Label-free protein quantitation of protein mixtures using Zeno SWATH data-independent acquisition (DIA)

Nanoflow quantitative proteomics using the ZenoTOF 7600 system

Takeshi Shibata[1], Ushio Takeda[1], Haruka Kumabe[2], Takeshi Masuda[3], Sumio Ohtsuki[2], Claudia P. Alvarez[4], Stephen Tate[4] and Patrick Pribil[4]

[1]SCIEX, Japan; [2]Kumamoto University, Japan; [3]Keio University, Japan; [4]SCIEX, Canada

This technical note describes the use of DIA for the label-free quantitation (LFQ) of trypsin-digested proteins from multiple organisms (human, yeast and *E. coli*) mixed at different ratios. Using nanoflow LC and variable-window Zeno SWATH DIA on a SCIEX ZenoTOF 7600 system, a high depth of protein detection and a high degree of quantitative fidelity are demonstrated. More than 14,000 total protein groups were identified in the mixtures and the quantitative ratios of protein abundances between the mixtures closely matched the expected theoretical ratios (Figure 1), thereby demonstrating the power of the ZenoTOF 7600 system for high-depth label-free quantitative proteomics research.

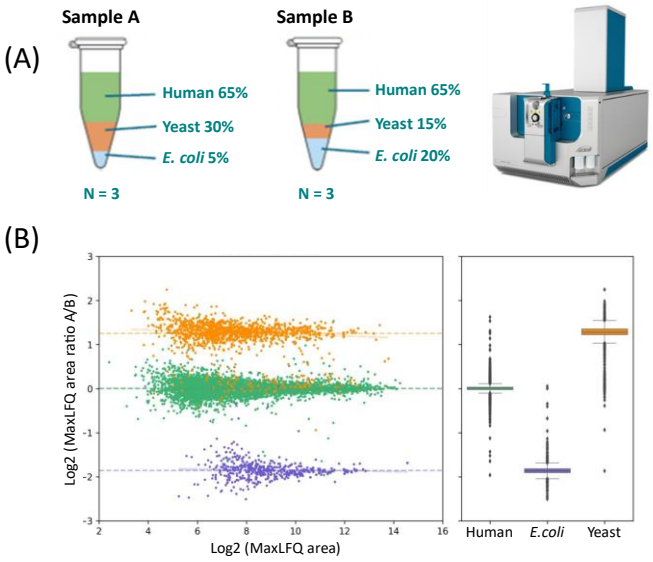

Figure 1. Label-free quantitation workflow using nanoflow LC with Zeno SWATH DIA on the ZenoTOF 7600 system. (A) Two sets of mixtures were prepared from human, yeast and *E. coli* tryptic digests, with different ratios of each digest. Samples were analyzed using nanoflow chromatography on a SCIEX ZenoTOF 7600 system, with a 100 variable-width window SWATH DIA method. Data were processed with a library-free approach using DIA-NN software, with a FASTA file combining the known protein sequences for the 3 organisms. (B) The Log2 of the protein group area ratios between the 2 samples were calculated using the MaxLFQ values from the DIA-NN result file. The ratios were plotted as a function of Log2 of the MaxLFQ areas for each protein group. The results show a very close match to the expected values for human (1:1), yeast (2:1) and *E. coli* (1:4), owing to the high degree of reproducibility and accuracy of Zeno SWATH DIA. The numbers of protein groups and precursors used in the ratio calculations for each organism are indicated in the plot.

The detection and quantitation of proteins are the cornerstone of proteomics. The ability to quantify changes in protein abundance in complex matrices is key to clinical and disease biomarker discovery. The main mass spectrometry-based techniques used for protein profiling consist of either data-dependent acquisition (DDA) or DIA methods. With DDA methods, sample labeling/tagging strategies can be used, in which samples are modified with either isobaric or non-isobaric labels, followed by MS and/or MS/MS quantitation. Although labeling strategies have advantages, such as the ability to multiplex samples, they also suffer from drawbacks such as dynamic range compression or added complexity of the MS1 space. In addition, the costs of labeling strategies can be prohibitive[1]. By comparison, LFQ strategies are attractive because of the simplified and cost-effective nature of the workflow. DIA methodologies have become increasingly popular and are often used for the LFQ analysis of peptides and proteins because of their sensitivity, simplicity, accuracy and reproducibility. This work shows the application of Zeno SWATH DIA for LFQ analysis of sets of protein mixtures.

Key features of label-free quantitation of protein mixtures using Zeno SWATH DIA on the ZenoTOF 7600 system

- More than 14,000 protein groups were identified in the mixtures of human, yeast and *E. coli* digests using a 70-minute nanoflow LC gradient with Zeno SWATH DIA and ≥95% of these protein groups were quantified with CVs <20%
- Zeno trap pulsing improved the sensitivity of MS/MS and resulted in significant gains in the detection and quantitation of protein groups and precursors
- The ratios of the protein abundances for each organism between the 2 mixtures closely matched the expected theoretical ratios (1:1 for human, 2:1 for yeast and 1:4 for *E. coli*), thereby highlighting the power of Zeno SWATH DIA for a label-free quantitation proteomics workflow

Methods

Sample preparation: Commercial yeast protein tryptic digest was purchased from Promega. Tryptic human and *E. coli* digests were prepared from human cultured cells (KMS-12PE) and DH5α, respectively, using a phase transfer surfactant method[2]. Digests were diluted in water containing 0.1% formic acid and mixed at the indicated ratios. Sample A contained 65% human, 30% yeast and 5% *E. coli* digests, whereas Sample B contained 65% human, 15% yeast and 20% *E. coli* digests. The final concentration of total combined protein in both samples was set to 200 ng/µL.

Chromatography: Nanoflow LC analysis was carried out using a Waters M-Class LC system. Direct-inject separation was performed using a Waters nanoEase M/Z Peptide CSH column (75 µm x 150 mm, 130 Å, 1.7 µm particle size). Mobile phase A was water with 0.1% formic acid and mobile phase B was acetonitrile with 0.1% formic acid. The LC flow rate was 300 nL/min and 2.5 µL injections (i.e. 500 ng on column) were performed. A 70-minute gradient profile (from 3% to 30% mobile phase B) was used with a total runtime of 120 minutes.

Mass spectrometry: Analyses were performed on a SCIEX ZenoTOF 7600 system, using the OptiFlow interface and OptiFlow Turbo V ion source with a nanoflow probe. A nebulizing gas setting of 10 psi and a curtain gas setting of 25 psi were used. The ionspray voltage was set to 3000 V and the source temperature was set to 200º C. Zeno SWATH DIA experiments consisted of a TOF MS scan from 400 – 1250 amu with an accumulation time of 100 ms, followed by 100 variable-width SWATH DIA windows spanning the Q1 mass range from 400 – 1000 amu. MS/MS scans were performed with a mass range of 100 – 1500 amu and an accumulation time of 25 ms, using dynamic collision energy and Zeno trap pulsing either on or off.

Data processing: Zeno SWATH DIA data were processed using DIA-NN software, version 1.8.1[3]. Library-free searches were performed using a combined FASTA database comprising human, yeast and *E. coli* protein sequences from UniProt (totaling 65,151 proteins and 34,343 genes)[4]. Default processing settings were used, with match between runs (MBR) checked and protein inference turned off. The total number of protein groups and precursors identified and quantified in the datasets were determined from the resulting output pg.matrix.tsv and pr.matrix.tsv files, respectively. The LFQ area ratio plots were generated from the MaxLFQ values (calculated from protein groups and precursors detected in all 3 replicates for each sample) using Python software.

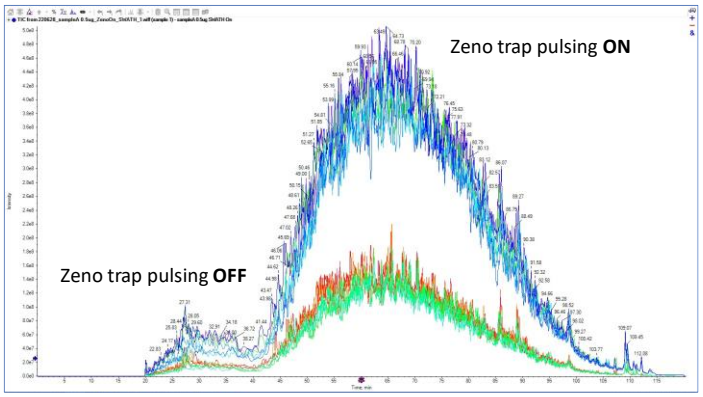

Figure 2. LC-MS reproducibility and sensitivity gains from Zeno trap pulsing in MS/MS with SWATH DIA for samples A and B. Overlays of the total ion chromatograms (TICs) for triplicate injections of each sample, using a 100 variable-width window SWATH DIA method either with or without Zeno trap pulsing for MS/MS. Zeno trap pulsing in MS/MS on the ZenoTOF 7600 system typically improves fragment ion signal intensities of tryptic peptides between 6- and 10-fold, as evidenced by the increase in overall TIC signal intensities.

Identification and quantitation of protein groups and precursors

Experiments were performed either with or without Zeno trap pulsing for MS/MS in SWATH DIA mode. Zeno trap pulsing increases peptide fragment ion intensities between 6- and 10-fold[5]. The gains in MS/MS sensitivity are shown at the total ion chromatogram (TIC) level in Figure 2, in which the overall TIC intensities are higher with Zeno trap pulsing. The numbers of protein groups and precursors identified and quantified (CV values <20%) are summarized in Figure 3. Zeno trap pulsing increased protein group identifications from 11,909 to 14,077 for sample A and from 11,834 to 14,216 for sample B. However, the respective gains were more evident in protein groups with CVs <20%, which increased from 10,452 to 13,540 for sample A and from 10,304 to 13,189 for sample B, highlighting the benefits of Zeno trap pulsing for quantitation. Gains in precursors identified and quantified were likewise evident with the activation of the Zeno trap. For sample A, the number of precursors identified increased from 65,738 to 85,918 and those with CV <20% increased from 63,465 to 84,683. Similarly, for sample B the precursors identified increased from 54,439 to 82,024 and those with CV <20% increased from 52,301 to 75,359. Together, these results reveal that Zeno trap pulsing increases the percentage of quantifiable protein groups between 5% and 9%, and increases the quantifiable precursors between 7% and 12% (Figure 3).

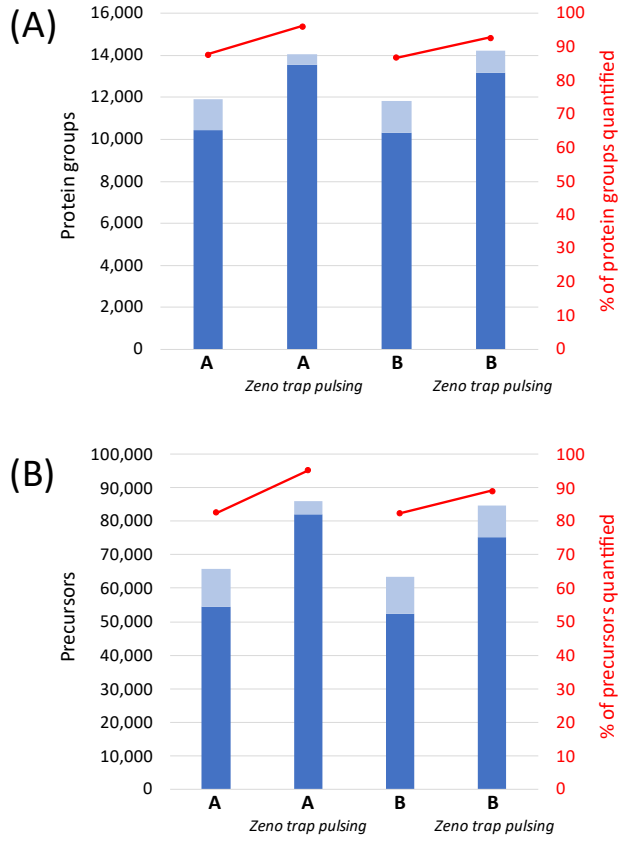

Figure 3. Protein groups and precursors detected and quantified in both samples. SWATH DIA data were processed with DIA-NN software using a library-free approach, against the combined FASTA sequences for human, yeast and *E. coli*. The numbers of protein groups (A) and precursors (B) identified and quantified are shown. Light bars represent the total numbers of protein groups and precursors detected at 1% FDR, while the numbers of protein groups and precursors quantified with CVs <20% are shown by the dark-shaded bars. The percentages of protein groups or precursors quantified out of the total numbers detected are shown in red. Values were extracted from the DIA-NN software output files for the protein groups (pg.matrix.tsv) and precursors (pr.matrix.tsv. Experiments were performed either with or without Zeno trap pulsing for SWATH DIA MS/MS. The results show that the added sensitivity of Zeno trap pulsing improves detection and quantitation rates in both samples.

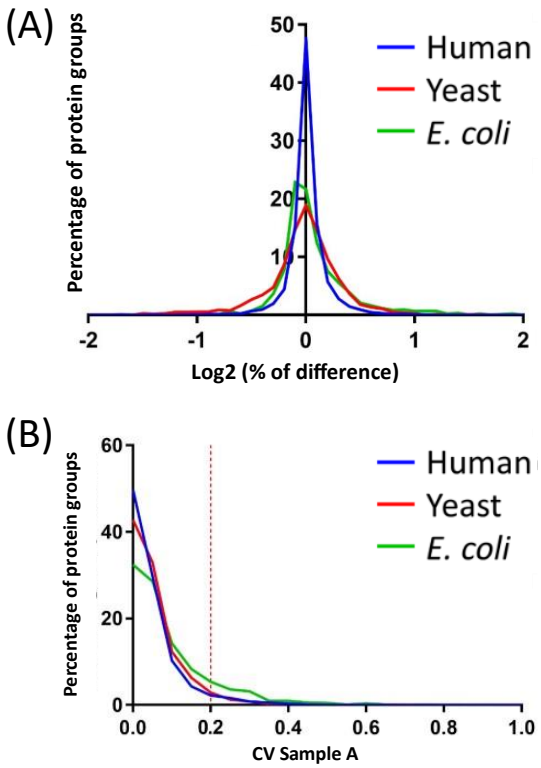

Figure 4. Quantitative statistics for the label-free quantitation analysis. (A) Distribution of the Log2 difference of the A/B ratio from the average for each species. (B) Distribution of the CVs for protein groups identified in sample A, showing the majority have CVs <20%.

Quantitation of organism-specific protein ratios

To calculate the protein group and precursor ratios for the different organisms between the two sets of mixtures, the MaxLFQ values generated from the DIA-NN software processing were used[6]. The Log2 of the MaxLFQ area ratios between samples A and B were plotted as a function of the Log2 of the areas. Figure 1B shows the results of the label-free quantitation of 7,414 total protein groups and 51,487 total precursors for which MaxLFQ area ratios could be calculated using values present in all 3 replicates per sample. This is also broken down by organism in Figure 1B. The graph indicates that the A/B ratios match closely with the theoretical ratios for each organism (1:1 for human, 2:1 for yeast, and 1:4 for *E. coli*), highlighting the quantitative precision of Zeno SWATH DIA. The high degree of precision is further demonstrated in Figure 4. Figure 4A shows the distribution of the Log2 difference in the A/B ratio across the dataset. The Log2 of the difference between the A/B ratio was calculated for a given protein, divided by the average A/B ratio High quantitative reproducibility is indicated by a narrow distribution. In this case, the distribution for each organism was ± 1 Log2 values. The high degree of reproducibility is further demonstrated by the plotted CV distribution for all protein groups identified. Figure 4B indicates that most of the protein group intensity values for the set of identified proteins in Sample A have CVs <20%.

Conclusions

- Zeno SWATH DIA was used on the ZenoTOF 7600 system to detect and quantify peptides and protein groups in 2 different mixtures of protein digests from 3 organisms. Zeno trap pulsing provided significant gains in the detection and quantitation of protein groups and precursors in the samples
- Using Zeno SWATH DIA, >14,000 protein groups and >85,000 precursors were identified in both samples with a 70-minute nanoflow LC gradient, with 95% of the protein groups quantified with CVs <20%
- The protein intensity ratios between samples A and B closely matched the expected values for each organism in the mixture, highlighting the depth and utility afforded by Zeno SWATH DIA for label-free protein quantitation

References

1. Rozanova, S et al. (2021) Quantitative mass spectrometry-based proteomics: an overview. Methods Mol. Biol. 2228, 85-116.
2. Masuda, T et al. (2008) Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. J. Proteome Research 7, 731-40.
3. Demichev V et al. (2019) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nature Methods, 17, 41-44.
4. Going library-free for protein identification using Zeno SWATH DIA and in silico-generated spectral libraries. SCIEX technical note, RUO-MKT-02-14675-A.
5. Zeno MS/MS with microflow chromatography powers the Zeno SWATH data-independent analysis (DIA) workflow to quantify more proteins, SCIEX technical note, RUO-MKT-02-14668-A.
6. Cox, J et al. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics 13, 2513-2526.

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