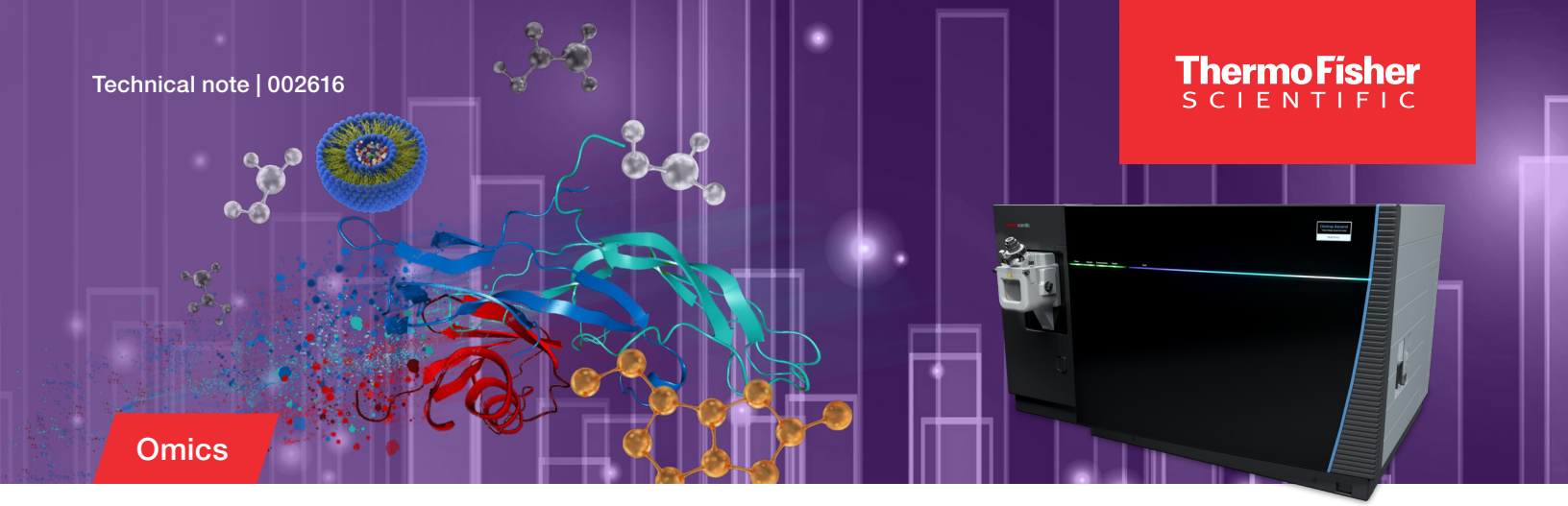
Technical note | 002616



Omics

High-throughput high-resolution data-independent acquisition workflow on an Orbitrap Ascend MultiOmics Tribrid mass spectrometer for accurate label-free quantitation

Authors

Kevin Yang[1], Julia Kraegenbring[2], Julian Saba[3], Sally Webb[1], and Amirmansoor Hakimi[1]

[1]Thermo Fisher Scientific, San Jose, California, USA

[2]Thermo Fisher Scientific, Bremen, Germany

[3]Thermo Fisher Scientific, Mississauga, ON, Canada

Keywords

Data-independent acquisition, DIA, Velocity DIA, Orbitrap Ascend MultiOmics mass spectrometer, µPAC Neo HPLC column, Vanquish Neo UHPLC, FAIMS, CHIMERYS, Spectronaut software, Proteome Discoverer software, Tribrid

Goal

To develop and assess qualitative and quantitative performance of label-free quantitation (LFQ) with an optimized data-independent acquisition (DIA) method on a Thermo Scientific™ Orbitrap™ Ascend MultiOmics Tribrid™ mass spectrometer using a long (60 min active gradient), short (30 min active gradient), and ultra-high throughput (9 min active gradient) gradient for large-scale proteomics analysis.

Introduction

Quantitative proteomics is an essential tool for understanding global protein expression and the mechanisms of biological processes and disease states. Accurately quantifying the abundances of proteins of interest in complex samples is a prerequisite for developing suitable statistical models to gain biological insights from experimental data sets. Statistical significance is improved by decreasing variability in measurements and/or increasing the sample set. However, increasing throughput means decreasing acquisition time, which often comes at a cost to measurement quality. Therefore, acquisition methods must be extensively optimized and validated to ensure that the data will produce meaningful biological insights. Traditional data-dependent analysis (DDA) approaches have been widely employed for LFQ experiments, but they suffer from run-to-run inconsistencies due to intensity-based stochastic triggering of precursors, often leading to under sampling especially of low-abundant proteins. Missing values become more likely as sample size increases, and so DIA has emerged as a popular technique for large scale quantitative analyses.

In contrast to DDA, DIA addresses missing value concerns by equally cycling through defined *m/z* windows along the survey scan range. The inherent tradeoff between measurement selectivity (isolation window size), frequency (cycle time), scope (mass range), and sensitivity (ion accumulation time) necessitates careful method optimization, and a suboptimal combination of parameters can have disastrous consequences. Additionally, 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 LFQ DIA-based workflows an obvious choice for ensuring quantitative analyses. A suitable analytical workflow addresses the need for reproducible sample preparation, robust separations, high-quality quantitative measurements, and reliable data analysis.

To this end, we present the Velocity LFQ DIA workflow, an end-to-end solution for quantitative proteomics (Figure 1 and Table 1). Briefly, sample preparation can be automated to increase throughput and decrease technical variability with the Thermo Scientific™ AccelerOme™ sample preparation platform. Robust LC-MS/MS analysis is performed on the Thermo Scientific™ Vanquish™ Neo UHPLC system with a Thermo Scientific™ µPAC™ Neo HPLC column coupled to an Orbitrap Ascend MultiOmics MS. The Vanquish Neo UHPLC system combined with the µPAC Neo HPLC column has been shown to increase sensitivity and retention time stability.[1] Additionally, the Orbitrap Ascend MultiOmics MS uses Orbitrap technology to deliver high resolution measurements with low background and high sensitivity. We compare multiple different analysis methods, including Thermo Scientific™ Proteome Discoverer™ software (3.1.0.638) with CHIMERYS™ intelligent search algorithm by MSAID. Overall, we demonstrate that the Velocity LFQ DIA workflow on the Orbitrap Ascend MultiOmics MS enables a high level of quantitative performance across short, medium, and long gradients to meet a variety of experimental needs.

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 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 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 not only by ensuring accurate and precise measurements, but also by careful data analysis and rigorous 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 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, and precision—while increasing sample throughput to advance biological insights. 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. Here, we present a robust and reproducible workflow on the Orbitrap Ascend MultiOmics MS for accurately quantifying and identifying hundreds to thousands of proteins from single cell line to complex sample mixtures with a high background of human peptides (Figure 1 and Table 1). The Velocity LFQ DIA workflow could be coupled with the optional AccelerOme automated sample preparation platform to improve throughput and minimize variations caused by manual sample handling (Figure 1). 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 long (60 min active gradient), short (30 min active gradient), and ultra-high throughput (9 min active gradient) gradient lengths without compromising identification rates at great quantitation accuracy and precision.

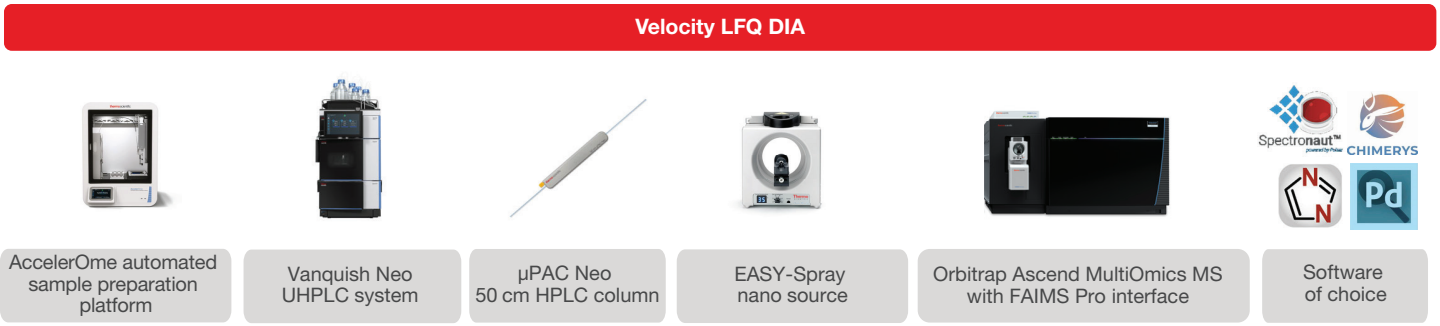


Figure 1. Graphical schematic of the Velocity LFQ DIA workflow for label-free quantitation on the Orbitrap Ascend MultiOmics MS together with the optional AccelerOme automated sample preparation platform

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	µPAC Neo HPLC column, 50 cm, 180 µm bed width, 16 µm pillar length (P/N COL-NANO050NEOB)
Emitter	• Thermo Scientific™ EASY-Spray™ Nano & Capillary adapter (P/N ES993) • EvoSep Thermo Scientific™ EASY-Spray™ adapter (P/N EV1072) • EvoSep Fused Silica Emitter (P/N EV1111)
Source	Thermo Scientific™ EASY-Spray™ ion source (P/N ES082)
Mass spectrometer	Orbitrap Ascend MultiOmics MS
Data analysis software	• Spectronaut™ 18 software (Biognosys) • DIA-NN software (v1.8.1) • Proteome Discoverer software using CHIMERYS (v3.1.0.638)

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)
- Evosep Thermo Scientific™ EASY-Spray™ Adapter (P/N EV1072)
- Evosep fused silica emitters 10 µm (P/N EV1111)
- 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 s of vortexing. 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 1:4 and a yeast peptide ratio of 1:0.5 (Figure 2A and 2B).

LC-MS method

HeLa digest and three-proteome mixtures were loaded onto a 50 cm µPAC Neo HPLC column and separated at a 350 nL/min flow rate in direct injection mode using a Vanquish Neo UHPLC system over 9 min, 30 min, and 60 min LC gradients, respectively, before being transferred into the Orbitrap Ascend MultiOmics MS (Figure 2C).

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.

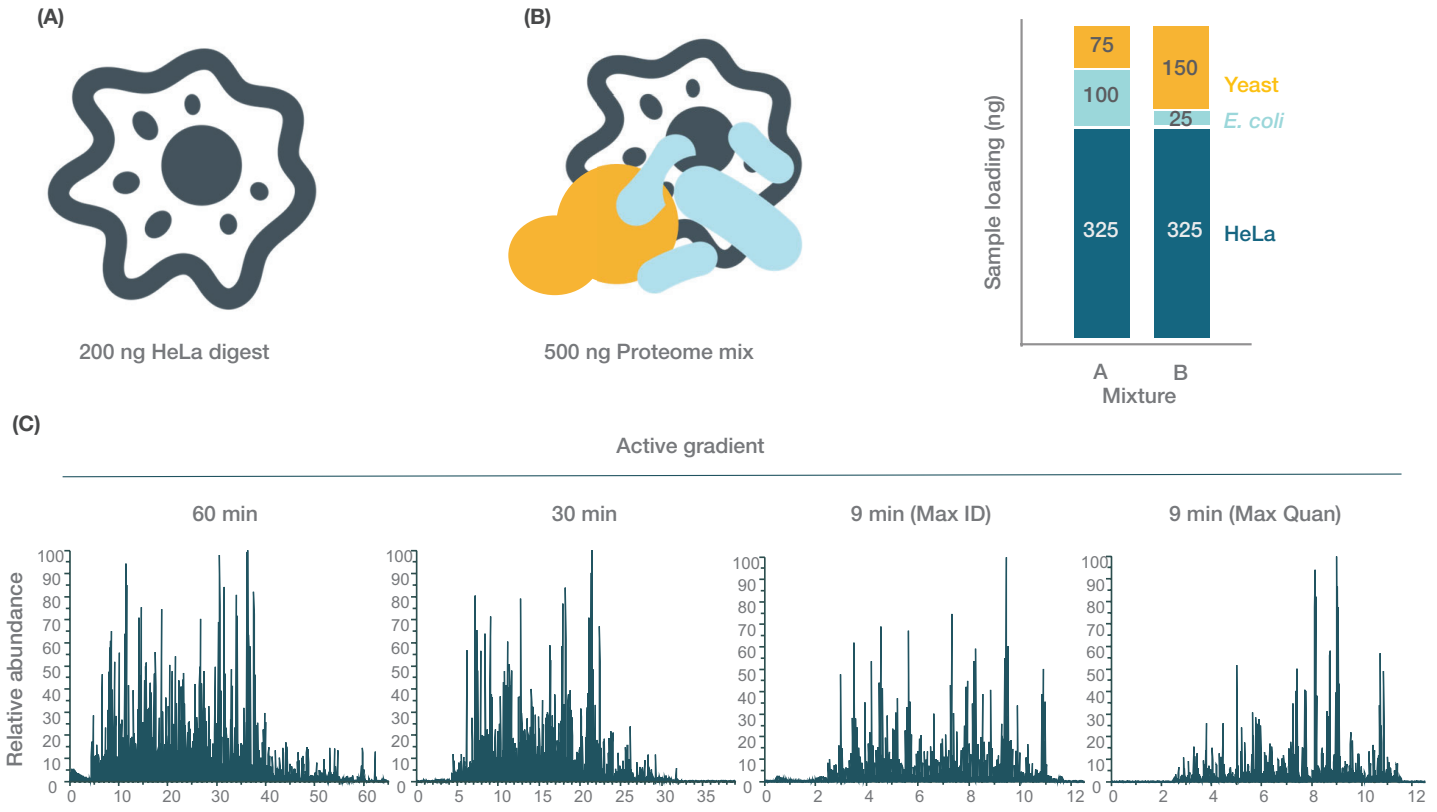


Figure 2. Experimental sample and active gradient design of the Velocity LFQ DIA workflow for label-free quantitation. Two different sample sets were used for assessing the identification and quantitative performance of the Velocity LFQ DIA workflow on the Orbitrap Ascend MultiOmics MS. (A) 200 ng HeLa digest to access quantitation precision and proteome coverage of human samples. (B) The three-proteome mix contains a medium human background of 325 ng HeLa peptides together with yeast and *E. coli* peptides digested in ratios of 1:0.5 and 1:4, respectively. The mixtures have been chosen because they closely mimic biological samples with larger or smaller protein expression changes. (C) Three different active gradient lengths, including 60 min, 30 min, and 9 min, were selected to develop the Velocity LFQ DIA workflow for different throughput needs. For the ultra-high throughput 9 min gradient setup, two methods were developed to obtain deep proteome coverage (Max ID) or excellent quantitation performance (Max Quan).

Data analysis and post-processing

Acquired data was processed by Spectronaut 18 software using a directDIA approach, DIA-NN (v1.8.1) or Proteome Discoverer software (v3.1.0.618) with CHIMERYS intelligent search algorithm by MSAID. For Spectronaut software, default settings were used 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. FASTA files for human, yeast, and *E. coli* were downloaded from UniProt™. 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.

For CHIMERYS in Proteome Discoverer software, default settings were used for both the processing and consensus workflow. All the PSM, peptides, and proteins were filtered at 1% FDR. For the three-proteome mix, species map and species names were set as True in the "Protein Marker" node. For DIA-NN software, default settings were used for either direct DIA or library search.

The resulting candidate tables and report files for data searched with either Spectronaut software and CHIMERYS in Proteome Discoverer software were exported to .csv or .tsv files. The ensuing tables were imported to Python™ or a spreadsheet for downstream data analysis and visualization.

Results and discussion

The Velocity LFQ DIA workflow was initially developed on the Thermo Scientific™ Orbitrap Exploris™ 240 mass spectrometer.[2] Here we expand this workflow to the Orbitrap Ascend MultiOmics MS (Figure 1). First, we characterized the benefits of Thermo Scientific™ FAIMS Pro™ interface in protein identification. Our results showed that the FAIMS Pro interface decreased peptide coverage but increased protein group identifications by an average of 5% (Figure 3). Along with the 5% increase in protein identifications, the FAIMS Pro interface helps to increase the robustness of the mass spectrometer by reducing interfering background ions and adds a gas-phase fractionation step, which allows for otherwise unidentifiable peptides to reach above the S/N threshold for identification. Consequently, the following experiments were all coupled with the FAIMS Pro interface.

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	0.7 µL/min
Maximum temperature	60 °C

LC method (60 min method duration)		
Gradient	Time (min)	%B
	0	4
	45	30
	60	45
	60.0–64.9	97.5
	64.9–65	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

LC method (30 min method duration)		
Gradient	Time (min)	%B
	0	4
	22.5	30
	30	45
	30.1–33.0	97.5
	33.0–33.1	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

LC method (9 min method duration)		
Gradient	Time (min)	%B
	0	4
	0–7.5	30
	7.5–9.0	45
	9.0–12	97.5
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 for 30 and 60 min (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/MS^2	60,000/15,000
	Scan range (*m/z*) MS^1	400–900
	Scan range (*m/z*) MS^2	145–1,450
	Normalized AGC target (%) MS^1/MS^2	300/800
	Maximum injection time mode MS^1/MS^2	Auto
	Isolation window (*m/z*)	12
	Window overlap (*m/z*)	1
	Window placement optimization	On
	Normalized HCD Collision Energy (%)	30

MS method for 9 min MaxID (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/MS^2	60,000/15,000
	Scan range (*m/z*) MS^1	400–800
	Scan range (*m/z*) MS^2	145–1,450
	Normalized AGC target (%) MS^1/MS^2	300/800
	Maximum injection time mode MS^1/MS^2	Auto
	Isolation window (*m/z*)	8
	Window overlap (*m/z*)	1
	Window placement optimization	On
	Normalized HCD Collision Energy (%)	30

MS method for 9 min MaxQuan (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/MS^2	60,000/30,000
	Scan range (*m/z*) MS^1	650–770
	Scan range (*m/z*) MS^2	145–1450
	Normalized AGC target (%) MS^1/MS^2	300/800
	Maximum injection time mode MS^1/MS^2	Auto
	Isolation window (*m/z*)	12
	Window overlap (*m/z*)	1
	Window placement optimization	On
	Normalized HCD Collision Energy (%)	30

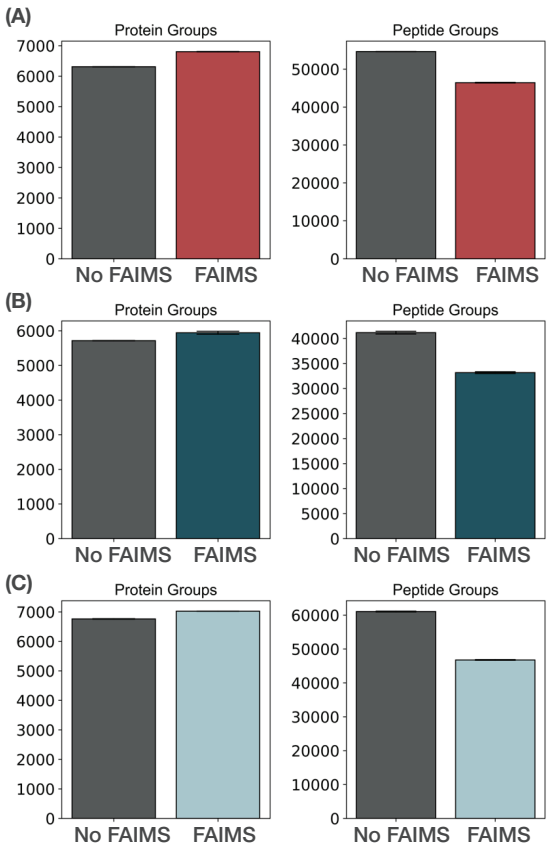


Figure 3. Evaluation of the FAIMS Pro interface in the Velocity LFQ DIA workflow. Box plots depicting the number of proteins and peptides identified from 200 ng of HeLa digest in the 30 min method with the FAIMS Pro interface at a CV of -45 V. Data were analyzed with (A) Spectronaut software, (B) Proteome Discoverer software using CHIMERYS, or (C) DIA-NN.

The initial set of experiments consisted of maximizing identification and quantitative performance for various run times of 9, 30, and 60 minutes for a HeLa digest standard (Figure 2C). With the 30 min active gradient, 7,000+ proteins and 47,000+ peptides were identified, along with a protein group CV of approximately 5% (Figure 4), suggesting that a 30 min active gradient method enables relatively high throughput while maximizing identification and quantitative performance. We extended this workflow to a 60 min active gradient and successfully identified close to 7,800 proteins and >76,000 peptides, highlighting that deeper proteome coverage can be achieved in the Velocity LFQ DIA workflow by using a longer gradient. We also evaluated a 9 min active gradient to enable ultra-high throughput. At such high throughput, there is a tradeoff between depth of coverage and quantitative measurement quality. Therefore, we developed two methods to meet different experimental needs: one for maximizing identifications (Max ID) and the other for maximizing quantitative (Max Quan) performance. In the Max ID method, we identified 5,400+ proteins in a 9 min active gradient, highlighting the sensitivity and scanning speed of the Orbitrap mass spectrometer in a high throughput setup. In contrast, the Max Quan method, although it compromised proteome coverage, provided better quantitation performance, as evidenced by ~6% protein group CV in such an ultra-high throughput setup (Figure 4).

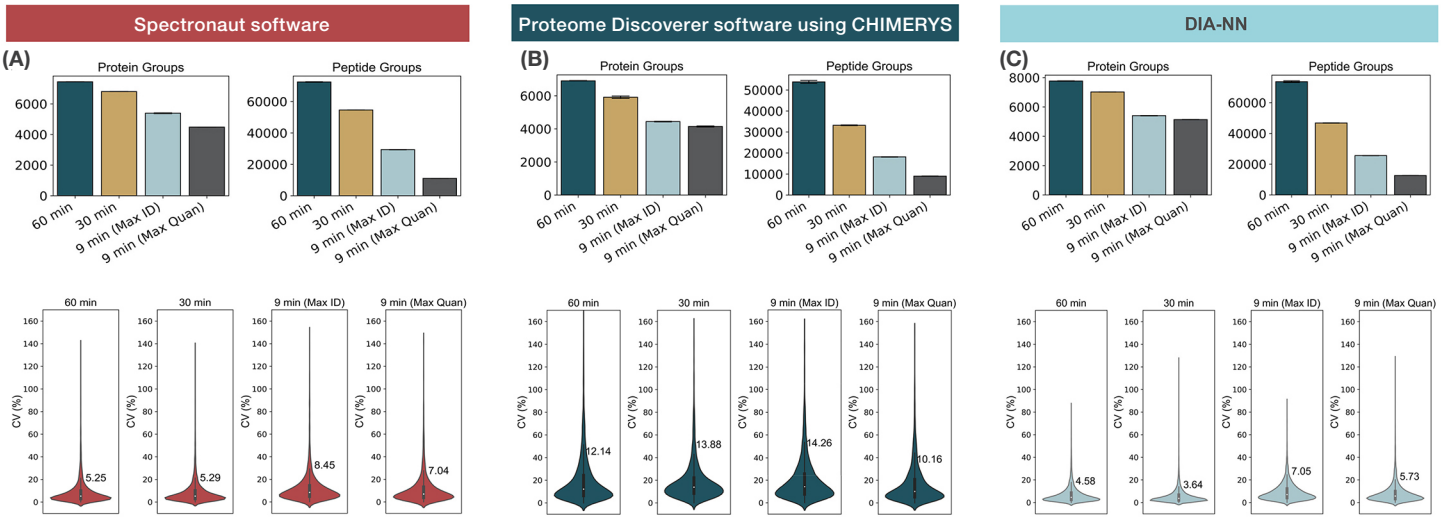


Figure 4. Determination of proteome coverage and quantitation precision of the human proteome. (A–C) Bar charts showing the number of proteins and peptides identified from 200 ng HeLa digests with different active gradients evaluated in the present study. (D–F) Violin plots of all four tested methods reveal high precision of protein quantities in technical replicates. Data was analyzed with Spectronaut software (left panel), Proteome Discoverer software using CHIMERYS (middle panel), or DIA-NN (right panel).

To understand if a higher load benefits the Velocity LFQ DIA workflow in maximizing identification, 500 ng of HeLa digest was analyzed with the 60 min active gradient experiments. The results indicated >8,100 protein groups were identified, with CV being ~4% (Figure 5). Thus, the ability of the Orbitrap Ascend MultiOmics MS to accommodate sample loads from picograms to micrograms can be leveraged to further increase the depth and quality of measurements.

High accuracy and precision of quantitation

In addition to protein identification, quantitative data is necessary to study biomarkers and get biological insights. The quantitative data must be both precise and accurate to reflect subtle changes in biological systems. Due to multiple statistical testing, using longer protein ID lists without ensuring quantitative performance of these IDs not only increases the risks of obtaining false positives, but also false negatives. Distorted quantitation will impede biomedical research and cause significant waste of resources down the road. Our results from the HeLa digest indicated a protein CV of approximately 4–5%, suggesting excellent quantitative precision.

However, CVs can be a misleading metric if the corresponding signals are below the limit of quantitation. To test the quantitative accuracy of the Velocity LFQ DIA workflow, we created a three-proteome mix to mimic biological samples where proteins might be up- or downregulated under different conditions. Here, *E. coli* and yeast proteomes were spiked into a high human background at different ratios and the quantitative accuracy of the resulting measurements was assessed for each run time (9, 30, and 60 min). The Velocity LFQ DIA workflow yielded excellent quantitative accuracy across a wide dynamic range 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 6).

Additionally, the three-proteome mix experiment further highlights the proteome depth that can be achieved with an Orbitrap mass spectrometer. In the 60 min active gradient experiment, >11,000 protein groups were identified. The numbers of quantified proteins differ by species, with close to 7,500 human protein groups, approximately 3,500 yeast protein groups, and nearly 890 *E. coli* protein groups (Figure 6). The data demonstrated deep proteome coverage and excellent quantitation accuracy afforded by the Orbitrap Ascend MultiOmics MS, both of which are critical for enabling biological discoveries.

Evaluating data processing software in DIA analysis

All protein measurements reported in these large-scale analyses are a product of the data produced by the workflow and the software used to identify and quantify proteins in the data. There are a large number of DIA software options available, all of which use different algorithms and can produce vastly different results. Therefore, it is important to benchmark software performance on your dataset to truly understand data quality.

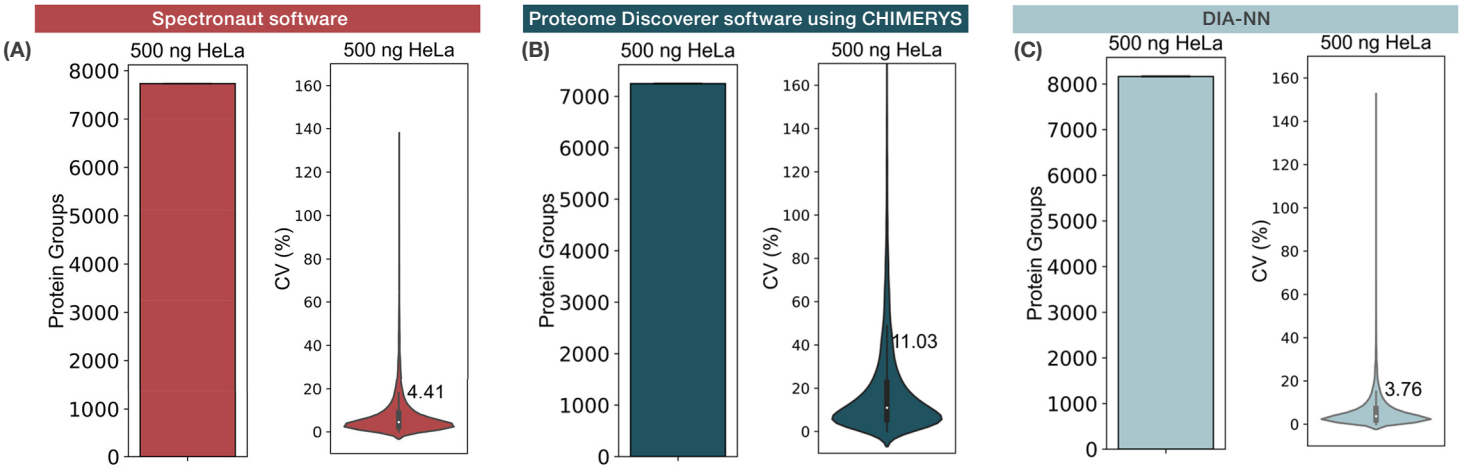


Figure 5. A long gradient coupled with a higher sample load affords the superior performance in the Velocity LFQ DIA workflow. Bar charts of protein groups identified from 500 ng of HeLa digest in the 60 min method and analyzed with different software demonstrate even deeper proteome coverage and violin plots reveal high precision of protein quantities in technical replicates. Data analyzed with Spectronaut software (A), CHIMERYS on Proteome Discoverer software (B), and DIA-NN (C) are shown.

To assess data analysis software, we compared the commercially available Spectronaut 18 software as a benchmark to also evaluate Proteome Discoverer software with CHIMERYS (v3.1.0.638) and DIA-NN (v1.8.1),[3] an academic software package (Figures 4-6). For the three-proteome mix, 8,200+ to 10,600+ protein groups were identified in the 30 min experiment across different software. For the 60 min gradient, the number of protein groups spans from 9,600+ to 1,1000+ (Figure 6). Thus, software should be taken into account when assessing the performance of mass spectrometers.

Library-based search in DIA may enhance proteome coverage and improve peptide identification in certain scenarios. To evaluate if Velocity LFQ DIA data can benefit from library search, we generated a general library by using Pulsar on Spectronaut 18 software from single shots of 12 cell lines acquired on a Thermo Scientific™ Orbitrap™ Astral™ mass spectrometer. The raw files were re-processed with library-based search on DIA-NN (v1.8.1). In line with the benefit of library-based search, we observed a 3–10% increase in protein identification with more benefits for shorter gradients (Figure 7).

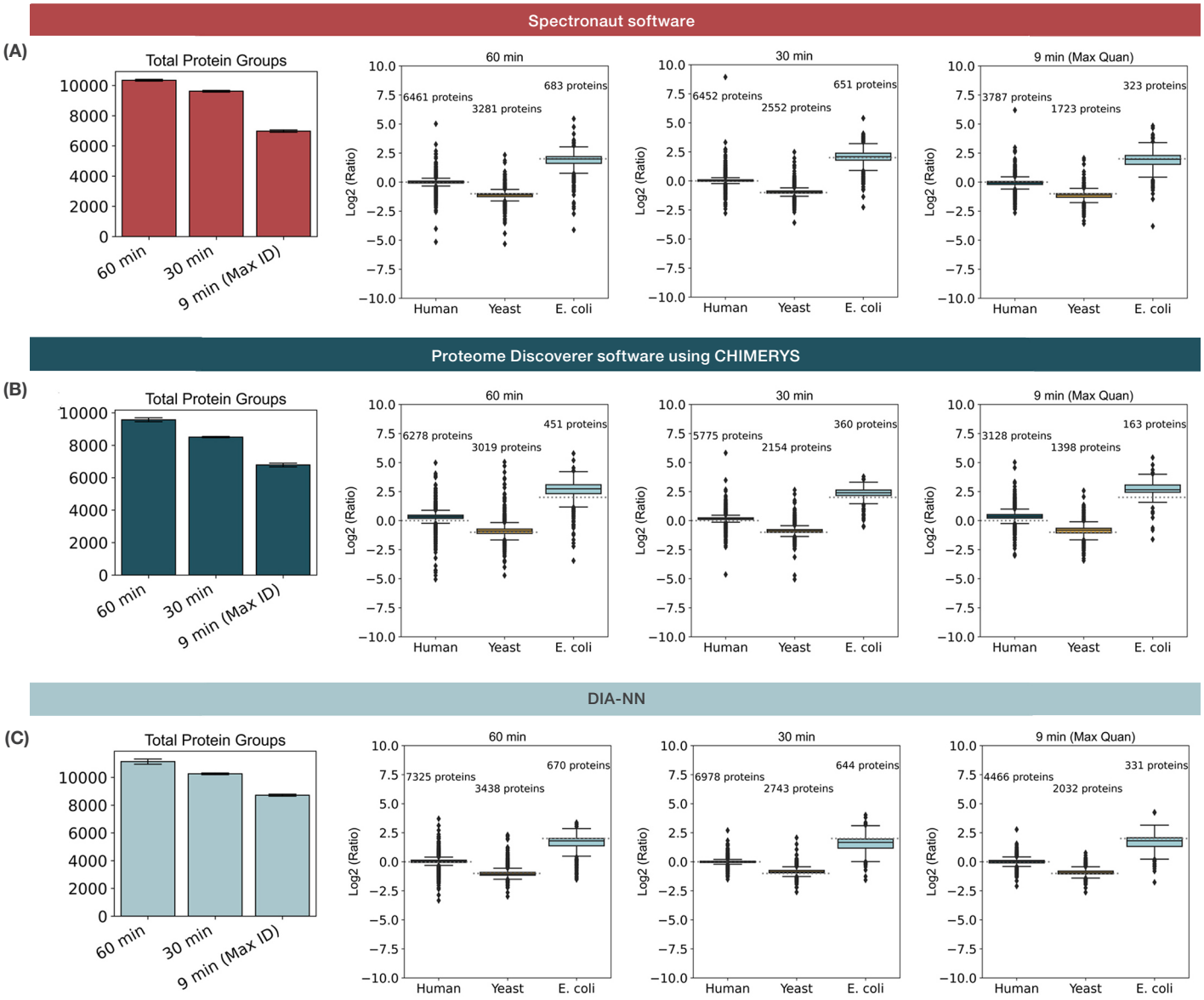


Figure 6. Determination of microbial and human protein abundance ratios in a three-proteome mixture. Bar plots showing the total protein groups identified from three-proteome mix in different active gradients and whisker box plots of protein abundance ratios of all three species demonstrate excellent quantitation accuracy by being consistent with the theoretical ratios (gray dotted line). Data was analyzed with Spectronaut software (A), Proteome Discoverer software using CHIMERYS (B), and DIA-NN (C).

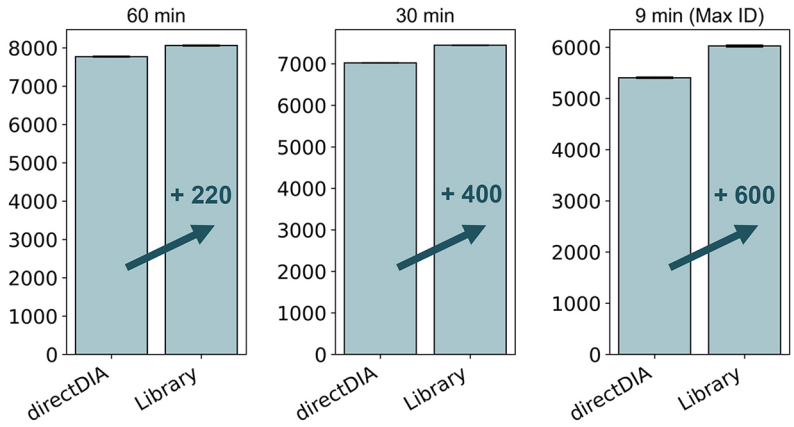


Figure 7. Library search improved proteome coverage. Bar charts showing the number of protein groups gained through library search by using DIA-NN (v1.8.1).

Conclusion

The high-resolution DIA workflow for LFQ setup on an Orbitrap Ascend MultiOmics MS coupled to a Vanquish Neo UHPLC system running with a 50 cm µPAC Neo HPLC column was shown to fulfill the following performance criteria:

- Deep (7,000+ protein groups) proteome coverage from 200 ng of HeLa standard
- Excellent quantitative 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

References

1. Stadlmann, J.; Hudecz, O.; Krššáková, G. et al. Improved Sensitivity in Low-Input Proteomics Using Micropillar Array-Based Chromatography. *Anal. Chem.* **2019,** *91*(22), 14203–14207. https://pubs.acs.org/doi/10.1021/acs.analchem.9b02899
2. Thermo Scientific Technical Note 1251 (TN1251): High-throughput high-resolution data-independent acquisition workflow for accurate label-free quantitation. https://assets.thermofisher.com/TFS-Assets/CMD/Technical-Notes/tn-001251-ms-lfq-dia-orbitrap-tn001251-en.pdf
3. Demichev, V.; Messner, C.B.; Vernardis, S.I. et al. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **2020,** *1,* 41–44. https://doi.org/10.1038/s41592-019-0638-x

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