

# Going library-free for protein identification using Zeno SWATH data-independent analysis (DIA) and *in silico*-generated spectral libraries

Using DIA-NN software and the ZenoTOF 7600 system

Christie Hunter and Alexandra Antonoplis  
SCIEX, USA

While obtaining accurate quantification information for proteins across sample sets is key to understanding biology, the proteins present in the sample must first be identified. Thus, the ability to identify proteins present in a biological sample remains an important workflow. There are 2 main workflows in which researchers can identify very large numbers of proteins. The traditional workflow, termed shotgun proteomics, is a data-dependent workflow (DDA) in which MS/MS data are acquired for detected precursors and then all MS/MS spectra are subjected to a database search to identify the peptides and proteins in the sample.<sup>1</sup>

Over the last 10 years, data-independent acquisition (DIA or SWATH DIA) has emerged as a powerful workflow for quantitative proteomics. In this workflow, MS/MS spectra are acquired for all detectable peptides and peptide peak areas are extracted using experiment-specific spectral libraries. More recently, multiple powerful algorithms have emerged that have

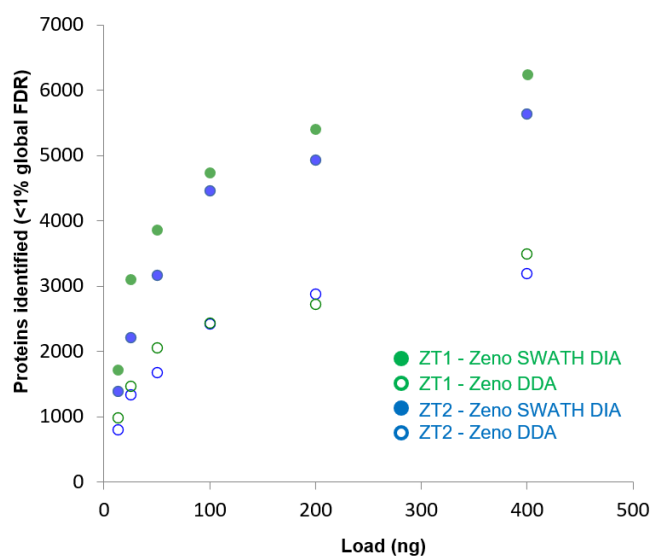


enabled proteins to be identified from DIA data without having to generate an experimental spectral library in advance. These advances have increased the ease of using DIA for protein identification workflows.<sup>2</sup>

Here, the traditional DDA workflow (using Zeno DDA) for protein identification was compared to the Zeno SWATH DIA workflow, in which proteins are identified using a library generated *in silico* from a FASTA file. The numbers of proteins and peptides identified with these approaches were compared for 45-min and 10-min microflow LC gradients. The results generated by the library-free Zeno SWATH DIA workflow were compared with other library-based approaches.

## Key features of Zeno SWATH DIA for protein identification

- The ZenoTOF 7600 system has the speed and sensitivity to acquire very high-quality, high-resolution MS/MS data
- Zeno trap activation provides 5-6x increases in peptide MS/MS sensitivity and can be applied to MRM<sup>HR</sup>, SWATH DIA and DDA workflows<sup>3</sup>
- *In silico* libraries generated with DIA-NN software reduce the need to generate experimental libraries for SWATH DIA processing<sup>4</sup>, streamlining the protein identification workflow
- Zeno SWATH DIA identified 2x more proteins than Zeno DDA through the use of *in silico*-generated libraries, providing an easy, rapid and comprehensive library-free workflow for protein identification (Figure 1)
- *In silico*-generated libraries used to process Zeno SWATH DIA data yielded similar results to other large, experimentally generated libraries



**Figure 1. Comparing proteins identified from Zeno SWATH DIA data vs. Zeno DDA data.** Zeno SWATH DIA data processed with the *in silico* generated library (filled circles) outperformed Zeno DDA data (open circles) for protein identifications using a 45-min microflow LC gradient on 2 ZenoTOF 7600 systems (green vs. blue) across a range of sample loadings on column.

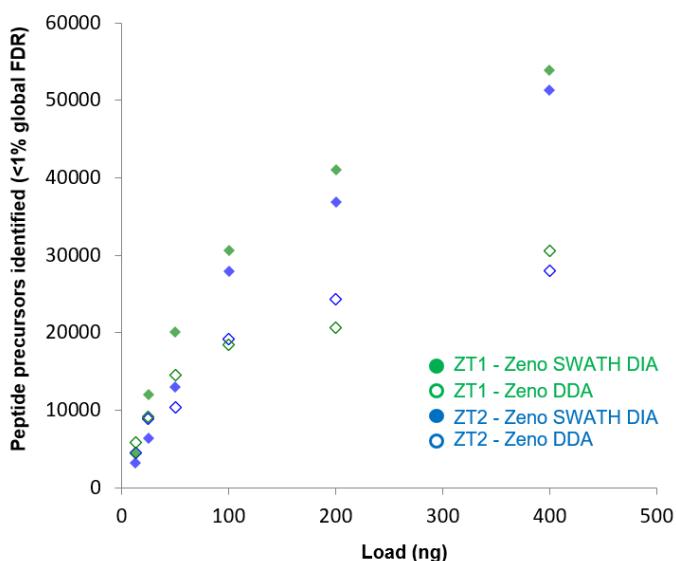
## Methods

**Sample preparation:** A digest of human K562 cell lysate from the SWATH performance kit (SCIEX) was used. Sample loadings ranged from 12.5–400 ng on column.

**Chromatography:** Separations were performed using a Waters ACQUITY UPLC M-class system plumbed for microflow chromatography (5  $\mu$ L/min), using a Phenomenex micro trap analytical column. All data were acquired using either a 45-min or a 10-min gradient. See previous technical note for more details.<sup>5</sup>

**Mass spectrometry:** Data were acquired on 2 ZenoTOF 7600 systems with either Zeno DDA or variable window Zeno SWATH DIA. Data were acquired in triplicate with the Zeno trap enabled in MS/MS mode for all acquisitions. See previous technical note for more details.<sup>5</sup>

**Data processing:** Zeno DDA data were processed using the ProteinPilot App in OneOmics suite using the canonical+isoforms version of a UniProt FASTA file based on human samples. Protein and peptide FDR numbers (<1% global FDR) were obtained from the Analytics dashboard. Zeno SWATH DIA data were processed using DIA-NN software and a library-free process.<sup>4</sup> First, a spectral library was generated from the canonical+isoforms version of a UniProt\_Human FASTA file, then this *in silico* library was used to process the DIA data to determine the number of proteins identified. A <1% precursor FDR filter was applied. Processing settings were previously described in detail in a community post.<sup>6</sup> Protein and peptide



**Figure 2. Comparing peptides identified from Zeno SWATH DIA data vs. Zeno DDA data.** At comparable sample loads, Zeno SWATH DIA data (solid diamonds) processed with the *in silico* generated library outperformed Zeno DDA data (open diamonds) for peptide identifications on both ZenoTOF 7600 systems (green vs. blue), using a 45-min microflow LC gradient.

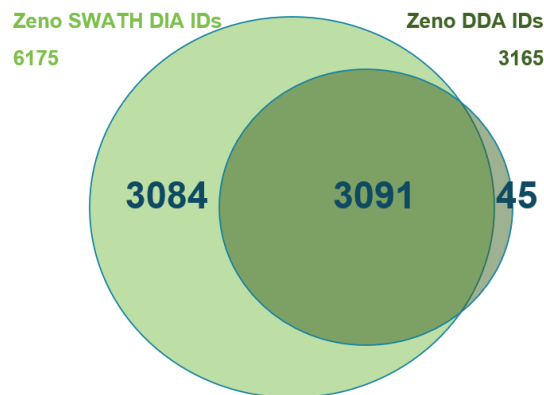
precursor areas from the \*.pr\_matrix.tsv and \*.pg\_matrix.tsv files were copied into Microsoft Excel and the numbers of proteins and peptides quantified at 20% CV were computed.

Zeno SWATH DIA data were later processed in DIA-NN software using the Pan Human Library<sup>5</sup> and a large library generated on the ZenoTOF 7600 system<sup>8</sup> to compare the performance of the library-free method. This large library was generated from two fractionation experiments of two human cell lines (Hela, K562) which 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.

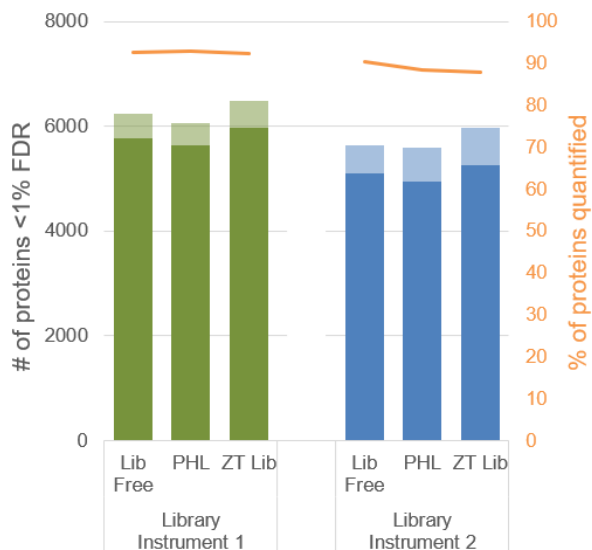
## Zeno DDA vs. Zeno SWATH DIA for protein identifications

When the Zeno trap is activated on the ZenoTOF 7600 system, MS/MS sensitivity is enhanced as much as 5- to 6-fold for peptides.<sup>4,6</sup> This MS/MS acquisition mode can be applied to both DDA and DIA workflows. Here, these workflows were compared on their ability to identify large numbers of proteins and peptides from a complex proteome digest that was separated with microflow chromatography. The gains in proteins identified and quantified using the Zeno SWATH DIA workflow were previously characterized.<sup>6</sup> Here, the data collected using the Zeno DDA workflow are compared to those acquired using the Zeno SWATH DIA workflow.

The Zeno DDA data were processed using the ProteinPilot app in OneOmics suite using a Thorough search to identify the maximum number of peptides and proteins (<1% global FDR) in the dataset. The Zeno SWATH DIA data were processed using a library generated *in silico* from a human FASTA file using DIA-



**Figure 3. Overlap of protein identifications between Zeno DDA and Zeno SWATH DIA.** Using the 45-min gradient data collected on instrument 1 for a 400-ng load, nearly all the proteins identified from the Zeno DDA dataset were also identified in the Zeno SWATH DIA dataset. The Zeno SWATH DIA approach yielded 2-fold more protein identifications.



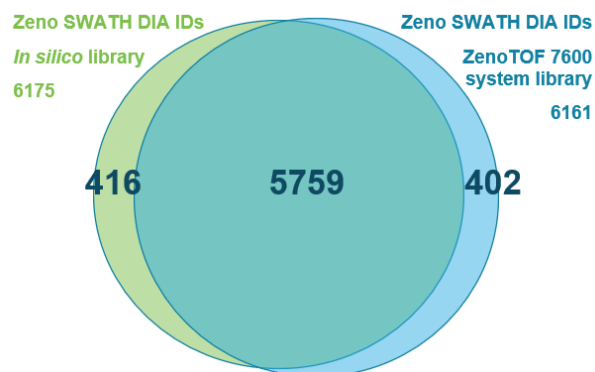
**Figure 4. Comparing protein identifications obtained with the library-free approach and 2 experimental libraries.** Zeno SWATH DIA data were processed with DIA-NN software using 3 different libraries (Lib Free, using an *in silico* library from a FASTA file; PHL, Pan Human library; ZT Lib, large library from Zeno DDA dataset collected on a ZenoTOF 7600 system). Using a 45-min gradient and a 400-ng sample load, the numbers of proteins identified at <1% FDR (transparent) and quantified at <20% CV (solid) were measured.

NN software. An *in silico* library can be generated in the span of a few hours, depending on the size of the FASTA file selected. It can then be saved to later process the Zeno SWATH DIA data in DIA-NN software, using the standard library extraction processing workflow.

The Zeno SWATH DIA data were used to identify more proteins and peptides than the Zeno DDA data, as shown in Figures 1 and 2, respectively. Gains in proteins identified were ~2-fold greater across the sample loading range tested and this result was consistent between the 2 ZenoTOF 7600 systems tested. The lists of identified proteins generated by the Zeno DDA and Zeno SWATH DIA workflows were compared and nearly all proteins identified from the Zeno DDA data were also identified using the Zeno SWATH DIA approach (Figure 3).

## Library-free approach vs. using experimental spectral libraries

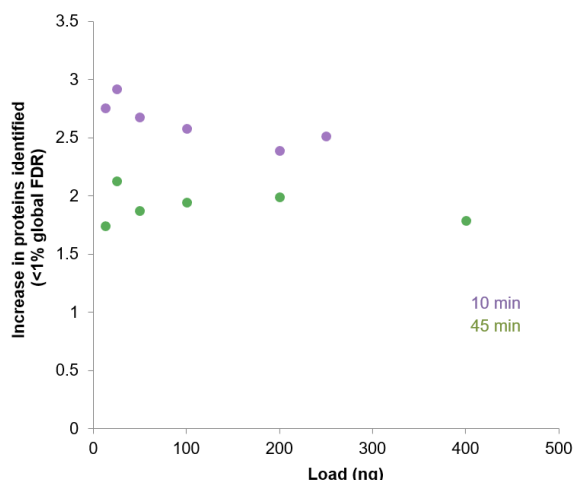
Next, the results generated using the emerging library-free approach were compared with those from a traditional SWATH DIA data processing strategy that uses experimentally generated spectral libraries to extract peptide and protein data from a SWATH DIA dataset. Two experimentally derived libraries were selected for comparison, including a large library previously generated using the ZenoTOF 7600 system and fast microflow chromatography<sup>7</sup> and the publicly available Pan Human library.<sup>3</sup>



**Figure 5. Overlap of protein identifications between Zeno SWATH DIA processed with 2 different libraries.** Using the 45-min gradient data collected on instrument 1, proteins identified from Zeno SWATH DIA using the *in silico* generated library were compared to those identified using the large ZenoTOF 7600 library. Strong overlap between the proteins identified was observed.

Zeno SWATH DIA data that were acquired in triplicate were processed using DIA-NN software. The number of proteins identified (<1%FDR) and quantified (<20% CV) using the library-free approach and experimentally derived libraries were compared (Figure 4).

The protein identification numbers from the 3 different data processing approaches were similar, highlighting the power of the library-free approach. The proteins identified using the *in silico* generated library and using the ZenoTOF 7600 system library were compared to determine whether similar proteins were found by each approach. The Venn diagram (Figure 5) shows that very similar lists of proteins were generated, adding further confidence to the approach.



**Figure 6. Fold increase in proteins identified using the Zeno SWATH DIA vs. Zeno DDA approach.** The number of proteins identified at <1% FDR was compared between the library-free Zeno SWATH DIA approach and the traditional Zeno DDA approach. The fold increase in proteins identified was computed for both the 45-min and 10-min gradient data.

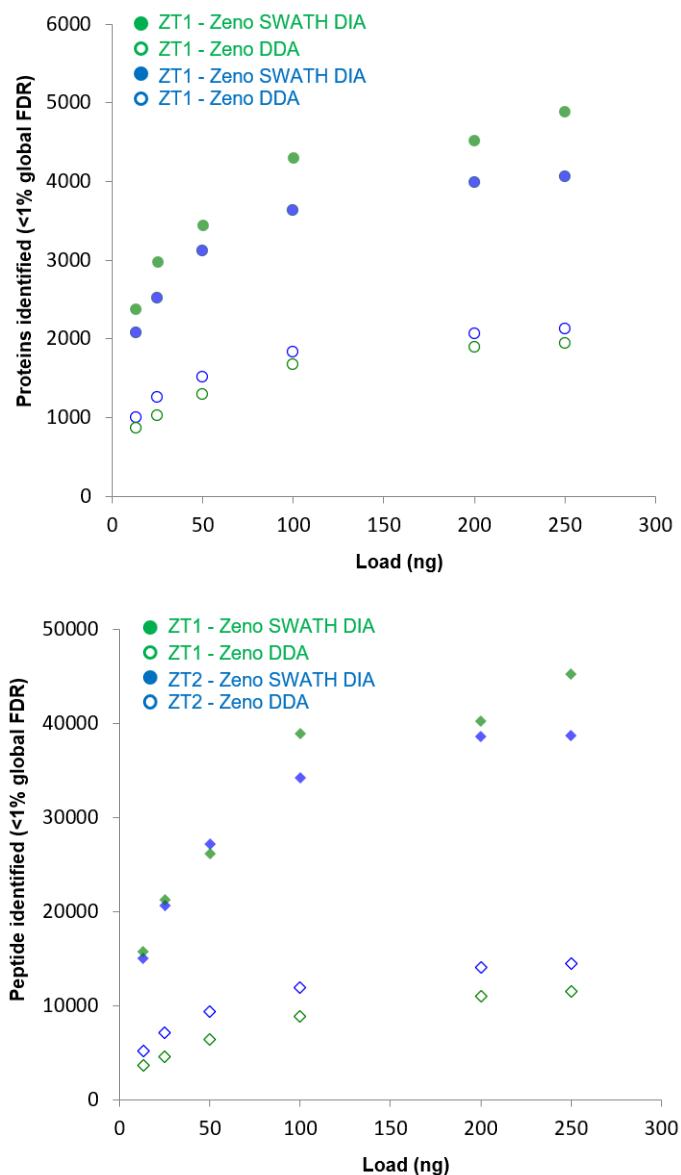
The Zeno SWATH DIA approach offers the additional advantage that the rich MS/MS dataset it generates can be used to derive more robust, specific, quantitative data, rather than using the MS1 data for quantification from Zeno DDA. Approximately 90% of the proteins identified were also quantified with <20% CV across triplicate analysis (Figure 4) for each library strategy when using the Zeno SWATH DIA workflow.

## Using faster chromatography

The study was replicated using a rapid 10-min LC gradient to confirm whether the Zeno SWATH DIA workflow would yield similar gains in identifications with faster chromatography. With the 45-min gradient, ~2-fold gains were observed across the load curve with the Zeno SWATH DIA approach, relative to the Zeno DDA approach (Figure 6). With the faster gradient, a greater increase in gains was observed using Zeno SWATH DIA, as 2.5- to 3-fold more proteins were identified from the Zeno SWATH DIA data under these conditions. This observation is expected, as DDA approaches are stochastic and therefore the acquisition strategy is not suitable to analyze large numbers of rapidly eluting precursors. The load curves for the proteins and peptides identified from the 10-min gradient data are shown in Figure 7, highlighting the improvements observed with the Zeno SWATH DIA approach over Zeno DDA.

## Conclusions

The Zeno SWATH DIA workflow described here was superior to the traditional shotgun proteomics approach using Zeno DDA for the identification of large numbers of proteins from a complex proteomic sample. Using an *in silico*-generated library for data processing with DIA-NN software, ~2-fold more proteins were identified from a cell lysate using a 45-min LC gradient and 2.5- to 3-fold more were identified using a 10-min gradient. Comparing the proteins identified using the *in silico*-generated library to 2 commonly used experimentally generated libraries demonstrated high agreement, lending confidence to the library-free workflow as an easy and comprehensive alternative for large-scale protein identification experiments.



**Figure 7. Comparing peptides identified for a 10-min gradient.** At a range of sample loadings, Zeno SWATH DIA data processed with the *in silico* generated library (solid) outperformed Zeno DDA data (open) for the identification of proteins (top) and peptides (bottom) on both ZenoTOF 7600 systems (green vs. blue).

## References

1. Wu CC, MacCoss MJ (2002) Shotgun proteomics: tools for the analysis of complex biological systems. *Curr Opin Mol Ther*, **4(3)**, 242-250.
2. Krasny L, Huang PH (2021) Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Mol. Omics* **17**, 29-42.
3. Large-scale, targeted, peptide quantification of 804 peptides with high reproducibility, using Zeno MS/MS. [SCIEX technical note, RUO-MKT-02-13346-A](#).
4. Demichev V *et al.* (2019) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods*, **17**, 41-44.
5. Rosenberger G *et al.* (2014) A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Scientific data*, **1**, 140031.
6. Zeno MS/MS with microflow chromatography powers the Zeno SWATH DIA workflow for more proteins quantified. SCIEX technical note, RUO-MKT-02-14668-A.
7. Processing ZenoTOF 7600 system data with DIA-NN software. [SCIEX community post, RUO-MKT-18-14611-A](#).
8. Large-scale protein identification using microflow chromatography on the ZenoTOF 7600 system. [SCIEX technical note, RUO-MKT-02-14415-A](#).
9. Creating a library from a FASTA file for library-free data analysis. [SCIEX community post, RUO-MKT-18-14611-A](#).

The SCIEX clinical diagnostic portfolio is For In Vitro Diagnostic Use. Rx Only. Product(s) not available in all countries. For information on availability, please contact your local sales representative or refer to <https://sciex.com/diagnostics>. All other products are For Research Use Only. Not for use in Diagnostic Procedures.

Trademarks and/or registered trademarks mentioned herein, including associated logos, are the property of AB Sciex Pte. Ltd. or their respective owners in the United States and/or certain other countries (see [www.sciex.com/trademarks](http://www.sciex.com/trademarks)).

© 2022 DH Tech. Dev. Pte. Ltd. RUO-MKT-02-14675-A.



**Headquarters**  
500 Old Connecticut Path | Framingham, MA 01701 USA  
Phone 508-383-7700  
[sciex.com](http://sciex.com)

**International Sales**  
For our office locations please call the division  
headquarters or refer to our website at  
[sciex.com/offices](http://sciex.com/offices)