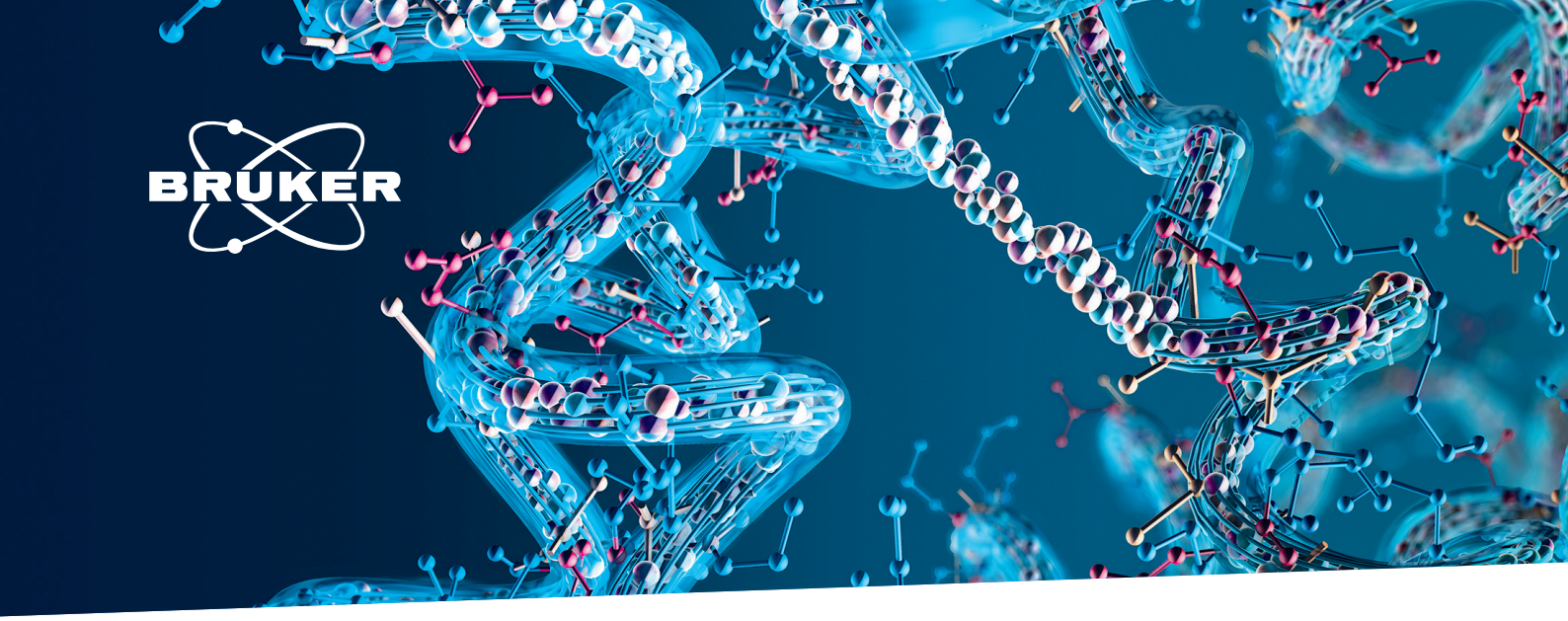

Rapid phosphoproteomics with dia-PASEF

Bruker's timsTOF technology including an additional dimension of separation in combination with dia-PASEF® enables highly sensitive PTM profiling.

Abstract

Post-translational modifications (PTMs) contribute to the regulation of protein activity and thereby partake in essentially all biological processes. Protein phosphorylation is one of the most studied PTMs due to its prominent role in cellular signaling. Mass spectrometry-based phosphoproteomics now routinely measures tens of thousands of phosphorylation sites on a proteome-wide scale. To disentangle their functional relevance, increasing the throughput and sensitivity of phosphoproteomics approaches is highly desirable. Here, we report the combination of dia-PASEF with short liquid chromatography gradients to analyze the phosphoproteome of a human cancer cell line. Injecting sample amounts equivalent to about 20 μg protein digest, we quantified >12,500 phosphopeptides (~9,200 class I phosphosites) in 21 min gradients without a spectral library. Increasing the achievable throughput with 7 min gradients, we still quantified about 80% of the class I sites, while maintaining a very high data completeness and quantitative accuracy in quadruplicates (median coefficient of variation <10%). Our results indicate that TIMS separation alleviates the higher spectral complexity resulting from the shorter gradients, even allowing the separation of some positional isomers of nearby phosphosites that were not resolved chromatographically. In conclusion, we demonstrate how the analytical merits of dia-PASEF facilitate rapid and sensitive phosphoproteomics.

Keywords: phosphoproteomics, short gradients, dia-PASEF, high-throughput proteomics

Introduction

Data-independent acquisition (DIA) has shown promise to increase quantitative reproducibility and accuracy in MS-based phosphoproteomics [1, 2, 3]. In dia-PASEF, DIA is combined with ion mobility separation which allows separation of signals from peptides that would otherwise be co-fragmentated [4]. It has already been described that phosphorylation can alter the gas phase structure of peptides [3]. In addition, the collisional cross section (CCS) allows for an unbiased alignment of precursor and fragment ion information adding confidence to the identification of phosphopeptides.

Here, we analyzed the HeLa cell line phosphoproteome with liquid chromatography gradients from 7- to 21-minutes to investigate whether the increased efficiency of dia-PASEF in combination with the TIMS ion mobility separation facilitates high-throughput phosphoproteomics.

Stephanie Kaspar-Schoenefeld[1], Gary Kruppa[1], Denys Oliinyk[2], Florian Meier[2]; [1]Bruker Daltonik GmbH & Co. KG, Bremen, Germany; [2]Functional Proteomics, Jena University Hospital, Jena, Germany

Methods

Human epithelial cervix carcinoma cell lysates were prepared following standard protocols and 200 μg proteome digest were used to enrich phosphorylated peptides according to the 'EasyPhos' protocol [1,6].

Peptides were separated within 7- and 21-minute acetonitrile (ACN) gradients at flow rates of 1.5 and 0.8 μL/min, respectively, on a 8 cm x 75 μm column with 1.9 μm ReproSil-Pur C18 - AQ particles (PepSep, Bruker Daltonik GmbH & Co KG) using a nanoElute LC. The LC system was connected via a CaptiveSpray nano-electrospray source to a timsTOF Pro (Bruker Daltonik GmbH & Co KG). The mass spectrometer was operated in dia-PASEF mode. The isolation scheme in *m/z* and ion mobility dimensions was adjusted to expected position and density of phosphopeptides with the py_diAID software ([4], https://github.com/MannLabs/pydiaid). The TIMS ion mobility range was set from $1/K_0$ = 1.43 to 0.6 Vs cm^{-2} with an ion accumulation and TIMS ramp time of 100 ms. The collision energy was lowered from 59 eV at $1/K_0$ = 1.4 Vs cm^{-2} to 20 eV at $1/K_0$ = 0.6 Vs cm^{-2}.

The dia-PASEF raw files were processed in Spectronaut using the default settings (v17, Biognosys) in library-free mode with the UniProt human reference proteome. Carbamidomethylation (C) was set as a fixed modification and Acetyl (Protein N-term), Oxidation (M) and Phospho (STY) as variable modifications. To analyze phosphorylation sites, we set a localization score threshold of 0 for all phosphopeptides and 0.75 for class I phosphopeptides. We set 'data filtering for quantification' to Q-value, normalization to Automatic and enabled Cross Run Normalization. To report unique phosphosites, we parsed the result files with the Peptide Collapse (v.1.4.4., [2]) plugin for Perseus.

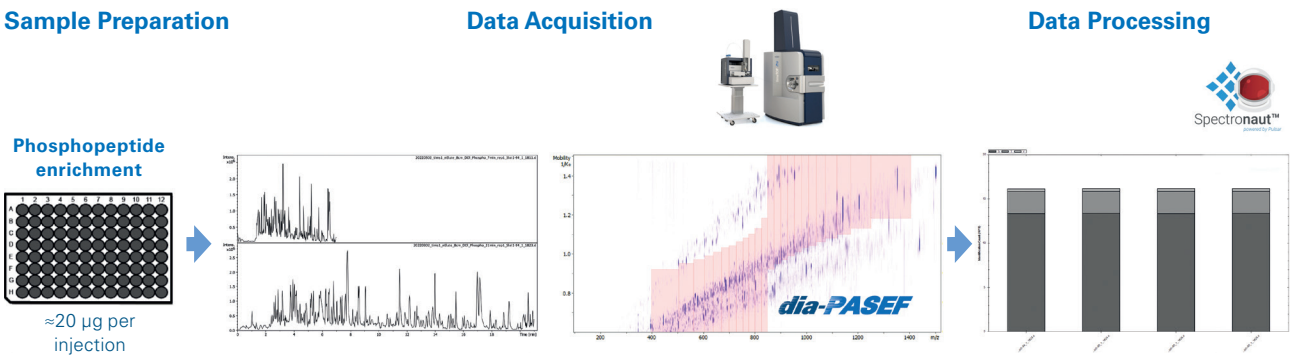

Figure 1
Phoshoproteomic workflow using dia-PASEF.
TiO_2 enriched phosphopeptides were separated on nanoLC system (7- and 21-minute gradient length) coupled to a timsTOF Pro using dia-PASEF. Data processing was done using Spectronaut (v17, Biognosys).

Results and Discussion

Enabling high-throughput phosphoproteomics experiments necessitates reducing the amount of input material and shortening the analysis time per sample. Here, we report results from improved LC and MS methods in combination with optimized data processing applied to low amounts of starting material for phosphoproteomics analysis.

The LC gradients were optimized so that the ion current is evenly distributed within the separation time. With a short 8 cm column it is possible to analyze up to 120 samples per day with nanoflow-sensitivity, including overhead times for sample handling and column preparation.

Using shorter gradients results in peptide signals being compressed into narrower elution peaks. While this is beneficial for increased signal intensities, it also requires faster MS methods for accurate peak quantitation. To achieve a short acquisition cycle of 1.4 s while optimally covering phosphopeptide precursor ions in the *m/z* vs. $1/K_0$ space, we used the py_diAID software to design a suitable isolation scheme.

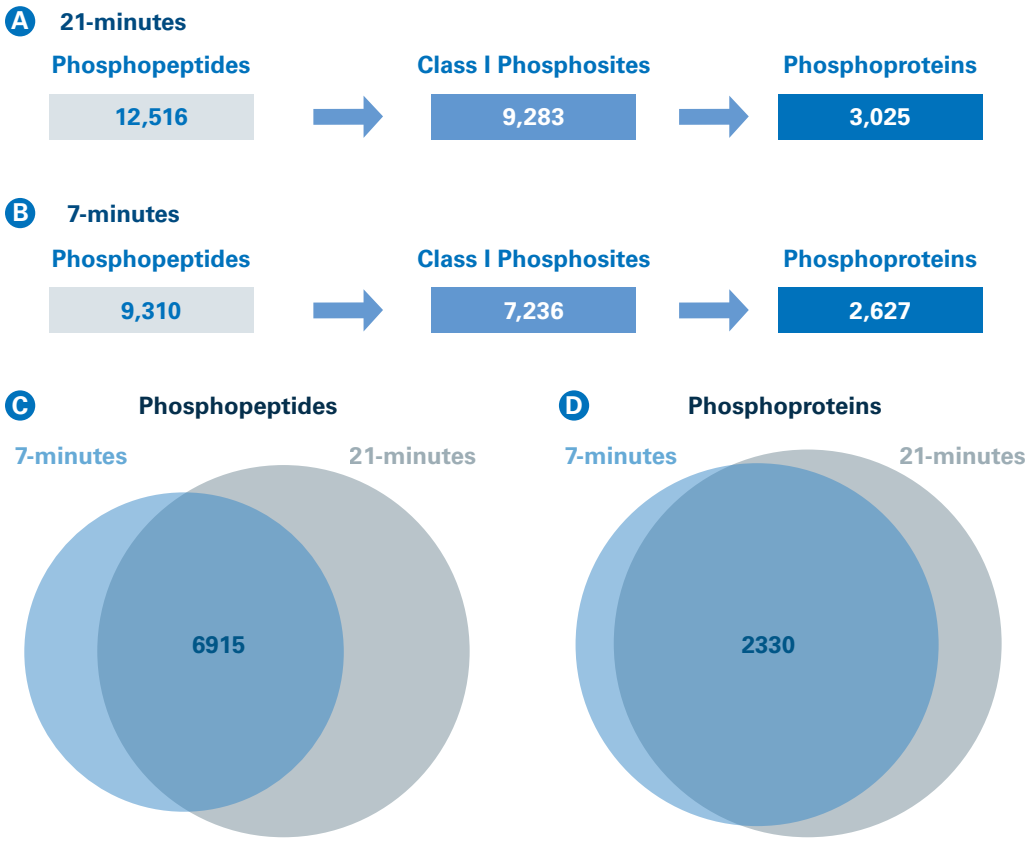

Figure 2

dia-PASEF phoshoproteomic with fast chromatographic separation.

(A) Total number of identified phosphopeptides and phosphorylated proteins in quadruplicate analysis with 21-minute LC gradients, injecting aliquots equivalent to 20 µg protein extract. (B) Same as A, but 7-minute LC gradients. (C) Overlap of identified phosphopeptides between 7- and 21-minute LC gradients. (D) Overlap of identified phosphoproteins between 7- and 21-minute LC gradients.

Additionally, high-throughput proteomics screening is not only dependent on the latest hardware but also on fast processing software. Traditionally, DIA approaches required the generation of spectral libraries from laborious measurements of fractionated samples based on DDA approaches, significantly increasing instrument and data processing time. Nowadays, library-free approaches are becoming a viable alternative. Here we applied directDIA, implemented in the Biognosys Spectronaut software, enabling analysis of DIA data without the need for experimental spectral libraries. In this approach precursor and fragment ions are correlated to generate pseudo-MS/MS spectra, which are then analyzed similar to classical DDA approaches. Using library-free data processing is especially advantageous for phosphoproteomics as it can detect rare or low-abundant phosphorylation events that might not be covered in an experimental library.

To assess the performance of dia-PASEF, we injected amounts of HeLa phosphopeptides equivalent to the enrichment from 20 µg starting material (the protein mass of ~100,000 cells) in short gradients of 21- and 7-minutes.

With directDIA, we identified in total ~12,500 phosphopeptides from quadruplicate injections with a 21-minute gradient. Out of these, ~9200 were localized by the Spectronaut software with a probability >0.75 to a specific site (class I sites) in one of 3025 protein sequences (Figure 2A). Remarkably, with the 7-minute gradient, we still identified ~9300 phosphopeptides (~7200 class I sites) and 2627 phosphoproteins (Figure 2B).

DIA approaches can result in a very high reproducibility as they are not relying on stochastic precursor selection. Reassuringly, our data also revealed a high overlap between the two gradients (Figure 2C, D) with about 78% of the protein group identifications in the 7-minute gradient being shared with the 21-minute gradient. Additionally, the proportion of identified serine, threonine and tyrosine phosphorylations were similar for the two gradients (Figure 3).

While proteome analysis typically aggregates multiple peptide data points per protein, in phosphoproteomics, in principle, each peptide represents an individual modification site. When assessing the reproducibility on the peptide level, the data matrices were virtually complete (>99%, Figure 4A, B) and almost all phosphopeptides were detected in four out of four replicates. Intensity values were also highly correlated between technical replicates for both 7-minutes and 21-minutes gradients (R >0.99, Figure 4C, D). We attribute this result to the combination of reproducible data acquisition with dia-PASEF and the highly specific library-free data analysis.

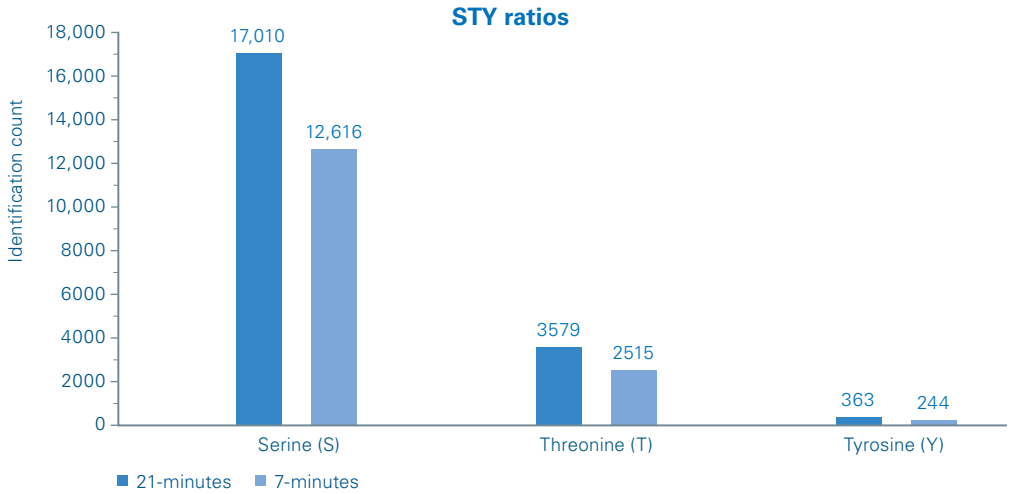

Figure 3
dia-PASEF phoshoproteomics with fast chromatographic separation.
Ratio of phosphopeptide identifications with modifications on serine, threonine, and tyrosine residues for the 7- and 21-minutes gradients.

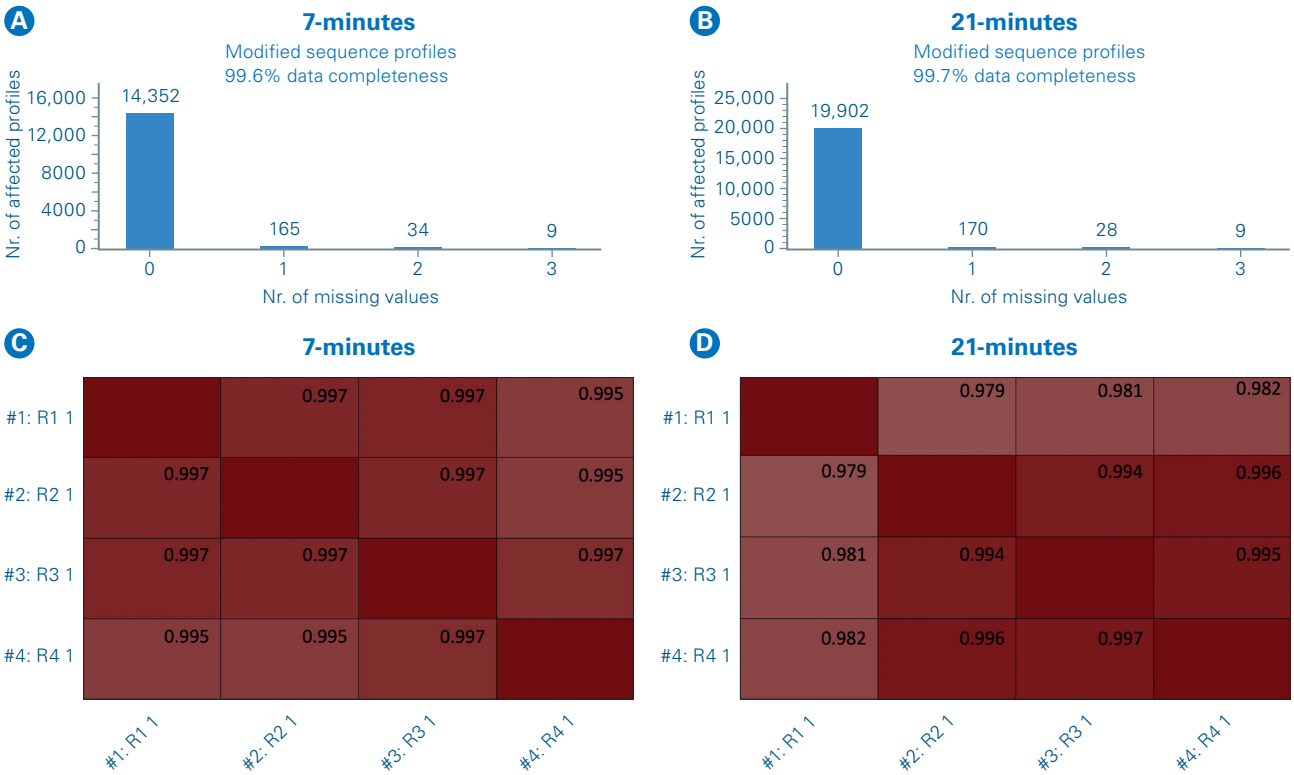

Figure 4
Technical reproducibility of dia-PASEF phoshoproteomics with fast chromatographic separation.
(A) Number of missing values as a measure of data completeness. Shown are the total numbers of identified phosphopeptides in quadruplicate analysis with 7-minute LC gradients. 99.6% of all modified peptides were detected in all 4 replicates. (B) Same as A, but 21-minute LC gradient with data completeness of 99.7%. (C) Sample correlation matrix for quadruplicate analysis with 7-minute LC gradients. (D) Same as C, but 21-minute gradient.

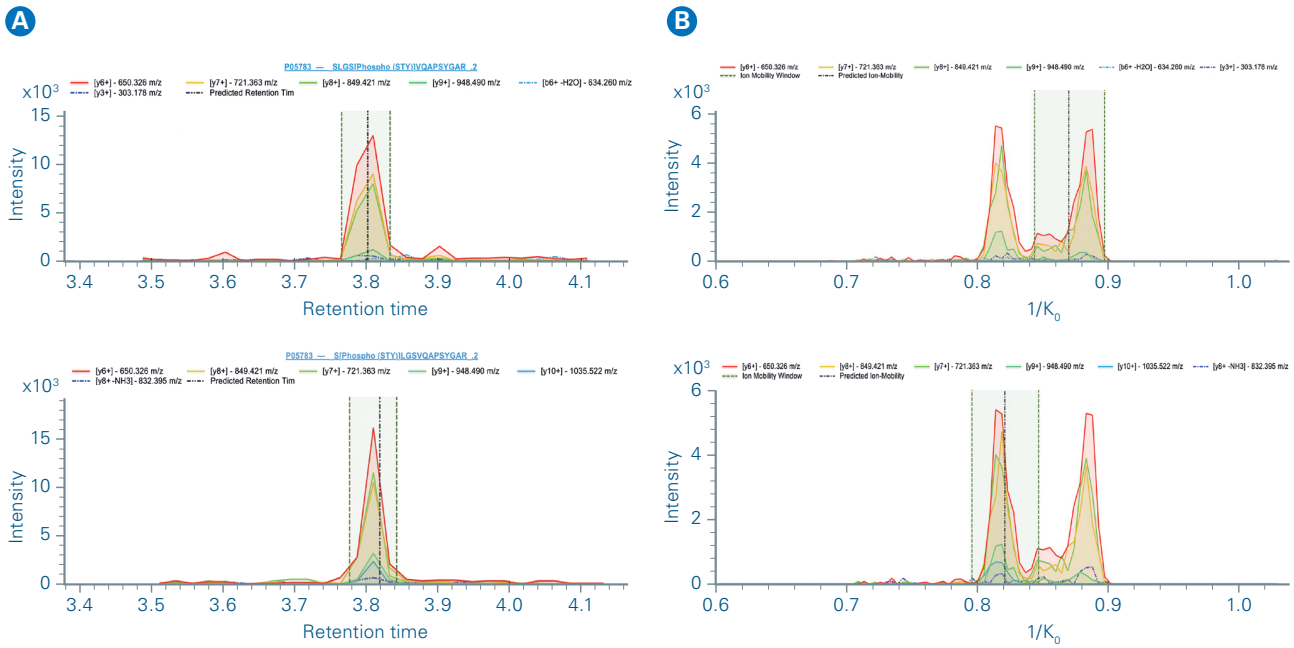

Figure 5

Two-dimensional separation of isobaric phosphopeptides.

Extracted ion chromatograms (A) and mobilograms (B) for an example isobaric positional isomers (SLGpSVQAPSYGAR (top) and pSLGSYQAPSYGAR (bottom)) not resolved by chromatography but ion mobility spectrometry.

Accurate identification and quantitation of phosphopeptides strongly depends on resolving peptides of similar mass and retention time. To this end, we investigated the analytical peak capacity in more detail. In the 21-minute gradient, the average chromatographic peak width was 8 seconds, equating a theoretical capacity of ~165 peaks. Conversely, the average peak width in the 7 min gradient was 5 s, resulting in a peak capacity of ~90. Mobility separation adds an additional dimension for separating isobaric phosphopeptides independent from chromatographic conditions. Oliinyk and Meier [1] have shown that gradually less peptides are resolved chromatographically from a 60-minute to a 7-minute gradient. Nonetheless, peptides are still effectively separated in the ion mobility dimension with comparable ion mobility peak widths independent of gradient length.

One advantage of ion mobility in phosphoproteomics is the possibility to separate isobaric positional isomers that have the exact same mass, similar elution behavior and share many fragment ions. We found 144 ion mobility-resolved positional isomers for the 7-minute gradient (applying a stringent localization score cut-off > 0.99), among them 24 pairs of isobaric phosphopeptides that were not resolved by chromatography. We extracted ion chromatograms for a representative pair of isobaric positional isomers (SLGpSVQAPSYGAR and pSLGSYQAPSYGAR, Figure 5) that was not separated by chromatography. However, as the two isomers differ in their size and shape in the gas phase as measured by their collisional cross section, ion mobility results in a clear separation.

Conclusion

dia-PASEF on the timsTOF platform is very well suited for high-throughput and high-sensitivity phosphoproteomics.

- More than 12,500 phosphopeptides, including ~9200 class I phosphosites could be reproducibly identified with short 21-minute gradients from low input material without the need for a project specific library.
- Even for very short 7-minute gradients more than 9300 phosphopeptides were detected when applying a dia-PASEF scheme that takes precursor density distribution into account for optimal window placement.
- Presented results show high degree of reproducibility and data completeness for short gradients using low sample input material, making dia-PASEF ideally suited for application to large sample cohorts.

References

[1] Yagnik G, Liu Z, Rothschild KJ, Lim JM, 2021. J Am Soc Mass Spectrom, **32**:977-988

[1] Oliink D, Meier F, 2022. PROTEOMICS, https://doi.org/10.1002/pmic.202200032.

[2] Bekker-Jensen DB, et al., 2020. Nat. Commun. **11**:787.

[3] Srinivasan A et al., 2022. J. Proteome Res. **21**:1789–1799.

[4] Meier, F. et al., 2020. Nat Methods **17**:1229–1236.

[5] Ruotolo et al., 2002. J. Proteome Res. **1**:303–306.

[6] Humphrey et al., 2015. Nat Biotechnlogy, **33**:990-995

[7] Skowronek et al., 2022. MCP, https://doi.org/10.1016/j.mcpro.100279

For Research Use Only. Not for use in clinical diagnostic procedures.

Bruker Switzerland AG

Fällanden · Switzerland
Phone +41 44 825 91 11

Bruker Scientific LLC

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

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

Learn more at **www.bruker.com/timstofpro**