Pushing the Limits of Bottom-Up Proteomics with State-Of-The-Art Capillary UHPLC and Orbitrap Mass Spectrometry for Reproducible Quantitation of Proteomes

Daniel Lopez-Ferrer[1], Michael Blank[1], Stephan Meding[2], Aran Paulus[1], Romain Huguet[1], Remco Swart[2], Andreas FR Huhmer[1]
[1]Thermo Fisher Scientific, San Jose, CA, USA; [2]Thermo Fisher Scientific, Germering, Germany

Application Note 639

Key Words
Orbitrap Fusion Lumos, EASY-nLC 1200, peptide, protein, high resolution, accurate mass, EASY-Spray, 75 µm i.d. x 75 cm PepMap C18 column, Proteome Discoverer 2.1, peptide retention time calibration mixture, HeLa protein digestion standard, Thermo Fisher Cloud

Goal
To describe a higher-performing quantitative proteomics workflow compared to current methods using a Thermo Scientific™ EASY-nLC™ 1200 system with a Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer with a 75 cm long, 75 µm i.d. capillary column using both 2 and 4 hour gradients.

Introduction
Since its inception, bottom-up proteomics has aimed to identify and quantify the complete proteome from a cell, tissue, or whole organism.[1] Although many advances have been made in the last 15 years, there are still three main challenges to overcome. The first is to obtain complete coverage of the proteome by identifying all the expressed proteins in a given time.[2] The second is working with samples of limited amounts such as like clinical biopsies[3], and the third is achieving sufficient analytical throughput.[4] Understanding the dynamics of the proteome requires analyzing it across different conditions and time points throughout the cellular life cycle. For many studies, this analysis needs to be performed in a high-throughput manner. To further complicate matters, being able to discriminate the most important proteins constituting a given cellular state requires accurate peptide measurement across several orders of magnitude.[5]

Peptide separation and MS/MS identification are pillars of modern proteomic analysis, and each has seen performance improvements with advances in instrumentation. Orbitrap mass spectrometers are now considered the gold standard for mass spectrometry-based proteomics.[6] The recently introduced Orbitrap Fusion Lumos Tribrid mass spectrometer is, at the time of this writing, the instrument with the best sensitivity, best mass resolution, and fastest scan rate. Despite being the most advanced mass spectrometer, use of the improved separations techniques brings even further gains to the performance of the Orbitrap Fusion Lumos MS. An efficient sample separation method has to be coupled with the best peptide separation system to characterize as many unique peptides and identify as many proteins in a given lysate as possible. To achieve the maximum efficiency in identifying and quantifying proteins, the separation efficiency of the chromatography system and detection speed of the mass spectrometer must match in resolution and speed, respectively. Nano chromatography, using 50 to 75 µm i.d. fused silica capillaries filled with 2 to 3 µm silica particles and reversed-phase stationary phases, has proven to be particularly successful in this respect. Typically, a 75 µm i.d. capillary is run at flow rates of 300 to 400 nL/min, using a water/acetonitrile gradient containing 0.1% formic acid. Early bottom-up proteomics experiments were done in 15 to 25 cm long capillaries with gradients of 15 to 60 minutes.[7]

In recent years, several efforts have been made to push separation efficiency even further. The performance of progressive separation techniques can be quantified in terms of peak capacity (Cp), defined as the maximum number of components that can be separated in a given separation time. Mechtler and coworkers demonstrated that for a packed reversed-phase liquid chromatography column, the maximum column peak capacity that can be obtained for a given gradient that aims to separate peptides is proportional to the square root of the length or by the particle size if the column length stays constant.[8] In any case, achieving better peptide separation, if the separation time stays constant, could be done using either longer capillaries (for example, 50 cm or above) or by using smaller sized particles. In both scenarios, at the required optimum flow rate of 300 to 400 nL/min, a UHPLC pump with augmented pressure capabilities is necessary.

Thermo SCIENTIFIC

Using the Thermo Scientific™ EASY-nLC™ 1000 UHPLC system with a maximum pressure limit of 1000 bar, it was possible to run columns of up to 50 cm at elevated temperatures of 40 to 50 °C. The newly introduced EASY-nLC 1200 system now allows for a maximum backpressure up to 1200 bar, enabling routine operation with columns of 50 cm and longer. In this technical application note, we outline the cooperative use of an EASY-nLC 1200 system with an Orbitrap Fusion Lumos mass spectrometer to separate a HeLa cell lysate in a 75 cm long, 75 µm i.d., Thermo Scientific™ Acclaim™ PepMap™ capillary column using both 2 and 4 hour gradients. The results are compared with those obtained under the same conditions with a 50 cm column, which was, until now, the longest commercially available high-performance nano LC column for bottom-up proteomics. In both cases, the columns were used in an EASY-Spray column format. The length increase resulted in the separation and detection of 10% more unique peptides and 7% more protein identification in a 4 hour gradient. Protein identifications exceeded 5700 proteins for a single injection of mammalian cell lysate. More importantly, longer columns showed better reproducibility as seen by increased correlation among technical replicates, higher numbers of quantifiable peptides, and a smaller coefficient of variance (CV), resulting in improved protein quantification for complex lysates by high-resolution, accurate-mass (HRAM) LC-MS.

Experimental

Reagents

All solvents were Fisher Chemical LC-MS grade solvents. Mixing of water and acetonitrile was performed volume-to-volume. Solvent A and B were degassed using an ultrasonic bath before use. Solvent A was 100% water with 0.1% formic acid. Solvent B was 80% acetonitrile, 20% water, and 0.1% formic acid. These solvents are also available premixed from Fisher Scientific.

Aliquots containing 500 ng/µL Thermo Scientific™ Pierce™ HeLa protein digest (P/N 88328) and 50 fmol/µL of Pierce peptide retention time calibration (PRTC) standards (P/N 88320) in water with 0.1% formic acid were prepared for the study. Either 1 or 2 µg of HeLa digest were loaded onto the column.

LC-MS/MS

All analyses were performed using an EASY-nLC 1200 UHPLC system. HeLa cell digest sample was loaded directly onto the column using the one-column (direct injection) mode, with either 2 or 4 µL injected onto the column, corresponding to 1 or 2 µg, respectively. The analytical columns used were 75 µm i.d. Acclaim PepMap columns with 2 µm particles, either 50 cm or 75 cm long, manufactured in the EASY-Spray format (Table 1). The columns were connected to the LC system by way of a Thermo Scientific™ Dionex™ nanoViper™ fingertight fitting. The column temperature was maintained at a constant 55 ˚C during all experiments. Injection, sample loading, column equilibration, and autosampler wash conditions were kept consistent between the gradient durations and column lengths (Table 2). The flow rate during the gradient was kept constant at 300 nL/min yielding a backpressure of approximately 900 bar for the 75 cm column and approximately 600 bar for the 50 cm.

Table 1. Columns used during the experiments.

Separation Column	Part Number
EASY-Spray column Acclaim PepMap C18 100 Å, 2 µm particle size, 75 µm i.d. x 50 cm	P/N ES803
EASY-Spray column Acclaim PepMap C18 100 Å, 2 µm particle size, 75 µm i.d. x 75 cm	P/N ES805

Table 2. 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

An Orbitrap Fusion Lumos MS was used for peptide MS/MS analysis. Survey scans of peptide precursors were performed from 375 to 1575 *m/z* at 120K FWHM resolution (at 200 *m/z*) with a 4 x 10^5 ion count target and a maximum injection time of 50 ms. The instrument was set to run in top speed mode with 3 s cycles for the survey and the MS/MS scans. After a survey scan, tandem MS was performed on the most abundant precursors exhibiting a charge state from 2 to 7 of greater than 5 x 10^3 intensity by isolating them 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 AGC target for MS/MS was set to 10^4 and the maximum injection time limited to 35 ms. The dynamic exclusion was set to 12 s with a 10 ppm mass tolerance around the precursor and its isotopes. Monoisotopic precursor selection was enabled.

Data Analysis

Raw data was processed using Thermo Scientific™ Proteome Discoverer™ software version 2.1.0.80. MS^2 spectra were searched with the SEQUEST® HT engine against a database of 42,085 human proteins, including proteoforms (Uniprot, July 14th, 2015). Peptides were generated from a tryptic digestion allowing for up to two missed cleavages, carbamidomethylation (+57.021 Da) of cysteine residues was set as 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. Precursor mass tolerance was 10 ppm and product ions were searched at 0.8 Da tolerances. Peptide spectral matches (PSM) were validated using the Percolator® algorithm[9], based on q-values at a 1% FDR. With Proteome Discoverer software, peptide identifications were grouped into proteins according to the law of parsimony and filtered to 1% FDR. The area of the precursor ion from the identified peptides was extracted using the Precursor Ions Area Detector plug-in. For further analysis, PSMs and Peptide Groups passing the FDR were exported to a text file and processed using DAnTE RDN.[10]

In addition, Skyline (3.1)[11] was used to extract ion chromatograms of the PRTC standards to calculate full width at half maximum (FWHM), coefficients of variation, retention time variation, and peptide peak capacity.

Results and Discussion

In recent years, shotgun proteomics has moved towards the use of longer gradients and columns with higher peak capacity to overcome the limitations of oversampling, ion suppression, and the need for laborious peptide fractionation.[12] To this purpose, we have developed a new UHPLC system that has a higher backpressure limit, increased robustness, and very accurate flow control at nanoflow regimes (Figure 1). Based on this new chromatography system, we have selected column lengths of 50 and 75 cm and operated using linear gradients of 120 and 240 minutes. With this configuration, we not only tested the performance of the system but also validated the use of longer columns in a routine environment.

Chromatographic Performance

Reproducibility of the chromatographic separation is the most important requisite for a reliable comparison among different runs and ultimately obtaining quantitative information about the proteome under analysis. Figure 2A shows representative chromatograms for each of the columns and gradients. As it can be seen, base peak chromatograms are very consistent among all the analyses with the highest degree of similarity among replicates for each of the setups. As is expected, since the same solid phase is used, few differences are noticeable among the two columns lengths. However, at the beginning of the chromatogram, a small shift in the retention time (RT) is observable due to the increased volume of the 75 cm column. The peak profiles among replicates were almost identical, and peptide retention time shifts of less than 1 minute were observed even when employing a 240 minute gradient. Figure 2B shows the significant chromatographic performance parameters for the 15 PRTC standards spiked in as quality controls in all of the runs. The 75 cm column performs significantly better than its shorter counterpart. Coefficients of variation for peptide peak areas, median full width half maximum values, and RT variation were always less for the longer column, irrespective of the gradient length. Furthermore, peak capacity, Cp, was used to evaluate the performance of each chromatography configuration. Cp, is 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.

Equation 1. $$C_p = 1 + \frac{T_G}{\frac{1}{n}\sum_1^n W_p}$$

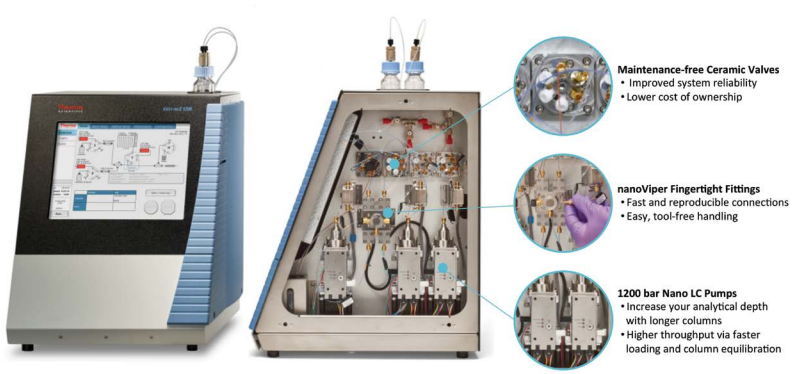

Figure 1. Front and side views of the EASY-nLC 1200 chromatography system and detailed view of its technical features.

Using data generated in this study, Cp increases linearly with increasing either gradient length or column length. The 75 cm column achieves a Cp of over 800 employing a 240 minute gradient, which almost doubles previous data reported recently by MacCoss and colleagues.[7] Interestingly, the 75 cm column achieves a higher peak capacity in 2 hours than that of the 50 cm column with a 4 hour gradient. With careful optimization of the LC and MS parameters, the 75 cm column could achieve very similar results with a shorter gradient than the 50 cm column with the 4 hour gradient. Since the 75 cm column does not approach the maximum pressure rated for the EASY-nLC 1200 system, the chromatography could potentially be optimized even further to maximize separation.

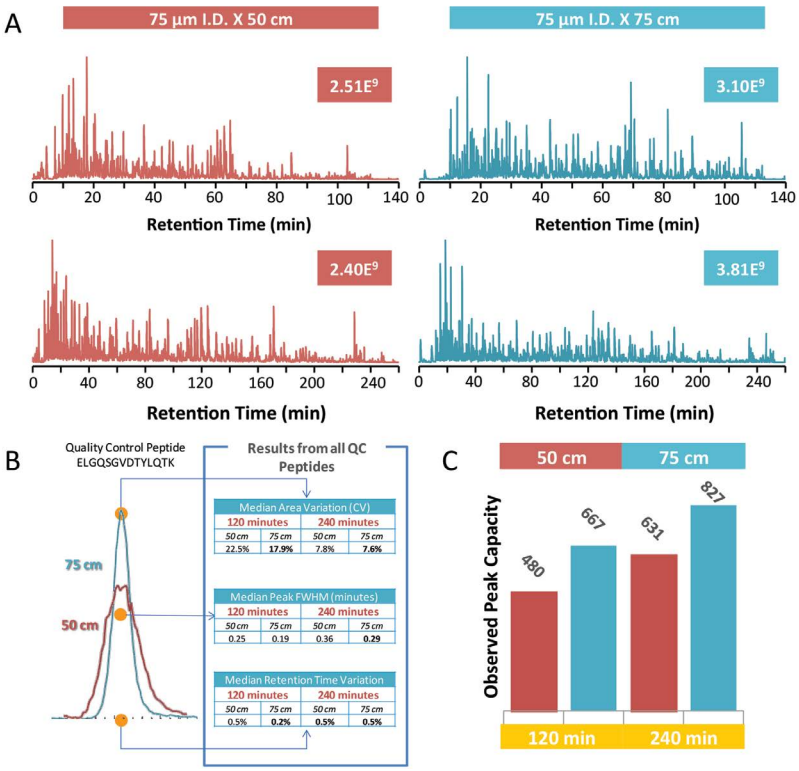

Figure 2. A) Representative chromatograms obtained using different gradients and column lengths. B) Extracted ion chromatogram for one of the 15 representative QC peptides and average chromatographic metrics of all 15 QC peptides obtained for different experimental configurations. C) Histogram comparing the peak capacity obtained for each of the experiments.

Peptide and Protein Identifications

From a proteomics perspective, researchers in the field are commonly interested in the number of peptides identified, either in terms of peptide spectral matches, unique peptides, or protein groups. As shown in Figure 3, the 75 cm column consistently resulted in the highest total number of peptides and protein identifications by at least a 7% margin. Whereas in the past, reproducibility among replicates was typically around 80%, in this study, the results are highly reproducible with less than 5% of the peptide/protein identifications for a given dataset not shared with any of the other replicates in all three replicates. Efficient chromatography allows not only for better separation of the analytes, but also improves ionization. Furthermore, the brighter source and higher transmission quadrupole of the Orbitrap Fusion Lumos MS, combined with the ultrafast (up to 20 Hz) and parallelized precursor ion isolation, fragmentation, and scanning in the dual linear ion trap, helped to ensure that most of the ions that were detected in the survey scan were subsequently efficiently fragmented, yielding a positive peptide identification. Therefore, a higher reproducibility in peptide and protein identifications was consistently achieved.

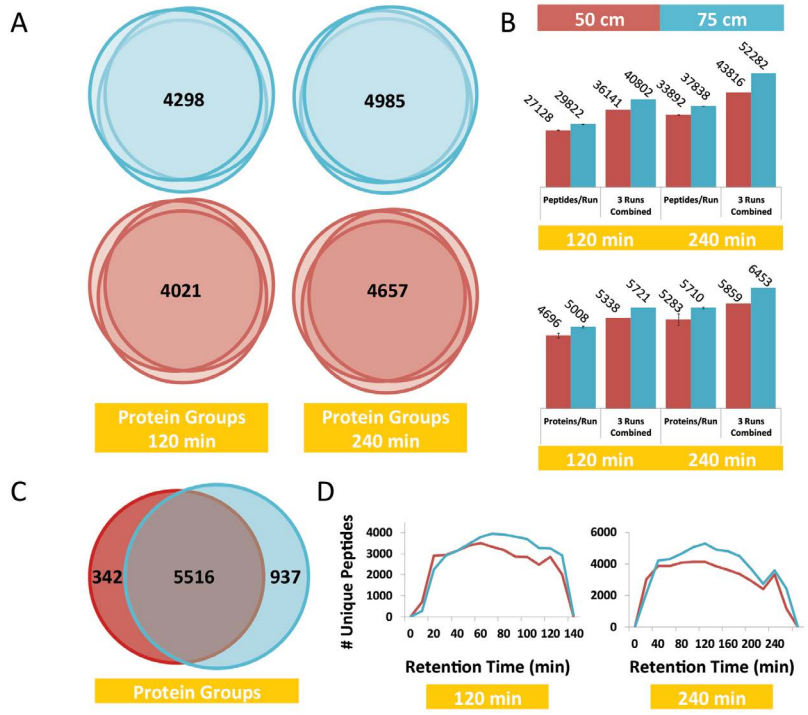

Figure 3. A) Venn diagrams showing the overlap among technical replicates for the identified proteins from varying column and gradient lengths. B) Number of identified peptide and protein groups. C) Venn diagram showing the total number of overlapped proteins for both column lengths. D) Line graph displaying the trend of identified peptides versus retention time during the LC-MS analysis.

We further investigated if the better performance in the peptide identification occurred only in certain parts of the LC-MS analysis or consistently across the entire gradient. As shown in Figure 3D, the 75 cm column provides better identifications across the entire gradient. This can be explained by the improved separation in the 75 cm column setup, which will cause a given peptide to elute at higher concentration and thus be more likely to yield a higher quality MS/MS spectrum. This in turn results in a positive identification. Figure 4A shows the rank of the identified and quantifiable proteins for the 4 hour gradient, as expected the longer column goes deeper into the proteome coverage.

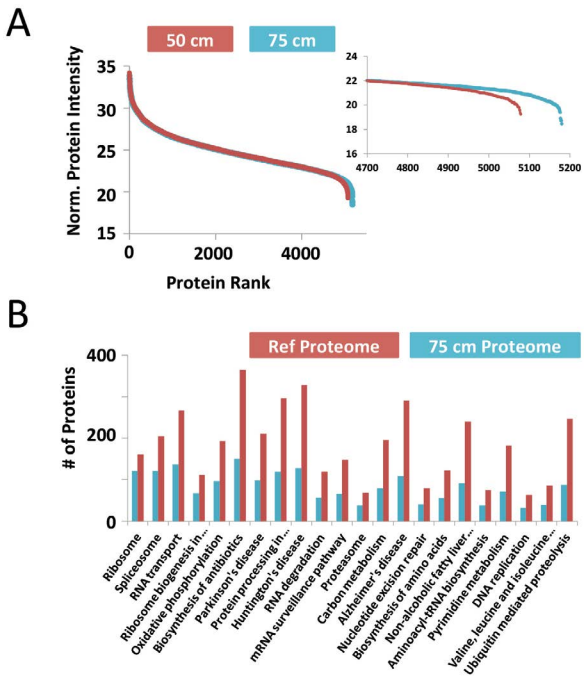

Figure 4. A) Protein rank of proteins over their normalized protein intensity for both column lengths. B) Overrepresented pathways for the 75 cm column length and 4 hour long gradient dataset.

It can be claimed that a 5% to 10% increase in peptide and protein identifications is not substantial. However, compared with other studies, these experiments represent a breakthrough in new levels of both peptide and protein identification and depth of coverage. Pathway analysis was then performed using Thermo Fisher Cloud. Results from both columns yielded the same profile of overrepresented pathways but with different degrees of coverage, demonstrating that the overall study was unbiased. Figure 4B shows that the data obtained for the 75 cm column with the 4 hour long gradient provides direct quantitation of almost 50% percent of the proteins in any of the 23 overrepresented pathways.

In addition, major improvements were achieved with regard to quantitation. Figure 5 shows that the 75 cm column increased the number of quantifiable peptides by 20%. This resulted in not only more peptides quantified but in a higher correlation among replicates (>85%), while at the same time those peptides and proteins had lower CVs allowing for more accurate quantitation. Finally, we examined the effect of the peptide loading amount on protein identification, quantitation, and impact in the retention time. Doubling the amount of peptide digest loaded onto the column did not significantly increase the number of protein identifications or quantifiable peptides. However, increasing the loading amount dramatically improved the correlation among runs up to 89% and doubled the number of proteins with CVs below 5%, allowing for more accurate proteome quantitation. The chromatography was not substantially affected, with observed retention time shifts between 1 µg and 2 µg loads of less than 1 min.

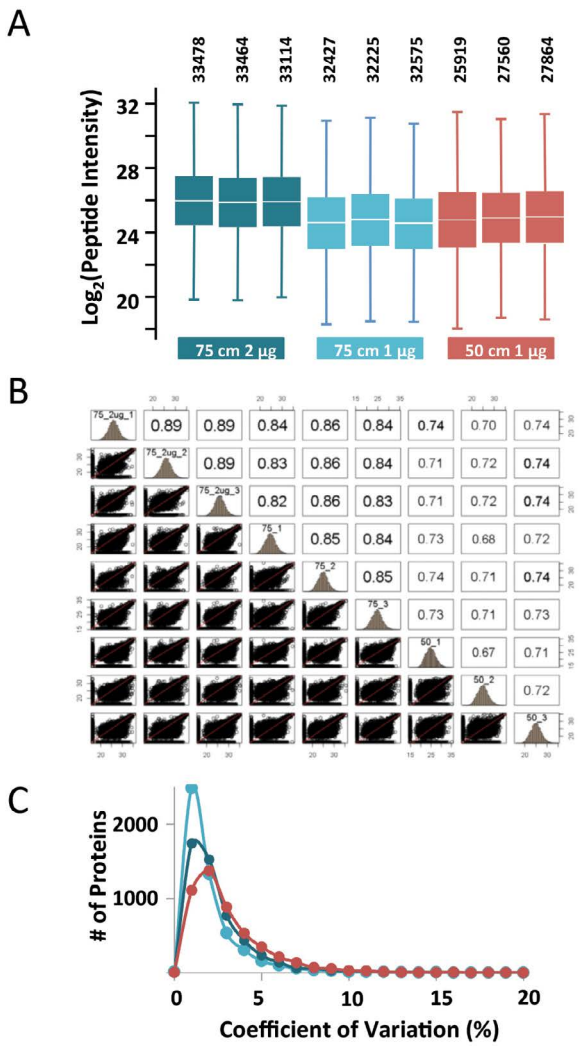

Figure 5. A) Box-and-whisker plot indicating the median peptide group intensity and distribution for increasing column length and load on column. B) Correlation plot showing the peptide peak area similarity between replicates and among other experimental conditions. C) Distribution of coefficients of variation (CV) for protein area quantitation among the different experimental conditions.

Accurate quantitation at the protein level enables profiling of critical pathways and systems, thereby allowing access to the underlying biological function information for the cell. To achieve this, the concentration of each protein has to be measured both accurately and reproducibly. To that end proteomic scientists have been developing more efficient mass spectrometric methods. A common method, known as data independent acquisition (DIA), relies on semi-targeted approaches that require building spectral libraries upon which to match ion fragmentation spectra in order to identify and extract quantitative information. In Figure 5, we show that very accurate quantitation can be obtained in a completely untargeted workflow using data dependent acquisition (DDA), which offers the same high-quality quantitative information. We demonstrated that the combination of long columns, state-of-the-art EASY-nLC systems, and state-of-the-art Orbitrap mass spectrometers result in the ideal platform for research and routine operations in protein quantitation, which is of critical value in emerging fields such as personalized medicine, clinical proteomics, and translational research.

Conclusion

The EASY-nLC 1200 system coupled with a high-performance Orbitrap mass spectrometer represents a very powerful platform for carrying out high-performance proteomic experiments:

- Increased number of peptide and protein identifications
- Increased identification rate with shorter gradients compared to 50 cm column analyses
- High sample loading capacity
- Increased number of proteins quantified
- Higher correlation of quantifiable peptides between injections and better run-to-run reproducibility

By systematically evaluating the most common gradients in the proteomic field, we have demonstrated that the use of longer columns employing 2 or 4 hour gradients represents a valuable alternative to perform quantitative proteomics compared to current methods. Moreover, we have identified approximately 6500 proteins without fractionation and reproducibly quantified over 5000 proteins based only on three technical replicate injections. These results clearly surpass the current standards in the proteomics paradigm. They rival quantitation results derived from DIA methods in terms of reproducibility and depth of analysis but are more efficient, as there is no need to first generate a spectral library.

References

1. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature*. **2014** May 29, **509**(7502), 582-7.
2. Hebert, A.S., et al. The one hour yeast proteome. *Mol Cell Proteomics*. **2014** Jan, *13*(1), 339-47.
3. Wu, X. et al. Global phosphotyrosine survey in triple-negative breast cancer reveals activation of multiple tyrosine kinase signaling pathways. *Oncotarget*. **2015** Sep 3, Epub ahead of print.
4. Livesay, E.A. et al. Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. *Anal Chem*. **2008** Jan 1, *80*(1), 294-302.
5. Cox, J.; Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem*. **2011**, *80*, 273-99.
6. Scigelova, M.; Hornshaw, M.; Giannakopulos, A.; Makarov, A. Fourier transform mass spectrometry. *Mol Cell Proteomics*. **2011** Jul, *10*(7), M111.009431.
7. 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** Jan, *24*(1), 148-53.
8. Köcher, T. et al. Development and performance evaluation of an ultralow flow nanoliquid chromatography-tandem mass spectrometry set-up. *Proteomics*. **2014** Sep, *14*(17-18), 1999-2007.
9. 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** Nov, 4(*11*), 923-5.
10. Polpitiya, A.D., et al. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics*. **2008** Jul 1, *24*(13), 1556-8.
11. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. **2010** Apr 1, *26*(7), 966-8.
12. Shen, Y. et al. Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics. *Anal Chem*. **2005** May 15, *77*(10), 3090-100.

AN64626-EN 0916S

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand