Rapidly advance quantitative proteomics with a high-throughput SWATH® Acquisition solution

Microflow LC with OptiFlow® Turbo V Source on the TripleTOF® 6600+ System and cloud processing

Alexandra Antonoplis, Christie Hunter
SCIEX, USA

The combination of microflow LC with SWATH Acquisition for accelerating quantitative proteomics studies is becoming increasingly more widespread, due to the improved robustness and throughput obtained relative to the traditional nanoflow LC approach. Previously an investigation was done to characterize the tradeoff between sensitivity and throughput/robustness when switching from nanoflow to microflow for global protein quantitation experiment using SWATH Acquisition.[1] It was demonstrated that with only 4x more total protein on column similar results could be achieved with much higher throughput and robustness.

The impact of shortening microflow gradient length on the number of proteins/peptides quantified from a SWATH Acquisition dataset was also studied, providing a good guidance for researchers when selecting the optimal chromatographic strategy for a study.[2]

Due to these method optimizations, microflow SWATH Acquisition datasets can be generated with high throughput (up to 100 samples per day, depending on the selected gradient length).[2] This will put high pressure on the downstream data processing tools and results generation, to keep pace with the generation of datasets. Thus, processing of Omics datasets in the cloud is becoming increasingly important, to handle the speed and scale of today's industrialized proteomics approaches.

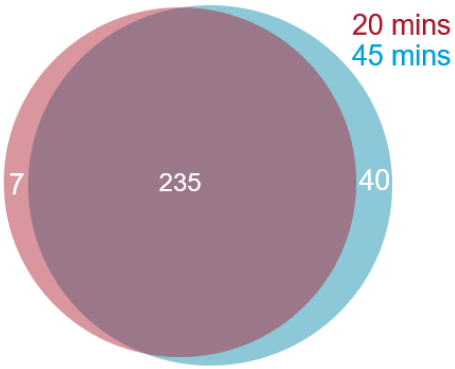

Figure 1. Differential proteins found between 20 and 45 min gradients. Focusing on two of the cell lines in the study, HEK vs HEK_MG132 cell lines, 235 proteins were found to be differentially expressed by both gradient lengths and similar numbers of proteins were quantified (Table 1). The measured fold changes were compared to assess quantitative accuracy (Figure 3).

In this work, using fast microflow gradients, a small set of cell lines were studied to further characterize the impact of fast gradients on quantitative accuracy. Data was collected using the OptiFlow Ion Source on the TripleTOF 6600+ System at two different gradient lengths, 20 and 45 mins. Using the tools in the OneOmics™ Suite in SCIEX Cloud, protein quantitation data was compared (Figure 1).

Key features for industrialized quantitative proteomics

- TripleTOF 6600+ System - High speed MS/MS enables the use of 100 variable sized Q1 window isolation[1] for SWATH Acquisition, for improved data quality through increased specificity, even at fastest gradient speeds [2]
- OptiFlow™ Interface [3] - Universal interface to switch between nanoflow and high flow sources in minutes
- OptiFlow Turbo V Source [4] - Single source for all low flow LC-MS experiments with enhanced usability and robustness
- NanoLC™ 425 System [5] – A versatile LC system that enables the researcher to easily switch between nanoflow to high microflow LC
- OneOmics Suite in SCIEX Cloud - a full data processing pipeline for SWATH Acquisition studies [6]

Methods

Sample preparation: Whole cell lysates (HEK293, Hela, MCF-7, A549 and MG132 treated Hela and MG132 treated HEK293) were obtained as standard cell lysates (Rockland Inc, USA). Each cell lysate was cleaned using S-Trap MS Sample Prep Kit (Protifi, USA), then digested with trypsin using standard protocols.[7] Sample loading of 3-5 µg of total protein (estimated) were used for each injection.

Chromatography: A NanoLC 425 System plumbed for microflow chromatography (5 µL/min) was used and operated in trap/elute mode. The analytical column used was a 0.3x150 Phenomenex Omega Polar column. Column temperature was controlled at 30°C. Gradients of 20, and 45 minutes were used, more information on gradients used can be found in previous work.[2]

Mass spectrometry: All data was acquired using a TripleTOF 6600+ System using the OptiFlow Turbo V Source equipped with the microflow probe and 25 µm electrodes. SWATH Acquisition data were collected using a TOF MS scan of 150 msec and 100 variable Q1 windows/cycle.[8] Each sample was acquired in triplicate at both gradient lengths.

Data processing: SWATH Acquisition data were processed using the OneOmics Suite in SCIEX Cloud. The ion library used for processing was the Pan Human Library.[9] Results were evaluated using the visualization applications in OneOmics Suite.[6]

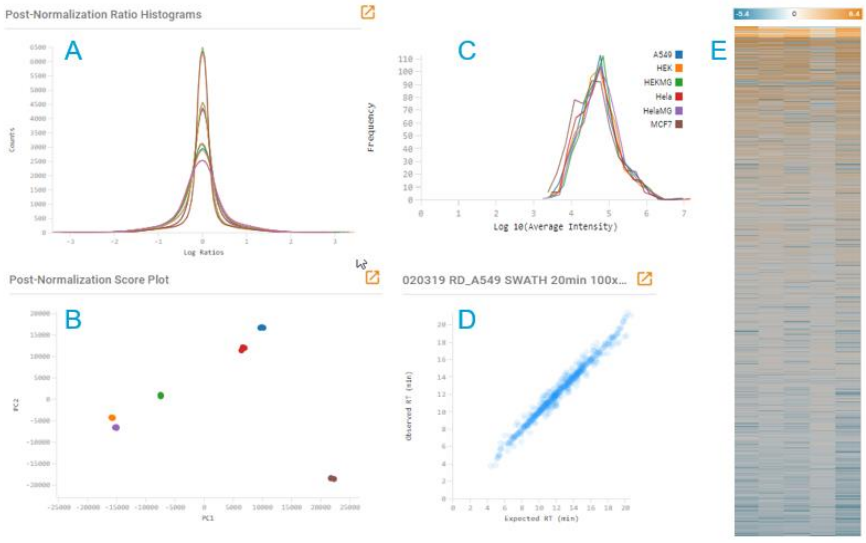

Figure 2. Assess data quality for 20 min gradient. Using the Analytics and Browser applications, the user can quickly assess the underlying MS data quality: A) alignment achieved after MLR normalization, and B) sample grouping on PCA Scores plot post-normalization shows quality of normalized data. C) Intensity histogram of peptide intensities across all samples confirms data consistency. D) Retention time alignment between library and current dataset. Right – heat map showing differential expression across cell lines relative to the HEK cell line.

Table 1. Numbers of proteins quantified. Using the Pan Human Library for data processing, and filtering the results based on <1% Peptide FDR and <20% CV on 3 replicates, similar # of proteins were quantified in the 20 min and 45 min gradients (within 5%). More peptides were quantified in the 45min gradient.

Cell Line	Proteins Quantified (<1%FDR, <20%CV)			Peptides Quantified (<1%FDR, <20%CV)	
	20 mins	45 mins	%	20 mins	45 mins
A549	*5119*	*5012*	*2.1*	*15686*	*20926*
MCF-7	*5150*	*5114*	*0.7*	*14342*	*19068*
HEK	*5404*	*5309*	*1.8*	*15893*	*21403*
HEK MG132 Stim	*5410*	*5250*	*3.0*	*16392*	*21593*
Hela	*5206*	*4970*	*4.7*	*15668*	*20376*
Hela MG132 Stim	*5464*	*5251*	*4.1*	*16505*	*21407*

Proteins quantified

SWATH Acquisition data were acquired in triplicate for all of the six cell lines, for both the 20 and 45 min gradients using microflow chromatography. With the OptiFlow source, the microflow column is attached directly on top of the microflow probe, minimizing post-column delays and dispersion for high-quality chromatography.

All the data were uploaded to the SCIEX Cloud and processed using the tools in OneOmics. After data extraction and normalization, the numbers of proteins quantified in each cell line was determined using the Analytics application, filtering the data at <1% FDR and <20% CV (Table 1). The numbers of proteins quantified using the 20 min gradient were very similar to that obtained with the 45 min gradient, although more peptides per protein were successfully quantified with the longer run time.

The MS data quality can also be assessed in Analytics (Figure 2). Visuals showing the reproducibility of replicates and the quality of the retention time alignment step can all be easily visualized. In the Browser application, the quality of the normalization step can be assessed by viewing the pre and post-normalization ratio histograms as well as the pre and post-normalization score plots.

In the Browser application, the differential proteins can be viewed in heat map form (Figure 2E) to allow investigation of specific proteins across the samples.

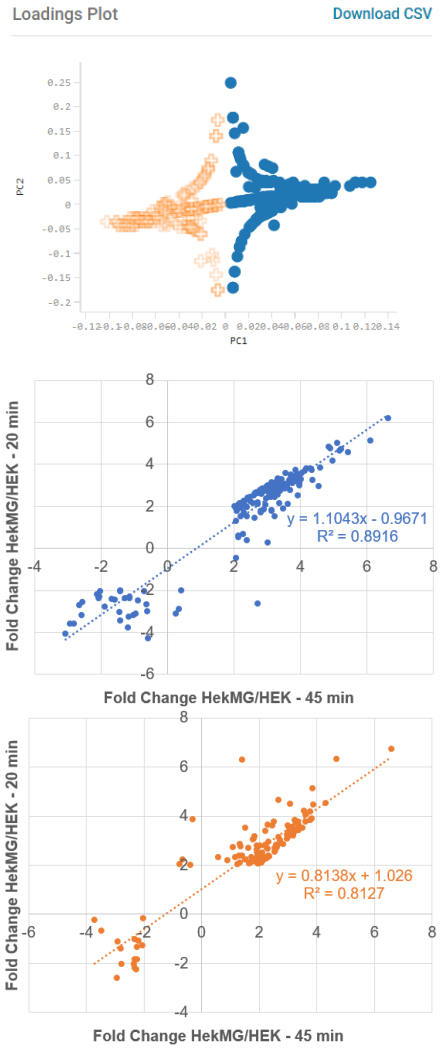

Figure 3. Accuracy of quantification. After alignment of the data and filtering by 80% confidence in fold change and 2-fold protein ratio, PCA analysis (top) showed two main clusters. Comparisons of the protein fold changes between the 20 and 45 min gradients for each cluster showed very good correlation, indicating that there was quantitative quality was maintained for the faster gradient.

Quantitative accuracy of fast gradients

To determine whether accurate protein quantitation is impacted when accelerated gradients are used, the protein quantitation ratios between two cell lines (MG132 treated HEK293 vs HEK293) were compared. The MarkerView™ App in SCIEX Cloud was used to align the two datasets, and the aligned dataset was filtered to include only the proteins that shows a 2-fold fold change and 80% confidence in fold change. Principal component analysis was performed and the protein results separated into two clusters (Figure 3). The fold change values for each cluster was plotted to assess if similar results were found between the two gradient conditions. Very good correlation was observed between the protein fold changes measured with each approach, suggesting that the quantitation results are not significantly degraded when using the accelerated microflow gradients with SWATH Acquisition.

Browsing results

There are a large variety of views that can be used to further delve into the biological results. One informative view is to select a cluster from the PCA analysis in the MarkerView App, then view it as a force directed graph (Figure 4). Here the proteins and their enriched gene ontologies (GO terms) are plotted together to show the relationships. Orange and blue circles show up and down-regulation of proteins and the increased depth of color shows a larger fold change. The same coloring is used for the ontologies squares. The size of the circle or square depicts the number of associations. Clicking on each will pull up a side bar that provides detailed information.

One can also look at proteins individually in the Browser app (Figure 5), here the expression differences of one protein are viewed across the 6 samples. Examples of some of the views are the Box and Whisker plot showing the protein changes. The area of the individual quantified peptides for the protein can also be viewed.

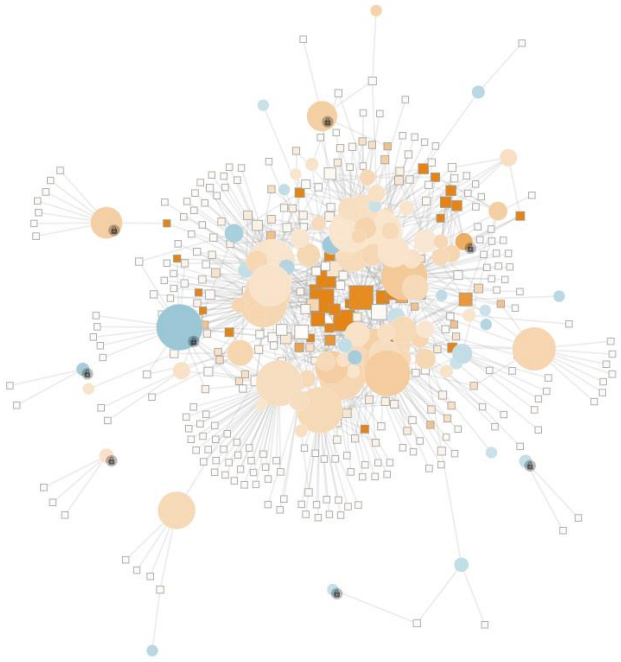

Figure 4. Force directed map showing biological changes between HEK_MG vs HEK. For the proteins in the orange cluster (Figure 3), GO term enrichment is done and drawn in a force directed graph to show the relationships. Many of the proteins are upregulated in this subset of proteins, as seen by the orange circles. GO ontologies are shown as squares and the direction of enrichment and significance are indicated by the depth of the color.

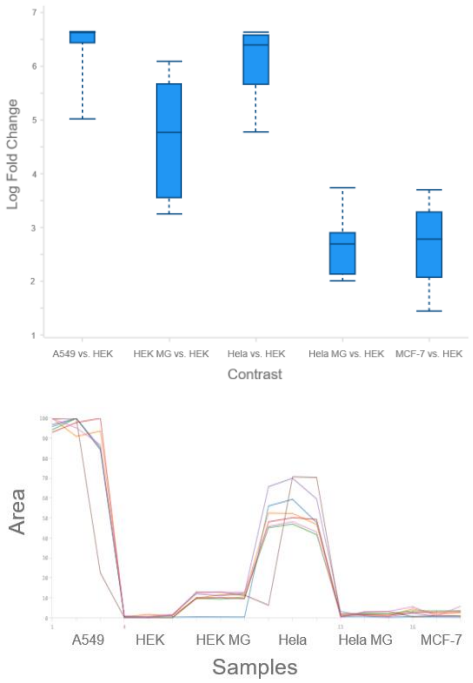

Figure 5. Viewing individual protein expression changes. The changes for Aldo-keto reductase family 1 member C2 (P52895) across the 6 samples are viewed as a Box and Whisker Plot (top) or views as peak areas across the samples for the individual peptides (bottom).

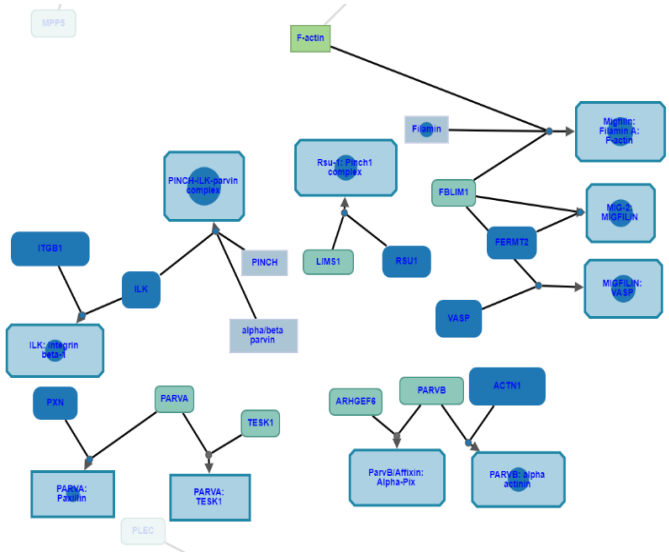

Figure 6. Finding perturbed pathways from the differentially expressed proteins. Pathway diagram representing cell-extracellular matrix interactions (Reactome ID R-HSA-446353) that was found to be enriched in a comparison of HEK_MG132 vs HEK. An upregulated set of proteins found in HEK_MG132 as compared to the HEK cell line was identified in MarkerView App and then mapped in Pathways to reveal enriched biochemical reactions. The proteins from the selected cluster are colored dark blue in the pathway diagram.

The enriched biochemical pathways present in the data set can be viewed with the Pathways application. Pathways uses a MarkerView App results session to map clusters of compounds onto pathways present in Reactome, a peer-reviewed and open source pathways database. Upon selecting a particular cluster, the associated compounds will be mapped onto pathway diagrams in the Pathways Diagram viewer. More information can be obtained by clicking the nodes in each pathway (Figure 6).

Conclusions

As the number of samples to be analyzed in the biological samples set steadily increases, so will the need for more robust and higher throughput protein quantitation strategies. Here, SWATH Acquisition on the TripleTOF 6600+ System with the OptiFlow Source is shown to provide accurate quantitation results when running shorter LC gradients.

- The numbers of proteins quantified is very similar between the 20 and 45 min gradients
- For the differentially expressed proteins, the fold change measured is also very similar showing quantitative accuracy is maintained in these faster experiments
- The OneOmics Suite in the SCIEX Cloud enable fast data processing and easy results interrogation for biological insight.

References

1. Microflow SWATH Acquisition for industrialized quantitative proteomics, SCIEX technical note RUO-MKT-02-3637-A.
2. Accelerating SWATH Acquisition for protein quantitation – up to 100 samples per day - microflow LC with TripleTOF 6600 System. SCIEX Technical note RUO-MKT-02-8432-A.
3. OptiFlow™ Interface for TripleTOF 6600 System - switch from nanoflow LC to microflow LC in minutes. SCIEX Technical note RUO-MKT-02-7219-A.
4. Single source solution for low flow chromatography – OptiFlow® Turbo V Source – always working in the sweet spot of sensitivity and robustness. SCIEX technical note RUO-MKT-02-9701-A.
5. Achieving very high reproducibility for quantitative proteomics with nanoflow LC-MS - NanoLC™ 400 System. SCIEX Technical note RUO-MKT-02-5755-A.
6. Fast-track proteomics data processing with the SCIEX Cloud - overview of the OneOmics™ Suite. SCIEX Technical note RUO-MKT-02-6969-B.
7. In-solution protein digestion for proteomic samples protocol. SCIEX Technical Note RUO-MKT-02-5438-A.
8. SWATH Performance Kit Standard Operating Protocol.
9. Rosenberger G *et al*. (2014) *Scientific Data*, **1**, 140031.

Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
sciex.com

International Sales
For our office locations please call the division headquarters or refer to our website at
sciex.com/offices