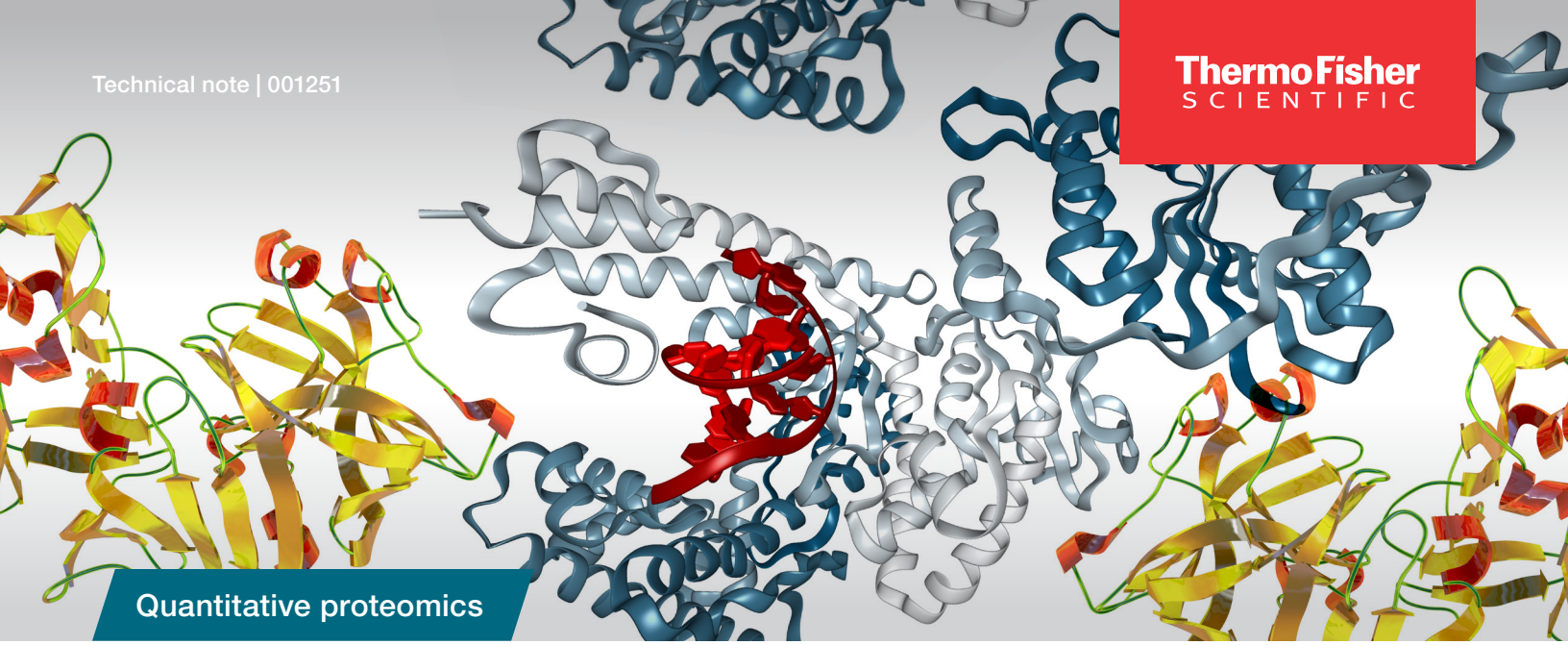
Technical note | 001251



Quantitative proteomics

High-throughput high-resolution data-independent acquisition workflow for accurate label-free quantitation

Authors

Julia Kraegenbring[1], Tabiwang Arrey[1], Jeff Op de Beeck[2], Maciej Bromirski[3], Alexander Harder[1]

[1]Thermo Fisher Scientific, Bremen, Germany; [2]Thermo Fisher Scientific, Zwijnaarde, Belgium; [3]Thermo Fisher Scientific, Warsaw, Poland

Keywords

Label-free quantitation (LFQ), high throughput, bottom-up proteomics, translational proteomics, cohort studies, Orbitrap Exploris 240 mass spectrometry, data-independent, Vanquish Neo UHPLC acquisition

Goal

To assess qualitative and quantitative performance of label-free quantitation (LFQ) with an optimized data-independent acquisition (DIA) method on a Thermo Scientific™ Orbitrap Exploris™ 240 mass spectrometer using a short 30 min gradient on a micropillar array-based µPAC Neo column for large-scale proteomics analysis.

Introduction

Quantitative proteomics is essential to understanding global protein expression and modifications that underlie the mechanisms of biological processes and disease states. Accurately quantifying abundances of proteins of interest in complex samples is a prerequisite for developing suitable predictive models and testing them against experimental data sets. Statistical significance is improved by increasing the sample set and ensuring reproducible results from run-to-run.

Traditional data-dependent analysis (DDA) approaches have been widely employed for LFQ experiments, but they suffer vastly from run-to-run inconsistencies due to intensity-based stochastic triggering of precursors, often leading to undersampling especially of low-abundant proteins. In contrast, DIA addresses missing value concerns by equally cycling through defined *m/z* windows along the survey scan range. The resulting spectral complexity of the mixed precursor fragmentation and mixed product ions is often addressed by employing large spectral libraries. However, recent developments in data analysis software, (e.g., using machine-learning approaches for *in silico* prediction of high-quality spectral libraries), have made library-free approaches a valid time- and cost-effective alternative.

The need for analyzing large numbers of samples, especially in clinical and biomarker discovery studies, makes DIA methods an obvious choice for ensuring reproducible quantitative analyses. In addition, in large cohort studies a robust setup (separation technology, column, and mass spectrometer) that can run stably for an extended period is a necessity. The Thermo Scientific™ Vanquish™ Neo UHPLC system delivers maximum performance for reproducible and versatile LC-MS experiments. New technologies in chromatographic separations also help achieve robustness: micropillar-array LC columns like the Thermo Scientific™ µPAC™ Neo HPLC column have been shown to deliver increased sensitivity and higher retention time stability[1] making them ideal candidates for setting up a robust and reproducible workflow.

Aside from robustness and reproducibility, confidence in identification and quantitation is an imperative for impactful proteomics research. Confidence in quantitative results is driven by ensuring accurate and precise measurements, but also by careful data analysis and rigid validation methods like controlled strict false discovery rates (FDR). Thermo Scientific™ Orbitrap™ technology fulfills the prerequisites for confident measurements by delivering both highly accurate mass as well as high resolution providing statistical sensitivity and specificity in an unparalleled manner. These two factors are key to reliable identifications and the ability to accurately and precisely detect and resolve ion species in complex DIA scans.

The challenge for applications in large-scale clinical studies is to achieve high levels of performance for all these factors: sensitivity, mass resolution, mass accuracy, quantitative accuracy, precision, reproducibility, and robustness—while increasing sample throughput in order to advance biological insight. In many cases, these performance factors are influenced by counteracting method parameters, and a fine balance must be found in an optimized setup to best fulfill all these criteria (Figure 1).

Here, we present a robust and reproducible workflow for accurately quantifying and identifying hundreds to thousands of proteins from complex sample mixtures with a high background of human peptides. The LC method and gradient were optimized to fully utilize the capabilities of micropillar array-based separation columns. Similarly, the mass spectrometric method was adopted to account for the short gradient lengths and high throughput without compromising identification rates at great quantitation accuracy and precision. The long-term robustness of the workflow is demonstrated by intermittent QC runs.

Workflow for high-throughput label-free quantitation and proteome

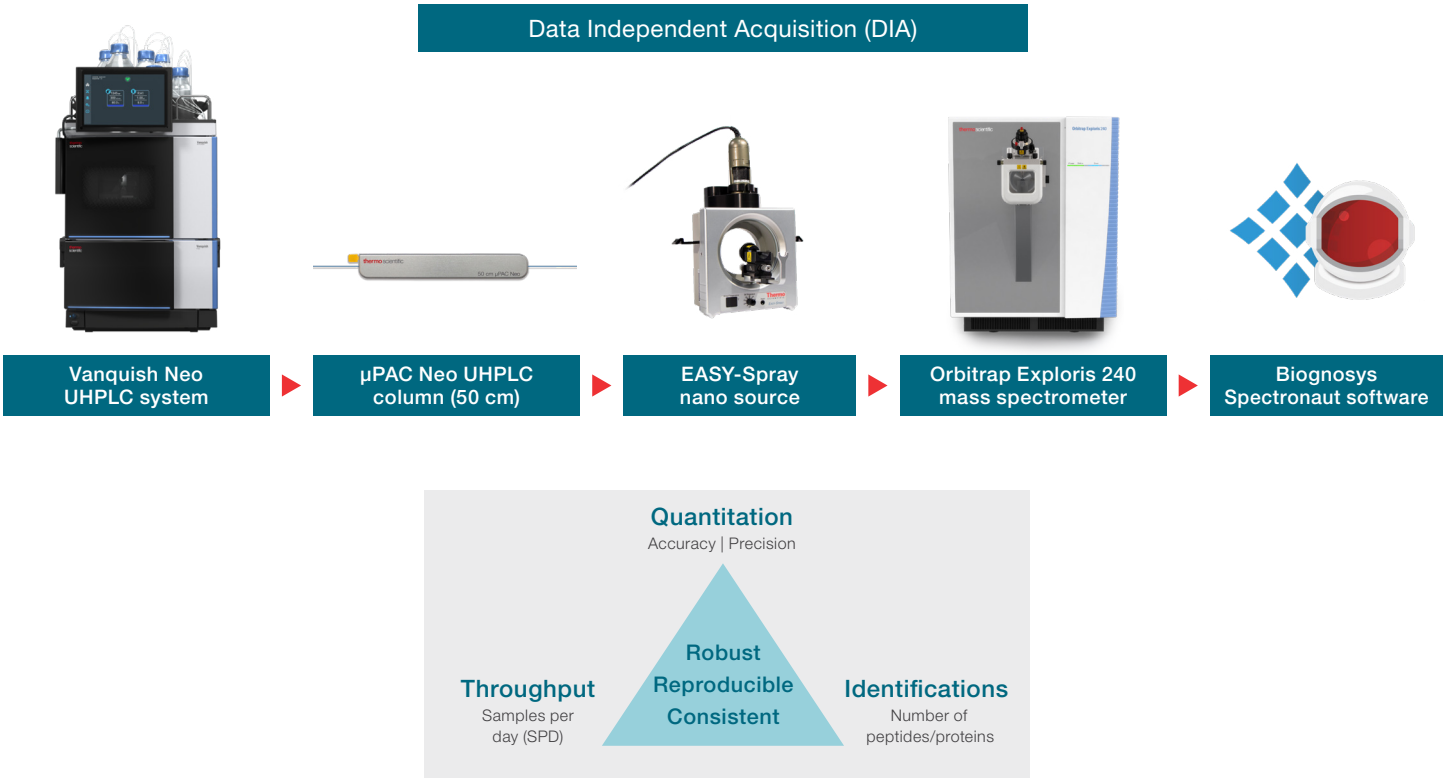


Figure 1. Graphical schematic of HR-DIA workflow for label-free quantitation of two- and three-proteome mixtures. The different components of the workflow are depicted on the top. The main goal of the setup is the quantitative performance at high sample throughput while delivering robust and reproducible results to make it a perfect fit for large scale clinical and biomarker discovery studies.

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 8829)
- Waters™ MassPREP *E. coli* Digest Standard (P/N 186003196)
- Promega™ Mass Spec-Compatible Yeast Protein Extracts (P/N V7461)

Fluidics and consumables used to set up the Vanquish Neo UHPLC system for direct injection are given in Table 1.

Sample preparation

Pierce HeLa Protein Digest Standard, Waters *E. coli* MassPREP Standard and Promega Mass Spec-Compatible Yeast Protein Digest were dissolved in 0.1% formic acid (FA) with 30 seconds of vortexing. For the two-proteome mix, *E. coli* peptide digest was added to a fixed amount of HeLa digest (800 ng) at amounts of 24 ng, 48 ng, 96 ng, and 192 ng, yielding *E. coli* peptide ratios of 1:2:4:8. For the three-proteome mix, *E. coli* peptide digest and yeast peptide digest was 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 1:4 and a yeast peptide ratio of 1:0.5 (Figure 2).

Materials and samples

Table 1. List of workflow components with part numbers.

Workflow components	Description
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, 100 µL (P/N 6252.1950), Vanquish Column Compartment N (P/N VN-C10-A-01)
Column	Thermo Scientific™ µPAC™ Neo HPLC column, 50 cm, 180 µm bed width, 16 µm pillar length (P/N COL-nano050NeoB)
Emitter	• PepSep Fused Silica emitter, 10 µm ID, 4 cm length, 150 µm OD (P/N 1893520) • PepSep EASY-Spray™ adapter (PSS2)
Source	Thermo Scientific™ EASY-Spray™ ion source (P/N ES082)
Mass spectrometer	Thermo Scientific™ Orbitrap Exploris™ 240 mass spectrometer (TNG Tune v4.2 SP1)
Data analysis software	• Spectronaut™ 16 software (Biognosys) • Spectronaut™ 17 software (preliminary test version) • DIA-NN universal software, v1.8.1, github release f34 from June 28, 2022

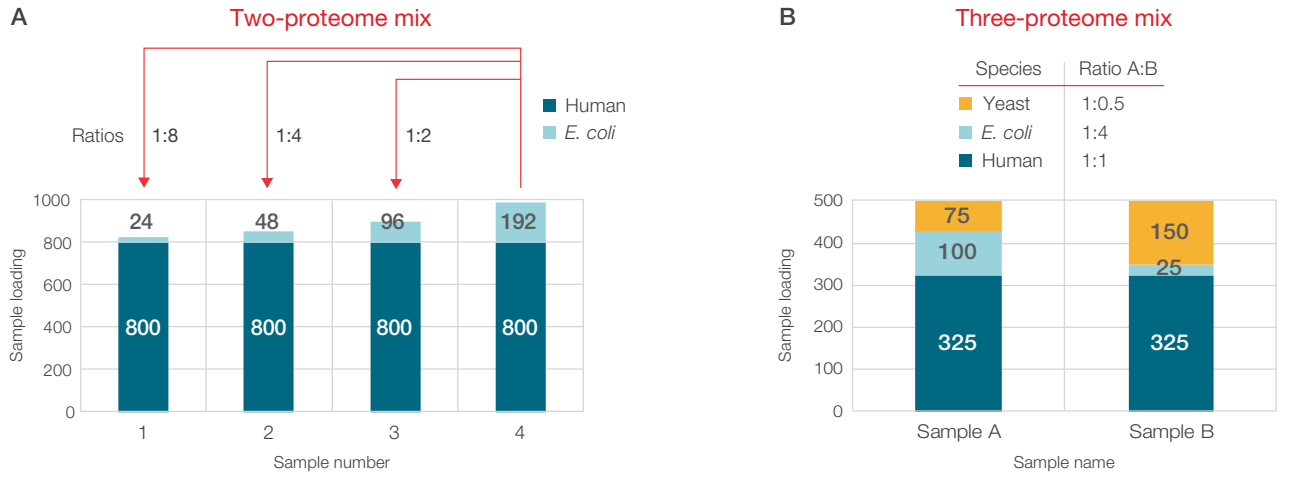


Figure 2. Experimental sample design of High-Resolution-DIA workflow for label-free quantitation. Two different sample sets were used for assessing quantitative performance of the workflow. (A) The two-proteome mix contains a high human background of 800 ng HeLa peptide digest, and low to medium amounts of spiked *E. coli* peptides in ratios of 1:2:4:8. (B) The three-proteome mix contains a medium human background of 325 ng HeLa peptides and 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.

Table 2: Summary of all LC and MS method parameters (parameters not mentioned in the table are set to default values).

Separation column specifications (set in the Vanquish Neo system)		
Inner diameter		75 µm
Length		50 cm
Maximum pressure		450 bar
Maximum flow		0.8 µL/min
Maximum temperature		60 °C
LC method (39 min method duration)		
Gradient	**Time**	**%B**
	0	4
	22.5	30
	30	45
	30.1–33.0	97.5
	33.1–39	4
LC parameters	Flow rate	350 nL/min
	Column temperature	50°C
	Fast loading/equilibration	PressureControl
	Pressure loading/equilibration	350 bar
	Equilibration factor	2.0
	Sampler temperature	7°C

MS method (39 min method duration, application mode "Peptide")		
Global parameters	Use ion source settings from Tune	True
	Expected peak width	10
	Advanced peak determination	True
	Default charge state	2
MS parameters	Resolution MS^1/DIA	60k/15k
	Scan range *(m/z)* MS^1	400–900
	RF lens (%) MS^1/DIA	70
	Normalized AGC target (%) MS^1/DIA	300/800
	Maximum injection time mode MS^1/DIA	Auto
	DIA precursor mass range/scan range *(m/z)*	400–900/ 145–1450
	Isolation window *(m/z)*	12
	Window overlap *(m/z)*	1
	Window placement optimization	On
	Normalized HCD Collision Energy (%)	30

LC-MS method

The two- and three-proteome mixtures were loaded on to a µPAC column and separated at 350 nL/min flow rate in direct injection mode using a Vanquish Neo UHPLC system over a 30 min LC gradient before being transferred into the Orbitrap Exploris 240 mass spectrometer (Figure 1).

Source parameters including spray voltage and ion transfer tube temperature are tunable parameters and must be optimized for the individual setup. The details of the LC gradient, LC parameters, and mass spectrometric method are reported in Table 2.

Data analysis and post-processing

Acquired data has been processed by Spectronaut 16 and 17 software using a directDIA approach. 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.

Resulting candidate tables and report files have been exported to .csv files. For analysis of the human protein group abundance ratios in the three-proteome mix, the default filter *"Absolute AVG Log2 Ratio"* in the candidate table was disabled. Results have been imported to a spreadsheet for data visualization and plotting (Origin 2022).

Results and discussion

Workflow performance in label-free quantitation

The goal of an LFQ approach is the reliable and reproducible quantitation of proteins or peptides at differential abundance levels. In this data set, two samples with different amounts of spiked microbial proteins have been used to mimic biological samples where proteins might be up- or downregulated under different conditions.

The HR-DIA workflow used for relative quantitation of *E.coli* proteins in a high amount of human peptides as background yields excellent quantitation accuracy across all three ratios with median values extremely close to the theoretical ratios as well as a narrow distribution of all data points around the median values indicating high quantitative accuracy and precision of the workflow (Figure 3A). In addition, peptide quantities can be repeatably determined from technical replicates leading to coefficients of variation (CV) well below 10% (Figure 3B).

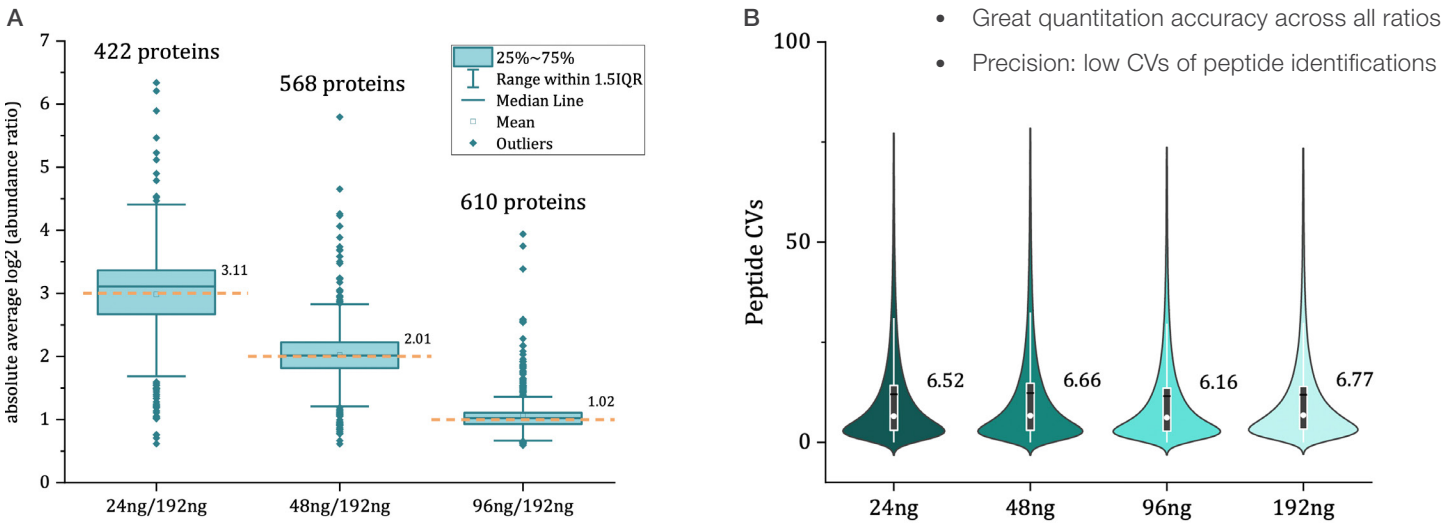


Figure 3. Determination of bacterial protein abundance ratios in two-proteome mixture and signal reproducibility within replicates. (A) Whisker boxplots of all three interrogated abundance ratios of *E. coli* proteins demonstrate excellent quantitation accuracy in being consistent with the theoretical ratios (orange dotted line). (B) Violin Plots of all four conditions reveal high precision of peptide quantities in technical replicates, with coefficients of variation well below 10%.

Comparable results could be obtained from the analysis of the three-proteome mixture: excellent quantitation accuracy was achieved for all three species with a narrow data distribution around the median abundance ratios, demonstrating highly precise measurement of protein quantities. Numbers of quantified proteins differ by species, with close to 3,000 human protein groups and nearly 500 *E. coli* protein groups (Figure 4A). Plotting corresponding peptide abundances of sample A to sample B yields good alignment of the three proteomes along their ratio-specific linear slope, as well as a good species differentiation over the whole abundance range (Figure 4B).

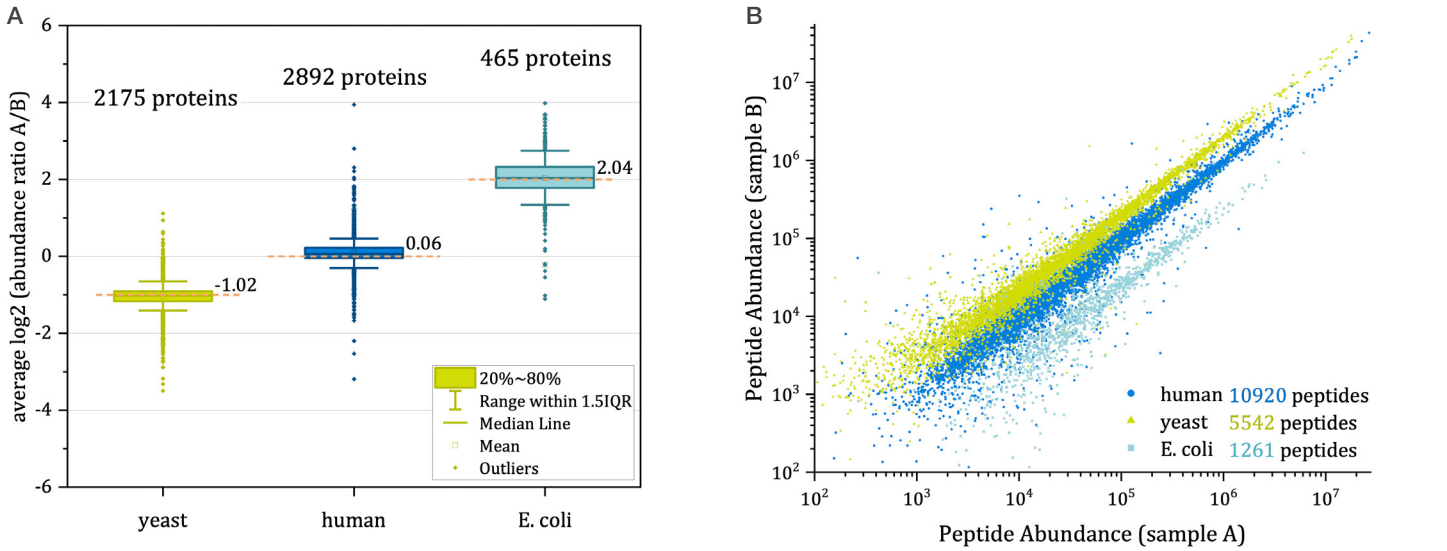


Figure 4. Determination of microbial and human protein abundance ratios in a three-proteome mixture. (A) Whisker boxplots of protein abundance ratios of all three species demonstrate excellent quantitation accuracy in being consistent with the theoretical ratios (orange dotted line). (B) Plotted peptide abundances of sample B to sample A align well on a ratio-species-specific linear slope and show good separation of the three proteomes down to the low abundance region.

Workflow robustness

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, meaning that the same column has been run throughout the experiments. System performance has been verified by intermittent defined QC runs of 200 ng HeLa from two different sample batches separated on a 67 min gradient in DDA mode. The number of identified peptides and protein groups was 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 5). This outstanding robustness of column, chromatography, and mass spectrometry is a valuable and important prerequisite for statistically meaningful, large-cohort studies.

DIA performance and data analysis

Being able to accurately quantify protein abundances and achieve a high proteome coverage are important prerequisites to investigate the underlying mechanisms and proteins of interest in biological processes. However, protein grouping and identification must be confident to avoid false positives as much as possible. The data analysis and post-processing of the analysis results are therefore essential for meaningful LFQ results.

Using the directDIA approach of Spectronaut 16 software results in high protein and precursor numbers in a 30-minute gradient with protein groups filtered to 1% FDR at the experimental level without sacrificing quantitative performance. Analyzing the same data set with an academic third-party software package, DIA-NN[2], delivers similar performance if comparable filter and analysis parameters are set (1% global protein group FDR, no heuristic, protein grouping by FASTA). In the recently released Spectronaut 17 software, the improved directDIA+ approach outperforms these results by as much as 30% on precursor and 5% on protein group level. These results clearly show that library-free approaches offer a valid alternative to time-consuming library building.

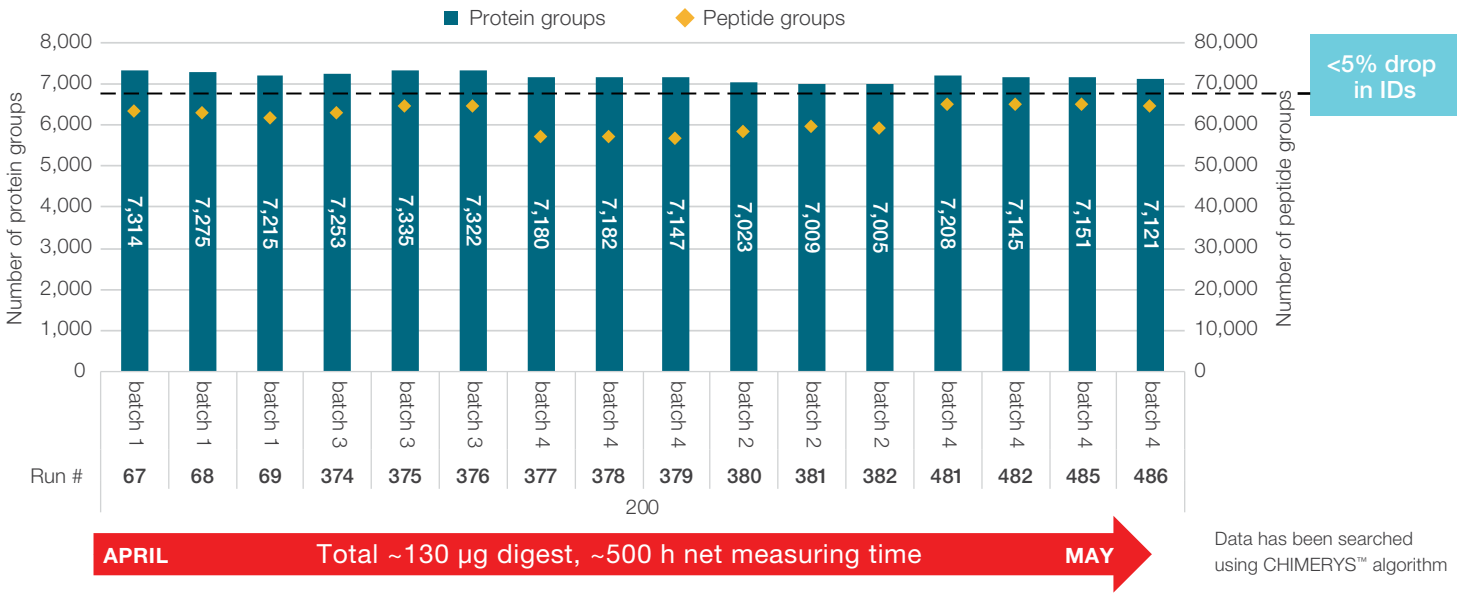


Figure 5. Workflow robustness demonstrated by reproducible peptide and protein identification over two months of operation time. Defined DDA QC experiments have been intermittently run throughout the whole experiment series. Protein and peptide IDs as obtained by a CHIMERYS search could be maintained over a time period of 2 months, including idle times, with a net measuring time of around 500 h and an estimated total of 130 µg digest.

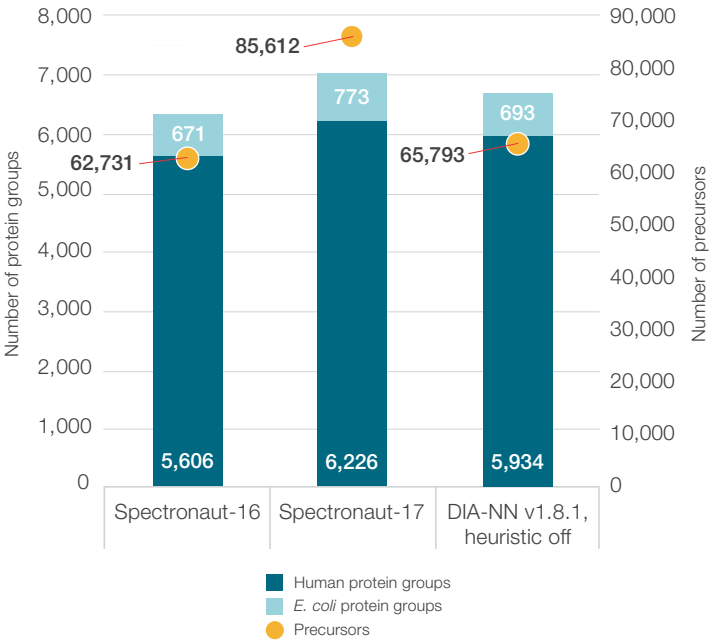


Figure 6. HR-DIA Workflow delivers confident proteome coverage utilizing next generation library-free analysis approaches. Bar graph comparison of protein group (human and *E. coli*) and precursor (total) numbers identified in 12 runs of two-proteome mix by use of three different software packages. Data analysis has been done by library-free analysis. All protein group results are filtered for 1% experiment-wide FDR.

Conclusion

The high-resolution DIA workflow for label-free quantitation setup on an Orbitrap Exploris 240 mass spectrometer coupled to a Vanquish Neo UHPLC system running with a µPAC Neo UHPLC column was shown to fulfill the following performance criteria:

- Excellent quantitation accuracy and precision for small amounts of bacterial and fungal proteomes from challenging sample matrices
- Sample throughput and quality of the obtained data while achieving high proteome coverage
- Long-term stability and reproducibility of the whole workflow which was demonstrated by excellent performance stability over a period of two months

These combined capabilities make the HR-DIA workflow suitable for the application in large-cohort studies.

References

1. Stadlmann J, Hudecz O, Krššáková G, et al. Improved Sensitivity in Low-Input Proteomics Using Micropillar Array-Based Chromatography. Anal Chem 91(22), 14203–14207 (2019). https://pubs.acs.org/doi/10.1021/acs.analchem.9b02899.
2. Demichev, V, Messner, CB, Vernardis, SI, et al. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods 17, 41–44 (2020). https://doi.org/10.1038/s41592-019-0638-x.

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