ds

High-throughput quantitative proteomics using Zeno SWATH data-independent analysis (DIA) and the Evosep One system

Using the ZenoTOF 7600 system

Ihor Batruch[1], Yves Le Blanc[1], Jose-Castro Perez[2] and Christie Hunter[2]
[1]SCIEX, Canada; [2]SCIEX, USA

Identifying and quantifying large numbers of proteins and peptides are important in translational research to understand biological functions. One example of this is the analysis of large sample cohorts to identify robust biomarkers of disease. The complexity of biological fluids, tissues and cell lines often overwhelms the capabilities of data-dependent acquisition (DDA) strategies using mass spectrometry due to the stochastic nature of this methodology. This limitation has driven the rapid adoption of data-independent acquisition (DIA or SWATH DIA) approaches, such as Zeno SWATH DIA. This method achieves 5-6x more MS/MS sensitivity,[1] resulting in this DIA approach surpassing DDA for protein identifications and quantification in complex matrices. This advantage is further observed at short acquisition times that are required to meet the throughput demands of current proteomics studies.[2,3]

To meet the demands of high-throughput proteomics, the Evosep One system (Evosep) provides standardized liquid chromatography with preset methods and disposable trap columns. The system is optimized such that the user selects the number of samples to run per day (SPD) to determine the column, flow rates and run times necessary for analysis. The system is highly reproducible, robust and offers high MS utilization. These system features are ideal for applications in high-throughput laboratories.

Figure 1. Evaluating the relationship between LC workflows and quantified proteins. Using Zeno SWATH DIA, protein quantification was evaluated across the 4 different preset SPD conditions. The number of proteins identified at <1% FDR (transparent) and quantified at <20% CV (solid) across the workflows for a 200 ng load of cell lysate digest are shown. At this sample load, ~90% of proteins identified were also reliably quantified.

Here, the combination of the Evosep One system and the ZenoTOF 7600 system was investigated for fast quantitative proteomics. Zeno SWATH DIA was used to investigate the numbers of proteins that can be detected and robustly quantified at varying SPD throughputs (Figure 1).

Key features of the Evosep One system with the ZenoTOF 7600 system

- Zeno SWATH DIA using the ZenoTOF 7600 system provides a large increase in peptide MS/MS sensitivity, enabling 50% more proteins to be quantified from a single run[3]
- The Evosep One system is a robust, standardized platform that is designed for high-throughput studies and has specific workflow modes to fit any quantitative proteomics study
- The high speed and sensitivity of MS/MS acquisition on the ZenoTOF 7600 system enable it to easily couple with Evosep One system workflows
- At a 30 SPD throughput and with 200 ng of sample loaded, ~6200 proteins were quantified with <20% CV (Figure 1)

- Data processing with DIA-NN software is optimized for fast chromatography and can provide quantitative information on many peptides and proteins

Methods

Sample preparation: Commercial human cell line tryptic digest (HeLa, Thermo Fisher Scientific) was reconstituted with 95% buffer A (water with 0.1% formic acid) and 5% buffer B (acetonitrile with 0.1% formic acid) to a working concentration of 0.1 µg/µL. Working stock solution was further diluted to 1.25 ng/µL for the 25 ng load, 2.5 ng/µL for the 50 ng load, 10 ng/µL for the 200 ng load and 25 ng/µL for the 500 ng load. A 20 µL sample of each concentration was loaded into Evotips for sample injection.

Chromatography: The Evosep One system (Evosep, Denmark) was operated using the preset SPD configurations, as described in Table 1. Flow rates and columns were adjusted based on vendor specifications. Examples of chromatography for the 50 ng load are shown in Figure 2. Control of the Evosep One system was fully integrated with SCIEX OS software, such that a single batch could be used to control both LC and MS.

Mass spectrometry: The ZenoTOF 7600 system was operated in Zeno SWATH DIA mode using the OptiFlow Turbo V ion source. The 200 SPD, 100 SPD and 60 SPD throughput methods were run in the microflow configuration, and the 30 SPD method was run in the nanoflow configuration, as described in Table 2. Analysis was performed in triplicate for each tested condition.

Data processing: Zeno SWATH DIA data were processed using DIA-NN software beta version 1.8.1.[3,4] Data were analyzed using an experimental library generated from the pH fractionation of 2 human cell lines,[6] a library-free approach using a library generated *in silico* from a human FASTA file[7] or the Pan Human Library.[8] Pg.matrix.tsv and pr.matrix.tsv reports were used to report the number of protein groups and peptide precursors identified at <1% FDR and the number of proteins and peptide precursors quantified at <20% CV.

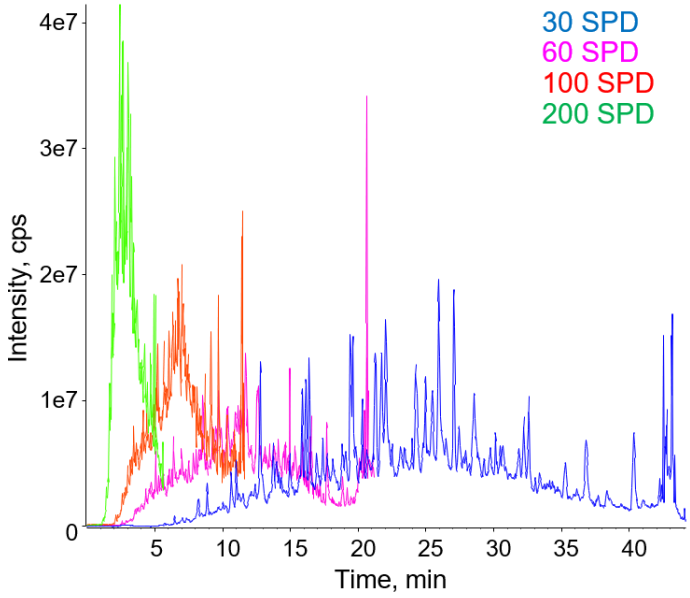

Figure 2. Total ion chromatograms (TIC) for the different Evosep workflows. With the different SPD workflows, the total run time varies, as illustrated by the TOF TIC for the 50 ng sample loading.

Table 1. Chromatographic conditions for the 4 different SPD experiments. Different run times required different columns and different flow rates.

Samples per day (SPD)	Gradient length (min)	Flow rate (µL/min)	Total run time (min)	Column
200	5.6	2.0	*7.1*	4 cm x 150 µm ID, 1.9 µm particle column (EV1107)
100	11.5	*1.5*	*14.4*	8 cm x 150 µm ID, 1.5 µm particle column (EV1109)
60	*21.0*	*1.0*	*24.0*	8 cm x 150 µm ID, 1.5 µm particle column (EV1109)
30	*44.0*	*0.5*	*48.0*	15 cm x 150 µm ID, 1.9 µm particle column (EV1106)

Mobile phase A – 100% water with 0.1% formic acid
Mobile phase B – 100% acetonitrile with 0.1% formic acid

Table 2. MS settings for the 4 different SPD experiments. As the flow rates changed for the different chromatographic configurations, the MS conditions were also adjusted to ensure a good cycle time was obtained. Curtain gas was 25 psi for all experiments.

Samples per day (SPD)	Ionization voltage (V)	Temperature (°C)	Nebulizing gas GS1 (psi)	Heater gas GS2 (psi)	TOF MS accumulation time (msec)	Number of Q1 windows	TOF MS accumulation time (msec)	Mass range (m/z)
200	4500	150	*12*	*60*	*25*	*56*	*11*	*400–750*
100	4500	150	*12*	*60*	*25*	*56*	*11*	*400–750*
60	4500	150	*12*	*60*	*50*	*60*	*11*	*400–900*
30	*3400*	*300*	*10*	*N/A*	*50*	*85*	*18*	*400–900*

Figure 3. Comparing protein results across different SPD throughputs. The numbers of proteins identified at <1% FDR (transparent) and quantified at <20% CV (solid) from Zeno SWATH DIA data at different SPDs were compared. Longer run times and higher loads (Figures 1 vs. 3) provided more proteins quantified for deeper proteome coverage.

Proteins quantified

Digested HeLa cell lysate was used to characterize the various SPD workflows using Zeno SWATH DIA on the ZenoTOF 7600 system. For these experiments, 200 and 50 ng sample loads were used (Figures 1 and 3, respectively). This characterization permits users to select the workflows and required sample amounts that best match the throughput and depth of proteome coverage for their experimental needs.

For both loading amounts, as the SPD was reduced and the sample analysis time increased, the number of proteins identified and quantified increased (Figure 3). When the throughput was reduced from 200 to 30 SPD, 102% and 82% more proteins were identified at <1% FDR at 50 and 200 ng loads, respectively. More importantly, the number of proteins quantified at <20% CV also greatly increased, with 77% and 80% more quantified at 50 and 200 ng loads, respectively.

For a 50 ng load of HeLa digest (Figure 3), the 30 SPD workflow provided maximal proteome coverage. With Zeno SWATH DIA, 5727 protein groups were detected, 4232 of which were quantified at <20% CV (Figure 3). With this load and SPD workflow, 74% of detected proteins were reliably quantified.

With a higher load, run times were easily accelerated. Using the 100 SPD workflow and a sample load of 200 ng of HeLa digest (Figure 1), 5499 protein groups were detected and 4870 (88%) were reproducibly quantified with <20% CV. This method therefore provided both proteome coverage and high sample throughput. When the throughput was reduced to 30 SPD and 200 ng of HeLa digest was loaded, 7014 protein groups were detected, 6189 (88%) of which were quantified with <20% CV.

Library-free protein identification from Zeno SWATH DIA data

An emerging workflow powered by improvements in software algorithms is the library-free approach for processing DIA data. Here, DIA-NN software was used to convert a FASTA file *in silico* to a spectral library to use for data processing. The Zeno SWATH DIA data from these experiments were processed with both the *in silico* library and a spectral library generated from experimental ZenoTOF 7600 system data.[6] These results were also compared with those generated by processing the data using the Pan Human library (Figure 4, top).[7]

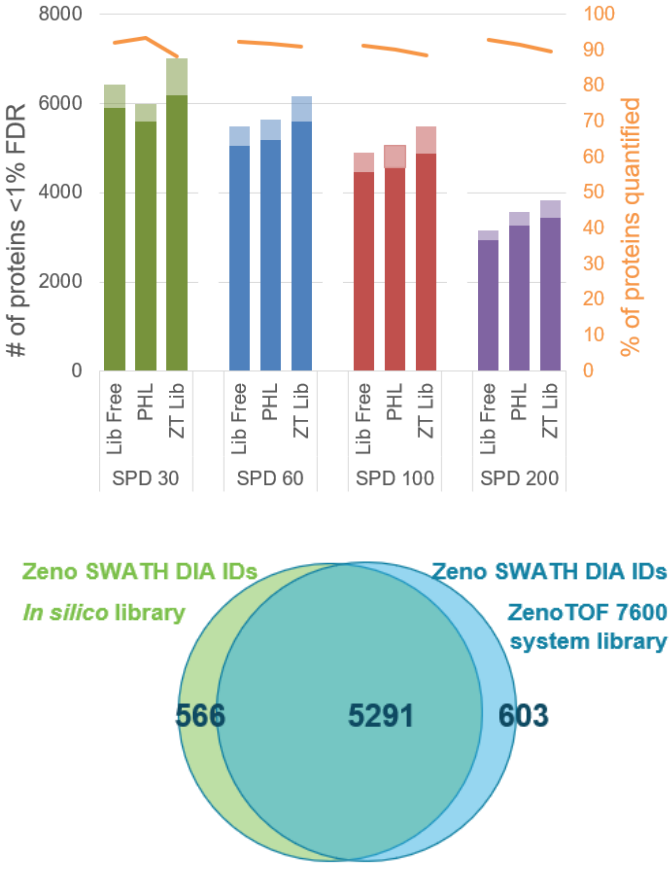

Figure 4. Comparing the library-free approach and 2 experimentally generated libraries at the 200 ng load. (Top) Zeno SWATH DIA data were processed with DIA-NN software using 3 different libraries (Lib Free, using an *in silico* library from a FASTA file; PHL, Pan Human library; ZT Lib, large library from Zeno DDA dataset collected on a ZenoTOF 7600 system). (Bottom) Using the 30 SPD data, the numbers of proteins quantified at <1% FDR and <20% CV between the data analyzed with the library-free and ZenoTOF 7600 system library approaches had very good overlap.

Analysis of the data generated with a sample load of 200 ng HeLa digest on column revealed good performance with the library-free approach. Analysis using the *in silico* generated library provided protein identification results similar to those yielded by analysis with the 2 experimentally generated libraries (Figure 4). At 30 SPD, 6417 proteins were identified at <1% FDR using the library-free approach, whereas 7014 proteins were identified with the ZenoTOF 7600 library.

A key advantage of the Zeno SWATH DIA approach for protein identification is the ability to also obtain quantification information from the Zeno MS/MS data, as shown in Figure 4. Typically, 90% of the proteins identified were also quantified with <20% CV when 200 ng of sample was loaded for all throughput workflows.

The library-free approach is a very effective search strategy, as it enables the identification and quantification of proteins without the prerequisite of generating sample and cohort-specific libraries.

Figure 5. Peptide precursors quantified across the different SPD conditions. Using Zeno SWATH DIA, the numbers of peptides quantified (<1% FDR, <20% CV) were evaluated at 2 different sample loads (50 and 200 ng).

Peptide precursors quantified

Finally, the impact of SPD and load on the numbers of peptide precursors quantified was examined (Figure 5). For a 200 ng load, ~15,000 precursors were quantified at 200 SPD and ~42,000 at 30 SPD (Figure 5). The most significant improvement occurred when reducing throughput from 100 to 60 SPD, as identifications increased 43%.

In terms of SPD throughput and different proteomic loads, lower loads favored higher SPDs. A throughput of 60 SPD offered balance between performance in detection and quantification. Lower SPD throughputs and higher loads led to increases in protein detection and quantification, as expected. The protein and peptide data presented here can be used as guidance to select appropriate sample loads and required throughputs to meet future study needs.

Conclusions

The ZenoTOF 7600 system using Zeno SWATH DIA provides excellent depth of proteome coverage with high quantitative reproducibility, due to the increase in MS/MS sensitivity with the Zeno trap enabled. Here, Zeno SWATH DIA was coupled with the Evosep One system to demonstrate compatibility with this system and to highlight various workflows for high-throughput, quantitative proteomics applications.

- Sample load and throughput were varied by testing 2 sample loads (50 and 200 ng on column) of HeLa digest at 30, 60, 100 and 200 SPD
- As expected, as run times increased and sample throughput decreased, the depth of proteome coverage increased
- When 200 ng of sample was loaded at 30 SPD, ~7000 proteins were detected at <1% FDR, ~6200 of which were quantified with <20% CV
- Analysis of Zeno SWATH DIA data, collected following separation on the Evosep One system, with the library-free workflow yielded comparable results to the typical library-based approaches

References

1. Large-scale, targeted, peptide quantification of 804 peptides with high reproducibility, using Zeno MS/MS. SCIEX technical note, RUO-MKT-02-13346-A.
2. Going library-free for protein identification using Zeno SWATH DIA and in silico-generated spectral libraries. SCIEX technical note, RUO-MKT-02-14675-A.
3. Zeno MS/MS with microflow chromatography powers the Zeno SWATH DIA workflow for more proteins quantified. SCIEX technical note, RUO-MKT-02-14668-A.
4. Processing ZenoTOF 7600 system data with DIA-NN software. SCIEX community post, RUO-MKT-18-14611-A.
5. Demichev V *et al.* (2019) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods*, **17**, 41-44.
6. Large-scale protein identification using microflow chromatography on the ZenoTOF 7600 system. SCIEX technical note, RUO-MKT-02-14415-A.
    - Data from two fractionation experiments of two human cell lines (Hela, K562) were each processed into a single search result in the ProteinPilot app in OneOmics suite. The search results for each cell line were then merged, and retention time aligned using the Extractor application to create a final ion library.
7. Creating a library from a FASTA file for library-free data analysis. SCIEX community post, RUO-MKT-18-14611-A.
8. Rosenberger G *et al* (2014) A repository of assays to quantify 10,000 human proteins by SWATH-MS. *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