BRUKER

dia-PASEF® applied on different gradient lengths

Data Independent Acquisition (DIA) workflows have gained in popularity as they overcome the issue of stochastic selection of peptide precursors encountered in typical data-dependent approaches (DDA) and thereby promise reproducible and accurate protein identification and quantification across large sample cohorts.

Abstract

In collaboration with groups at the ETH IMSB in Zurich, the Max Planck Institute in Martinsried, and the University of Toronto, we have previously combined the advantages of DIA with the inherent efficiency of PASEF (Parallel Accumulation Serial Fragmentation) to create a new acquisition mode, called dia-PASEF® [1]. In this application note, we wanted to further investigate the benefits of applying dia-PASEF to different proteomics samples, namely HeLa and Yeast digest, using different gradient lengths down to 30 minutes. In brief, more than 7600 and 4400 protein groups could be reliably identified and quantified from HeLa and Yeast protein digests, respectively, in triplicate 90 minute gradients each, at a false discovery rate of 1%. When reducing the gradient time to 30 minute we were still able to identify on average 6395 HeLa and 4032 Yeast protein groups.

Authors: Stephanie Kaspar-Schoenefeld[1], Thomas Kosinski[1], Markus Lubeck[1], Oliver Raether[1], Gary Kruppa[1];
[1] Bruker Daltonik GmbH, Bremen, Germany

Introduction

Mass spectrometry (MS)-based proteomics has become a powerful technology for the identification and quantification of thousands of proteins from biological samples, e.g. a single cell type. However, even state-of-the-art mass spectrometers use only a small fraction of the ion beam for mass analysis. In principle, a 100% duty cycle could be achieved by parallel ion storage and sequential release from a trapped ion mobility (TIMS) device into a quadrupole time of flight mass analyzer. Synchronizing the release of ions from the TIMS device with the quadrupole (PASEF) increases the MS/MS sequencing speed more than 10-fold (> 100 Hz) resulting in a sensitivity gain [2]. However, while analyzing sample cohorts, and because of the high complexity of most proteomics samples, the same peptides are typically not targeted in all runs, leading to the so-called "missing value problem". This can be partially overcome with "match between runs" approaches.

Data-independent acquisition (DIA) is an alternative approach which promises reproducible and accurate protein quantification across large sample cohorts by using wider selection windows than used in DDA methods. The mass spectrometer typically cycles through many isolation windows covering a broader *m/z* range of interest. Narrower isolation windows increase specificity, but also entail longer cycle times when sampling the same mass range. As the high reproducibility of the CCS values make library-based approaches very attractive, we simply transferred the PASEF principle to DIA resulting in a new acquisition method, called dia-PASEF. This method also benefits from the sensitivity of the PASEF method and improves efficiency of ion usage for MS/MS with up to 100% of the ion beam being fragmented for MS/MS analysis.

We previously showed the identification of more than 7000 proteins in single, 2 h runs from human cell lysate peptides [3] and the reproducible and accurate quantification of more than 8000 proteins from a complex mixture of three proteomes [4].

Here we explore the potential of dia-PASEF on different gradient lengths (30, 60, 90 min) analyzing samples of different complexity, namely a human cancer cell line (HeLa) and Yeast.
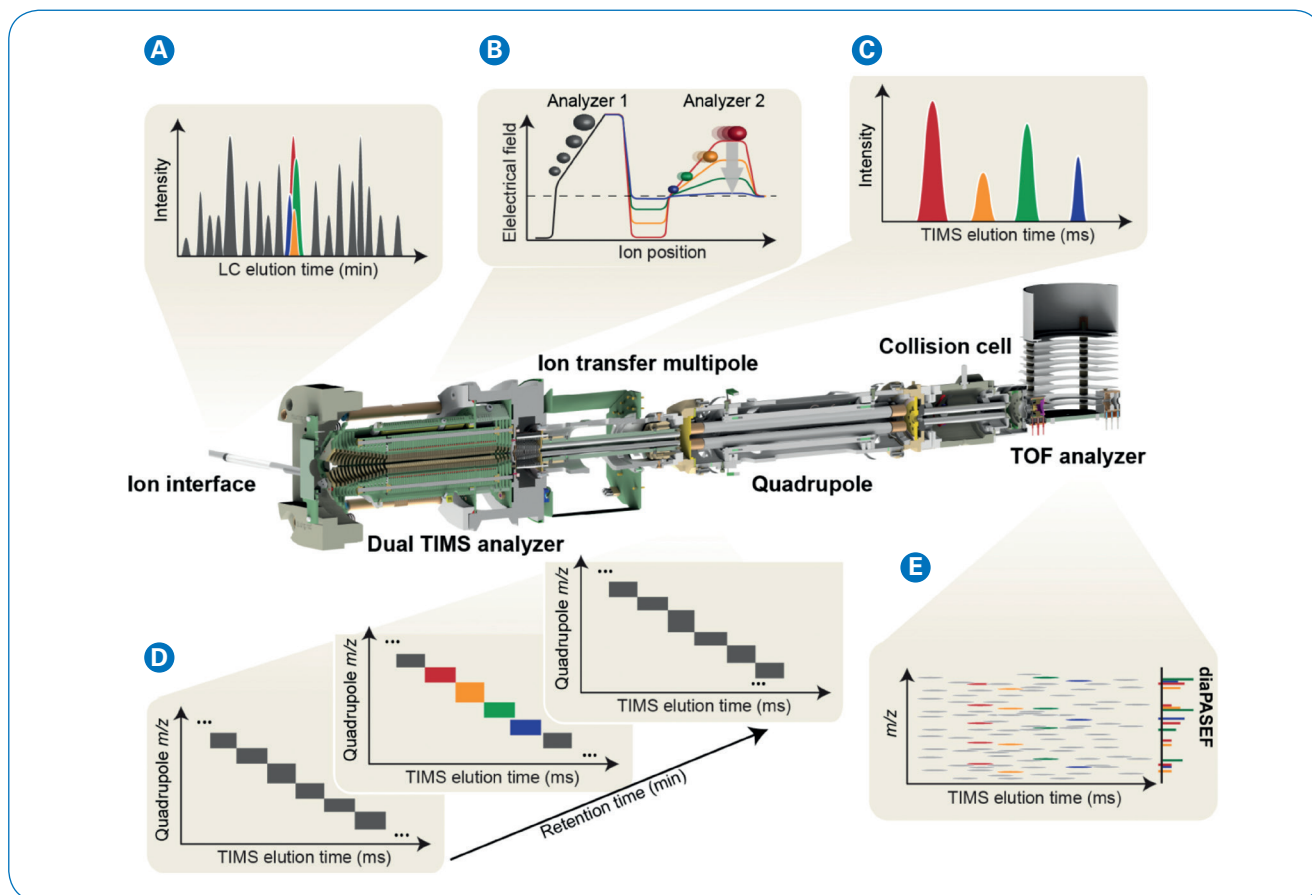

Figure 1: Overview dia-PASEF workflow. **(A)** *Peptides eluting from the chromatographic column are ionized and enter the mass spectrometer through a glass capillary.* **(B)** *In the dual TIMS analyzer, the first TIMS section traps and stores ion packets, and the second resolves them by ion mobility.* **(C)** *Ion mobility separated ions are released sequentially from the second TIMS analyzer as function of decreasing electrical field strength.* **(D)**, **(E)** *For dia-PASEF we couple DIA isolation windows to the precise ion mobility elution of the corresponding ions. Within a single TIMS separation multiple precursor windows are set*
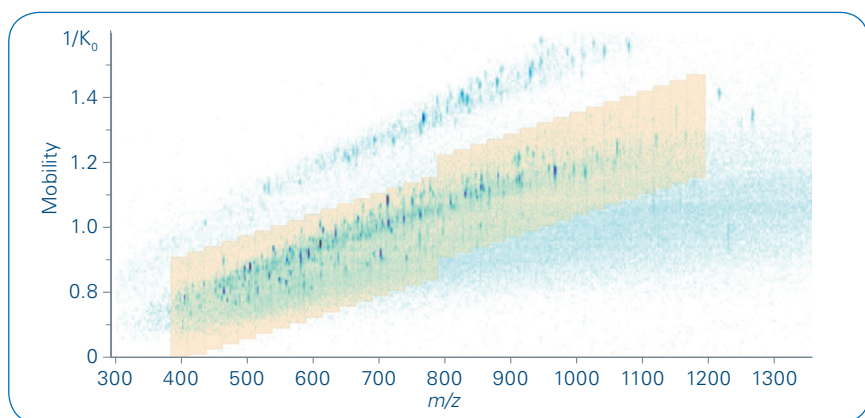
Figure 2: Method scheme for dia-PASEF. The applied method consists of two windows in each 100 ms dia-PASEF scans. Sixteen of these scans cover the diagonal scan line for doubly and triply charged peptides in the m/z – *ion mobility plane with narrow 25* m/z *precursor windows.*

Methods

Purified and predigested Yeast standard was purchased from Promega (Madison, USA) and resuspended in 0.1% formic acid. Whole HeLa cell pellets were purchased from CIL Biotech (Mons, Belgium). Cell lysis was performed using trifluoroethanol (TFE) according to [5]. Proteins were enzymatically cleaved by adding trypsin in a 1:100 (wt/wt) enzyme:protein ratio. Samples were analyzed by coupling a nanoflow liquid chromatography system (nanoElute, Bruker Daltonics) online to a trapped ion mobility spectrometry – quadrupole time of flight mass spectrometer (timsTOF Pro, Bruker Daltonics). Peptides were separated on a reversed-phase C 18 column (25 cm x 75 µm i.d., 1.6 µm, IonOpticks, Australia) using a 30-, 60-, or 90-minute gradient of 2 – 37% B (0.1% formic acid in acetonitrile) at a constant flow rate of 400 nl/min. Column temperature was controlled a 50°C.

For dia-PASEF, we adapted the instrument firmware to perform data-independent isolation of multiple precursor windows within a single TIMS separation (100 ms). We used a method with two windows in each 100 ms dia-PASEF scan. Sixteen of these scans covered the diagonal scan line for doubly and triply charged

peptides in the *m/z* - ion mobility plane with narrow 25 *m/z* precursor windows.

For targeted data extraction, we used a project-specific library generated from 48 high-pH reverse peptide fractions. Peptides were separated using a 100-minute gradient of 2 – 37% B (0.1% formic acid in acetonitrile. DDA MS data for the library was acquired with the PASEF method [2] as follows: MS data were collected over an *m/z* range of 100 to 1700, during each MS/MS data collection each TIMS cycle time was 1.1 seconds; each cycle included 1 MS and 10 MS/MS 100 msec TIMS scans; in each of the 10 PASEF MS/MS scans an average of 12 precursors were selected, resulting in an MS/MS data acquisition rate of 109 Hz.

Raw files were processed using a developmental version of Spectronaut (Biognosys). The ion mobility enhanced library was generated using Spectronaut's Pulsar database search engine with 1% FDR control at PSM, peptide and protein level. We only generated the library using the 90-minute gradients on the fractionated samples, and then used the extensive retention time capabilities available in Spectronaut for targeted extraction from the 60- and 30-minute gradient data. For targeted analysis of dia-PASEF

data we used a 1% FDR at peptide and protein level.

Results and Discussion

dia-PASEF combines the well-known PASEF method, which exploits trapped ion mobility spectrometry to achieve high duty cycle, efficient ion usage and improved peptide identification rates with DIA (Figure 1). By coupling DIA isolation windows to the precise ion mobility elution of the corresponding ions, the combination of windowed DIA with the PASEF principle allows multiplexing of DIA windows in a single 100 ms ion mobility separation of precursor ions.

Here we applied the dia-PASEF method for the analysis of human cell lysate as well as for Yeast using three different gradient lengths (30, 60 and 90 min gradients). For the study presented here we used a sample optimized dia-PASEF method with 2 windows in each 100 ms dia-PASEF scan.Sixteen of these scans (resulting in 32 windows) covered the diagonal scan line in the *m/z* – ion mobility pane to ensure coverage of doubly and triply charged species with narrow 25 *m/z* isolation windows (Figure 2). Each dia-PASEF cycle was started with one MS1 survey scan. This setup results in a total cycle time of 1.7 s (1x 100 ms MS1 survey scan, 16x 100 ms dia-PASEF scans), which allowed us to obtain sufficient data points over the chromatographic peak while maintaining best compromise between specificity and sensitivity.

Processing of the four-dimensional data was done using a developmental version of Spectronaut. The Spectronaut workflows generate spectral libraries using the integrated Pulsar database search engine directly from fractionated DDA PASEF runs, which are used for targeted data extraction. We created a project-specific library from high-pH using 90-minute gradient

time runs. As deeper ion libraries increase the depth of coverage with DIA in general, we decided to use the 90-minute library for targeted data extraction not only from the 90-minute, but also from the 60- and 30-minute gradient time data. The library comprised 220,628 unique target peptides sequences and 11,578 target protein groups for HeLa sample and 80,874 unique target peptides and 5127 target protein groups for Yeast, respectively.

We used these comprehensive libraries for targeted data extraction in Spectronaut, which includes fully automated in-run calibration (retention time, mass and mobility), automatic interference correction, targeted analysis and automatic quality control. We could identify and quantify a total 111,706 of peptides and on average 97,363 peptide precursors per run (with 87% of them identified in all runs) from triplicate injections of HeLa cell digests using our standard 90 min gradient (Figure 3). On protein level on average 7678 protein groups could be identified in single PASEF runs at a 1% FDR. For yeast digest we identified a total 60,116 of peptides and on average 52,370 peptide precursors and 4410 protein groups per run.

DIA-based approaches benefit from sampling all the peptides present in the selected mass ranges rather than being dependent on precursor selection algorithms. This leads to very reproducible identification rates in replicate injections. When evaluating the triplicate injections the cumulative increase in identifications we detected only a slight increase of 31 proteins for HeLa and 74 for Yeast sample (Figure 4) underlining the outstanding reproducibility of identified proteins using dia-PASEF.

We further decreased gradient time to improve sample throughput down to 60 and 30 min. For HeLa we identified on average 86,821
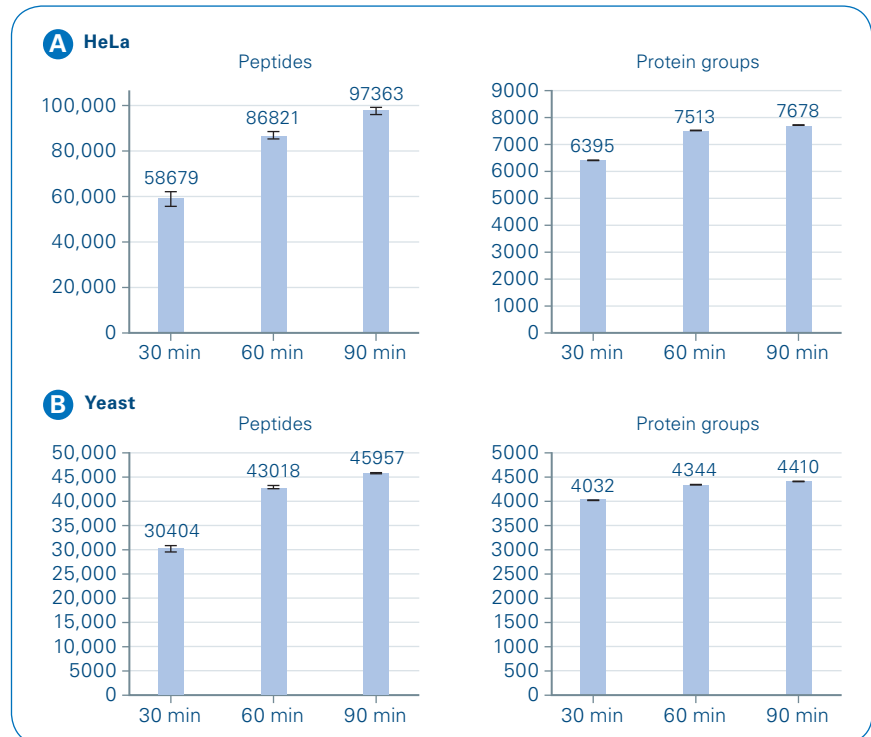

Figure 3: Number of identified peptides and proteins at an FDR of 1% for the different gradient length for (A) HeLa and (B) Yeast sample. Shown are the average values from the three runs together with the standard deviation

peptide sequences and 7513 protein groups with a 60 minute gradient, and 58,679 peptide sequences and 6395 protein groups with a 30 minute gradient. For Yeast using a 60 min gradient led to the identification of 43,018 peptide sequence and 4344 protein groups, and using a 30 min gradient to 30,404 and 4032 protein groups on average.

In total, we covered 66% and remarkable 86% for HeLa and Yeast, respectively, of all protein groups in our comprehensive libraries in our 90 min gradient single runs without fractionation (Figure 5). Using 30 min gradients we can still cover 55% and 79% of all protein groups present in the library created using 90 min gradients for HeLa and Yeast, respectively, in single runs.

For proteomics approaches it is not only important to increase the proteome coverage but also to precisely quantify detected proteins. To test the quantitative accuracy of

dia-PASEF, we analyzed the reported protein intensities of the triplicate injections. The average median coefficient of variation (CV) for the reported data sets was at 8.6% on protein group level. In detail, using a 90 min gradient 89% (6915) of the identified HeLa proteins were reproducibly quantified with a CV below 20% and 80% (6159) with a CV below 10% illustrating the excellent reproducibility of dia-PASEF.

These results clearly underline the advantage of dia-PASEF for deeper proteome coverage and precise quantification even for shorter gradients down to 30 minutes. In future, we will investigate the potential of dia-PASEF for even higher throughput using shorter gradients delivered from the Evosep one system.
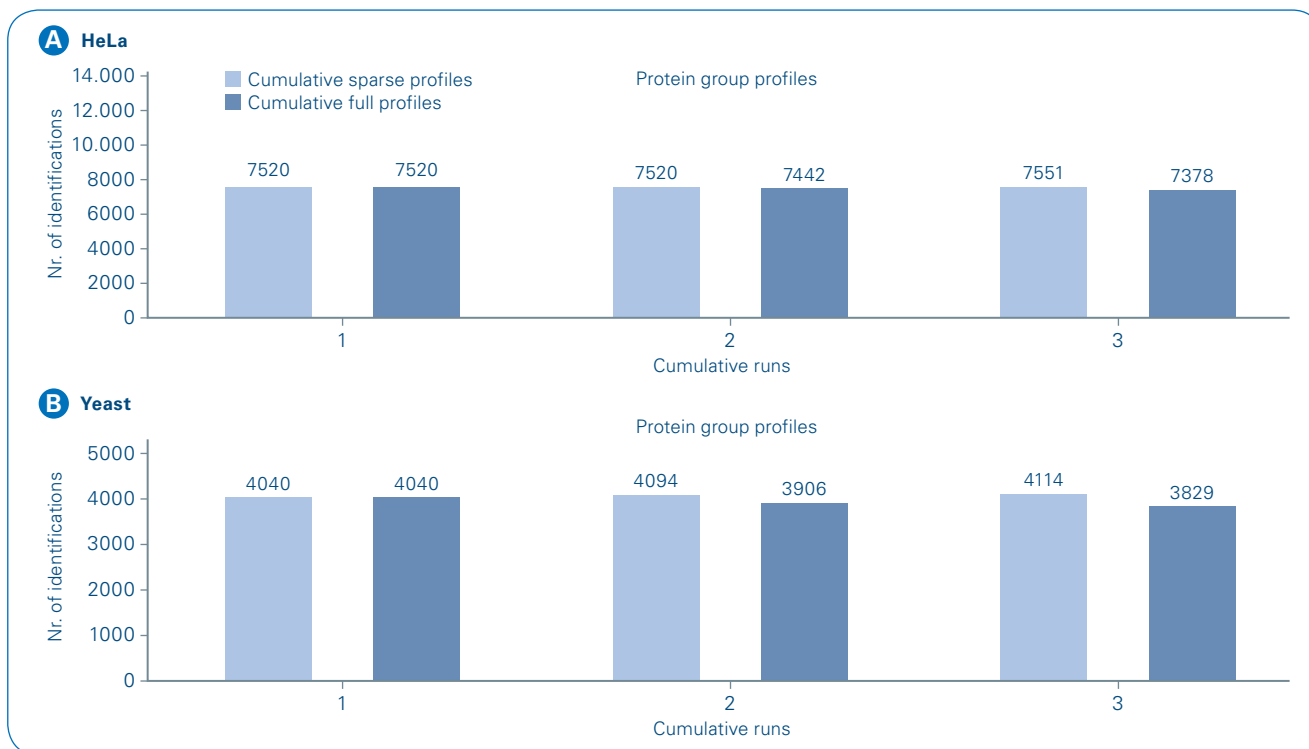
Figure 4: Overview cumulative protein group profiles for Ⓐ HeLa and Ⓑ Yeast sample. Shown are the results obtained from the 90 min gradient runs
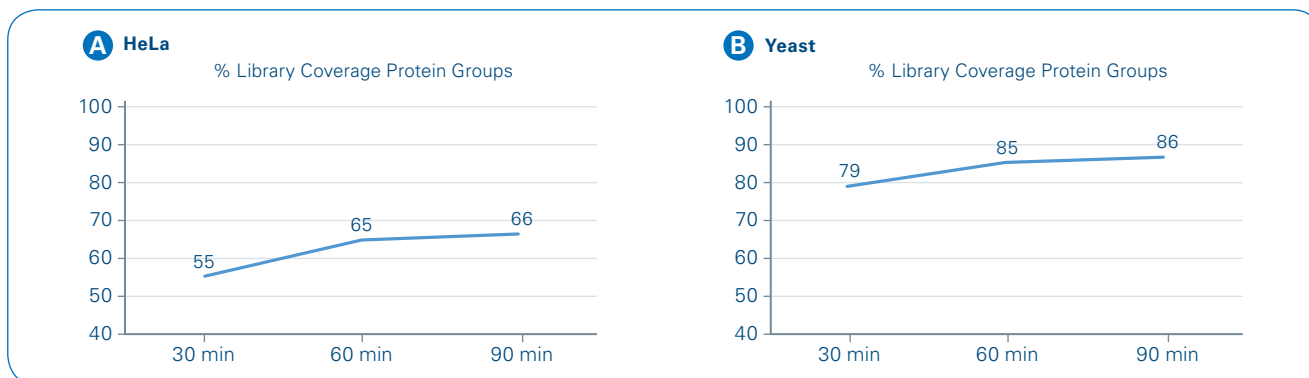

Figure 5: Library coverage for Ⓐ HeLa and Ⓑ Yeast sample on protein group level. Shown are the results for three different gradient lengths (30, 60, 90 min)

Conclusion

- dia-PASEF results in excellent results for different gradient lenghts.

- More than 4000 Yeast protein groups could be identified in a single-shot 30 min gradient resulting in a near-complete proteome coverage.

- Excellent identification rates have been reproducibly achieved for complex HeLa digest with 7600 protein groups identified at 90 min gradient time and close to 6400 protein groups using a 30 min gradient.

- With the excellent identification rates in combination with robust and reproducible quantification and improved ion usage efficiency for MS/MS, dia-PASEF is one further step toward the development of the ideal mass analyzer.

Learn More

You are looking for further Information?
Check out the link or scan the QR code for more details.

www.bruker.com/timstofpro

References

[1] Meier F. et al. (2019). bioRxiv, doi: https://doi.org/10.1101/656207

[2] Meier F. et al. (2018). Mol Cell Proteomics, doi: 10.1074/mcp.TIR118.000900

[3] AppNote LC-MS 157, https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/ApplicationNotes/1869385_LCMS-157_diaPASEF_ebook.pdf

[4] AppNote LC-MS 160, https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/ApplicationNotes/1871458_LCMS-160_diaPASEF_label-free_quantification_09_2019_ebook.pdf

[5] Wange et al. (2015). J Proteome Res, **4**(6): 2397-2403