests of human, yeast, and E.coli proteins mixed in specific ratios (325ng HeLa peptides plus yeast and E.coli peptide digest in ratios of 1:0.5 and 1:4, respectively). The mixtures have been chosen because they closely mimic biological samples with up- and downregulated protein expression.

The quantification precision was evaluated by measuring the coefficients of variation (CV) across different technical replicates both at the protein and peptide levels (Figure 3). Results clearly show that all the three HT-DIA workflows tested have peptide and precursor CVs well below 10% and a very limited number of missing values, indicating high technical reproducibility.

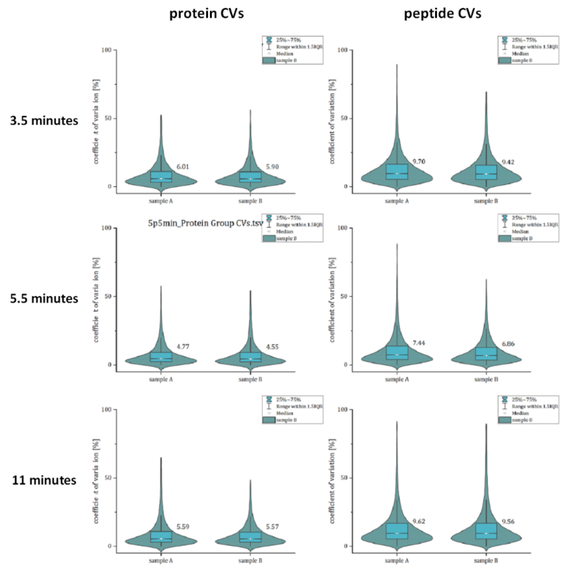


Figure 3: Determination of protein and peptide reproducibility within replicates. Plot of the two different conditions in the 170 SPD methods demonstrate high precision of protein (A) and peptide (B) quantities, with coefficients of variation well below 5% and 10% at protein and peptide level, respectively.

To assess quantification accuracy, the experimental ratios of the MS2-based protein quantification results for human, yeast, and E.coli proteins were compared to the expected ones, 1:1, 1:0.5 and 1:4, respectively. In all the three different gradients/workflows (3.5min/260SPD, 5.5min/170SPD, and 10min/100SPD), the median values of the experimental ratios measured are very close to the theoretical ones and all data points have a narrow distribution of around the median (Figure 4). These ratios are based on great identifications on the protein and peptide level, as shown in Table 1 reaching close to 6000 protein identifications for the 100SPD method.

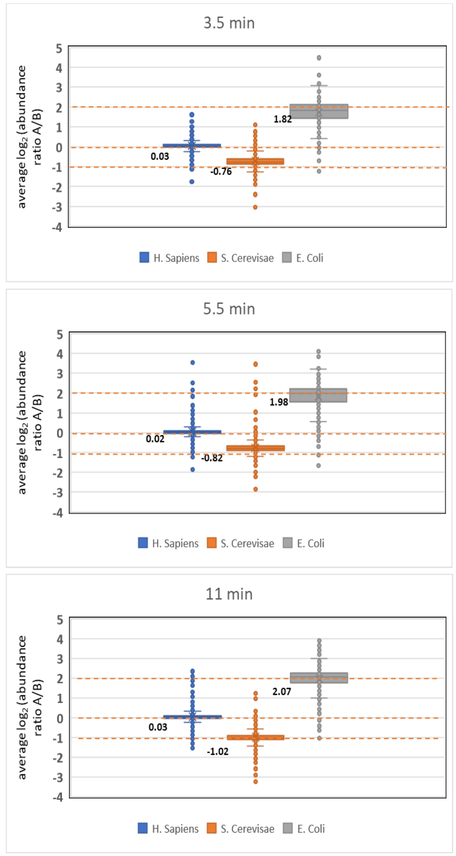


Figure 4: Determination of protein abundance ratios in a three-proteome mixture. Whisker boxplots show that protein abundance ratios of all three species (Human, yeast, and E.Coli) are consistent with the theoretical ratios (dotted line), thus indicating excellent quantitation accuracy across all different gradients / DIA methods (A-C) evaluated. Overall, these results clearly point out that the Orbitrap Exploris 240 MS provides the best quantification performance for high-throughput LFQ proteomic studies of large cohort of clinical samples.

Workflow robustness

The reproducibility and robustness of the workflow becomes even more imperative when hundreds of samples are acquired per day. Therefore, we set up the described workflow at three different laboratories to showcase the reproducibility across different sites: Ghent (Belgium), Reinach (Switzerland), and Germering (Germany). In Figure 5, you can see the protein identifications at different sites, showing the reproducibility of the workflow with RSDs of 5.5%, 6.7% and 4.9% for 3.5-, 5.5-, and 11-minute gradient length, respectively.

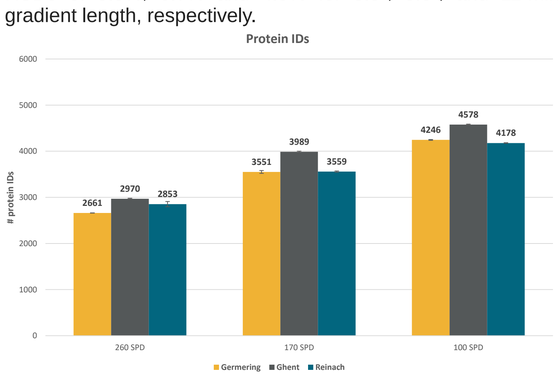


Figure 5: Inter-laboratory reproducibility of high-throughput LC-MS analysis. Proteome coverage obtained for 200 ng Human protein digest standard across three sites and with three high−throughput methods, data processed with Spectronaut 18 software.

To demonstrate the longevity of the µPAC Neo high throughput column, we tested the workflow over 11 days with a turnover of 100 SPD, yielding us more than 1000 consecutive runs without changing the hardware (Figure 6). Each run showed exceptional reproducibility of protein identifications with RSD of ~1.37%.

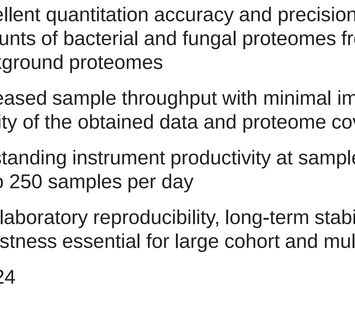


Figure 6: Workflow robustness demonstrated by reproducible protein identification of over 1000 consecutive runs and 11 days of operation time. 200 ng Human protein digest QC experiments have been run back-to-back at a sample turnover rate of 100 samples per day. Protein IDs as obtained by a CHIMERYS search could be maintained over a period of 11 days, resulting in an RSD on proteome coverage of 1.37% and a verified total of over 200 µg digest loaded.

Conclusions

The performance of a novel high throughput (HT) DIA workflow for label-free quantitation was demonstrated by combining an Orbitrap Exploris 240 mass spectrometer with fast capillary flow chromatography on a high throughput µPAC Neo HPLC column and Vanquish Neo UHPLC system. Key performance criteria of the setup/workflow include:

- Excellent quantitation accuracy and precision for small amounts of bacterial and fungal proteomes from mammalian background proteomes
- Increased sample throughput with minimal impact on the quality of the obtained data and proteome coverage
- Outstanding instrument productivity at sample turnover rates up to 250 samples per day
- Interlaboratory reproducibility, long-term stability, and robustness essential for large cohort and multi center studies

PO002624

Acknowledgements

We would like to thank Runsheng Zheng, Chris Pynn, and Robert van Ling from Thermo Fisher Scientific for the fruitful discussions.

Trademarks/licensing

Learn more at thermofisher.com/OrbitrapVelocityDIA