

Increasing Translational Proteomics Workflow Efficiency: Ultra-High Performance Liquid Chromatography with pSMART Data Acquisition and Processing

Scott Peterman, Thermo Fisher Scientific BRIMS Center, Cambridge, MA

Key Words

Nanoflow, Vanquish, UHPLC, Orbitrap Fusion Tribrid, HRAM, pSMART, translational proteomics, biomarker, diabetes research

Goal

To develop an efficient analytical method to perform robust and reproducible translational proteomics research. The method described here includes UHPLC separations, data acquisition using DIA strategies, and efficient, automated data processing designed to rapidly and confidently credential proteins for translation into subsequent, more targeted experimental methods for verification and validation.

Introduction

Translational proteomics research is designed to identify protein features differentially expressed between well-defined groups and/or donors.¹ Data acquisition methods are designed to perform global profiling to accurately characterize the sub-proteome. The results of the initial set of experiments are used to create efficient experimental methods for secondary verification, validation, and ultimately clinical assays after additional development and any necessary regulatory authorizations. Protein stratification is based on statistical and biological relevancy and requires increased cohort sizes to accurately assess biological variance and confidently identify those proteins exhibiting differential expression. As such, efficient experimental strategies are needed to accommodate large-scale studies while maintaining standardized and comprehensive protein/peptide profiling.^{2,3}

For these reasons, experimental efficiency is becoming a crucial factor in determining the desirability and usability of a translational proteomics research workflow. The increased sample size per study and nature of the samples requires implementation of more robust analytical methods that balance the amount of information obtained (accurate protein/peptide detection and quantification data) with throughput. Researchers have already begun to

investigate experimental efficiency using UHPLC separations at analytical flow rates.^{3,4} Comparative results show increased detection and quantitation capabilities using analytical flow rates by leveraging the significant increase in loading capacity.⁴

However, the analytical flow rates typically used with UHPLC systems compress peak widths to approximately 10–15 seconds or even less, challenging mass spectrometer data acquisition in translational proteomics workflows. To address this concern, pSMART data acquisition and processing was used to maximize automated and reproducible detection of components in narrow UHPLC peaks.⁵ The speed, sensitivity, dynamic range, and high-resolution accurate-mass (HRAM) capabilities of Thermo Scientific™ Orbitrap™ mass spectrometers are ideally suited to these analyses.

Experimental

Overview

The UHPLC research workflow applied the following:

- Sample preparation routines based on experimental requirements
- UHPLC system and columns chosen to increase experimental efficiency at analytical flow rates
- Study-specific spectral libraries created using data dependent acquisition (DDA)
- pSMART data acquisition to maximize peptide information obtained from chromatographic peaks
- Automated data processing software
- Statistical and biological analysis to identify protein groups

Sample Preparation

A small set of whole plasma samples were used for the comparative analysis. A 100 μ L aliquot was extracted from each plasma sample and digested using a standard trypsin protocol. A load amount of approximately 40 μ g was used for each analysis.

Liquid Chromatography

The Thermo Scientific™ Vanquish™ UHPLC system (System Base, Binary Pump H, Split Sampler FT, and Column Compartment H) was used for all UHPLC experiments at analytical flow rates. The UHPLC column was comprised of three individual 250 x 2.1 mm Thermo Scientific™ Acclaim™ 120 Å C18 columns of 2.2 µm particle size connected in series using the Thermo Scientific™ Viper™ fingertight fittings system. The columns were housed in the Vanquish column compartment and heated to 55 °C. A binary solvent system consisting of A) water with 0.2% formic acid and B) acetonitrile (MeCN) with 0.2% formic acid was used for sample analysis. A 20 µL injection volume was used to load 40 µg of the digested plasma directly onto the analytical column at 3% of solvent B.

The UHPLC method consisted of three steps: (i) sample loading and cleanup, (ii) the analytical gradient, (iii) and column washing and equilibration. The sample was cleaned at a flow rate of 400 µL/min for 10 minutes with 3% B. During this time the flow was directed into waste with a divert valve. For the second step, the flow path was diverted to the mass spectrometer and the solvent composition stepped to 8% B for 1 minute prior to the start of an analytical gradient of 0.49% per minute for 85 minutes at a flow rate of 200 µL/min. At the end of the gradient, the flow path was diverted to waste and the flow rate increased to 600 µL/min with 90% B for 3 minutes prior to changing the solvent composition back to 3% B for the last 4 minutes (still at 600 µL/min). Determined by retention time stability, 4 minutes was the shortest possible period of time for re-equilibration. The total injection cycle time was 110 minutes.

The nano LC method has been previously described and is briefly recapitulated here. Similar solvents were used for separations flowing at 700 nL/min. Samples were loaded onto a 50 x 0.15 mm trapping column packed with 5 µm PS-dvb particles (Agilent®) at 5% B. A 0.1 x 120 mm column packed in-house with C18 Aq particles (Bischoff) was used for analytical separations with a linear gradient (5–45% B) over 180 minutes before increasing the organic solvent to 90% to clean the column prior to re-equilibration for the next sample injection.

Mass Spectrometry

For peptide detection during the UHPLC experiments at analytical flow rates, a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer was used. Two different data acquisition experiments were performed. The first used DDA for the analysis of five different samples selected from each group with a top 10 MS/MS acquisition setting and with a resolving power of 120,000 and 15,000 for MS and MS/MS scans, respectively. A maximum ion fill time of 125 msec was used for both MS and MS/MS data acquisition. The data obtained were used to build a consolidated spectral library.

The second acquisition method used pSMART, which has been described previously.⁵ Similar Orbitrap mass spectrometer parameters were used for both the MS and data independent acquisition (DIA). MS/MS events and a loop count of ten (ten DIA spectra acquired after every

MS spectra) were used to meter DIA events in between MS scans. The total acquisition cycle acquired at least nine full-scan MS spectra. The goal was to acquire nine HRAM full-scan spectra and at least one DIA (MS/MS) spectrum for each peptide eluted.

Data Analysis

Using standard search parameters, an unbiased search of the human RefSeq NCBI Reference Sequence database was performed for all UHPLC DDA data using Thermo Scientific™ Proteome Discoverer™ software (revision 1.4). The search results were consolidated and used to create a spectral library. DDA uses the narrowest precursor isolation strategies, resulting in the highest probability that the product ion spectrum is representative of the isolated peptide. The DDA spectral library was imported into Pinnacle software loaded on a Velocity server (both from Optys Technologies, Philadelphia, PA). The DDA library was then used to search all DIA spectra using a spectral matching routine, as well as the measured retention time, precursor charge state, possible modifications, and product ion distribution information (product ion *m/z* values, charge states, and relative abundance values). Pinnacle software was used for all statistical analyses.

Comparison of UHPLC and Nano LC

The results of the UHPLC research workflow were compared to the data previously obtained using the pSMART data acquisition method.⁵ Other than the chromatographic method (LC system and column) applied, the primary differences between the workflows compared were the amount loaded on column (1 µg of digest for nano LC). For consistency, Pinnacle software was used to search the nano LC data against the same spectral library used to search the UHPLC data.

Results and Discussion

The goal when performing large-scale global protein profiling experiments is to significantly increase throughput without decreasing the robustness of global qualitative and quantitative analysis. Higher throughput enables data analysis of larger sample cohorts, which in turn, provides the statistical power needed to stratify various proteins and to identify putative panels that can be rapidly transitioned from global profiling to targeted methods.

Throughput and Chromatographic Performance

Chromatography is the most important parameter that affects sample throughput. To significantly reduce the injection cycle time, thus increasing the number of samples that can be measured per day, higher flow rates were used for sample analysis. Figure 1 shows comparative base peak chromatograms for the analysis of digested plasma. The UHPLC research workflow resulted in much higher throughput compared to the nano LC workflow. The total experimental cycle time for the UHPLC method was 110 minutes, of which 95 minutes was data acquisition time. In the nano LC experiments described previously, the total experimental cycle time was 210 minutes, of which data acquisition time was about 160 minutes. Comparatively, the UHPLC method reduced the individual sample analysis time by 100 minutes.

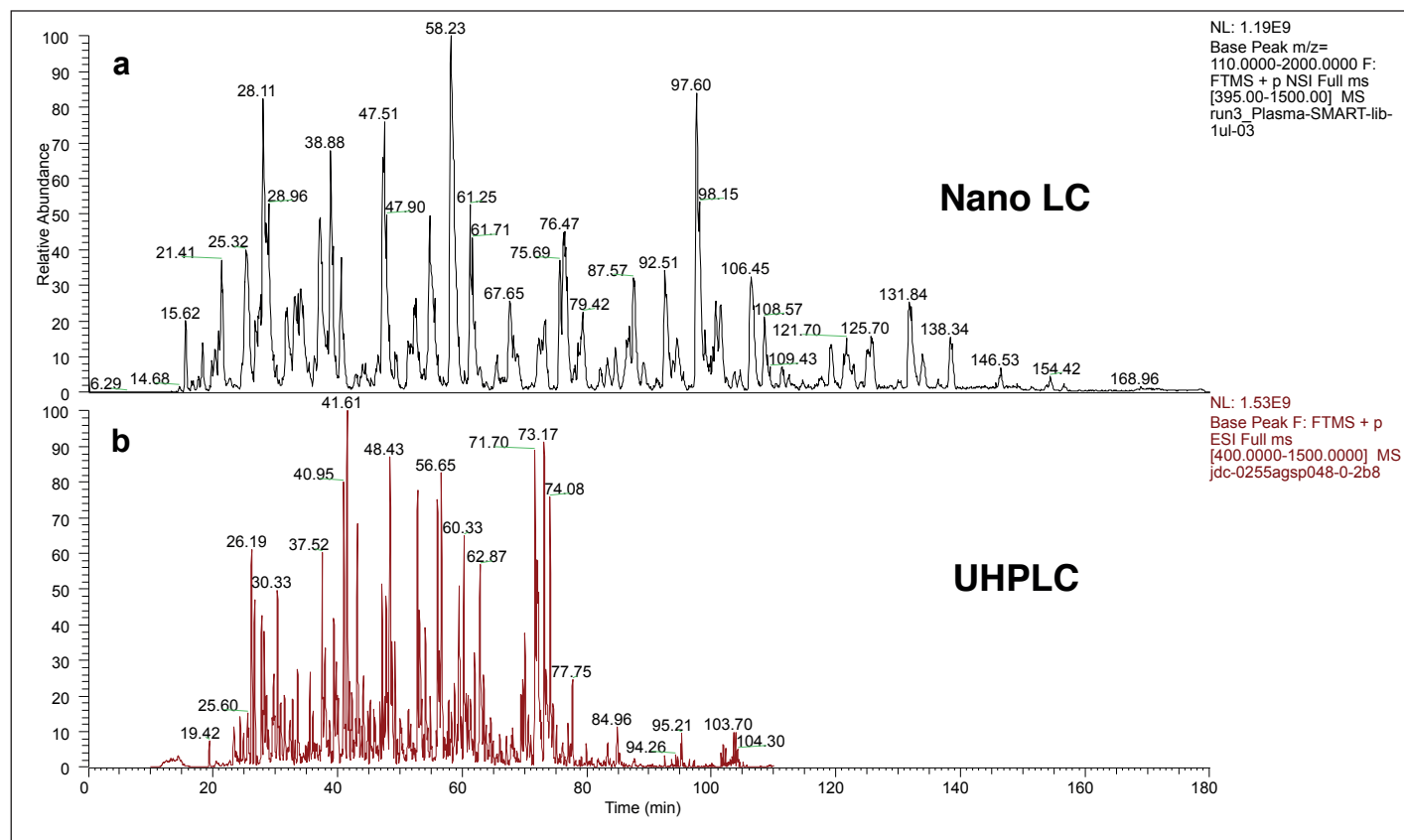


Figure 1. Base peak chromatograms of plasma tryptic digests using (A) nano LC and (B) UHPLC at analytical flow rates. A total of 1 μ g of plasma digest was injected on the 0.1 mm ID column for the nano LC experiments compared to 40 μ g of plasma digest for the UHPLC experiments.

UHPLC separations of peptides at analytical flow rates provide many benefits. Higher flow rates ($> 50 \mu\text{L}/\text{min}$) overwhelm gradient delay volumes in the flow path, enable faster loading and washing steps, as well as column cleaning and equilibration prior to injecting the next sample. The Vanquish UHPLC system pump has a maximum backpressure of 1500 bar, thus higher flow rates can be implemented to reduce the time needed for pre- and post-gradient steps despite the long column arrangement. For the UHPLC experiments, 400 and 600 $\mu\text{L}/\text{min}$ flow rates were used for non-gradient steps, reducing sample injection cycle time by approximately 30 minutes as compared to previous nano LC methods.

The analytical flow rates also increased gradient performance for robust qualitative and quantitative peptide analysis. The fast column response and reduced gradient delay enabled the use of flatter gradient profiles over shorter timeframes to maximize peak capacity and retention time stabilities. The average retention time variance for all measured peptides was 0.15% over the entire study. The change in chromatography also significantly reduced chromatographic peak widths of approximately 1 minute measured for the nano LC experiments to an average of 0.18 minutes (approximately 11 seconds) for the UHPLC experiments (Figure 2). The comparative distribution showed that the UHPLC research method results in a more uniform peak width for all detected peptides (Figure 2). The higher chromatographic resolution also significantly reduced co-elution despite loading 40 μ g of plasma digest on column compared to 1 μ g for the nano LC workflow.

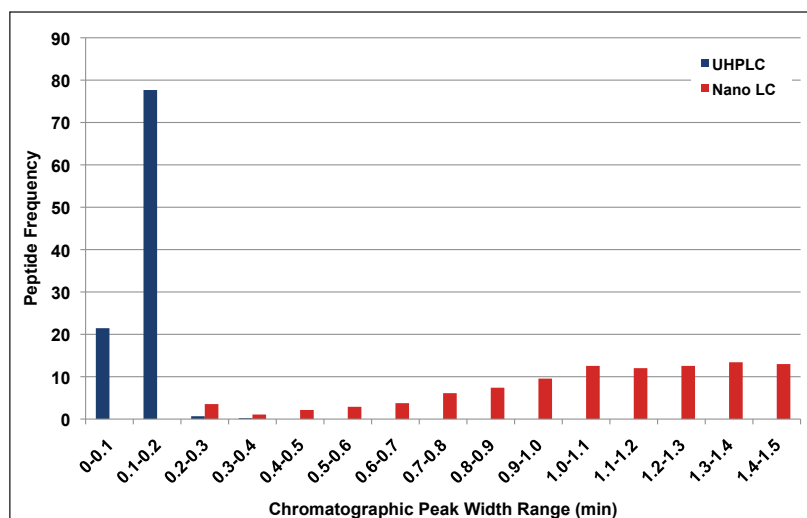


Figure 2. Histogram showing the distribution of chromatographic peptide peak widths obtained using UHPLC and nano LC. The peak widths were measured at 10% peak height to reduce the effects asymmetrical peak shapes obtained using nano LC.

Mass Spectral Acquisition

The benefits realized using UHPLC separations must be balanced with robust mass spectral data acquisition. Global peptide profiling must be maintained to provide an opportunity to identify putative biomarkers, which requires confident peptide sequencing and quantitation. Thus, for a 12-second wide peak (at the baseline), the overall cycle time would have to be approximately 1.7 seconds to acquire the minimum number of 7 data points per precursor, regardless of whether the data are MS or DIA data. The pSMART acquisition was used due to its flexibility in performing global qualitative and quantitative analysis despite narrow peak widths.

The pSMART research method decouples the data used for quantitation (HRAM MS) and sequencing (narrow DIA) to maintain required acquisition cycle times. In addition, the pSMART method takes advantage of the unique data acquisition capabilities of the Orbitrap Fusion Tribrid mass spectrometer (as well as the Q Exactive mass spectrometer) to maintain accurate global quantitation and reproducible sequencing across a wide precursor mass range. Specifically, the trapping function and high charge density provided by the C-trap facilitates large intra-scan dynamic range for both MS and DIA detection. In addition, the high resolution used for MS detection provides sufficient selectivity for most precursors, as well as for complex DIA spectra.

Figure 3 presents the pSMART data acquisition results for the targeted peptide LWAYLTINQLLAER across ten sample injections. Despite systematic acquisition of the narrow DIA scan events containing each precursor (triangles), only those DIA spectra containing matched fragment ions for the peptide LWAYLTINQLLAER are displayed within the blue “up-triangles,” marking the target peptide retention time. The extracted ion chromatographic traces (XICs) represent at least four precursor isotopes per detected charge state and were used for qualitative and quantitative analysis of each peptide.

Despite the narrow peak width, all injections have at least two matched DIA spectra for sequence confirmation due

to sampling the +2 and +3 charge states. Note the relative retention time values measured across the ten injections. The retention time variance was approximately 2.4 seconds across 134 injections, which was in agreement with the average variance calculated for all peptides.

Increased chromatographic resolution also improves DIA mass spectral quality. The narrower chromatographic peak width reduces the probability of co-eluting peaks in any one narrow DIA precursor isolation window. Figures 4 and 5 demonstrate the effects of peak width on detection and sequencing of targeted peptides. The results compared are for the targeted peptide LWAYLTINQLLAER where Figures 4 and 5 show the results obtained from the nano LC and UHPLC experiments, respectively. The averaged HRAM MS spectrum shows a typical spectrum of targeted peptides from a plasma digest with generally one abundant peptide and numerous low-level peptides. The spectrum in Figure 4A shows the base peak having an averaged measured intensity of $1.3e7$ (for the targeted peptide LWAYLTINQLLAER at 129.14 minutes shown in Figure 4B) compared to the averaged measured intensity for the targeted peptide of $5.7e4$ (Figure 4C). In addition to the wide dynamic range, any of the co-eluting peptides in the immediate precursor m/z range would be co-isolated and fragmented during the DIA spectral acquisition.

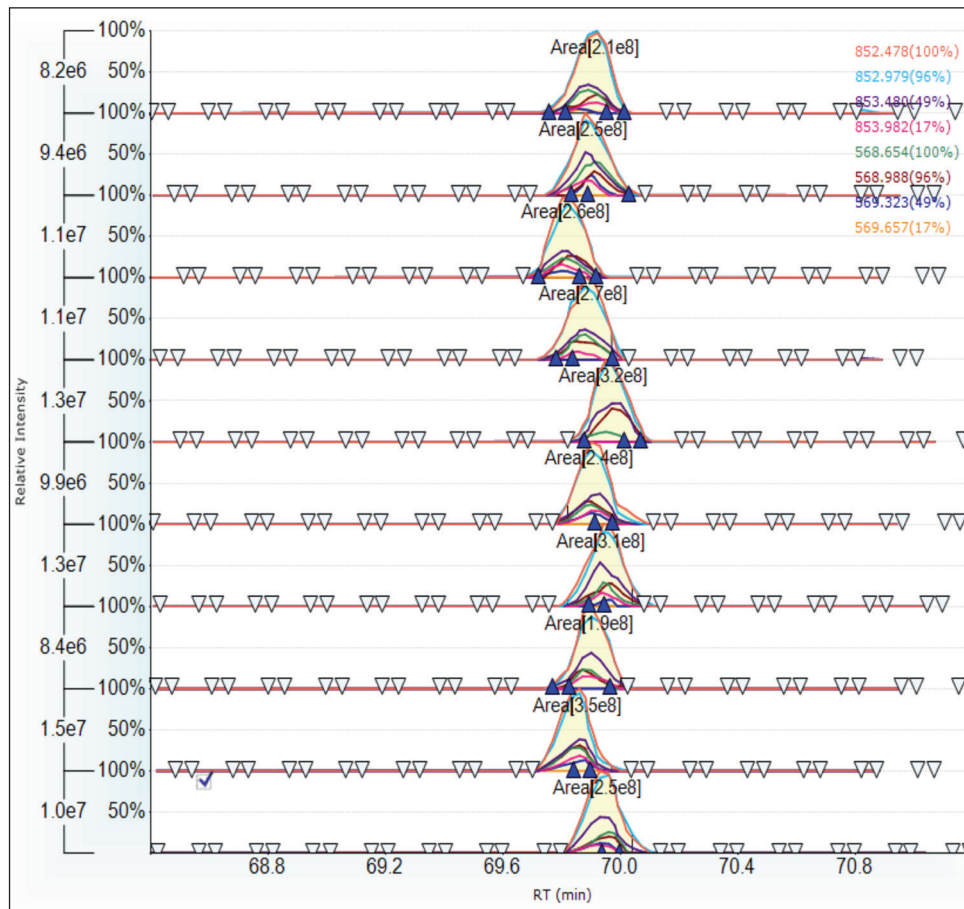


Figure 3. Repetitive analysis of the peptide LWAYLTINQLLAER using pSMART data acquisition. The ten XIC profiles represent ten different biological samples separated using the UHPLC method. The legend at top right matches the precursor m/z value with XIC line color for the three most abundant isotopes for the +2 and +3 charge states. The triangles represent DIA scan events resulting from specific precursor isolation.

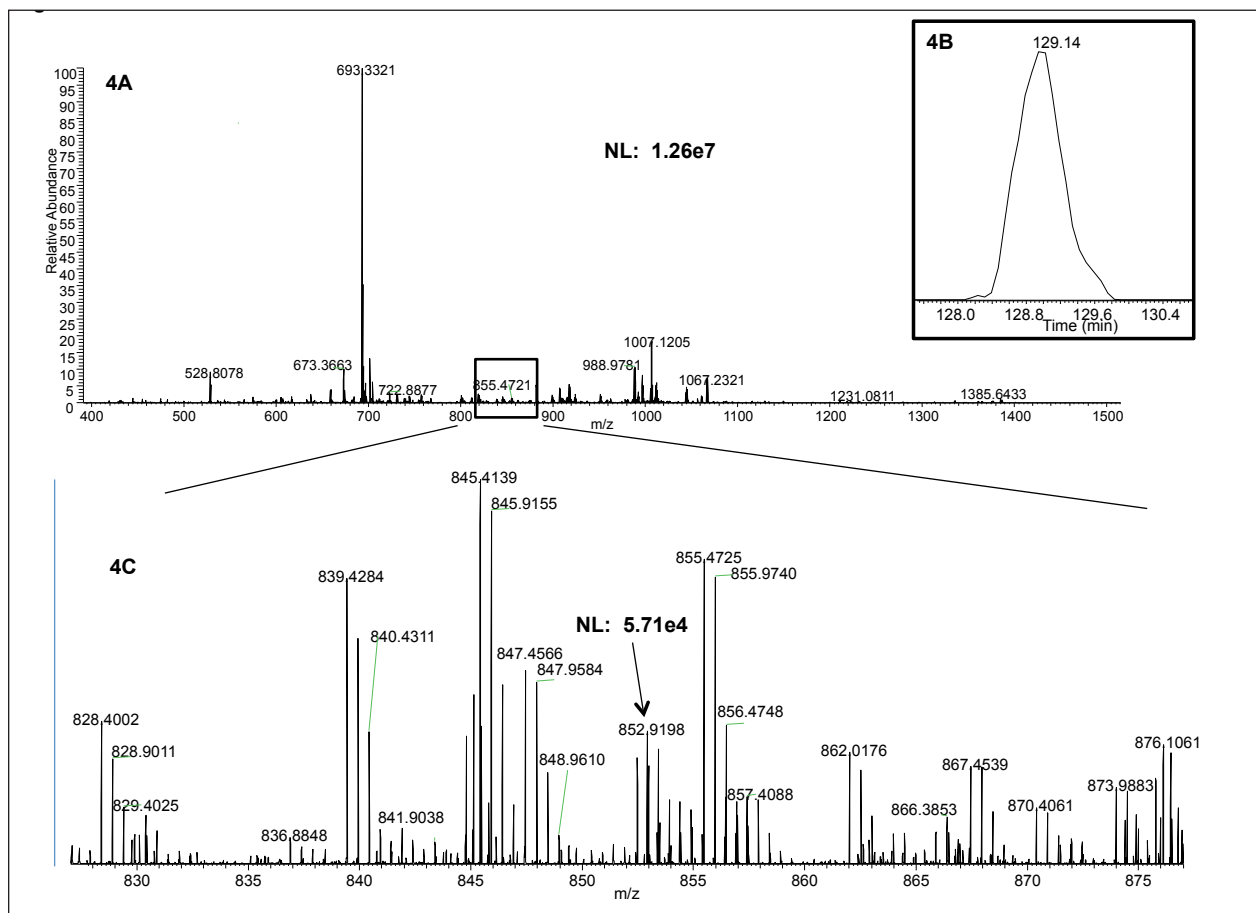


Figure 4. Effects of peak capacity on sequencing targeted peptides using nano LC methods. (A) The full-scan HRAM MS was averaged across the 14 spectra acquired under the precursor elution profile (B) for the targeted peptide LWAYLTINQLLAER. (C) The narrow mass range surrounding the targeted precursor isotopic cluster.

Figure 5 shows the precursor analysis of the same targeted peptide (LWAYLTINQLLAER) using analytical flow rates. With respect to the wide dynamic range, the full-scan MS observed using UHPLC showed a similar profile to that observed in the nano LC experiments, but there were fewer peptides measured across the entire mass range. In addition, the measured ion intensity difference between the base peak and targeted peptide was approximately 100-fold. In comparison, the difference observed for the nano LC experiments was approximately 300-fold. The resulting narrow mass window around the targeted precursor also showed significantly fewer co-eluting peptides.

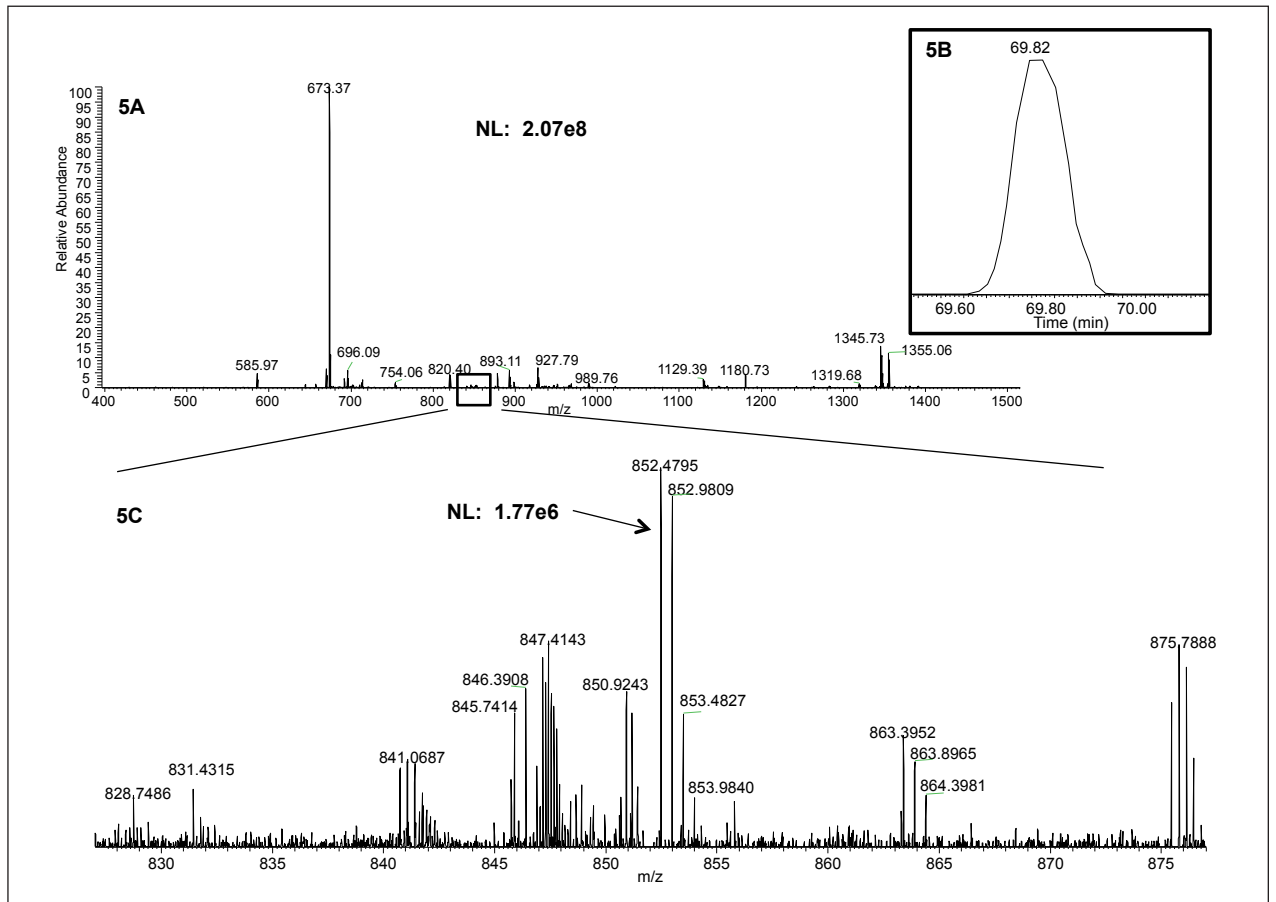


Figure 5. Effects of peak capacity on sequencing targeted peptides using analytical UHPLC methods. (A) The full-scan HRAM MS is averaged across the 10 spectra acquired under the precursor elution profile (B) for the targeted peptide LWAYLTINQLLAER. (C) The narrow mass range surrounding the targeted precursor isotopic cluster.

Figures 6 and 7 compare three full-scan DIA spectra collected using the two different chromatographic research methods. One spectrum was taken per injection to ensure reproducibility. The blue triangles represent matched fragment ions consistently measured in each spectrum. The red triangles represent fragment ions not measured in every spectrum. Figure 6 shows the nano LC

experiment results. Here the fragment ions attributed to the targeted peptide were still measured, but generally their relative abundance was approximately 1 to 5% of the other fragment ions. The calculated dot product correlation coefficient was greater than 0.7, indicating excellent matching to the spectral library entry confirming the peptide identification.

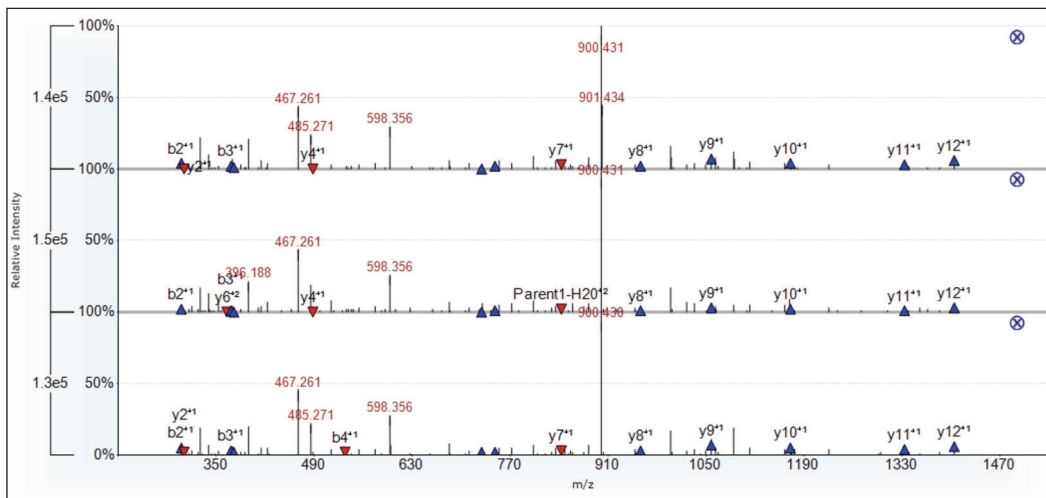


Figure 6. Full-scan HRAM DIA spectra for the targeted peptide LWAYLTINQLLAER acquired for three replicates using nano LC.

In comparison, the product ions were routinely the most abundant fragments obtained in the UHPLC experiment (Figure 7). As a result, the UHPLC research method yielded much better product ion distribution correlation with the spectral library (0.94 vs. 0.73 where a value of 1.0 indicate the best fit), with more fragments identified and measured for all injections. In addition, both precursor charge states were confidently sequenced across most samples, providing increased confidence in results despite the narrow peak widths.

Comparative Protein Coverage

Analytical columns provide greater loading capacities compared to capillary columns, which are ideal for the analysis of plasma/serum samples. In the UHPLC experiments, 40 μ g of plasma digest was loaded on column compared to only 1 μ g for the nano LC experiments. The column arrangement maintained high peak capacities, resulting in an increased dynamic range and coverage per protein as determined by the comparative number of proteins confidently profiled. Specifically, 219 proteins and 2349 peptides were profiled using the UHPLC research method compared to 181 proteins and 1834 peptides using the nano LC method. Considering the very high dynamic range (complexity) of plasma, the UHPLC separations provided a significant benefit in overcoming the signal suppression often produced by highly abundant peptides.

Table 1 shows that the UHPLC method produced much greater coverage per high abundance protein⁴ as well as similar AUC responses, although the rank orders were slightly different. The increased quantitative coverage enables improved evaluation of the protein response; that is the uniformity of the measured expression for all confidently identified peptides per protein, across all biological groups. Secondly, the increased capacity for confident detection and quantification facilitates secondary data processing for peptide variants (truncation, SNPs, PTMs, etc.) without having to alter the sample preparation routine.

Similarly, Table 2 compares the protein coverage and measured response found for low abundance proteins. The AUC values were calculated per peptide using precursor isotopic AUC values and then summed per protein for all confidently identified peptides. Even though the ratio of flow rates and column ID using the analytical flow rate should have led to approximately a 400-fold sensitivity decrease, each protein showed significantly more peptides confidently measured and the summed AUC values were significantly higher. The most likely reason is the increased peak capacities reduced the probability of co-elution with more abundant peptides, resulting in ion suppression.

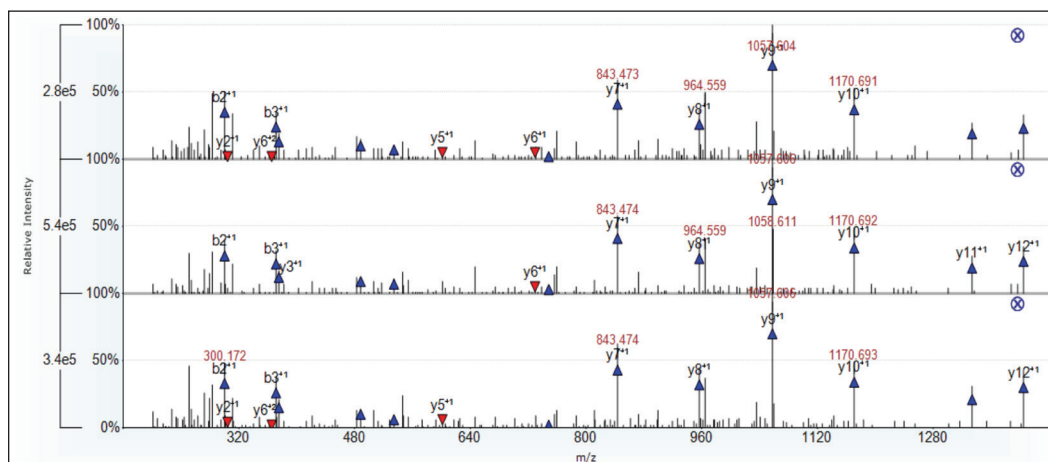


Figure 7. Full-scan HRAM DIA spectra for the targeted peptide LWAYLTINQLLAER acquired for three replicates using UHPLC.

Table 1. Comparison of peptides reproducibly measured for abundant proteins. All peptides considered were scored in the highest category. The summed AUC values were obtained using pSMART data acquisition with nano LC and UHPLC.

Protein	Nano LC		UHPLC	
	Number of Peptides	Summed AUC Values	Number of Peptides	Summed AUC Values
Serum albumin	107	1.2e12	195	1.1e12
Serotransferrin	49	1.6e11	95	1.6e11
Apolipoprotein A-1	17	1.2e11	48	6.6e10
Alpha-2-macroglobulin	73	1.1e11	100	1.4e11
Complement C3	71	9.9e10	181	1.3e11
Fibrinogen Gamma Chain	30	7.0e10	40	4.0e10
Alpha-1-antitrypsin	28	6.5e10	56	8.9e10
Fibrinogen Beta Chain	28	5.8e10	62	4.3e10
Fibrinogen Alpha Chain	29	3.7e10	59	3.9e10
Apolipoprotein B-100	60	2.3e10	246	4.5e10

Table 2. Comparison of peptides found for low abundance proteins in plasma digests and summed AUC values obtained using pSMART data acquisition with nano LC and UHPLC.

Protein	Nano LC		UHPLC	
	Number of Peptides	Summed AUC Values	Number of Peptides	Summed AUC Values
Leucine-rich α -2 Glycoprotein	4	2.7e8	15	1.5e9
Complement C2	4	2.6e8	27	6.2e9
Lumican	2	2.2e8	12	6.1e8
Coagulation Factor XIII B	4	2.1e8	12	2.5e8
Serum Amyloid P-Component	2	2.0e8	10	2.5e8
Kallistatin	2	2.0e8	15	9.9e8
Fibulin-1	3	1.8e8	11	2.4e8
Complement Factor H	2	1.6e8	15	1.3e9
Thyroxine-binding Globulin	2	1.3e8	14	7.3e8
Ficolin-3	2	7.7e7	7	3.0e8

Conclusion

The UHPLC research workflow presented addresses many of the requirements of efficient translational proteomics experiments. The 1500 bar Vanquish UHPLC system operated using normal flow rates, three coupled high-resolution columns, in combination with the pSMART data acquisition method, and Pinnacle software resulted in reproducible protein/peptide profiling across large numbers of samples. Incorporation of analytical flow rates and wide-bore UHPLC columns significantly increased chromatographic performance based on consistent, narrow peak widths across the entire gradient, low retention time variance across the entire study, and peak capacities comparable to nano LC experiments using much longer gradients. In addition, the columns used for this study facilitated much greater loading capacities, which increased dynamic range, protein detection, and coverage as compared to previous nano LC experiments. In particular, the UHPLC separations provided a significant benefit in overcoming signal suppression often produced by highly abundant peptides in complex plasma samples.

The pSMART data acquisition method demonstrated accurate global quantitation while reproducibly collecting narrow DIA spectra across a wide precursor mass range despite 9 to 12 s wide peaks. The large amount of data produced (approximately 92 GB) was efficiently and confidently processed in approximately 90 minutes using Pinnacle software. Robust peak picking, scoring, and integration significantly reduced the time needed for data analysis. The statistical analysis performed by Pinnacle software stratified proteins across the multiple groups and automatically credentialed a subset of proteins for export to auxiliary programs to determine biological relevancy and to automatically create more targeted analytical methods.

For research use only. Not for use in diagnostic procedures.

To find a local representative, visit:

thermofisher.com/ProteinBiomarkers

References

1. Yang, P.; Patrick, E.; Tan, S-X.; Fazakerley, D. J.; Burchfield, J.; Gribben, C.; Prior, M. J.; James, D. E.; Yang, Y. H. Direction pathway of large-scale proteomics data reveals novel features of the insulin action pathway. *Bioinformatics*. 2014, 30(6), 808-814.
2. Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics*. 2012, Jun;11(6).O111.016717.
3. Lopez, M. F.; Peterman, S.; Sarracino, D.; Vogelsang, M.; Prakash, A.; Krastins, B.; Byram, G.; Vadali, G.; Mohiuddin, M.; Muraca, P.; Goldfine, A. B.; Patti, M. E. Rapid discovery of differentially abundant proteins in T2D plasma samples using improved UHPLC chromatography and pSMART data acquisition. US HUPO, Tempe, AZ. 2015.
4. Percy, A. J.; Chambers, A. G.; Yang, J.; Domanski, D.; Borchers, C. H. Comparison of standard- and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. *Anal. Bioanal. Chem*. 2012, 404, 1089-1101.
5. Prakash, A., Peterman, S., Ahmad, S., Sarracino, D., Frewen, B., Vogelsang, M., Byram, G., Krastins, B., Vadali, G., Lopez, M. Hybrid data acquisition and processing strategies with increased throughput and selectivity: pSMART analysis for global qualitative and quantitative analysis. *Journal of Proteome Research*. 2014, 13(12), 5415-5430.

Thermo
SCIENTIFIC

A Thermo Fisher Scientific Brand