Thermo Fisher
SCIENTIFIC

Proteomics

Unleashing the power of DIA acquisition on an Orbitrap Exploris 240 mass spectrometer – precise and accurate quantitation at 260 SPD

Authors

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

[1]Thermo Fisher Scientific, Reinach, Switzerland

[2]Thermo Fisher Scientific, Ghent, Belgium

[3]Thermo Fisher Scientific, Bremen, Germany

[4]Thermo Fisher Scientific, Germering, Germany

Keywords

Data-independent acquisition, DIA, label-free quantitation, LFQ, high-throughput, bottom-up proteomics, Orbitrap Exploris mass spectrometer, Vanquish Neo UHPLC, 5.5 cm μPAC Neo high-throughput HPLC column

Goal

Assessing and demonstrating the qualitative and quantitative performance of fast DIA gradients (<270 SPD) on a Thermo Scientific™ Orbitrap Exploris™ 240 mass spectrometer, 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 to the understanding of biological processes. Biological systems operate in equilibrium states, so meaningful insights can be drawn from quantitative information on identified, impacted proteins. Therefore, the mere identification of the proteins present in a biological sample is rarely informative. In developing suitable workflows, the aim is always to achieve the highest possible amount of high-quality, confident quantitative information on proteins of interest.

Nowadays, thanks to the increasing capabilities of modern mass spectrometers in terms of throughput, depth of analysis, and consistency, Mass spectromtery-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. Driven by the aim of obtaining sufficient statistical significance, study cohorts typically comprise several hundreds of biological samples and technical replicates. A suitable proteomics workflow must accommodate the need to process, measure, and analyze a large number of samples with reproducible data and maintain the same performance over the duration of the whole study.

thermo scientific

Aside from a rugged sample preparation process, the instrumentation is exposed to high amounts of analyte material for a significant amount of time. When high throughput is required, both the liquid chromatograph (LC) as well as the mass spectrometer are constantly acquiring data. An optimal LC-MS workflow should be capable of delivering reproducible and consistent data. The Vanquish Neo UHPLC system delivers maximum performance for reproducible and versatile LC-MS experiments. Chromatographic columns are equally important to achieve robustness and performance. Conventional packed bed columns have been applied successfully for proteomics experiments; however, with the onset of high-throughput workflows that benefit from higher flow rates, only shorter columns can be used without creating too much strain in the fluidic system. Micropillar-array based LC columns can run at higher flow rates with longer separation paths (radially elongated pillar shape) while maintaining back pressure below 450 bar. The lower back pressure enables faster sample loading and equilibration of the column, which significantly reduces the overhead between injections and increases the MS duty cycle. These characteristics, combined with the improved peak shapes and signal intensities, make the 5.5 cm µPAC Neo high-throughput column the ideal choice for a proteomics workflow where reproducibility and consistency are essential requirements of the study.

The mass spectrometer is usually considered the key part of the data acquisition instrumentation. Additionally, careful evaluation is needed when preparing a quantitative MS method. For label-free quantification (LFQ) experiments, data-independent acquisition (DIA) schemes have surpassed the more traditional data-dependent acquisition (DDA) approaches. In DIA, the exact same information is recorded from run to run because all ions within a pre-defined set of fixed isolation windows within the interrogated mass ranges are sequentially isolated and fragmented. This ensures that in each run there is theoretical information relative to the presence and abundance of every peptide, thus providing consistent coverage and accuracy across multiple samples. In contrast to DDA, where the choice of precursors to be isolated and fragmented is stochastic, eventually leading to missing precursor information in between runs, DIA does not suffer from these missing values. As a result, DIA is considered the ideal acquisition scheme for the analysis of large cohort clinical studies, which require accurate and reproducible results.

The performance and reliability of the MS analysis must be additionally controlled by adjusting the critical method parameters, which should be tailored based on the sample type, experimental setup, and gradient length. These parameters include the interrogated mass range, the width of the isolation window, and the resolution of MS1 and MS2 scans.

Finally, data analysis also plays a crucial role in every quantitative proteomics workflow. This step should be carefully evaluated based on the setup and the goal of the study. Modern software algorithms can process complex DIA data derived from high-throughput workflows in a reasonable amount of time without prior knowledge of the sample in the form of spectral libraries. However, careful and rigid filtering and post-processing must be applied to all data output to ensure high-quality quantitative information and minimize false positives. Poor curation of data could lead to misleading results that, at the very least, dilute the statistical power of the data set.

Here, we present a rugged workflow for LC-MS based high-throughput LFQ using a Vanquish Neo UHPLC system equipped with a 5.5 cm µPAC Neo high-throughput column coupled to 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 methods by using 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 a cross-site comparison of data collected at three different laboratories in Europe.

Experimental
Consumables
- Fisher Scientific™ LC-MS grade water with 0.1% formic acid (P/N LS118-500)

- Fisher Scientific™ Optima™ LC-MS grade 80% acetonitrile with 0.1% formic acid (P/N LS122-500)

- Fisher Scientific™ Optima™ LC-MS grade 100% acetonitrile with 0.1% formic acid (P/N LS120-212)

- Thermo Scientific™ Pierce™ HeLa Protein Digest Standard (P/N 88329)

- Waters™ MassPREP™ *E. coli* Digest Standard (P/N 186003196)

- Promega™ Mass Spec-Compatible Yeast Protein Extracts (P/N V7461)

Sample preparation

Pierce HeLa Protein Digest Standard, Waters *E. coli* MassPREP Standard, and Promega Mass Spec-Compatible Yeast Protein Digest were dissolved in 1% acetonitrile, 0.1% formic acid (FA) and sonicated for 5 minutes in a sonication bath.

For the three-proteome mix, *E. coli* peptide digest and yeast peptide digest were added to a fixed amount of HeLa digest (325 ng) at amounts of 100 ng to 25 ng, and 75 ng to 150 ng, respectively, yielding an *E. coli* peptide ratio of 4:1 and a yeast peptide ratio of 0.5:1.

Instrumentation

Table 1. Workflow components

Component	Parameter
Liquid chromatography	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	• µPAC Neo high-throughput HPLC column, 5.5 cm (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™ software (Biognosys, v18) • DIA-NN 1.8.1 • Thermo Scientific™ Proteome Discoverer™ 3.1 software with CHIMERYS 2.0

LC-MS method

Pierce HeLa digest and the three-proteome mix at 200 ng total amount were loaded onto the µPAC Neo high-throughput column and separated using each of the three gradients in direct injection mode with a Vanquish Neo UHPLC system coupled to an Orbitrap Exploris 240 mass spectrometer. 0.1% formic acid in water was used as solvent A and 80% acetonitrile with 0.1% formic acid as solvent B.

Source parameters, including spray voltage and ion transfer tube temperature defined in the table, serve as starting points and are subject to optimization to ensure a stable spray. All other details of each gradient are reported in Table 2. Fast loading was enabled under pressure control (400 bar) with a loading volume of 1 µL to ensure efficient loading on the µPAC Neo HPLC column. Due to the direct-injection workflow, the overall runtime is dependent on the injection volume of the sample. With an injection volume of 1 µL, the overall delivery volume on the separation column is 2 µL (1 µL injection volume + 1 µL loading volume), which will take approximately 1.25 minutes at 400 bar. With an additional duration of 0.75 minutes for sample uptake into the sample loop, the total overhead time to be added to each run is about 2 minutes.

Data analysis and post-processing

Data was processed using Spectronaut 18 software using directDIA+ (Deep), DIA-NN 1.8.1 software, or 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. Chart data from Spectronaut software was exported as a .csv file and imported to Microsoft™ Excel™ for data visualization and plotting.

Table 2. LC-MS method

Separation column specifications (set in in the Vanquish Neo system)	
Inner diameter	75 µm
Length	50 cm
Maximum pressure	450 bar
Maximum flow	5 µL/min
Maximum temperature	60 °C

LC method		260 SPD			170 SPD			100 SPD		
Gradient		Time (min)	Flow rate (µL/min)	%B	Time (min)	Flow rate (µL/min)	%B	Time (min)	Flow rate (µL/min)	%B
		0	3	6	0	2.5	1	0	2.5	1
		0.1	3	10	0.1	2.5	8	0.1	2.5	4
		2.1	3	22.5	0.7	2.5	12.5	0.7	2.5	7.5
		3.4	3	45	0.8	1.25	12.6	0.8	1.0	7.6
		3.5–4	3	99	3.6	1.25	22.5	7.8	1.0	22.5
					5	1.25	45	10.5	1.0	45
					5.5–6.7	1.25	99	11–12.5	1.0	99
LC parameters	Column temperature	60 °C			50 °C			50 °C		
	Fast loading	PressureControl			PressureControl			PressureControl		
	Pressure loading	400 bar			400 bar			400 bar		
	Loading volume	1 µL			1 µL			1 µL		
	Equilibration factor	0			0			0		
	Sampler temperature	7 °C			7 °C			7 °C		

MS method (application mode "Peptide")		4 min method duration	6.7 min method duration	12.5 min method duration
Global parameters	Spray voltage	1,900 V	1,900 V	1,900 V
	ITT temperature	305 °C	305 °C	305 °C
	Expected peak width	10 s	10 s	10 s
	Advanced peak determination	True	True	True
	Default charge state	2	2	2
MS parameters	Resolution MS1/DIA	30k/15k	30k/15k	30k/15k
	Scan range (m/z) MS1	525–800	525–825	525-825
	RF lens (%) MS1/DIA	70	70	70
	Normalized AGC target (%) MS1/DIA	300/800	300/800	300/800
	DIA precursor mass range (m/z)	525–800 /145–1,450	525–825/145–1,450	525–825/145–1,450
	Isolation width (m/z)	16	10	8
	Window overlap (m/z)	1	1	1
	Window placement optimization	On	On	On
	Normalized HCD collision energy (%)	28	28	28

Results and discussion

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 minutes, 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 200 ng of HeLa digests were measured (Figure 1). Data show that the directDIA approach of Spectronaut 18 software as well as the library-free search within DIA-NN enable high protein and peptide numbers, considering the length of the runs and the exceptional productivity of this workflow. Similarly, the same data set was also analyzed with Proteome Discoverer 3.1 software with CHIMERYS 2.0, which

provided slightly lower performance in terms of protein group and peptide identifications compared to Spectronaut 18 software and DIA-NN. As expected, running longer gradients (11 minutes, 100 SPD) resulted in the highest number of proteins (~5,000) and peptides (~30,000) identified. These numbers decreased as the capillary flow gradient duration was shortened, but even the shortest gradient (3.5 minutes, 260 SPD) provided sufficient proteome depth (~3,000 protein groups) to unravel the underlying mechanisms of protein dynamics in biological processes. The median CVs on the protein, peptide, and precursor levels were altogether well below 5%, 10%, and 10%, respectively (data not shown). These results clearly indicate that the HT-DIA methods described here on the Orbitrap Exploris 240 mass spectrometer offer a solid solution for proteomic studies in which high proteome coverage and throughput are essential requirements.
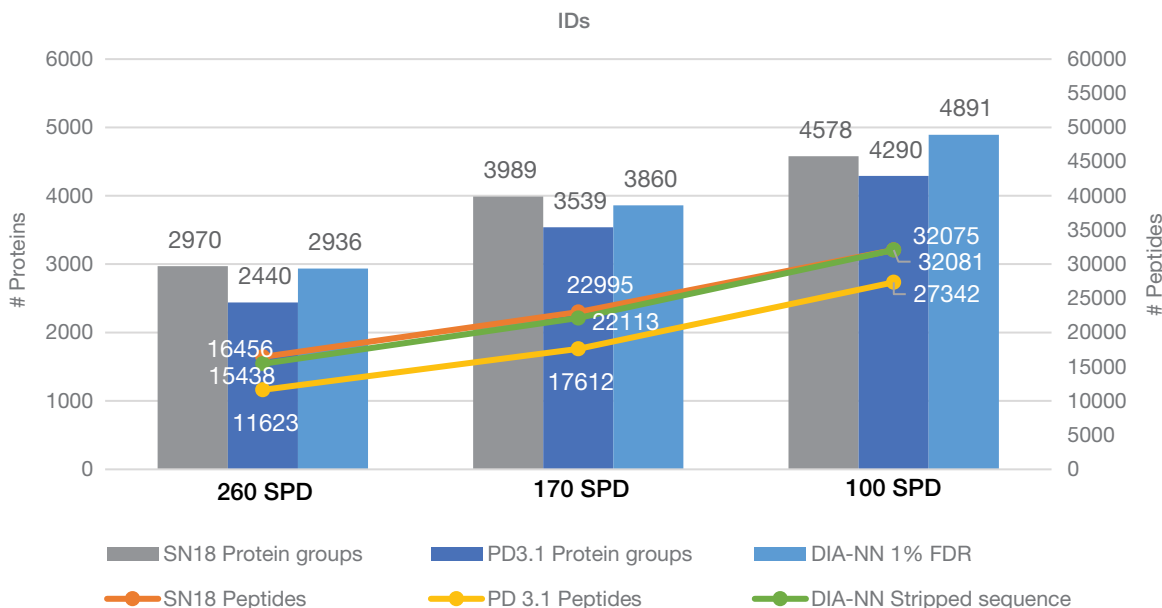

Figure 1. HT-DIA delivers exceptional proteome coverage depth combined with unprecedented productivity. Bar graph comparison of protein group and peptide numbers identified in 200 ng of HeLa digest (three technical replicates per condition). Data analysis has been conducted with Spectronaut 18 software (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

In DIA experiments, to be able to accurately quantify protein abundances, not only is depth in proteome coverage important but also the quality of the extracted ion chromatograms (XICs) of the fragment ions, which is a parameter tightly correlated with the number of points across the peak. To evaluate this aspect, 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, the 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 for each of the three DIA methods tested.

In LFQ experiments, the primary goal is the accurate and precise quantitation of proteins and 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 (HeLa peptides plus yeast and E. coli peptide digests in ratios of 0.5:1 and 4:1, respectively). These mixtures have been chosen because they closely mimic biological samples with up- and down-regulated protein expression.

First, 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 three HT-DIA workflows have protein and peptide CVs well below 10%, indicating high technical reproducibility.

Next, to assess the quantification accuracy, the experimental ratios of the measured protein abundances for human, yeast, and E. coli proteins were compared to the expected ones: 1:1, 0.5:1, and 4:1, respectively. In all three different gradients/workflows (3.5 min/260 SPD, 5.5 min/170 SPD, and 11 min/100 SPD) the median values of the calculated ratios closely resemble the theoretical ones, and all data points have a narrow distribution around the median (Figure 4). These ratios are based on excellent identifications at the protein and peptide level, reaching close to 6,000 proteins identified in total with the 100 SPD method.

These results clearly demonstrate that the quantification of the low abundant yeast and E. coli proteins in the high amount of "background" human proteins yields exquisite accuracy in the three different HT-DIA methods tested.
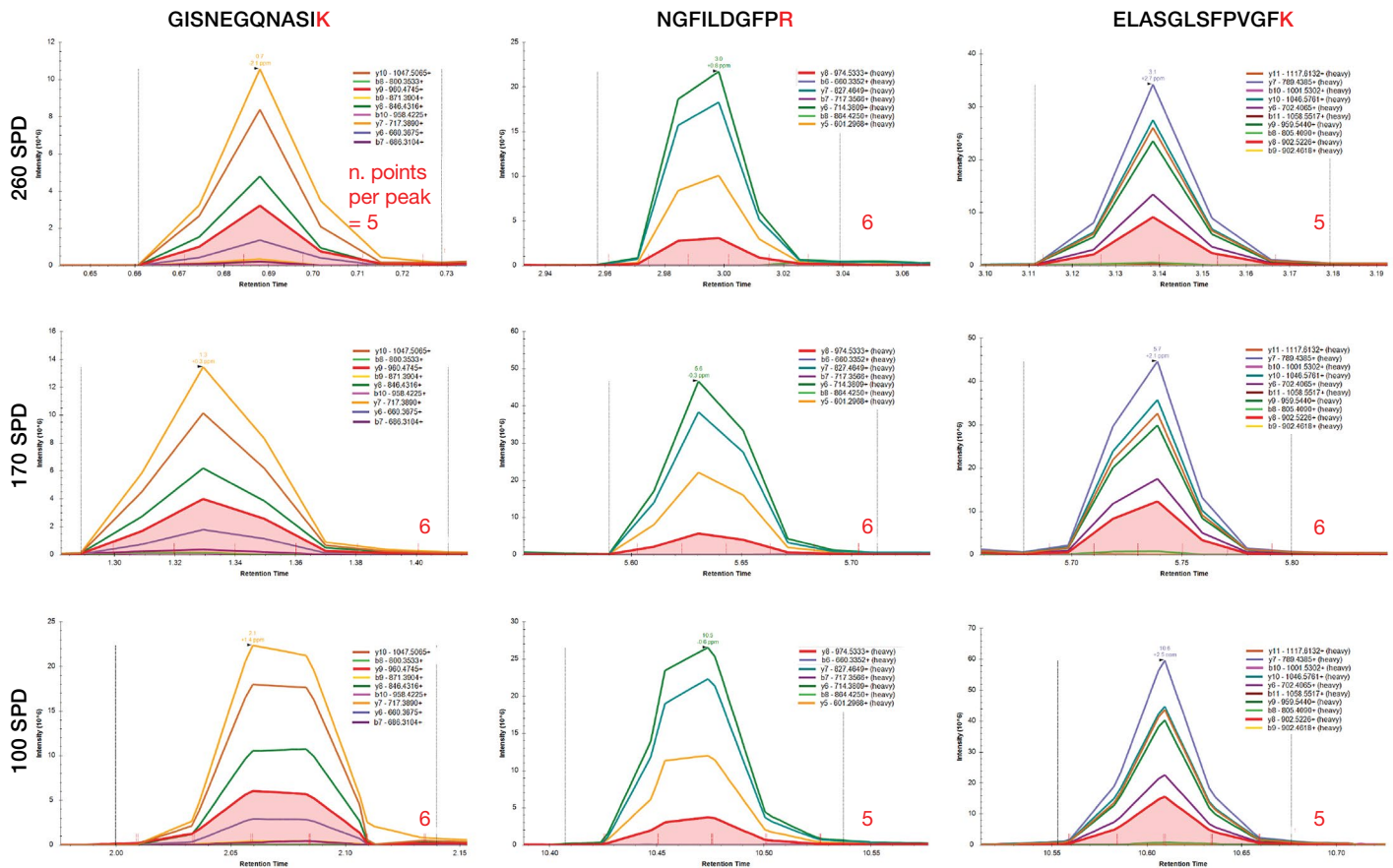

Figure 2. MS2 fragment profiling of PRTC peptides allows for the evaluation of the quantitation performance of the HT-DIA methods.
Extracted ion chromatograms (XICs) of PRTC peptides GISNEGQNASIK*, NGFILDGFPR*, and ELASGLSFPVGFK*. Skyline output shows the numbers of data points across the chromatographic peaks. MS/MS spectra of selected PRTC peptides, with the six most abundant b- and y-fragments highlighted.
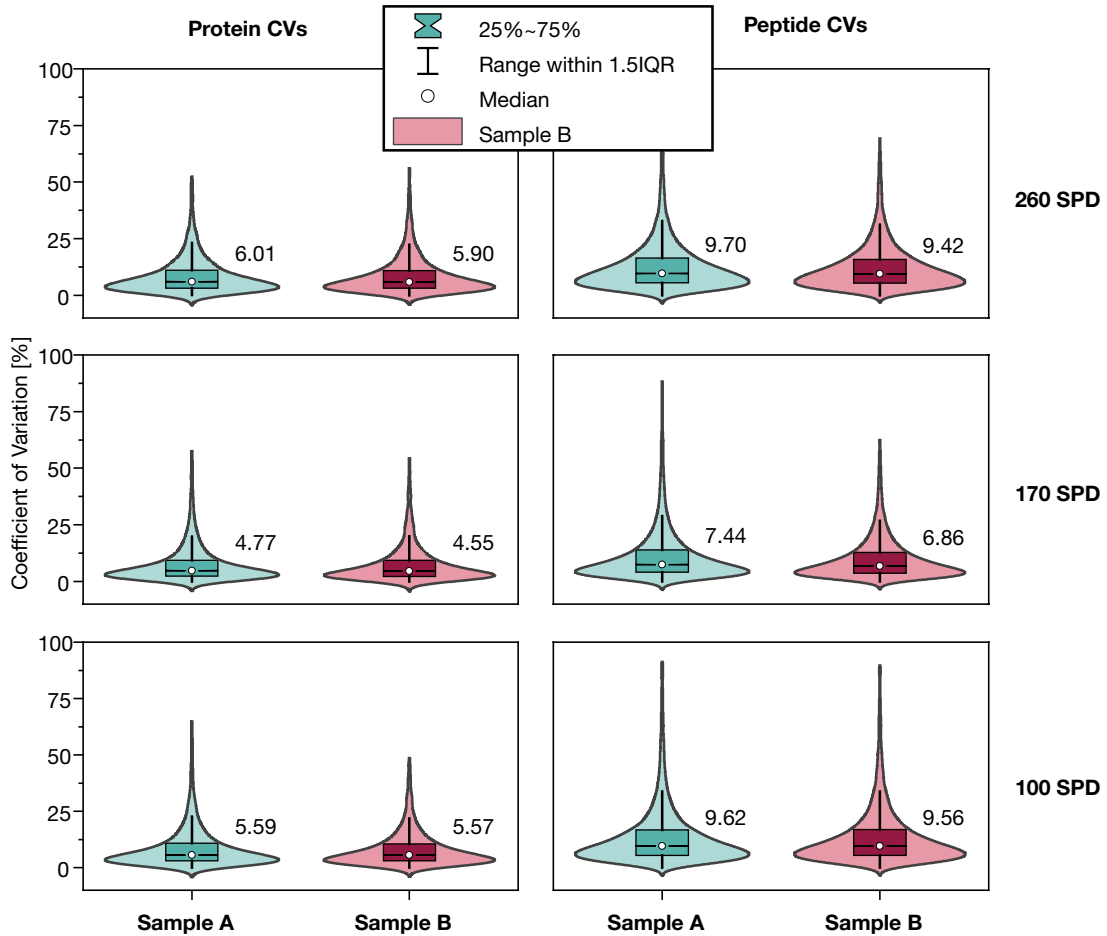
Figure 3. Determination of protein and peptide reproducibility within replicates. Plots of the two different conditions in the 260, 170, and 100 SPD methods demonstrate high precision of protein and peptide quantities, with coefficients of variation well below 5% and 10% at protein and peptide levels, respectively.
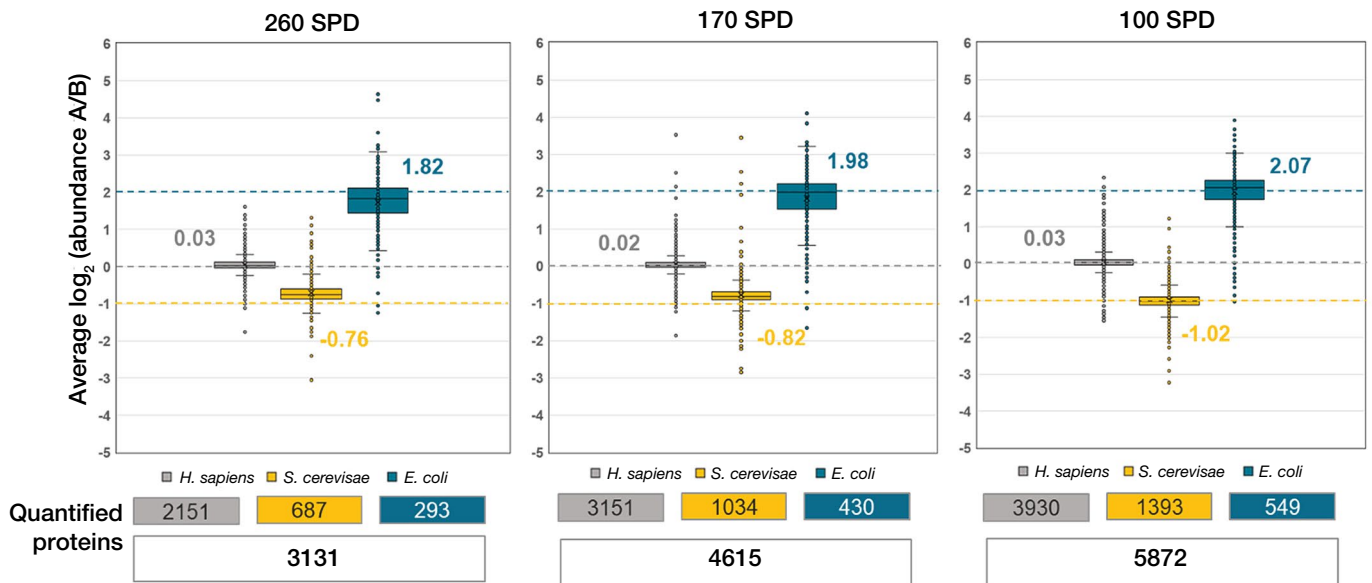

Figure 4. Determination of protein abundance ratios in a three-proteome mixture. Whisker boxplots show that protein abundance ratios of all three species (human, yeast, and *E. coli*) are consistent with the theoretical ratios (dotted line), thus indicating excellent quantitation accuracy across all different gradients / DIA methods evaluated. Overall, these results clearly point out that the Orbitrap Exploris 240 mass spectrometer 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 become even more imperative when hundreds of samples are acquired per day. Therefore, we set up the described workflow at three different laboratories to highlight the reproducibility across different sites: Ghent (Belgium), Reinach (Switzerland), and Germering (Germany). In Figure 5, the total number of protein identifications obtained at different sites is reported, 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.

To demonstrate the longevity of the µPAC Neo high-throughput HPLC column, we tested the workflow with a turnover of 100 SPD over more than 1,000 consecutive runs without maintenance (Figure 6). Each run showed exceptional reproducibility of protein identifications, with RSD of ~1.37% and a verified total load of over 200 µg digest.

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.
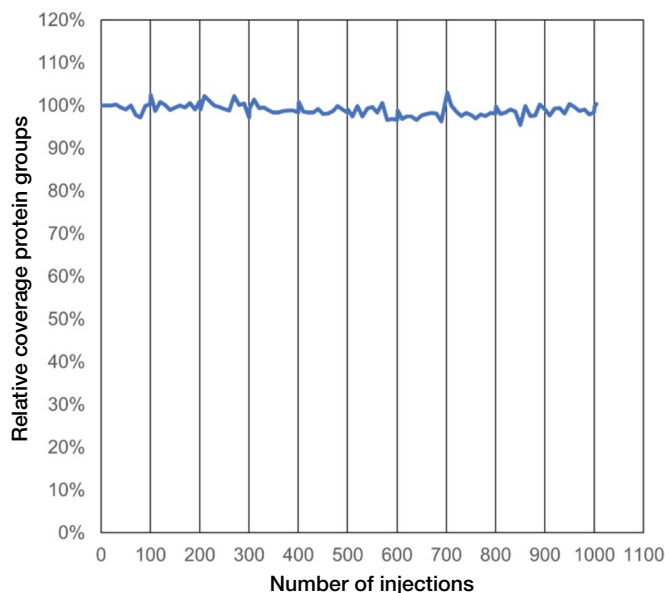

Figure 6. Workflow robustness is demonstrated by reproducible protein identification in over 1,000 consecutive runs. 200 ng human protein digest QC experiments have been run back-to-back at a sample turnover rate of 100 samples per day. Protein IDs as obtained by a CHIMERYS search could be maintained over a period of 11 days, resulting in an RSD on proteome coverage of 1.37%.

Conclusion

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

• Excellent quantitation accuracy and precision of small amounts of bacterial and yeast proteomes from mammalian background proteome

• Exceptional sample throughput with optimal performance in terms of data quality and proteome coverage

• Outstanding instrument productivity, with a sample turnover of up to 260 SPD

• Interlaboratory reproducibility, long-term stability, and robustness are essential for large cohort and multicenter studies

Learn more at **thermofisher.com/OrbitrapVelocityDIA**

thermo scientific