Spectral library generation for SWATH® Acquisition in less than 20 hours

Using the OneOmics™ Suite and the TripleTOF® 6600+ LC-MS/MS System

Nick Morrice[1], and Christie Hunter[2]
[1]SCIEX UK, [2]SCIEX, USA

The generation of spectral libraries is often thought of as a time consuming procedure, but in reality it can be performed in less than 1 day. It has been shown previously that utilizing rapid gradients and microflow chromatography can provide very good protein and peptide identification numbers.[1] It is also a common understanding in the field of proteomics that many more protein identifications can be generated by first performing high pH reversed phase fractionation of complex digested samples, running data dependent analysis of each fraction and then searching all the LC-MS/MS data together to generate a very large ion library.[2] The database searches for protein identifications can often be the rate limiting factor even when using powerful desktop computers as the datasets get very large.

Here, the combination of in-depth sample fractionation with microflow LC on a TripleTOF 6600+ System and cloud computing was explored for the rapid generation of bespoke libraries. Two sample types were tested, plasma and a tissue lysate.

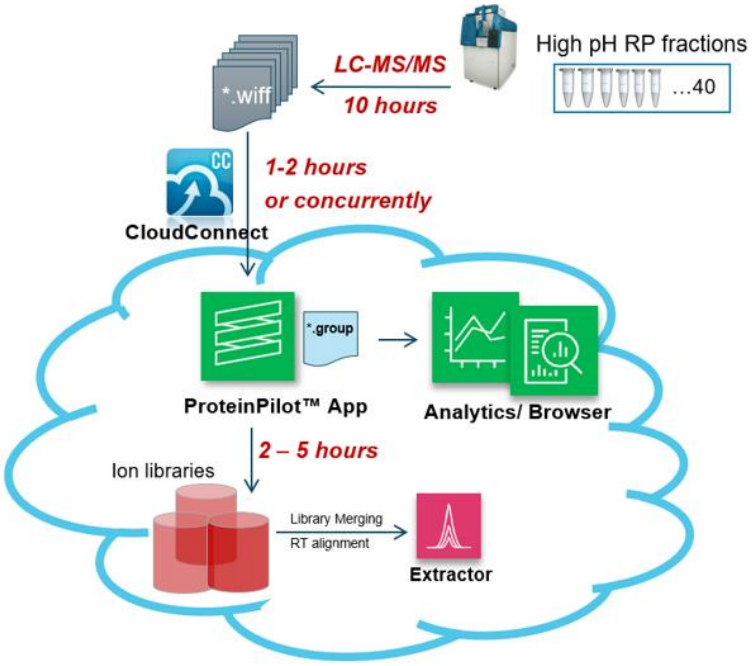

Figure 1. Workflow for fast ion library generation on proteomic samples. After using high pH fractionation to generate 40 fractions, they were rapidly analyzed by LC-MS/MS using microflow LC. Data were uploaded to the cloud for processing in order to generate a large ion library for SWATH® Acquisition data processing.

Key features of the fast library generation strategy

- Full workflow from sample to ion library took less than 24 hours (Figure 1)
- Forty high pH fractions were used for sub-fractionation to deeply interrogate the samples in order to generate large ion libraries
- Fast microflow LC gradients (10 mins) were used to then rapidly analyze the 40 fractions
- The TripleTOF 6600+ System was used and operated at 50 Hz to generate as much high-quality, high-resolution MS/MS data as possible
- Data were uploaded to SCIEX Cloud Platform and searched with the ProteinPilot™ App to identify a large number of proteins from the fractions and generate an ion library for SWATH® Acquisition processing

Methods

Sample preparation: 108 FFPE colon cancer biopsies were digested with trypsin, aliquots pooled and fractionated by high pH reversed phase HPLC, as described previously.[3] Plasma from 30 patient samples was digested with trypsin in the presence of 1M urea, pooled and fractionated as described above. The fractions were dried down, reconstituted to 0.5 µg/µL in 2% acetonitrile/98% water/0.1% formic acid. iRT peptides were added to each fraction according to vendor instructions (Biognosys).

Chromatography: Fractions from samples were separated using a Phenomenex Luna Omega Polar C18, 3 µm, 0.3 x 150 mm column connected directly to the OptiFlow Source. Using a trap-elute configuration on the nanoLC™ 425 System (SCIEX), a 10 minute gradient from 5-40% acetonitrile in 0.1% formic acid was used to separate the peptides as described previously.[1] Column was heated to 30 ºC using the column heater integrated with the source.

Mass spectrometry: The TripleTOF 6600+ System was equipped with the OptiFlow Source with the 1-10 µL/min probe and electrode. IDA acquisition was performed using a TOF MS survey scan (100 msec) followed by 50 MS/MS scans (20 msec), as described previously.[1] Total run time for each fraction including trap loading was 15 min.

Data processing: Data files were uploaded to SCIEX Cloud Platform using CloudConnect in PeakView® Software 2.2. Data were then searched using the multi-file option in the ProteinPilot App in OneOmics Suite using a UniProt human FASTA file. The search effort used in the cloud app was “Thorough” and “biological modifications” was selected to ensure the broadest search space. Search results were visualized using the Analytics and Browser apps to ensure the quality of the generated ion library. The ion libraries could then be used to directly process SWATH® Acquisition data using Extractor (Figure 1).

Data processing in SCIEX Cloud Platform

LC-MS data were generated on the TripleTOF 6600+ System, then the data were immediately uploaded to the cloud using CloudConnect. Once data upload was complete, an experiment was defined within Experiment Manager to define the files to be included in the data processing. Here, a multi-file search was done in the ProteinPilot App in order to create a single large search result. This search result could then be used for processing of SWATH® Acquisition data in the proteomics pipeline (Figure 2). The full workflow is described in the SCIEX technical note and the OneOmics Suite community.[3,4]

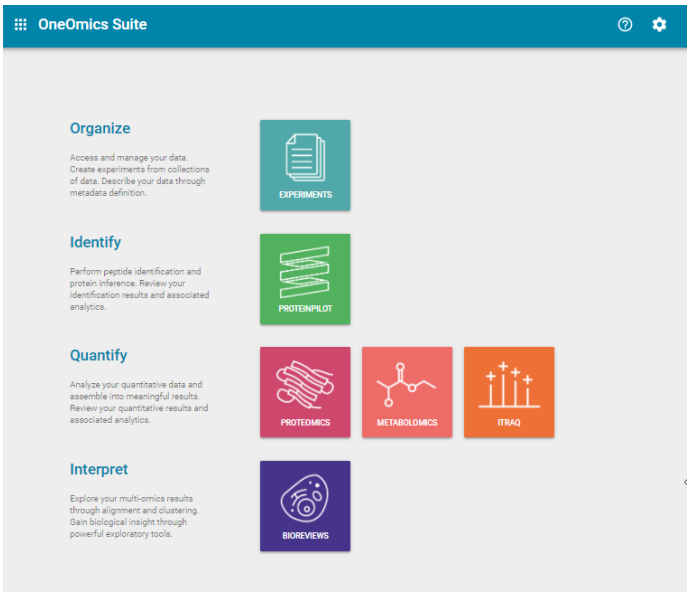

Figure 2. OneOmics landing page. After data upload, studies and experiments are built within Experiments. Once defined, the data are processed using ProteinPilot App.

Fast protein identification

In this study, the use of larger numbers of high pH fractionation with fast microflow LC-MS/MS was explored for the rapid generation of protein identification experiments and the generation of large SWATH Acquisition ion libraries. Two sample types were analyzed. Colon cancer FFPE biopsies and human plasma were digested and subjected to high pH fractionation where 40 fractions were collected for each. Then using a 10 min LC gradient, each fraction was analyzed by data dependent acquisition. After data upload, protein identification was performed using the ProteinPilot App. Time from start of data acquisition to obtaining protein identification results was under 20 hours.

For the undepleted plasma sample, the combined search of the 40 fractions identified over 1100 proteins at <1% FDR. For the 40 fractions from the colon cancer FFPE samples, over 7200 proteins were identified at <1% FDR. The search parameters for these samples are shown in Figure 3. The same searches were performed on the desktop computer provided for ProteinPilot Software searches to compare search speed (Table 1). As expected for cloud-based searches, the larger the search, the larger the speed gain by moving the search to the cloud. In addition, larger searches can be performed in parallel, further accelerating time to results.

FAST_FFPE_library Settings

Job type	Multi-File ProteinPilot Search
Input	Fast_FFPElibrary (Experiment)
Elapsed Time	4hr 17min 18sec
Analysis name	FAST_FFPE_library
Project	Biognosys_HPRPfractions
Cys alkylation	Iodoacetamide
Digestion	Trypsin
Species	Homo sapiens
Id focus	Biological modifications
Sample type	Identification
Bias correction	Yes
Background correction	Yes
Project data provider	DataStore
Fasta file data provider	BaseSpace
Fasta file	uniprot-sprot can iso 20140709 contam.fasta

Figure 3. ProteinPilot App search parameters. The search parameters were used to create a library from 40 high pH RP-HPLC fractions of FFPE colon cancer tryptic digest. Total elapsed time is 4.25 hrs for 40 IDA runs.

The resulting *.group file can now be used for processing of SWATH Acquisition data using the Extractor and Assembler Apps in the OneOmics Suite, or downloaded to the desktop and opened using ProteinPIlot Software 5.3.

Conclusions

Using OneOmics Suite in SCIEX Cloud Platform, protein identification experiments, or generation of ion libraries for SWATH Acquisition data, can be rapidly generated. Here the use of a higher number of fractions from digested samples combined with fast microflow LC on the TripleTOF 6600+ System was explored. In under 24 hours, LC-MS data were acquired, uploaded to the cloud and searched using the ProteinPilot App (Figure 1). Very good protein identification results were found for both samples tested, plasma and tissue lysates (Table 1).

This approach is highly scalable, more fractions can be generated to dig deeper into complex samples, rapidly interrogated and searched in the cloud, which can more easily accommodate larger and larger studies than a desktop solution.

Table 1. Comparison of protein identification results. Protein and peptide identifications at 1% FDR are shown, as well as the search time observed on both the desktop processing station and on SCIEX Cloud Platform.

			Data processing time	
Fractionated sample	**# proteins at <1% FDR**	**# peptides at <1% FDR**	**Cloud (hours)**	**Desktop (hours)**
Plasma (40 fractions) – fast gradient	*1178*	*23981*	*1.25*	*1.75*
Colon cancer FFPE (40 frs) – fast gradient	*7282*	*94411*	*4.7*	*9.6*
*Human cell line (65 fractions) – long gradient**	*--*	*--*	*12.3*	*33*

** Included for speed comparison*

References

1. Fast protein identification experiments with microflow LC – up to 100 samples per day. SCIEX technical note, RUO-MKT-02-8312-A.
2. Extending depth of coverage with SWATH® Acquisition using deeper ion libraries. SCIEX technical note, RUO-MKT-02-3247-A.
3. Fast-track proteomics data processing with the SCIEX Cloud Platform. SCIEX technical note, RUO-MKT-02-6969-B.
4. OneOmics Suite user community.
5. Digitizing the proteomes from big tissue biobanks - analyzing 24 proteomes per day by microflow SWATH® Acquisition and Spectronaut Pulsar analysis. SCIEX technical note, RUO-MKT-02-8868-A.

SCIEX

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