P-II-0442

Approaching close-to-complete proteome coverage in single-shot deep dive DIA workflows

Robert Van Ling[1], Amirmansoor Hakimi[2], Santosh Renuse[2], Paul Jacobs[3] and Jeff Op de Beeck[3]
[1]Thermo Fisher Scientific, Breda, The Netherlands; [2] Thermo Fisher Scientific, San Jose, USA; [3]Thermo Fisher Scientific, Ghent, Belgium

Abstract

For single-shot deep dive proteomic experiments, it is crucial to use LC columns that deliver the highest possible peak capacity to reduce sample complexity as much as possible prior to MS/MS analysis. By using a 110 cm long microfluidic channel uniformly filled with 2.5 µm pillars, we report on significant increases in separation performance, delivering peak capacities well above 1600 in single-shot analyses. We subsequently evaluated the effect of increased separation performance on proteome depth and quantitation accuracy for single and multiple species tryptic digests.

Introduction

Throughout the history of Liquid Chromatography Mass Spectrometry (LC-MS) based proteomics, comprehensive identification and quantitation of all proteoforms present in biological samples has been the ultimate ambition. Several approaches have been close to achieving this goal, with the deepest coverage commonly obtained by fractionation of digested proteins using a separation mechanism orthogonal to low-pH reversed phase separation (SCX, high-pH C18, HILIC) and subsequent LC-MS analysis of isobaric labeled peptide samples. As recent innovations in LC-MS instrumentation have leveraged significant increases in analysis speed, resolution and sensitivity, we investigate the possibility to achieve similar coverage using single-shot label-free analyses with ultra-high peak capacity LC-MS. Hereby significantly reducing the labor-intensiveness, overall cost and required analytical skill needed to achieve close-to-complete proteome coverage.

Materials and methods

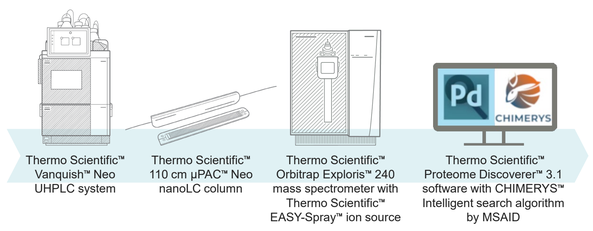
Sample preparation

Thermo Scientific™ Pierce™ HeLa Protein Digest Standard, Waters E. coli MassPREP Standard and Promega Mass Spec-Compatible Yeast Protein Digest were dissolved in 1% acetonitrile and 0.1% trifluoroacetic acid (TFA) to a final concentration of 1µg/µL with 30 seconds of vortexing. For the three-proteome mix, E. coli peptide digest and yeast peptide digest was added to a fixed amount of HeLa digest (2000 ng) at amounts of 400 ng to 800 ng, and 600 ng to 200 ng, respectively, yielding an E. coli peptide ratio of 1:0.5 and a yeast peptide ratio of 1:3 (Figure 7). For both sample series, 3 µL was injected using a direct injection procedure, hereby injecting a total of 3 µg onto the column.

LC-MS settings and data processing

Active solvent gradients of 180 min were evaluated at a constant flow rate of 250 nL/min and using variable flow rate from 750 to 250 nL/min. Electrospray ionization voltage was set at 1,9 kV. MS data were collected either in data-dependent acquisition mode (Top25) with MS1 resolution at 60,000 and MS2 resolution at 15,000 or in data-independent acquisition mode at different isolation window sizes (4, 6 and 8 Th), varying scan ranges (between 300 and 1100 m/z) and 2 resolution settings (MS1 60,000 – MS2 15,000 and MS1 120,000 – MS2 30,000). Both DDA and DIA acquired raw data files were processed with Proteome Discoverer software with the CHIMERYS intelligent search algorithm.

Figure 1. Experimental setup

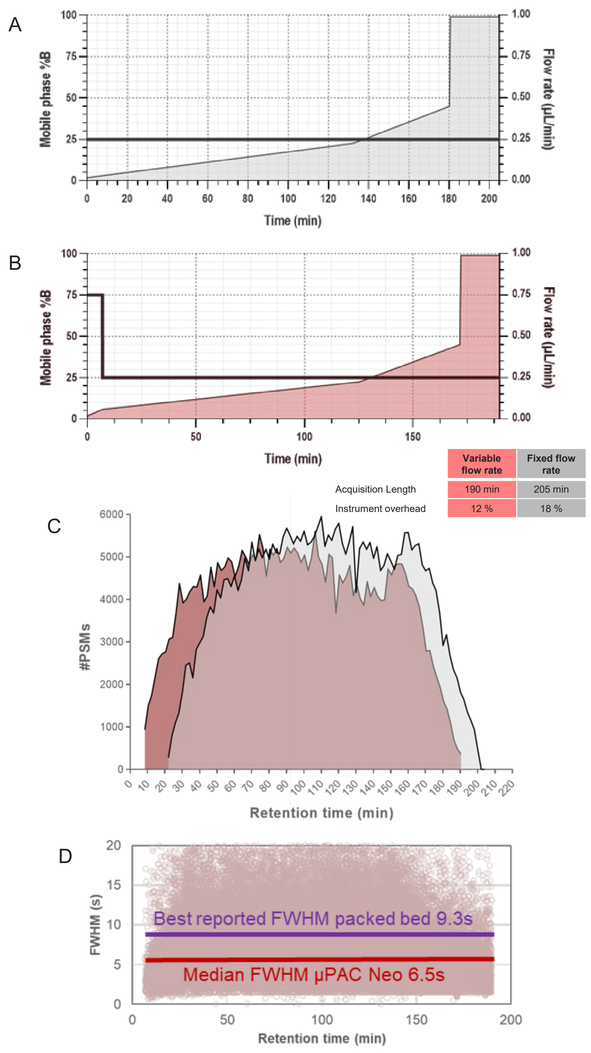


Results

Elution profile optimization

With a column volume of 4.5 µL, void volume, sample loading and column equilibration can have a large impact on instrument overhead. By using variable flow rate methods with high flow rate (750 nL/min) applied during the first 7 min of the gradient, and low flow rate (250 nL/min) applied as soon as analytes start eluting, excellent performance can be achieved at increased sample throughput. Chromatographic performance metrics reported here are superior to any other reported long gradient separation with pulled-tip packed bed columns, improving peak width by 30% for 180 min gradient separations.

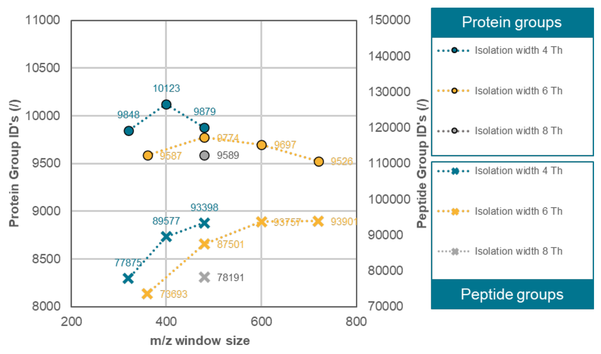
Figure 2. Solvent gradient and flow rate optimization for 180 min active gradient separations (A) Constant flow rate of 250 nL/min. (B) Variable flow rate from 750 to 250 nL/min. (C) number of PSMs per 30s for both gradient profiles. (D) FWHM distribution of all identified peptides.



DIA method optimization

When varying isolation window size (4, 6, and 8 Th) and combining with different scan ranges to yield at total of respectively 60, 80, 100 or 120 scan events, significantly higher coverage was achieved with the narrowest window size of 4 Th. With 100 scan events between m/z 400 and 800, more than 10,000 protein groups could successfully be identified with high confidence.

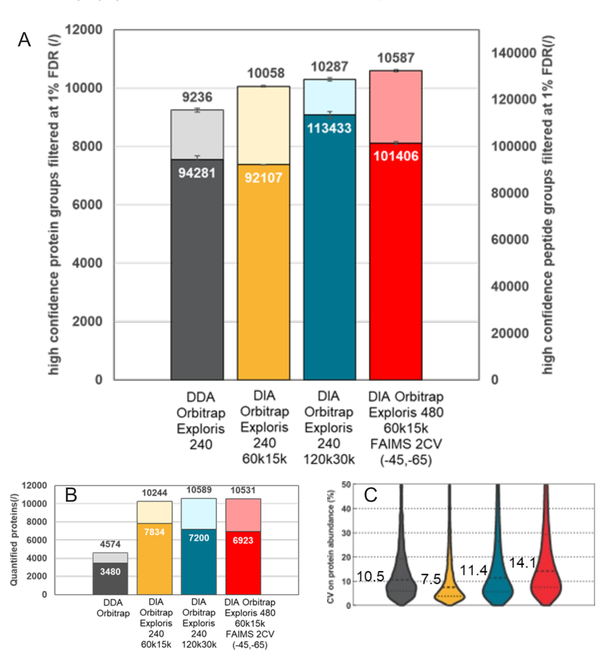
Figure 3. DIA method optimization for single-shot deep proteome coverage



Close-to-complete proteome coverage for single species protein digest

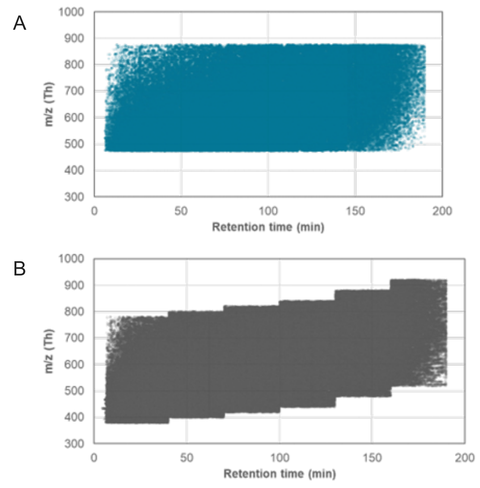
Compared to a conventional DDA acquisition strategy, up to 11% more protein groups could be identified in a single shot LC-MS analysis by using narrow window DIA. The highest absolute identification numbers were obtained using an MS1 resolution of 120,000 and MS2 resolution of 30,000. These settings do however increase MS cycle time and affect quantification precision. Even though the highest absolute number of quantified proteins was again achieved at these settings, variation on protein abundance was significantly lower when reducing the instrument resolution settings and MS cycle time (4s).

Figure 4. Proteome coverage and protein quantitation obtained for high load (3 µg) single species protein digest separations. (A) Protein and peptide groups identified using different acquisition strategies and instruments (n=3, no MBR). (B) Number of quantified proteins (total versus ≤ 20%CV). (C) Coefficient of variation on protein abundance.



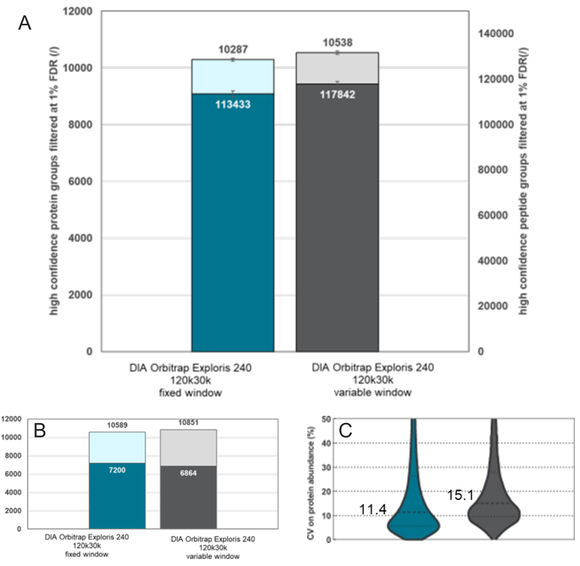
As precursor ion mass to charge tends to increase slightly along the course of a reversed phase LC gradient, slight increases in coverage can be achieved by using an optimized DIA method where the position of 400 m/z wide scanning windows is varied along the 180 min gradient. Up to 2% more protein groups and 4% more peptide groups were identified as compared to a method with a fixed scanning window from 475 to 875 m/z.

Figure 5. Scanning window used for (A) fixed window method, (B) variable window method.



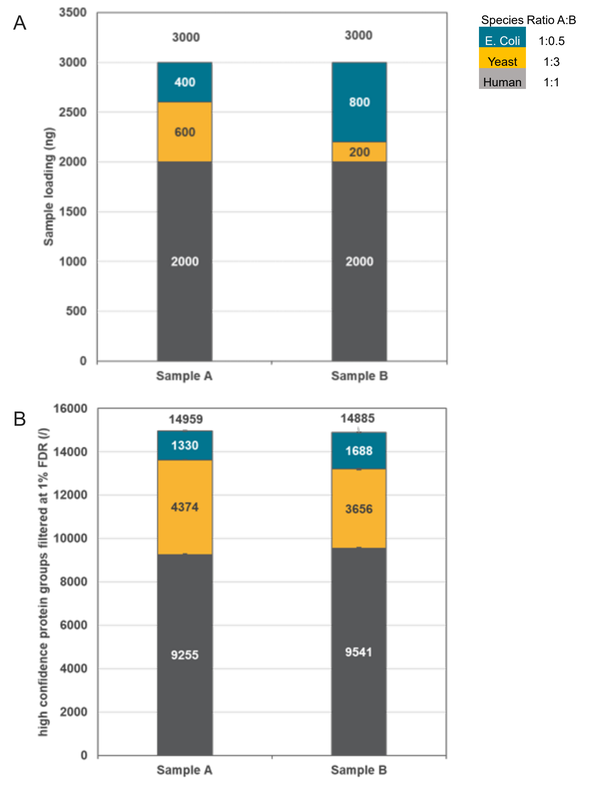
As precursor ion mass to charge tends to increase slightly along the course of a reversed phase LC gradient, slight increases in coverage can be achieved by using an optimized DIA method where the position of 400 m/z wide scanning windows is varied along the 180 min gradient. Up to 2% more protein groups and 4% more peptide groups were identified as compared to a method with a fixed scanning window from 475 to 875 m/z.

Figure 6. Proteome coverage and protein quantitation obtained for high load (3µg) single species protein digest separations using fixed and variable DIA scanning window positions. (A) Protein and peptide groups identified (n=3, no MBR) (B) Number of quantified proteins (total versus ≤ 20%CV). (C) Coefficient of variation on protein abundance.



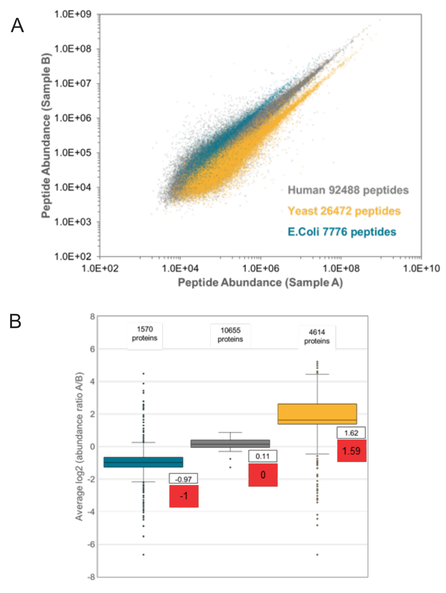
Workflow performance in label-free quantitation

Figure 7. (A) Experimental sample design of label free quantitation three-proteome mix (B) Proteome coverage obtained for single shot 3 species protein digest mix (n=3, no MBR).



By injecting a three proteome mixture with a total sample load of 3 µg, we were able to identify nearly 3 complete proteomes in a single run. Nearly 15,000 protein groups were consistently identified from both artificial samples, with up to 9541, 4374 and 1688 protein groups identified from respectively Human, yeast and *E.coli.* When plotting the corresponding peptide abundances of sample A to sample B, the three proteomes align well along their ratio specific linear slope (Figure 8A). and excellent quantitation accuracy could be achieved for all three species (Figure 8B).

Figure 8. Determination of abundance ratios in a three-proteome mixture. A) Relative peptide abundances of sample B to sample A. (B) Whisker boxplots of protein abundance ratios of all three species.



Conclusions

- 110 cm long µPAC Neo columns are an ideal match for deep-dive proteomics, delivering the sharpest peak width for long gradient separations.
- Flow rate and gradient optimization can reduce instrument overhead significantly and increase sample throughput.
- Data independent acquisition strategies outperform conventional data dependent acquisition, both in absolute identification as well as in quantitation.
- Using narrow window DIA and 180 min LC gradients, near full proteome coverage can be obtained with Orbitrap MS and 110 cm µPAC Neo columns.

Trademarks/licensing