

# Enhanced Peptide Identification Using Capillary UHPLC and Orbitrap Mass Spectrometry

by: Daniel Lopez-Ferrer<sup>1</sup>, Michael Blank<sup>1</sup>, Stephan Meding<sup>2</sup>, Aran Paulus<sup>1</sup>, Romain Huguet<sup>1</sup>, Remco Swart<sup>2</sup>, Andreas FR Hubmer<sup>1</sup>

<sup>1</sup>Thermo Fisher Scientific, San Jose, CA, USA

<sup>2</sup>Thermo Fisher Scientific, Germering, Germany

Chromatographic separation and tandem mass spectrometry (MS/MS) identification underpin 'bottom-up' proteomic analysis, with both essential steps witnessing significant performance improvements in recent years as a result of advances in instrumentation. The high backpressures offered by the latest ultra-high performance liquid chromatography (UHPLC) systems allow the use of longer and more efficient capillary columns that, when used in combination with high-resolution Orbitrap mass spectrometer technology, can achieve superior levels of peptide identification, protein quantitation and analytical throughput. Here, we demonstrate enhanced proteomic analysis of a HeLa cell lysate using a 75 cm long capillary column and compare these results to those obtained using a 50 cm column. The increase in column length resulted in the separation and detection of more unique peptides, improved protein quantitation, as well as better reproducibility and increased throughput, highlighting the suitability of this LC-MS/MS method for the quantitation of complex lysates.

## Introduction

Unlike 'top-down' approaches to proteomic analysis, which are based on direct fragmentation and analysis of intact proteins by tandem mass spectrometry (MS/MS), conventional 'bottom-up' proteomic techniques are underpinned by three principal steps: protein digestion, chromatographic separation, and MS/MS identification. Despite the significant advances in technology and methodology that have been made over the past two decades, "bottom-up" approaches to proteomic analysis face a number of key challenges.

One challenge relates to the depth of proteomic identification, with only a small proportion of the total peptide population typically recovered using current liquid chromatography (LC)-MS/MS techniques [1]. A second challenge concerns the sensitivity and dynamic range of these approaches, given the limited size of samples such as clinical biopsies [2] and the need for accurate peptide quantitation over concentrations spanning several orders of magnitude [3]. Finally, as many studies involve proteomic analysis of large numbers of samples across multiple conditions to reach meaningful conclusions, achieving high levels of analytical throughput are essential [4].

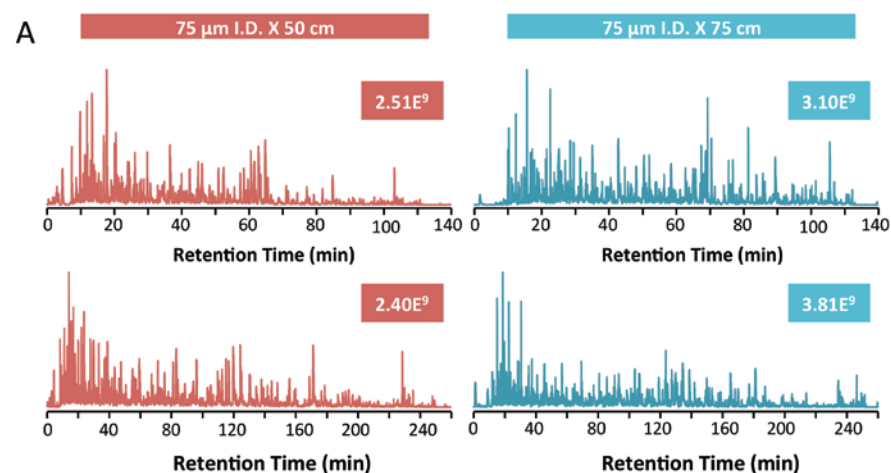


Figure 1. Representative chromatograms for each gradient and column length variation.

Recent advances in instrumentation are making significant progress towards these challenges. Orbitrap mass spectrometers are widely considered to be the gold standard for mass spectrometry-based proteomics, with best-in-class sensitivity, mass resolution and scan rates [5]. However, to achieve the optimum efficiency in protein identification and quantitation, this powerful detection technology must be paired with equally effective separation methods. Nano chromatography, using 50 to 75 µm i.d. fused silica capillaries based on a reversed-phase stationary phase supported on silica particles 2 to 3 µm in size, has proven to be particularly effective for this purpose. Typically, a 75 µm i.d. capillary column is run

at flow rates of up to 400 nL/min, with early bottom-up proteomics experiments typically employing 15 to 25 cm long capillaries with gradients of between 15 and 60 minutes [6].

In recent years, much progress has been made in pushing the limits of peptide separation efficiency. Separation performance can be quantified in terms of peak capacity ( $C_p$ ), which is defined as the maximum number of components that can be separated in a given time. Mechtler and colleagues have shown that for a packed reversed-phase liquid chromatography column, the maximum column peak capacity that can be obtained for a given gradient is proportional to the square root of the length, or alternatively, the particle size if the

Table 1. Gradient conditions.

Composition	120 min Gradient	240 min Gradient
5-28%B	0-105 min	0-210 min
28-40%B	105-120 min	210-240 min
40-95%B	120-130 min	240-250 min
95-95%B	130-140 min	250-260 min

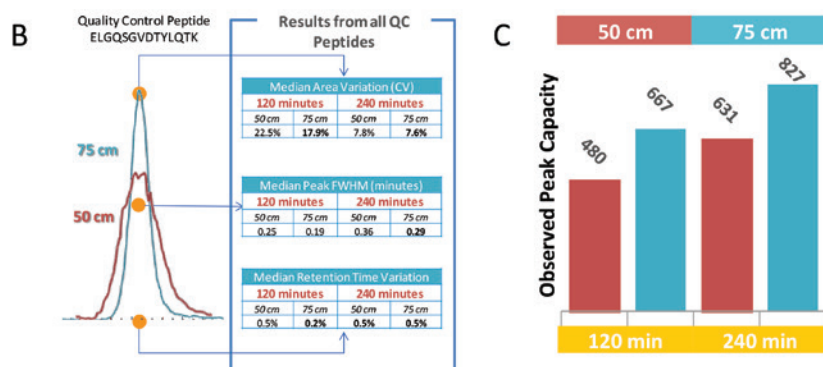


Figure 2. (A) Extracted ion chromatogram for one of the QC peptides and average chromatographic metrics of all 15 QC peptides obtained for each experimental configuration; (B) comparison of peak capacity for each experimental configuration.

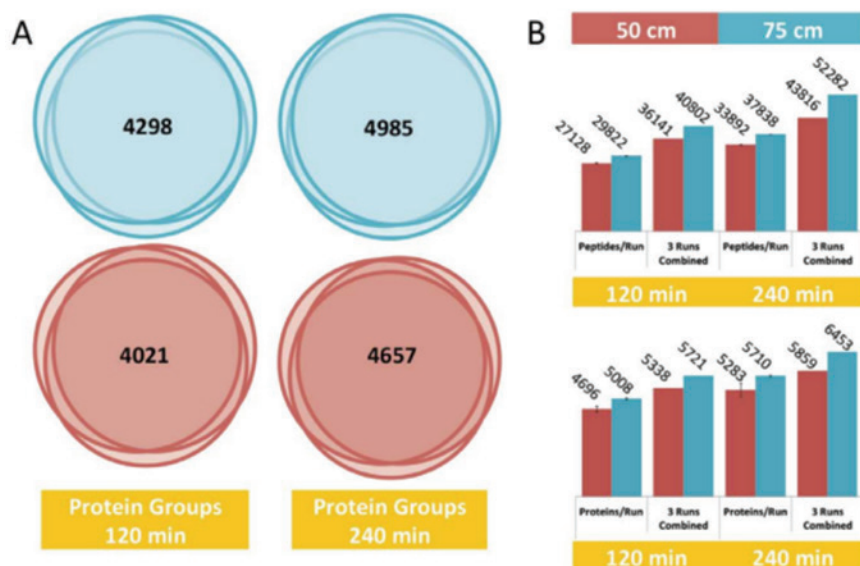


Figure 3. (A) Venn diagrams showing the overlap between technical replicates for the identified proteins for each experimental configuration; (B) number of identified peptide and protein groups.

column length remains constant [7]. In other words, within a given separation time, better peptide separation can be achieved using either longer capillaries or by using smaller particle sizes.

The latest ultra-high performance liquid chromatography (UHPLC) systems now allow for backpressures of up to 1200 bar, enabling the routine use of longer, more efficient columns. Here, we demonstrate the improved proteomic analysis of a HeLa cell lysate using a 75 cm long capillary column and a 2 hour and 4 hour chromatography gradient, and compare these results to those obtained using a 50 cm column, which until

recently was the longest high-performance nano LC column available on the market.

## Experimental

**Materials:** All solvents were LC-MS grade and purchased from Fisher Chemical. Protein digests and calibration standards were purchased from Thermo Fisher. Aliquots containing 500 ng/ $\mu$ L of Pierce HeLa protein digest (Thermo Fisher) (catalogue number: 88328) and 50 fmol/ $\mu$ L of Pierce peptide retention time calibration standards (catalogue number: 88321) in water with 0.1% formic acid were prepared

for the study. 2 or 4  $\mu$ L (equivalent to 1 or 2  $\mu$ g) of the HeLa cell digest was injected directly onto the column.

**Chromatography:** Analyses were performed using an EASY-nLC 1200 UHPLC system (Thermo Fisher). Two Acclaim PepMap C18 EASY-Spray columns (2  $\mu$ m particle size, 75  $\mu$ m i.d.), (Thermo Fisher) of length 50 cm and 75 cm, were used in this comparative study. The columns were connected to the LC system using a Dionex nanoViper fingertight fitting (Thermo Fisher). Solvent A was water containing 0.1% formic acid, and solvent B was an 80:20 mixture of acetonitrile-water, also containing 0.1% formic acid. The gradient methods used are summarised in Table 1. A constant column temperature of 55°C was maintained during all experiments. Injection, sample loading, column equilibration, and autosampler wash conditions were consistent between the gradient durations and column lengths. A flow rate of 300 nL/min was used for both columns, resulting in a backpressure of approximately 900 bar for the 75 cm column and approximately 600 bar for the 50 cm column.

**Mass spectrometry:** An Orbitrap Fusion Lumos MS (Thermo Fisher) was used for peptide MS/MS analysis. Survey scans of peptide precursors were performed from 375 to 1575  $m/z$  at 120K full width half maximum (FWHM) resolution (at 200  $m/z$ ) with a  $4 \times 10^5$  ion count target and maximum injection time of 50 ms. The instrument was run in top speed mode with 3 s cycles for the survey and MS/MS scans. Following a survey scan, tandem MS was performed on the most abundant precursors exhibiting a charge state from 2 to 7 with an intensity greater than  $5 \times 10^3$ . Precursors were isolated in the quadrupole at 1.2 Th. CID fragmentation was applied with 35% collision energy and resulting fragments were detected using the rapid scan rate in the ion trap. The automatic gain control target for MS/MS was set to  $10^4$  and the maximum injection time was limited to 35 ms. A dynamic exclusion time of 12 s was used with a 10 ppm mass tolerance around the precursor and its isotopes, and monoisotopic precursor selection was enabled.

**Data analysis:** The raw data was processed using Proteome Discoverer software (Thermo Fisher). MS<sup>2</sup> spectra were searched with the SEQUEST HT engine against a database of 42,085 human proteins, including proteoforms (Uniprot). In silico peptide sequences were generated using tryptic cleavages, allowing for up to two

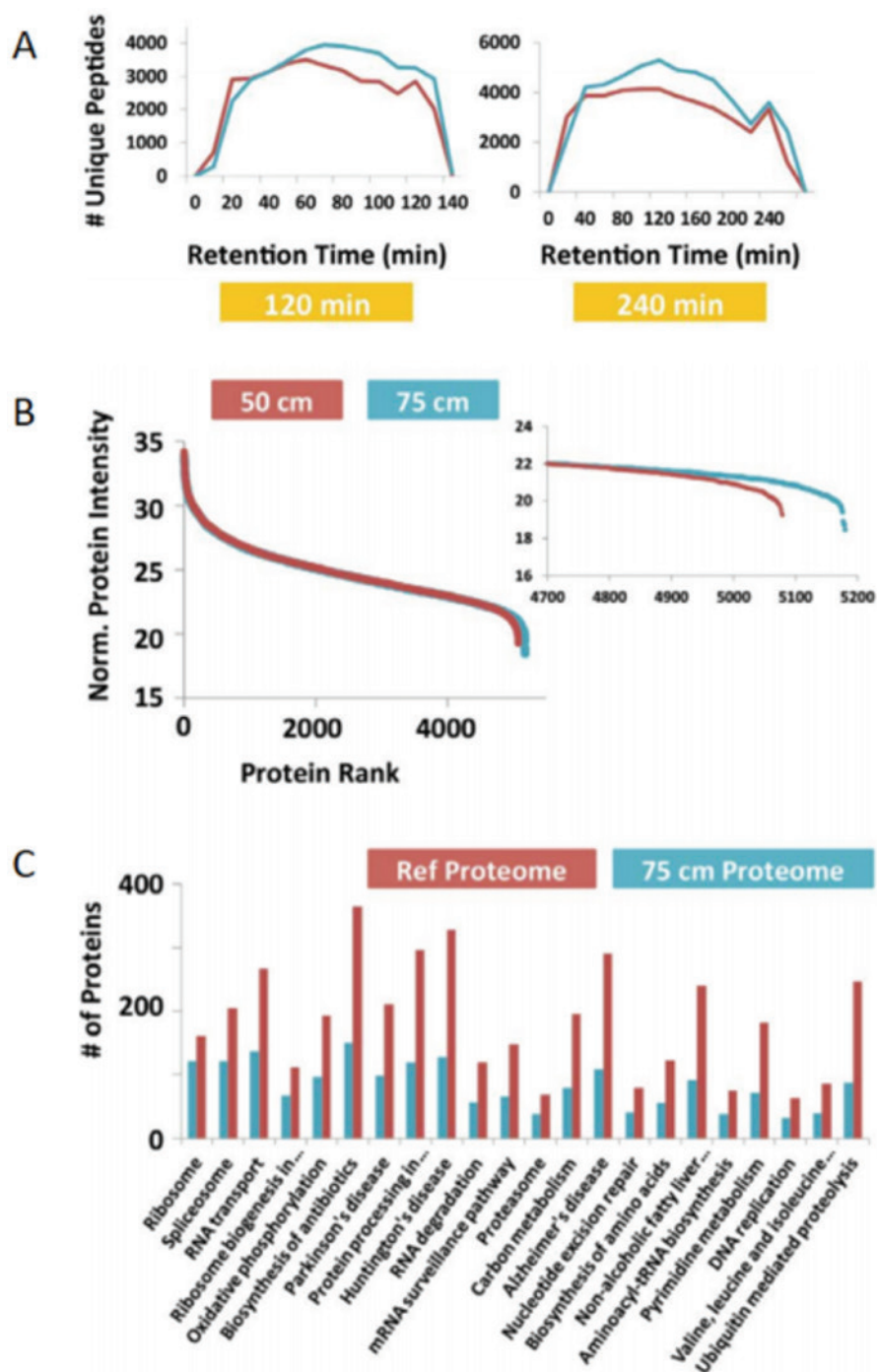


Figure 4. (A) Line graph displaying the trend of identified peptides against retention time; (B) protein rank of proteins over their normalised intensity for both columns; (C) overrepresented pathways for the 75cm column length and 4 hour gradient dataset.

missed cleavages; carbamidomethylation (+57.021 Da) of cysteine residues was set as a fixed modification. Oxidation of methionine residues (+15.9949 Da), acetylation of the protein N-terminus (+42.0106 Da), and deamidation of asparagine and glutamine (+0.984 Da) were treated as variable modifications. The precursor mass tolerance was 10 ppm and product ions were searched at 0.8 Da tolerances. Peptide spectral matches were validated using the Percolator algorithm [8], based on q-values at a 1% false discovery rate (FDR). Using Proteome Discoverer,

peptide identifications were grouped into proteins according to the law of parsimony and filtered to 1% FDR. The precursor ion area from the identified peptides was extracted using the Precursor Ions Area Detector plug-in. For further analysis, spectral matches and peptide groups passing the FDR were exported and processed using DAnTE RDN [9]. Skyline software was used to extract the ion chromatograms of the Pierce retention time calibration standards to calculate FWHM values, coefficients of variation, retention time variation, and peptide peak capacity.

## Results and Discussion

The EASY-nLC 1200 UHPLC system and Orbitrap Fusion Lumos mass spectrometer were used to separate a HeLa cell lysate using both a 50 cm and a 75 cm long, 75  $\mu$ m i.d. capillary column, under both 2 and 4 hour gradients. Representative chromatograms for each of the columns and gradients studied are shown in Figure 1.

### High analytical reproducibility and throughput

To obtain a reliable comparison between samples measured at different times and under different conditions, and ultimately obtain quantitative information about the proteome, it is essential that chromatographic separation is reproducible. For each experimental set-up, the peak profiles between replicates were found to be almost identical, and peptide retention time shifts of less than 1 minute were observed even when employing a 240-minute gradient.

The base peak chromatograms were also found to be consistent among the four experimental conditions. Very few differences were observed between the two columns, which could be expected as the columns contained the same packing material. However, a small shift in the retention times obtained using the longer column was observed due to the increased column length and longer column travel time.

The chromatographic performance parameters for the 15 PRTC standards added to the samples as quality controls are reported in Figure 2A. The 75 cm column resulted in improved performance over the 50 cm column, with the coefficients of variation for median peptide peak areas, FWHM values, and retention times equal to or lower than those obtained for the longer column, using both gradient lengths.

The peak capacity ( $C_p$ ) [10], defined in Equation 1 (where  $n$  is the number of peaks used for the calculation,  $T_G$  is the gradient length, and  $W_p$  is the width at half maximum of the peak height), was also used to evaluate the analytical capability of each experimental set-up. Using this analysis, longer gradient methods and longer column lengths improved performance, as shown in Figure 2B. Using a 240 minute gradient, the 75 cm column results in a  $C_p$  of 827, almost double that obtained by Hsieh et al. [11].

$$\text{Equation 1. } C_p = 1 + \frac{T_G}{\frac{1}{n} \sum W_p}$$

For many proteomics applications that involve the sampling of large sample volumes, high analytical throughput is essential. The fact that the 75 cm column, using a shorter gradient time, achieved a  $C_p$  that exceeded that of the 50 cm column with a longer gradient time, highlights the improved efficiency of using longer column lengths. Furthermore, as the backpressure using the 75 cm column does not approach maximum rated for the EASY-nLC 1200 system, with further optimisation of LC-MS/MS parameters the separation achieved using this column could potentially be improved further.

### Digging deeper into the proteome

The number of peptides identified in a given analysis is an important parameter in proteomic investigations and can directly affect confidence in results and the impact of conclusions. As shown in Figure 3, the 75-cm column consistently resulted in the largest number of peptides and protein identifications by a margin of at least 7%. While the use of smaller columns typically results in a reproducibility of around 80%, in this study more than 95% of the peptide and protein identifications for a given dataset were shared with one other replicate across three repeat analyses.

The brighter source and higher transmission quadrupole of the Orbitrap Fusion Lumos instrument, as well as its ultrafast scan speed of up to 20 Hz and ability to perform parallelised precursor ion isolation, fragmentation, and scanning in the dual linear ion trap, ensured that the majority of the detected ions in the survey scan were efficiently fragmented, resulting in superior peptide and protein identification.

As shown in Figure 4A, the 75 cm column produced consistently better peptide identification across the entire LC-MS/MS gradient. This enhanced performance can be explained by the improved separation achieved by the 75 cm column, which allows given peptides to elute at higher concentrations, yield higher quality MS/MS spectra and in turn, result in more positive identifications. Figure 4B shows the rank of the identified and quantifiable proteins for the 4 hour gradient, and as expected, the longer column goes much deeper into the proteome coverage.

The results of pathway analysis yielded the same profile of overrepresented pathways for both columns, but with different degrees of

coverage, demonstrating that the overall study was unbiased. Figure 4C shows that the data obtained for the 75 cm column with the 4-hour long gradient provides direct quantitation of approximately half of the proteins in each of the 23 overrepresented pathways.

### Improved Quantitation

Substantial improvements in quantitation were also achieved. The number of quantifiable peptides increased by 20% using the 75 cm column, which not only resulted in a greater number of quantified peptides but also a higher correlation among replicates (>85%). The lower coefficients of variation for these peptides and proteins also resulted in more accurate quantitation.

Doubling the amount of peptide digest loaded onto the column, from 1  $\mu$ g to 2  $\mu$ g, did not have a significant impact on the number of quantifiable peptides or protein identifications. However, increasing the loading amount did improve the correlation among runs to 89%, and resulted in twice as many proteins with coefficients of variation below 5%. This improvement in performance highlights the potential for allowing for more accurate proteome quantitation using larger loadings.

### Conclusions

The high backpressures offered by the latest UHPLC systems allow the use of longer columns that, when used in conjunction with state-of-the-art Orbitrap mass spectrometers, can achieve superior levels of peptide and protein identification while maximising throughput. These technologies represent a very powerful platform for conducting high-performance proteomics investigations, as they offer exceptional run-to-run reproducibility and depth of analysis, and also permit high sample loading capacity.

Increasing the column length from 50 cm to 75 cm resulted in the separation and detection of 10% more unique peptides and 7% more protein identification. Using this set-up we identified approximately 6500 proteins without the need for fractionation, and reproducibly quantified over 5000 proteins based on just three technical replicate injections, highlighting the potential of this method as a highly efficient and reliable alternative approach for quantitative proteomics.

### References

- [1] Hebert, A.S., et al. The one hour yeast proteome. *Mol Cell Proteomics*. 2014, 13(1), 339-347.
- [2] Wu, X. et al. Global phosphotyrosine survey in triple-negative breast cancer reveals activation of multiple tyrosine kinase signaling pathways. *Oncotarget*. 2015, 6(30) 29143-29160.
- [3] Cox, J.; Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem*. 2011, 80, 273-299.
- [4] Livesay, E.A. et al. Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. *Anal Chem*. 2008, 80(1), 294-302.
- [5] Scigelova, M.; Hornshaw, M.; Giannakopoulos, A.; Makarov, A. Fourier transform mass spectrometry. *Mol Cell Proteomics*. 2011, 10(7), M111.009431.
- [6] Hsieh, E.J.; Bereman, M.S.; Durand, S.; Valaskovic, G.A.; MacCoss, M.J. Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *J Am Soc Mass Spectrom*. 2013, 24(1), 148-153.
- [7] Köcher, T. et al. Development and performance evaluation of an ultralow flow nanoliquid chromatography-tandem mass spectrometry set-up. *Proteomics*. 2014, 14(17), 1999-2007.
- [8] Käll, L.; Canterbury, J.D.; Weston, J.; Noble, W.S.; MacCoss, M.J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*. 2007, 4(11), 923-925.
- [9] Polpitiya, A.D., et al. DANTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics*. 2008, 24(13), 1556-1558.
- [10] Shen, Y and Lee, M.L. General Equation for Peak Capacity in Column Chromatography. *Anal. Chem.*, 1998, 70 (18), 3853-3856
- [11] Hsieh, E.J.; Bereman, M.S.; Durand, S.; Valaskovic, G.A.; MacCoss, M.J. Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *J Am Soc Mass Spectrom*. 2013, 24(1), 148-53.