Biomarkers and Omics

SCIEX
The Power of Precision

Accelerating Protein ID for Deep Proteome Profiling

Fractionation with Microflow LC on TripleTOF® 6600 Systems

Nick Morrice[1], Joerg Dojahn[2], Christie Hunter[3]
[1]SCIEX, UK, [2]SCIEX, Germany, [3]SCIEX, USA

The complexity of a proteomics sample after digestion is extremely high requiring that extensive fractionation is done to deeply interrogate the proteome. The key goal is to spread the peptides out across fractions such that when each is analyzed by LC-MS/MS, the mass spectrometer has time to collect high quality MS/MS spectra on as many peptides as possible. Typically when more fractions are collected, more protein identifications are obtained. The downside is that the more fractions collected to increase depth of coverage, the more instrument time is needed to analyze all the fractions.

Typically, fractions are analyzed using nanoflow LC with 2-3 hour gradients making the total instrument time required very long. However moving to microflow LC provides a big increase in ease of use and robustness, as well as faster chromatography. Here, the use of fractionation coupled with microflow LC for analysis was explored to determine whether protein identification experiments could be accelerated while still obtained good proteome coverage.

Microflow DDA for Fast Protein Identification

- Combining high speed acquisition with high quality separation and deep database searching enabled the confident identification of **over 7100 proteins (over 106 000 peptides) in just 15 hours of MS time**

- TripleTOF® 6600 System
  - High sensitivity, resolution and speed of MS/MS acquisition enables up to 100 Hz MS/MS acquisition rates for data dependent acquisition

- NanoLC™ 425 System provides robust, high quality separations from nanoflow to microflow rates
  - 1 hour total run time per fraction with microflow LC – 15 hours of MS time generate a large number of protein identifications

- ProteinPilot™ Software 5.0[1]
  - Powerful Paragon™ and Pro Group™ Algorithms for protein identification with enhanced scalability for large protein datasets
  - ProteinPilot Report[2] for detailed analysis of identification metrics

Summary of Identification Yields

Data Level	FDR Type	FDR	ID Yield
Protein	Global	1%	7164
		5%	7606
		10%	8056
Distinct peptide	Global	1%	106849
		5%	124257
		10%	136225
Spectral	Global	1%	189476
		5%	220926
		10%	242776

Figure 1. Protein Identifications Results for a Hela Cell Lysate. 15 high pH fractions of a digested Hela cell lysate were analyzed using microflow LC and DDA on a TripleTOF® 6600 system. Using ProteinPilot™ Software 5.0, a combined database search of all 15 DDA fractions identified over 7100 proteins and over 106 000 peptides at a 1% global FDR.

Methods

2D LC Fractionation: A digest of a HeLa human cell lysate (1mg, gift from the Beatson Institute for Cancer Research, Glasgow UK) was fractionated using high pH reverse phase chromatography using a Shimadzu Nexera system (UV detection at 214nm). A Durashell RP column (250 x 4.6mm, 5µm, Agela Technologies) operating at 1 mL/min was used, running a gradient of 2 – 90% acetonitrile in 2mM ammonium hydroxide. 15 fractions were collected every 2 mins.

1D LC-MS/MS Chromatography: Each subsequent peptide fraction (5% of digested fraction, ~3-5 µg on column) was then separated using low pH reverse phase gradient on the NanoLC™ 425 system operating in microflow mode. A Triart C_{18} 150 x 0.3mm column (YMC) was used at 5 µL/min flow rate with a 45 min gradient from 2-40% acetonitrile in 0.1% formic acid for a total run time of 1 hour per fraction.

Mass Spectrometry: The eluent was analyzed using the TripleTOF® 6600 system equipped with a DuoSpray™ Source and a 25 µm ID hybrid electrode. Data dependent acquisition (IDA) was performed with 30 MS/MS per cycle, each with 40 msec accumulation.

Data Processing: All IDA data was processed using ProteinPilot™ Software 5.0, both individually and all in one large search. Identification results were visualized from the full ProteinPilot Reports that were automatically created.
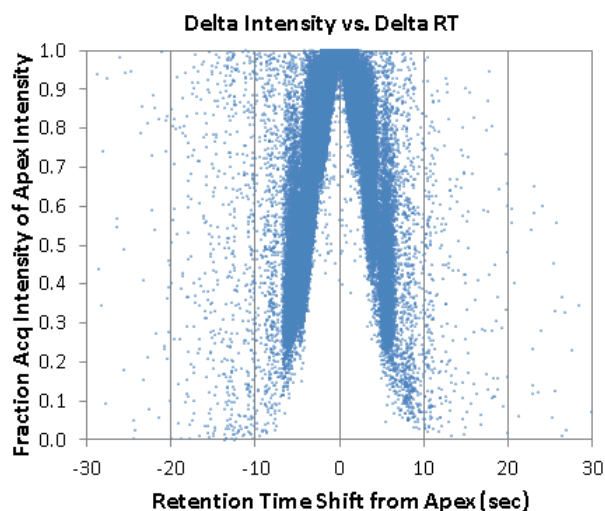

Figure 2. First Dimension Fractionation Provide Good Distribution of Peptides Across Fractions. On average, 9000 peptides were confidently identified from each of the 15 fractions (using 1% global FDR as a cutoff) as seen in the inset. The total number of peptides that were identified across all fractions was 106849 (red filled circle). Taking into account the # in each fraction, this suggests that only 25% of peptides were redundant across fractions highlighting good fidelity in the first dimension of fractionation.

High Quality Separation is Key to Protein Identification

The complexity of proteomic digests is extremely high therefore good fractionation in the first dimension is critical when the goal is deep protein identification. Using high pH fractionation in the first dimension, good even peptide separation was obtained (Figure 2) with ~9000 peptides identified in each fraction. Peak separation was also good as only about 25% of peptides were found in multiple fractions, providing 106849 distinct peptides in total.

High quality analytical separation is also important for good identification rates and this is easily obtained using microflow LC. The median peak with on the analytical column was 7.2 seconds at half height. As this is data dependent acquisition, the point at which MS/MS is triggered on each LC peak can be variable and this will impact the resulting intensity / quality of each spectrum. However when peak shape is sharp it is more likely that an MS/MS will be collected closer to the LC peak apex where intensity is best, this was the case in this study as indicated in Figure 3. Here the point of MS/MS acquisition relative to the LC peak apex is plotted and shows that 90% of MS/MS spectra are collected at 50% peak height or greater.

Figure 3. Acquiring MS/MS Spectra Near the LC Peak Apex. Plotting the retention time distance from each LC peak apex vs the fraction of total intensity at each trigger provides an interesting look at how the quality of chromatography affects data depending acquisition. A large majority of peptides are sent for MS/MS during the upper half of their elution profile, where ion statistics for the collected MS/MS spectrum will be better. Improved MS/MS quality means more identifications and a higher quality SWATH® acquisition library.
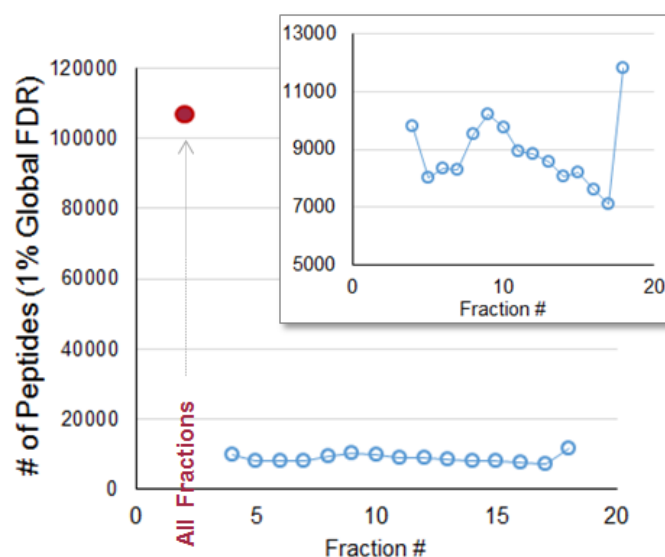
The quantity and the quality of the MS/MS spectra acquired is also key for identifying as many peptides and hence proteins from the sample. The quality of the MS/MS spectrum acquired can be reflected in the score from the subsequent database search. Here, the score distribution (Figure 4, top) indicated that the majority of MS/MS spectra acquired were of high quality, score of 8 or more (blue). When the distribution for the MS/MS spectra that resulted in a confident peptide ID is plotted you can see a slight shift of the distribution to higher score. This shows that the overall quality of most spectra is very good. Combined with the large search space covered by ProteinPilot™ Software, 83% of spectra collected was matched to the proteins / peptides at 1% global FDR. This means many confident peptides are identified for each protein as reflected in Figure 4 (bottom). The majority of proteins have 2 or more peptides identified for higher confidence.

Conclusions

While most proteomics experiments today are quantitative in nature in order to deeper understand complex biology, it is still important to have good protein identification workflows for the generation of spectral libraries or to quickly confirm the identity of the major components. Here, a strategy for accelerated protein identification is demonstrated using microflow LC on the TripleTOF® 6600 System. The throughput and ease of use of the higher flow LC approach reduces the instrument time required to perform these experiments, while maintaining high results quality. Using just 15 hours of instrument time, the high quality separation combined with the high speed acquisition of the TripleTOF 6600 system, enabled the identification of over 7100 proteins (over 106 000 peptides) from a digested Hela cell lysate.

References

1. ProteinPilot™ Software Overview, SCIEX Technical note RUO-MKT-02-1777-A.

2. ProteinPilot™ Report for ProteinPilot™ Software - Detailed Analysis of Protein Identification / Quantitation Results Automatically, SCIEX Technical note RUO-MKT-02-1778-A.
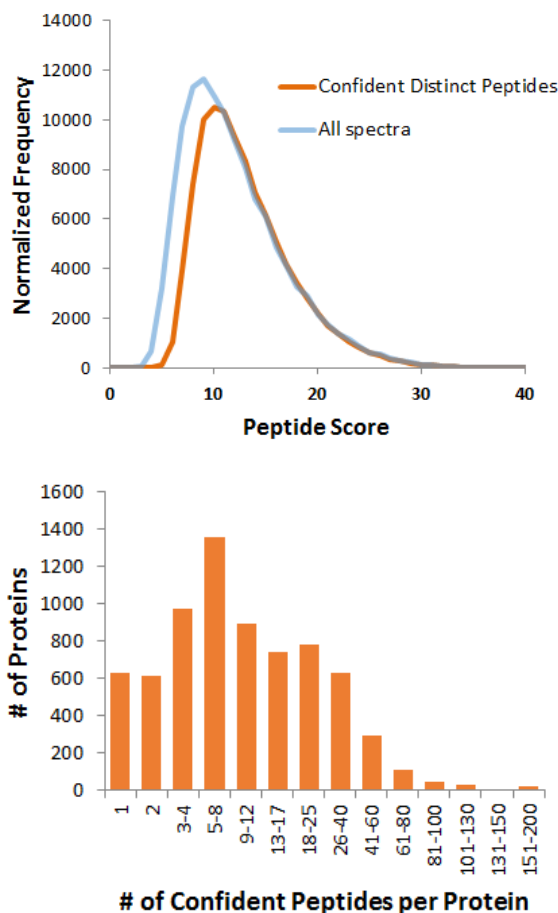

Figure 4. Quality of Data. (Top) The peptide score is a count of the fragment ions in the MS/MS spectrum that match the theoretical ions, the more matching ions loosely correlates with spectral quality. Plotting the frequency of the scores for confident peptides (orange) shows that the majority of spectra that resulted in a good ID had good scores of 10 or more from the Paragon™ Algorithm. The distribution of all spectra (blue) shows only a slightly lower distribution highlighting the overall good quality of the MS/MS collected. (Bottom) Histogram of the number of confident peptides (<95%) identified per protein shows that the majority of proteins are identified with 2 or more peptides.