P-I-0139

Assessment of high-resolution DIA methods and short gradients on a high-throughput µPAC column for maximum proteome coverage and quantitative performance

Julia Kraegenbring[1]; Dominic G. Hoch[2]; Jeff Op De Beeck[3]; Riccardo Stucchi[2]; Maciej Bromirski[4]; Robert Van Ling[3]; Heiner Koch[1]
[1]Thermo Fisher Scientific, Bremen, Germany; [2]Thermo Fisher Scientific, Reinach, Switzerland; [3]Thermo Fisher Scientific, Ghent, Belgium, [4]Thermo Fisher Scientific, Warsaw, Poland

Abstract

Purpose: Developing a workflow that is suitable for large study cohorts with high sample throughput and robustness and reproducibility over extended periods maintaining quantitative performance and deep proteome coverage.

Methods: The unique architecture of Thermo Scientific™ µPAC™ Neo columns with a 5.5cm flow path coupled to an easy-to-use Thermo Scientific™ Vanquish™ Neo LC system and a reliable Thermo Scientific™ Orbitrap Exploris™ 480 mass spectrometer operated under maximized duty cycle with optimized high-throughput LC methods.

Results: The presented workflow delivers robust results of >800 LC injections, with proteome coverage of more than 5,000 protein groups at 96samples per day (SPD) for 200 ng HeLa.

Introduction

Understanding biological processes on a global proteome-wide scale is crucial in scientific research. Achieving statistically significant quantitative data often involves studying large cohorts comprising hundreds of biological samples and replicates. An effective workflow must be capable of processing, measuring, and analyzing a high volume of samples while ensuring data reproducibility and consistent performance throughout the study. The primary aim of this study was to introduce a robust workflow for high-throughput label-free quantification using a Vanquish Neo LC system equipped with a 5.5 cm µPAC Neo High Throughput column and an Orbitrap Exploris 480 mass spectrometer.

Materials and methods

Sample preparation

For sample preparation, 200 µL of LC-MS grade water with 0.1% FA were added to 20 µg Pierce HeLa protein digest standard. The resulting concentration was 100 ng/µL of peptide digest and different peptide amounts were loaded onto the column by using appropriate injection volumes.

Test method(s)

Three different gradients were optimized to allow for high sample throughputs of up to 240 samples per day (SPD). The Vanquish Neo LC system was configured in both direct-injection and trap-&-elute setup. The highest throughputs were achieved with direct injection and the effect on identification numbers was assessed in this data set. Two different sites with different operators were participating in the study to ensure the adoptability of the workflow in different labs.

Table 1. Three Different Gradients for 100 SPD up to 200 SPD on µPAC Neo 5.5cm Columns Using Optimized LC Settings on the Vanquish Neo System and Mass Spectrometric Method on Orbitrap Exploris 480.

200 to 240 samples per day			160 to 160 samples per day			96 to 100 samples per day		
Time (min)	*Flow rate (µL/min)*	*%B*	*Time (min)*	*Flow rate (µL/min)*	*%B*	*Time (min)*	*Flow rate (µL/min)*	*%B*
0	3.0	6	0	2.5	1	0	2.5	1
0.1	3.0	10	0.1	2.5	8	0.1	2.5	4
2.1	3.0	22.5	0.7	2.5	12.5	0.7	2.5	7.5
3.4	3.0	45	0.8	1.25	12.6	0.8	1.0	7.6
3.5-4	3.0	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

Mass Spectrometric Method for All Gradients			
Global Parameters		***MS parameters***	
Spray voltage	1,900V	Resolution MS1/DIA	30k/15k
Ion transfer tube	305° C	Scan range MS1 (m/z)	525-800 (825*)
Expected LC peak width	10 s	RF lens (%)	50
APD	True	AGC target MS1/DIA (%)	300/800
Default charge state	2	Isolation width (240/200/96 SPD)	16Th/10Th/8Th**
		Window overlap	1 Th
		Window placement optimization	On
		NCE (%)	28

* Higher column oven temperature for shortest gradient ensures short loading and equilibration times
** these are the windows for optimal quantitation per gradient length; other data is shown for proof-of-principle purposes

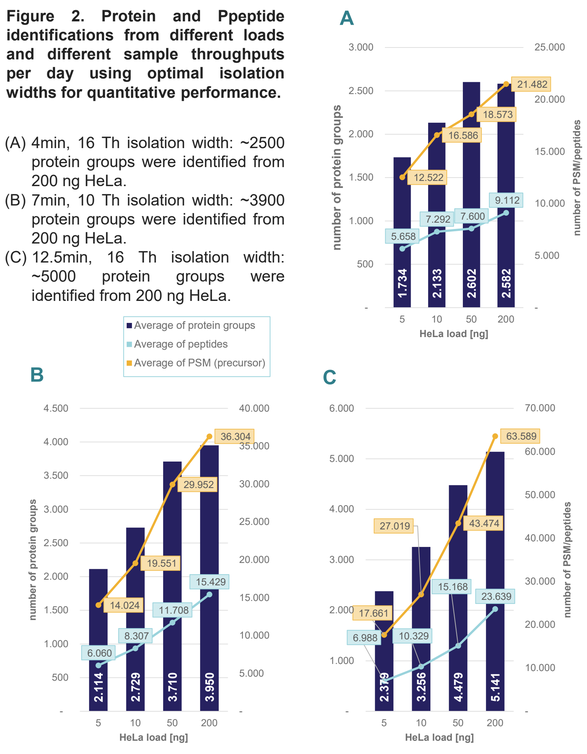
The settings of the high-throughput µPAC Neo column were set in the LC as follows: ID – 75µm, length – 50cm, max. pressure – 450 bar, max. flow – 5µL/min, max. temperature 60° C. The LC parameters in all methods were set as follows: column temperature 60° C for the shortest gradient, 50° C for the other gradients; loading was controlled solely by pressure; loading volume was set to fixed 1µL; equilibration factor was set to 0; sampler temperature was set to 7° C.The details on gradient and MS parameters are shown in Table 1.

Data analysis

All data were analyzed with Thermo Scientific™ Proteome Discoverer™ 1 software using Chimerys, an intelligent search algorithm by MSAID, for spectrum-centric search. Peptides and protein groups were controlled for false-discovery rate 1%.

Figure 2. Protein and Ppeptide identifications from different loads and different sample throughputs per day using optimal isolation widths for quantitative performance.

(A) 4min, 16 Th isolation width: ~2500 protein groups were identified from 200 ng HeLa.
(B) 7min, 10 Th isolation width: ~3900 protein groups were identified from 200 ng HeLa.
(C) 12.5min, 16 Th isolation width: ~5000 protein groups were identified from 200 ng HeLa.



Results

Proteome coverage for different throughput strategies and sample loads

The high-throughput workflow presented here delivers exceptional proteome coverage of close to 2,600 protein groups in only 4 min run time from 200 ng HeLa (see Figure 2). Even for loads as low as 5 ng HeLa, 1,700 protein groups and 5,600 peptides can be identified in the shortest runs, and 2,400 protein groups and 7,000 peptides from the longer 12.5 min runs. label-free quantitation.

Figure 3. Comparison of protein and peptide identifications from 200 ng HeLa with different sample throughputs, different isolation widths and different LC configuration.

Wider isolation windows in short run times allow for good quantitation while sacrificing overall IDs. The trap-and-elute setup was tested in two different sites on different instruments and gives similar IDs with less than 5% variation. The direct injection setup benefits the throughput but does not yield much benefit for increasing protein group numbers, though peptide numbers are slightly higher.

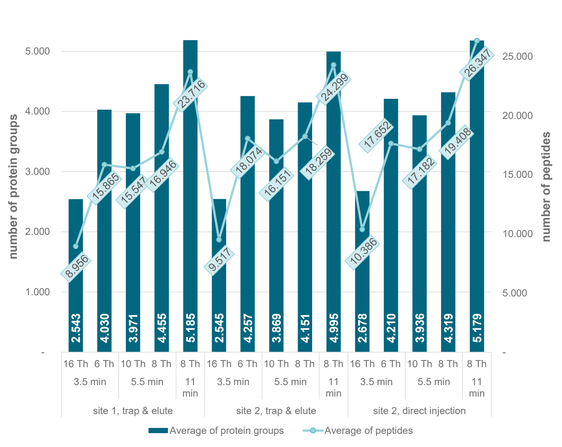


Figure 1. Schematic of timings for three different gradient lengths including sample loading, column wash and column equilibration on the Vanquish Neo LC system in direct injection configuration.

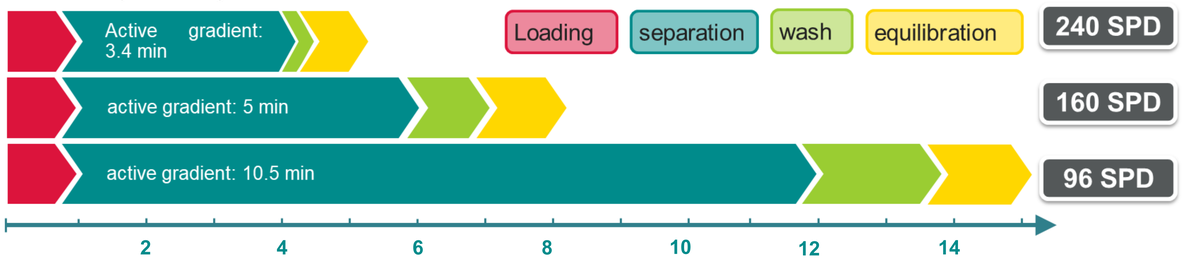
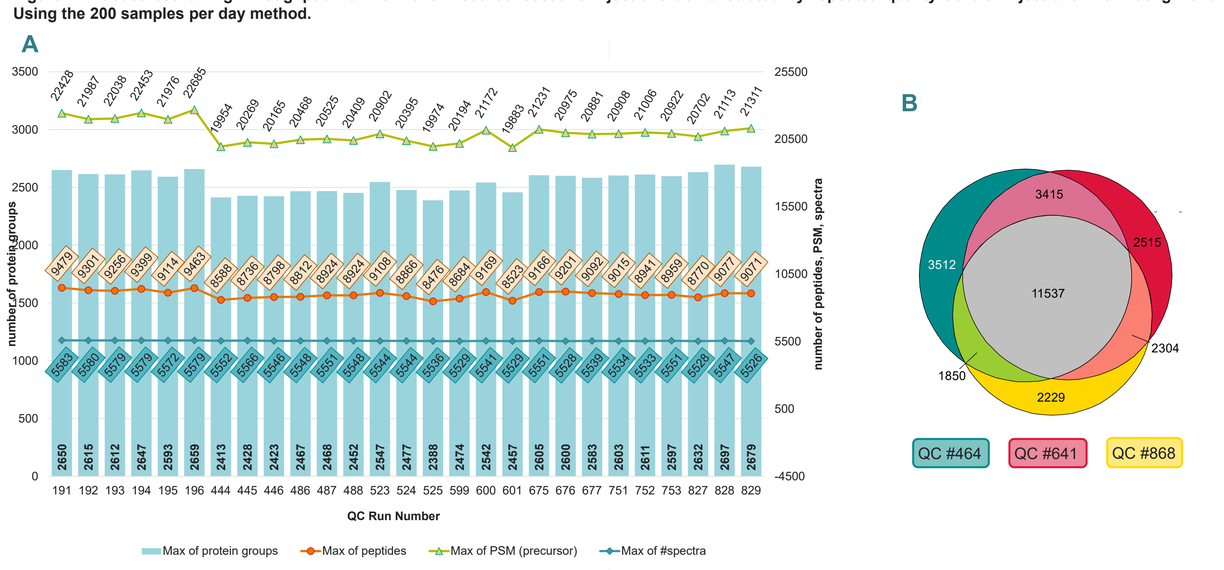


Figure 4. Robustness of high-hroughput workflow over >800 consecutive injections demonstrated by repeated quality control injections with 200ng HeLa Using the 200 samples per day method.

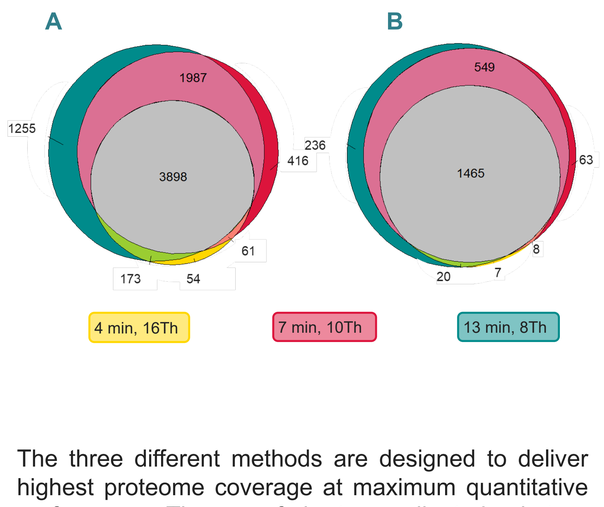


These results can be reproduced in two different sites with different instrument operators and even using different LC configurations with lower than 5% variation in ID numbers (see Figure 3). The comparison of wide and narrow isolation widths reveals major boosts in IDs for the shorter gradients enabling for even higher proteome coverage. However, if quantitative performance is needed, the narrower isolation windows ensure sufficient points across the chromatographic peak for ultimate precision and quantitation accuracy in label-free quantitation.

Quantitative performance, reproducibility, and robustness of the workflow

High-throughput workflows are especially applicable and adoptable for any large-cohort studies and therefore are required to run highly reproducibly over long time periods. In Figure 4, data is shown from 4 min runs with 200 ng HeLa that have been acquired throughput the whole length of the study. The results in terms of protein and peptide IDs (Fig. 4A) are within 5% variation over more than 800 injections, and the overlap of peptide IDs (Fig. 4B) is more than 60%.

Figure 5. Venn Diagrams of Overlaps of Identified Protein Groups (A) and Peptides (B) from 5 ng HeLa Digests in Different Gradient Lengths with Different Isolation Windows



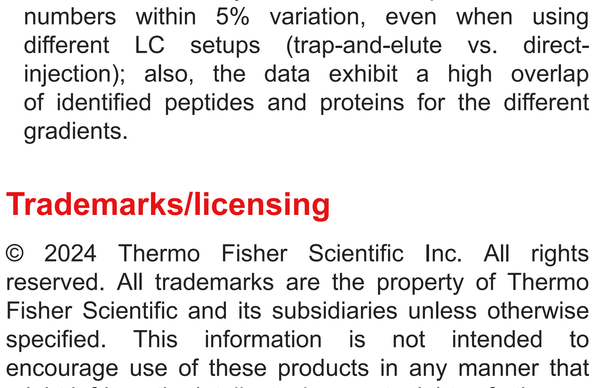
The three different methods are designed to deliver highest proteome coverage at maximum quantitative performance. The use of shorter gradients leads to a reduced proteome coverage but should not introduce a bias for peptides that cannot be identified from longer runs. The identified peptides of a shorter run should be ideally a subset of peptides identified from a longer run yielding higher proteome coverage. In fact, the overlap of peptides and proteins identified from a 4 min run is almost 100 % with identifications from longer runs, as can be seen in Figure 4 A and B.

Judging from the coefficients of variation (CV) of the peptide abundances, shown in Figure 6A, all methods perform well with low loads like 10 ng HeLa, 50% of all CVs are in the range of 5% to 20%, with median CVs align well around 10% or lower. The precision of peptide quantitation benefits in general from longer gradients, especially in these low loads. However, even in 4min runs, median CVs are 11% if a sufficiently wide isolation window is chosen to get appropriate number of points across peak. In the 4 min runs with 6Th, while largely increasing the number identified proteins, the CVs suffer from less points per peak.

Apart from coefficients of variation demonstrating high reproducibility within technical replicates, the retention times from run to run are extremely stable for all identified peptides in the first and third of three technical replicates, as seen in Fig. 6B. This makes the presented workflow suitable for large study sizes, ensuring reliable and reproducible results over the entire sample cohort.

Figure 6. Precision and Repeatability of High-Throughput Workflow in With Different Gradients and Methods

(A) Violin plots of coefficients of variation for all three gradients and different methods. Plots shown in blue indicate the use of optimized quantification method.
(B) Retention time repeatability in two injections of 10 ng HeLa separated on 3.5 min active gradient. All identified peptides/PSM correlate well over the whole retention time range.



Conclusions

- The presented workflow yields high proteome coverage for a wide range of sample loads
- The data show high robustness of results over more than 800 injections both in term of stable identification rates as well as high overlap of identified peptides in different runs over time.
- The data shown have high precision when using the methods optimized for quantitation (i.e. yielding sufficient points across the chromatographic peak). The CV of peptide abundances are generally below or around 10% even in low loads and short run times.
- The data show high repeatability and reproducibility: this is showcased by the two-site comparison with ID numbers within 5% variation, even when using different LC setups (trap-and-elute vs. direct-injection); also, the data exhibit a high overlap of identified peptides and proteins for the different gradients.

Trademarks/licensing

Science at a scan
Scan the QR code on the right with your mobile device to download this and many more scientific posters.

Learn more at thermofisher.com