High throughput 4D-Proteomics – Application of dia-PASEF® and the Evosep One for short gradients

The timsTOF Pro offers a combination of two unique technologies, namely a 4th dimension provided by Trapped Ion Mobility Spectrometry (TIMS) to enhance ion separation and sensitivity and Parallel Accumulation Serial Fragmentation (PASEF [1]) to improve ion utilization efficiency and data acquisition speed.

Abstract

In this application note, we demonstrate the benefits of dia-PASEF technology on the timsTOF Pro platform coupled to an Evosep One for high-throughput in-depth proteome analysis of up to 300 samples per day (SPD). We quantify about 5200 protein groups in only 21-minutes run time. In 4.8-minutes ultra-high throughput runs (300 SPD) we are still able to quantify more than 2000 protein groups and 8500 peptides.

Keywords: Proteomics, timsTOF Pro, dia-PASEF, high throughput, Evosep One

Authors: Stephanie Kaspar-Schoenefeld[1], Thomas Kosinski[1], Markus Lubeck[1], Oliver Rather[1], Gary Kruppa[1], Nicolai Bache[2], Dorte B. Bekker-Jensen[2]; [1] Bruker Daltonik GmbH, Bremen, Germany; [2] Evosep Biosystems, Odense, Danmark.

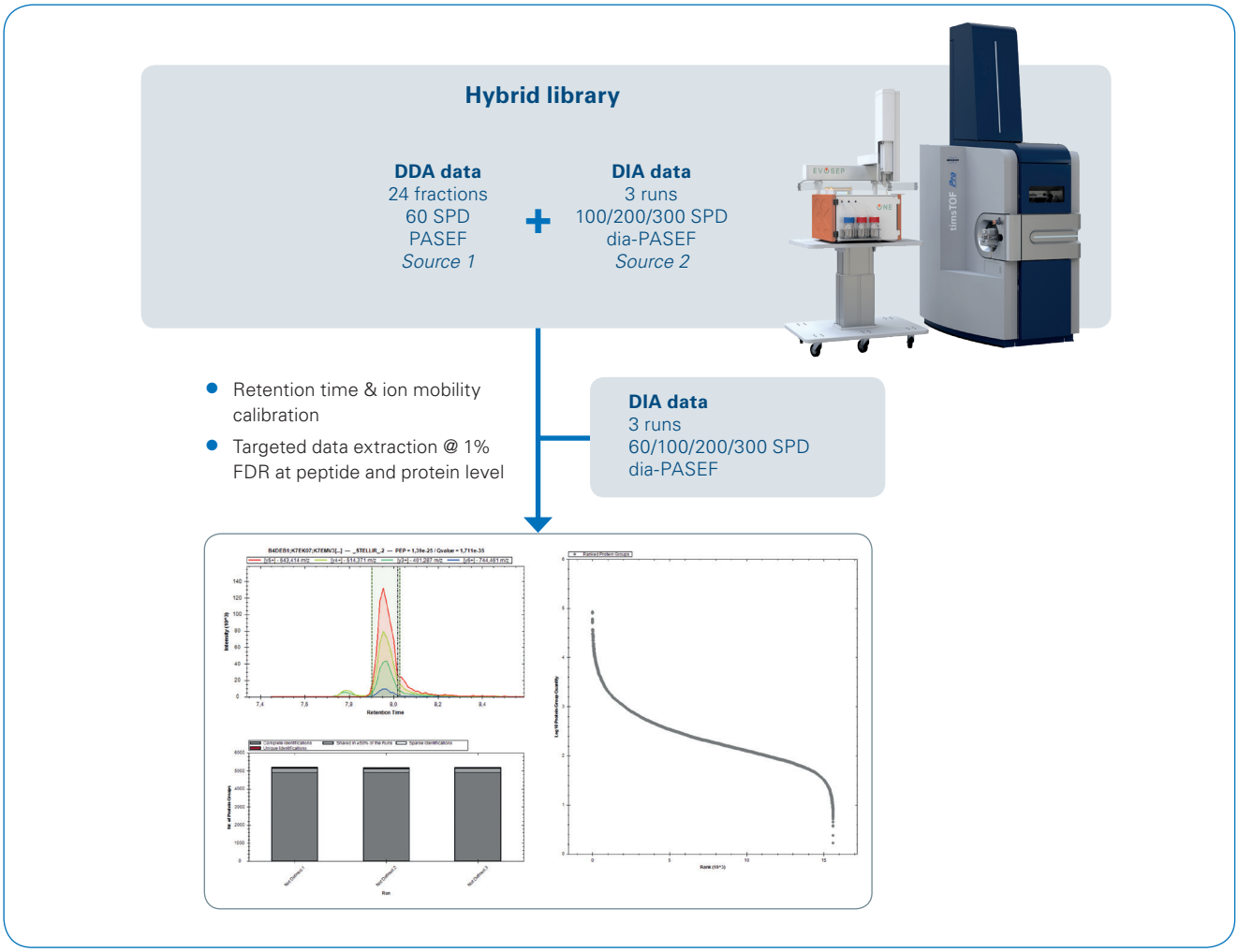

Figure 1: dia-PASEF data processing workflow using Spectronaut (v14). A resource-specific library was created from 24 fractions acquired using the 60 SPD Evosep method in combination with PASEF technology. Additionally, project-specific libraries were created for the 100, 200 and 300 SPD methods using dia-PASEF data to allow for source-specific retention time alignment. Both libraries have been combined into "Hybrid Libraries" used for targeted data extraction from the dia-PASEF data.

Introduction

Data-independent acquisition (DIA) facilitates reproducible and accurate protein identification and quantification across large sample cohorts. This is achieved by using isolation of wide quadrupole windows, rather than selecting individual precursors, to ensure that all precursor ions are fragmented in every sample. Ion mobility provides an additional dimension of separation for complex samples, which can also be used for aligning precursor and fragment ions. By making use of reproducible mobility values from the timsTOF Pro, we extend the PASEF principle

Table 1: LC conditions

Parameter	Settings
LC-System	Evosep One (Evosep Biosystems) Plugin version RC.Net driver 1.3
Separation Columns	Reversed-phase C18, 8 cm x 150 µm i.d., 1.5 µm, Evosep (for 100 and 60 SPD) Reversed-phase C18, 4 cm x 150 µm i.d., 1.9 µm, Evosep (for 200 and 300 SPD)
Mobile Phases	A: 0.1% formic acid (FA) in water B: 0.1% FA in Acetonitrile
Gradient	60 SPD 100 SPD 200 SPD 300 SPD
Column Temperature	50°C

to DIA resulting in a new acquisition mode, called dia-PASEF [2]. This approach benefits from the sensitivity of PASEF and improves efficiency of ion usage for MS/MS with up to 100% of the ion population being fragmented for MS/MS analysis.

We previously showed the applicability of dia-PASEF for different gradient lengths, with about 6400 protein groups identified in single 30-minute runs from human cell lines [3].

For true clinical proteomics, robust analysis of several hundreds of samples per day is highly desirable which in turn requires fast and robust instrumentation. The Evosep One is a conceptionally new HPLC that dramatically increases robustness and sample throughput while maintaining sensitivity inherent to nanoflow LC [4]. The timsTOF Pro with very high sensitivity and speed and the Evosep One turn out to be a perfect combination that could meet these requirements for analysis of clinical samples in a high-throughput robust manner.

In this application note, we demonstrate the performance of dia-PASEF on the timsTOF Pro mass spectrometer (Bruker Daltonics) in combination with the Evosep One HPLC system (Evosep) using ultra-short gradients for analysis of up to 300 samples per day.

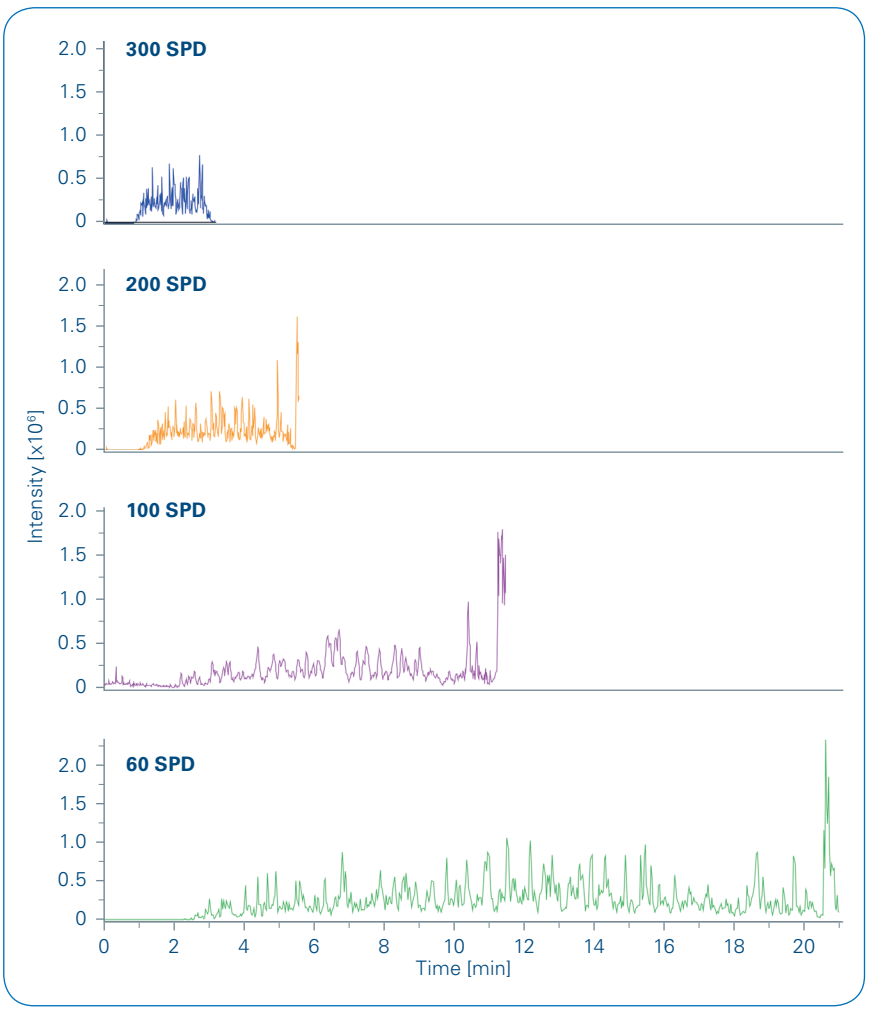

Figure 2: Example base peak chromatogram traces for the different Evosep One gradients for high throughput proteomics (300, 200, 100, 60 SPD).

Methods

Sample preparation

Whole HeLa cell pellets were purchased from CIL Biotech (Mons, Belgium). Cell lysis was performed using trifluoroethanol (TFE) according to [5]. Briefly, the suspension was kept 10 minutes on ice and then incubated at 56°C for 20 min. 200 mM dithiothreitol (DTT) was used to reduce the cysteine residues at 90°C for 20 min and 200 mM iodoacetamide (IAA) for alkylation of reduced cysteine residues (90 min at room temperature). Proteins were enzymatically cleaved overnight by adding trypsin in a 1:100 (wt/wt) enzyme:protein ratio. De-salting and purification were performed using a solid phase extraction cartridge (SepPak C18, Waters, USA) by diluting and washing protein digest with 0.1% formic acid (ACN) and subsequent elution with 50% (w/w) ACN in 0.1% FA. Purified and dried peptides were reconstituted in 0.1% FA.

We created a resource-specific library generated from 24 high-pH reversed phase peptide fractions. Samples were fractionated according to [6]. Briefly, 100 µg of peptides from the in-house HeLa digest was fractionated at pH 10 on a reversed phase column (Waters Acquity CSH C18 1.7 µm 1 x 150 mm) on a Dionex Ultimate 3000 system (Thermo Fisher Scientific). The fractions were freeze-dried and reconstituted in 0.1% FA.

The samples were loaded onto Evotips according to the manufacturer's instructions and analyzed on the Evosep One system (Evosep Biosystems, details see Table 1). The resource-specific library was generated using Evosep's 60 samples per day (SPD) method and dia-PASEF data has been acquired with the new and improved 60, 100, 200, and 300 SPD methods and associated new column recommendations.

Mass Spectrometry Settings

Eluting peptides (200 ng) were analyzed on the timsTOF Pro mass spectrometer (Bruker Daltonics). The timsTOF platform is uniquely equipped with state-of-the-art dual-TIMS funnel ion optics that sorts and time-focuses ions before they enter the quadrupole-time-of-flight (Q-TOF) mass analyzer, enabling 4D-Proteomics.

For dia-PASEF, the instrument firmware was adapted to perform data-independent isolation of multiple precursor windows within a single TIMS frame. An optimized dia-PASEF scheme for the short gradient methods was applied, targeting +2 and +3 ions in a three-window method (per 100 ms). Eight of these scans (resulting in 24 windows, each using 25 Da window size) covered an *m/z* range from 400 to 1000. Each dia-PASEF cycle includes 1 MS frame (100 ms) resulting in a total cycle time of 900 ms per dia-PASEF cycle.

For targeted data extraction, we used the resource-specific library generated form 24 high-pH reverse-phase fractions acquired with PASEF using a default proteomics method (see [3] for details). PASEF synchronizes MS/MS precursor selection with TIMS separation. This allows fragmentation of more than one precursor per TIMS scan and increases the sequencing speed several-fold without loss of sensitivity. The precursor selection engine dynamically selects precursors on intensity, *m/z*, and ion mobility.

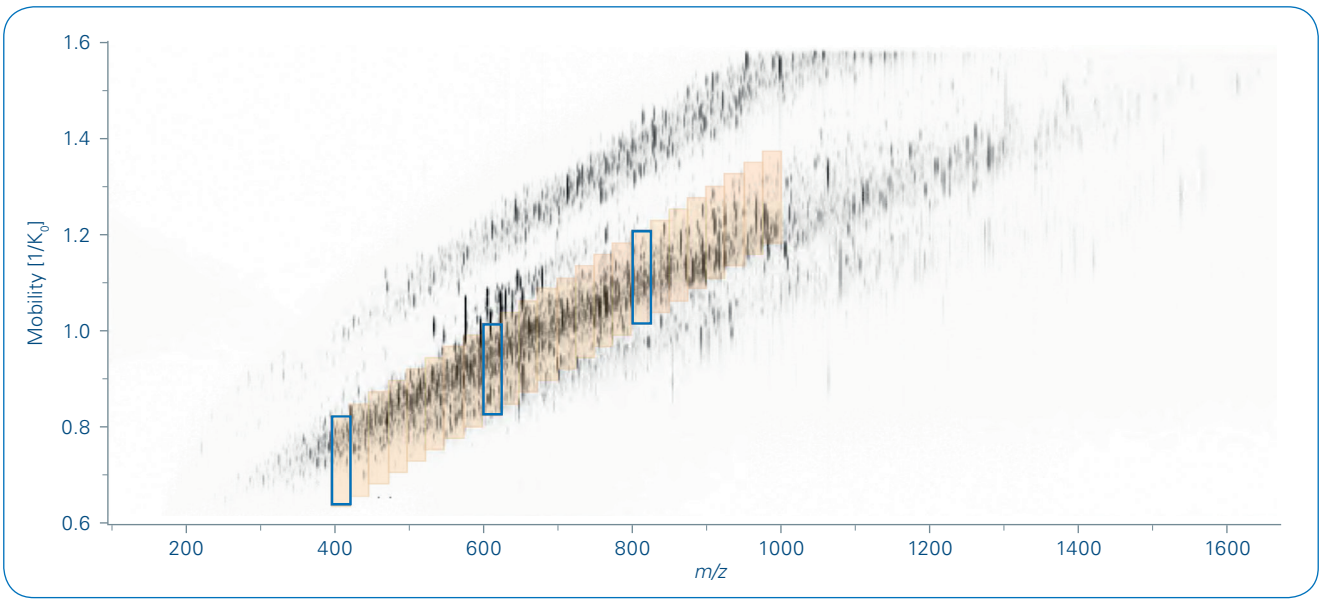

Figure 3: 4D-Proteomics method scheme for dia-PASEF. The applied method consists of three windows in each 100 ms dia-PASEF scan. Eight of these scans cover the diagonal scan line for doubly and triply charged peptides in the m/z *– ion mobility pane with narrow 25* m/z *precursor windows.*

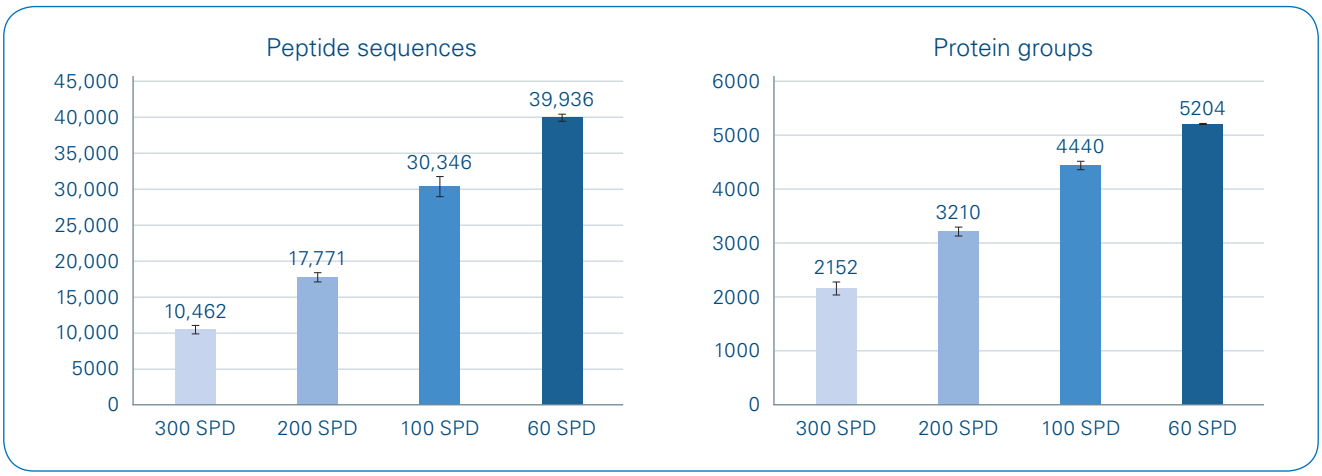

Figure 4: Average number of identified peptide sequences and protein groups (@ 1% peptide and protein FDR) from triplicate injections of 200 ng HeLa sample.

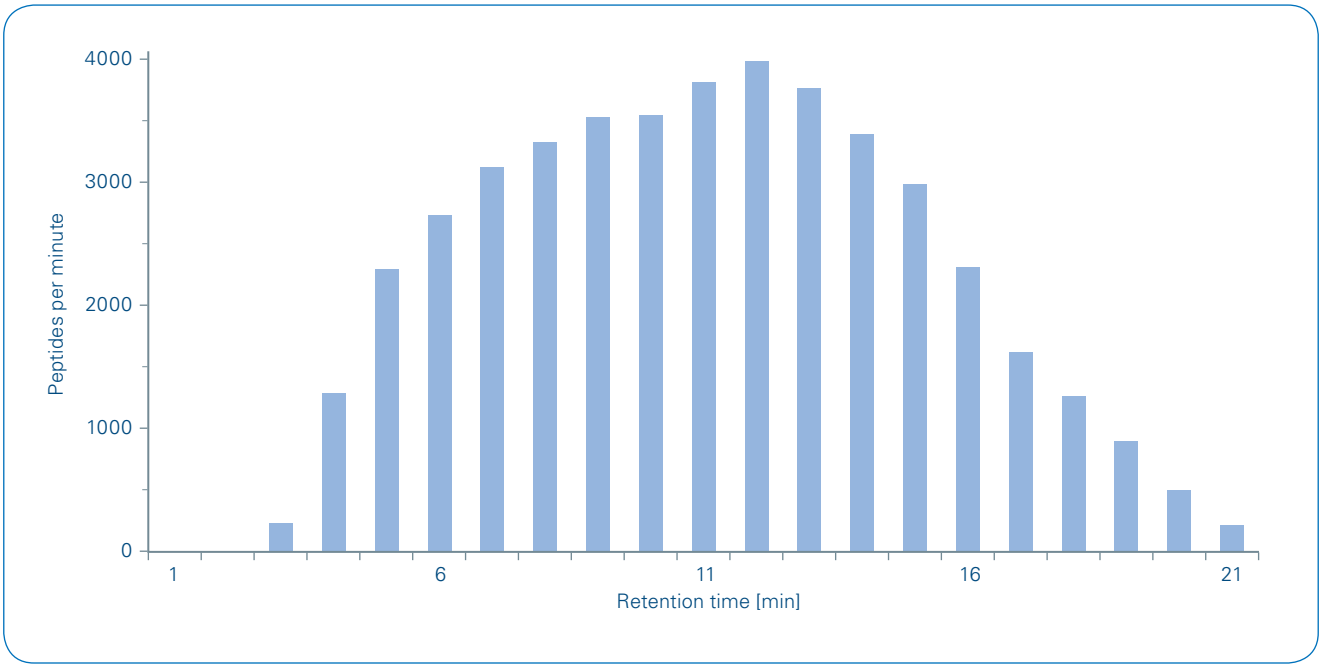

Figure 5: Numbers of peptides identified by 4-D data extraction across the 60 SPD method (21 min run time) for dia-PASEF using 200 ng sample load.

Data processing

The four-dimensional dia-PASEF data was processed using Spectronaut (version 14, Biognosys, Figure 1). The ion mobility enhanced libraries were generated using Spectronaut's Pulsar database search engine with 1% FDR control at PSM, peptide and protein level. We generated a resource-specific library using the 60 SPD LC-method on fractionated samples, which was directly used for targeted data extraction from the 60 SPD dia-PASEF data. For the shorter gradients (100, 200, 300 SPD) project-specific libraries were created by processing the dia-PASEF runs using the directDIA approach available in Spectronaut. By combining the resource- and project-specific libraries into hybrid libraries [7], retention-time precision can be preserved for the different gradients while still achieving completeness from the in-depth resource-specific library. For targeted analysis of dia-PASEF data we set a 1% FDR at the peptide and protein level.

Results and Discussion

dia-PASEF combines the PASEF method with DIA which allows multiplexing of DIA windows in a single 100 ms ion mobility separation frame. Here we applied the dia-PASEF method for the analysis of human cell lines using very short gradients on the Evosep One system (Figure 2).

We used an optimized dia-PASEF method with 3 windows in each 100 ms dia-PASEF scan. Eight of these scans (resulting in 24 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 3). Each dia-PASEF cycle contains a MS1 survey frame followed by 8 dia-PASEF frames (see Figure 3). This setup results in a total cycle time of 900 ms (1x 100 ms MS1 survey frame, 8x 100 ms dia-PASEF frames), which allowed us to obtain sufficient data points over the chromatographic peak while maintaining maximum coverage of +2 and +3 precursors.

Currently DIA workflows mainly rely on spectral libraries for correlating extracted fragment ion spectra with the available peptide spectrum. The resource-specific library data were acquired using DDA PASEF and processed in Spectronaut which resulted in a library of 8381 protein groups and 93,301 peptide sequences. Data acquisition time for library generation was just ~10 hours using the 60 SPD method. Targeted 4-dimensional extraction from the 60 SPD dia-PASEF data using this comprehensive resource-specific library resulted in the identification and quantification of on average 5204 protein groups and 39,936 peptide sequences at a 1% FDR (Figure 4). This translates into an identification rate of up to nearly 4000 peptides per minute of gradient time (Figure 5), underlying the exceptional sensitivity of the timsTOF Pro at high acquisition speed. The identified proteins cover a dynamic range of nearly 5 orders of magnitude. The median coefficient of variation was 8.2% at the peptide level and 5.9% at the protein group level for triplicate injections.

We further shortened the run time to increase sample throughput using the 100, 200, and 300 SPD methods on the Evosep One system.

When analyzing the HeLa sample using 4-D data extraction and the the highest throughput method (300 SPD), we identified more than 2100 protein groups and 10,462 peptide sequences on average in just 3-minutes gradient time (4.8-minutes total run time, Figure 4).

For large scale proteomic studies it is not only important to increase the proteome coverage in single runs, but also to reproducibly identify and quantify proteins and peptides to mitigate the missing value problem. DIA-based approaches benefit from sampling all the precursors present in the selected mass ranges rather than being dependent on precursor selection algorithms leading to very reproducible peptide and protein identification. The timsTOF Pro provides accurate and highly reproducible collisional cross section (CCS) values that adds a 4th dimension for feature matching and alignment during data processing, thereby further improving data completeness. When evaluating the triplicate injections of the 60 SPD method, we detected only a slight cumulative increase of 102 protein groups (Figure 6). More than 94% of the protein groups (4915 out of 5213) were identified in all three runs (Figure 7). This underlines the outstanding reproducibility of identified proteins using dia-PASEF.

In addition to excellent data completeness in dia-PASEF, quantitative accuracy strongly benefits from the 4th mobility dimension. Analyzing the reproducibility of protein intensities between the technical replicates revealed that 81% of the 5315 identified proteins could be reproducibly quantified with a CV below 20% and 65% with a CV below 10% (Figure 7).

In summary, these results highlight that the combination of the Evosep One with dia-PASEF and 4D-Proteomics is ideally suited for high-throughput proteomic profiling of hundreds of samples per day.

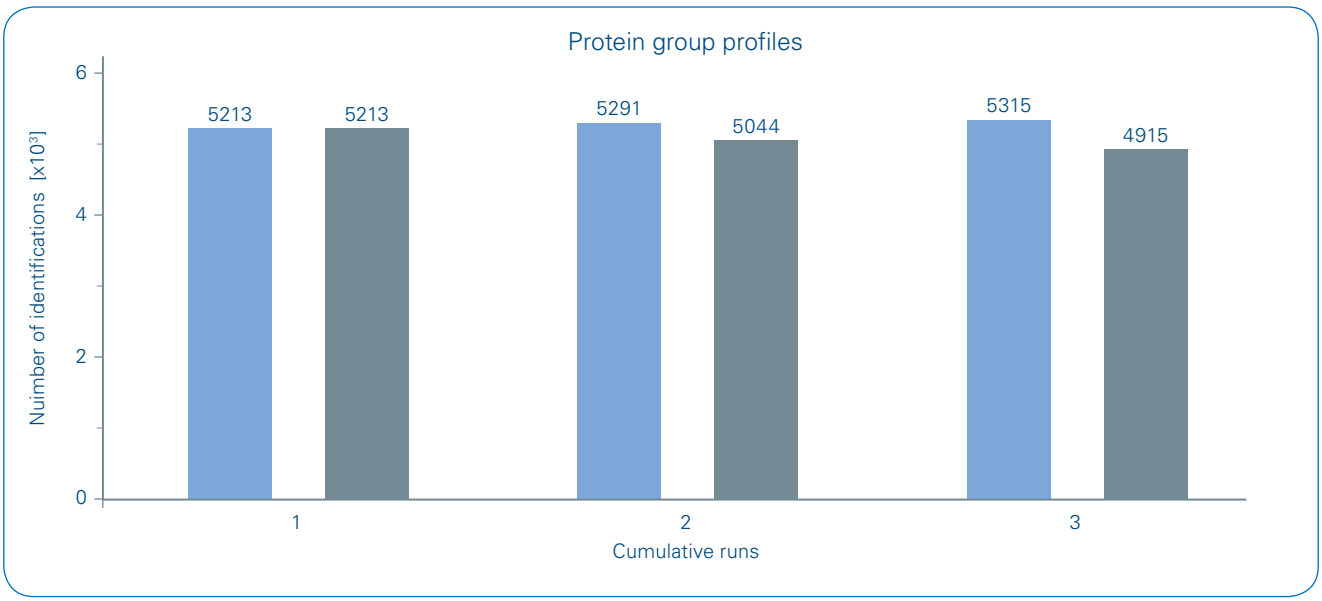

Figure 6: Overview cumulative protein groups profiles for the triplicate injections using the 60 SPD method. Sparse profiles are shown in blue and full profiles are shown in grey.

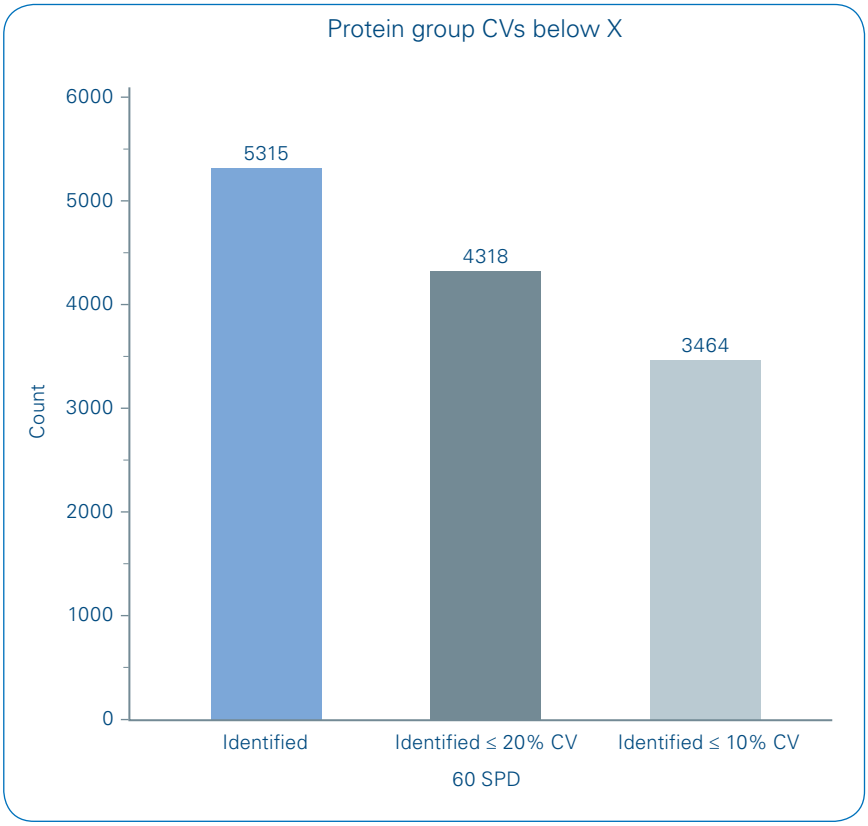

Figure 7: Number of all protein identifications (blue), proteins with CV < 20% (gray) and < 10% (light grey) for triplicate injections using the 60 SPD.

Conclusion

timsTOF Pro using dia-PASEF and 4D-Proteomics in combination with the Evosep One system is ideally suited for very high-throughput proteomics akin to clinical settings:

- On average 5200 protein groups can be identified and quantified in single-shot analysis using a short LC gradient (60 samples per day)
- Using a 300 SPD method (4.8-minutes total run time) dia-PASEF identifies and quantifies more than 2000 protein groups per run.
- dia-PASEF delivers extremely reproducible identification and quantitation information making the approach perfectly suited for the challenges of quantitative proteomics.

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 (2018). Mol Cell proteomics, doi: 10.1074/mcp.TIR118.000900

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

[3] Application Note LCMS 167: dia-PASEF®applied on different gradient lengths, https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/ApplicationNotes/1874392_LCMS-167_dia-PASEF_applied_on_different_gradient_lengths_ebook.pdf

[4] Bache N. et al. (2018). Mol Cell proteomcis, doi: 10.1074/mcp.TIR118.000853

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

[6] Kelstrup et al. (2018). J. Proteome Res., **17**, 1, 727-738

[7] Muntel et al. (2019). Mol. Omics, **15**, 348

For Research Use Only. Not for Use in Clinical Diagnostic Procedures.

Bruker Daltonik GmbH

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com – www.bruker.com