

Mass Spectrometry & Spectroscopy

dia-PASEF: A Result of the Co-evolution of Mass Spectrometry and Proteomics

Gary Kruppa, PhD, Vice President of Proteomics at Bruker Daltonics, Billerica, MA, and Executive Director, Bruker S.R.O., Life Science Mass Spectrometry, Brno, Czech Rep; gary.kruppa@bruker.com

Mass spectrometry (MS) has long been the technology of choice for a range of proteomics applications, and is a powerful tool for protein scientists, biologists, and clinical researchers. Ongoing developments have improved the capabilities of MS instruments over the years to uncover previously unexplored parts of the proteome, ultimately aiding in the discovery of new drugs, and pushing the boundaries of personalised medicine approaches. Despite such developments, coverage of the proteome remains challenging. Detection and quantitation of proteins direct from tissue or biofluids is difficult due to the large expected range of concentrations, and protein expression varies depending on both genetic and environmental factors. Moreover, because of the combined effects of alternative splicing, point mutation, post-translational modifications (PTMs) and endogenous proteolysis, a given protein (gene expression product) can be expressed as many different proteoforms, with each potentially having a dedicated biological activity.

Given these challenges, one area of innovation in MS-based proteomics methods has focused on combining accurate, sensitive, high-throughput protein identification and quantification, using bottom-up MS with orthogonal approaches. Ion mobility spectrometry (IMS) is a powerful analytical technique that has been widely applied over the last five decades, primarily in chemical physics and analytical chemistry applications. Only relatively recently has the potential of IMS coupled to MS (IMS-MS) been explored for the separation, identification and quantification of peptides and proteins. IMS is a gas phase ion separation technique utilising differences in the mobilities of ions through a gas under the influence of an electric field, with significant potential in proteomics applications to increase peak capacity, dynamic range, and improve signal-to-noise (S/N) ratio.

There are several types of IMS that extend the capabilities of MS alone, such as trapped ion mobility spectrometry (TIMS), which separates ions trapped in a flowing gas stream based on a non-uniform electric field that increases towards the exit of the device. TIMS are relatively small devices that can be easily integrated into a mass spectrometer without a noticeable loss in ion transmission and even an increase in sensitivity [1], increasing confidence in compound characterisation.

Expanding the possibilities of proteomics

In contrast with top-down proteomics, which analyses intact proteins by fragmentation of the intact protein, shotgun (or 'bottom-up') proteomics analyses peptides created by proteolytic digestion of a protein mixture. Shotgun proteomics approaches rapidly generate a profile of many proteins within a complex mixture. Protein mixtures are first digested by protease, and the resulting peptides are separated in high performance liquid chromatography (HPLC), followed by tandem MS/MS analysis to generate fragmentation information (Figure 1A) that is related to the sequence of the peptide. The fragmentation information for each peptide is compared with a protein database, to search for the proteolytic peptides that would match the fragmentation and thus identify the protein from which they were released by proteolysis. Although recent advances in proteomic technologies have made MS-based proteomics a central research tool, comprehensive proteomes coverage remains elusive, mainly due to the high complexity and dynamic range of biological samples with their array of PTMs. More than 10,000 proteins are typically present in each biological sample at one specific condition and after digestion, the complexity of analytes increases by at least two orders of magnitude because each protein generates tens to hundreds of peptides [2].

The last ten years of MS-based proteomics developments have endeavoured to increase the depth and breadth of proteome coverage, improve data quality and identification confidence, and increase sample throughput necessary to enable population-scale proteome measurements. Despite a variety of innovations that bring this goal closer, throughput and detection sensitivity still limit large sample cohort studies. Recent MS technology developments have therefore focused on increasing fragmentation speed without losing sensitivity.

By incorporating a TIMS device at the front of a quadrupole time-of-flight (Q-TOF) mass spectrometer, ions can be accumulated for a specific amount of time before being released for MS analysis (Figures 1B and 1C). Additionally, the ability to separate ions by mobility boosts sensitivity and provides additional selectivity, because the ions are separated by a fourth parameter - their collisional cross section (CCS). Compact ions with a small CCS drift faster than extended ions with a large CCS, effectively allowing ions to be separated by shape [3], as well as retention time, mass-to-charge (m/z) ratio, and intensity.

TIMS, also enables the parallel accumulation - serial fragmentation (PASEF) method, adding an additional dimension of separation, superior speed, and improved sensitivity to proteomics workflows. The PASEF acquisition method - based on using the position of the quadrupole for selecting peptides by their m/z , and moving it in sync with the mobility elution profile - exploits trapped ion mobility to achieve up to a 10-fold increase in sequencing speed. This increase in sequencing speed is crucial for delving deeper into complex proteomes to obtain quantitative data in a short amount of time. The combined power of TIMS and PASEF allows for greater proteome or sub-proteome coverage from small sample volumes taken from complex mixtures.

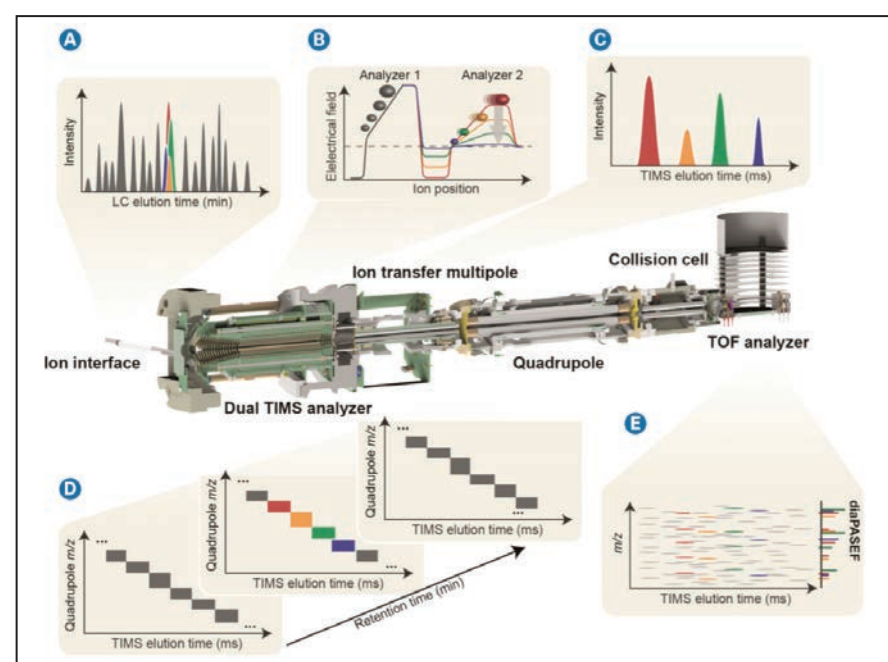


Figure 1. Overview dia-PASEF workflow. (A) Peptides eluting from the chromatographic column are ionised and enter the mass spectrometer through a glass capillary. (B) In the dual TIMS analyser, the first TIMS section traps and stores ion packets, and the second resolves them by ion mobility. (C) Ion mobility separated ions are released sequentially from the second TIMS analyser as function of decreasing electrical field strength. (D), (E) For dia-PASEF, DIA isolation windows are coupled to the precise ion mobility elution of the corresponding ions. Within a single TIMS separation multiple precursor windows are set. Figure reproduced from [6].

Increasing throughput with dia-PASEF

CCS, as a fourth dimension of separation, significantly increases selectivity, regardless of the acquisition strategy. Data independent acquisition (DIA) workflows have gained in popularity in recent years as they overcome the issue of stochastic selection of peptide precursors encountered in typical data dependent acquisition (DDA) approaches, and thereby promise reproducible and accurate protein identification and quantification across large sample cohorts.

DIA produces a quantitative answer for a specific analyte in every sample, whereas DDA workflows suffer from the need to select individual precursors. Due to variations from sample to sample, the same precursors may not be selected in each DDA run, meaning not all the identified peptides in one run will be comparable to other runs. This is known as the 'data completeness problem', and DIA methods were developed to solve this.

In DIA methods, all ions in a given m/z window are fragmented in every run, in principle allowing full identification and comparison from run-to-run. The PASEF principle has now been combined with DIA in a new acquisition method known as dia-PASEF, which sets new selectivity and sensitivity standards for DIA and achieves unprecedented peptide identification rates with reproducible quantification [7]. This DIA approach using PASEF takes advantage of the correlation between ion mobility and m/z , and uses windows that start at high m/z as larger species elute at the beginning of the TIMS scan, moving down to lower m/z windows (Figure 1D and 1E). dia-PASEF window schemes are at least 4x more efficient for ion utilisation than DIA with other acquisition methods and, depending on the sample and window scheme, can achieve up to 100% efficiency with unparalleled sensitivity.

Previous experiments have shown the identification of more than 7000 proteins in single 2-hour runs from human cell lysates [4], and the reproducible and accurate quantification of more than 8000 proteins from a complex mixture of three proteomes [5]. To further investigate the benefits of applying dia-PASEF to different proteomics samples of varying complexity, the method was applied to the analysis of a human cancer cell line (HeLa) and yeast digest, using different gradient lengths (30, 60, and 90 min gradients). 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.

A dia-PASEF method with 2 windows in each 100 ms dia-PASEF scan was used. Sixteen of these scans (resulting in 32 windows) covered the diagonal scan line in the m/z - ion mobility plane 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 scan), which enabled sufficient data points to be obtained over the chromatographic peak while maintaining the best combination of specificity and sensitivity.

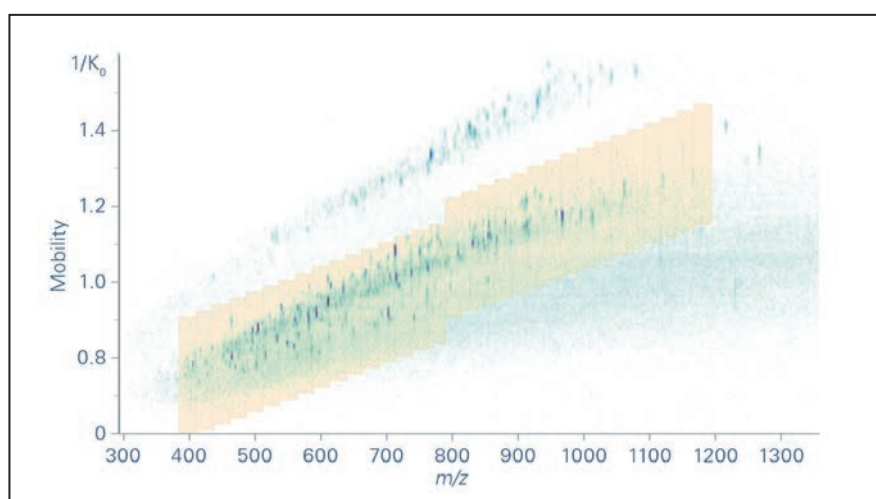


Figure 2. Method scheme for dia-PASEF. The applied method consists of two windows in each 100 ms dia-PASEF scans. Sixteen of these scans cover the diagonal scan line for doubly and triply charged peptides in the m/z - ion mobility plane with narrow 25 m/z precursor windows.

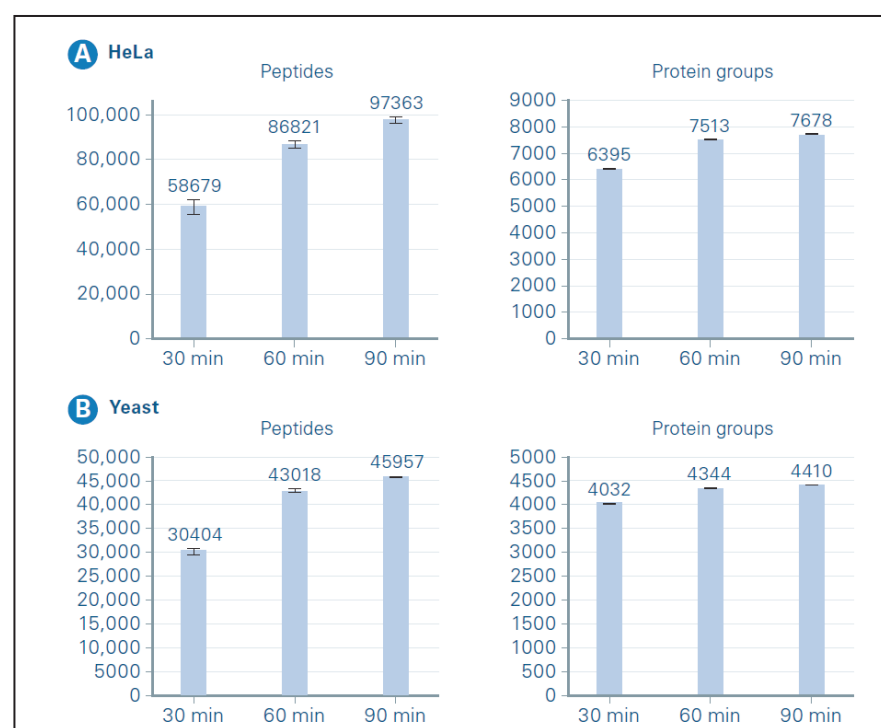


Figure 3. Number of identified peptides and proteins at a false discovery rate (FDR) of 1% for the different gradient lengths for (A) HeLa and (B) Yeast samples. Shown are the average values from the three runs together with the standard deviation.

The four-dimensional (4D) data was processed using a developmental version of Spectronaut (Biognosys AG, Switzerland). 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. A project-specific library was created from high-pH fractionated samples. The library was comprised of 220,628 unique target peptide sequences and 11,578 target protein groups for HeLa samples, and 80,874 unique target peptides and 5127 target protein groups for yeast. These libraries were used 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.

Figure 3 shows the number of peptide sequence and protein group identifications for 90, 60, and 30 minute gradients, which highlights the advantage of dia-PASEF for deeper proteome coverage and precise quantification even for shorter gradients down to 30 minutes. In total, 66% and 86% coverage were achieved for HeLa (Figure 3A) and yeast (Figure 3B), respectively, of all protein groups in the comprehensive library in the 90 min gradient single runs without fractionation. Using 30 min gradients still achieved 55% and 79% coverage for HeLa (Figure 4A) and yeast (Figure 4B), respectively.

The results of this study show that excellent identification rates can be reproducibly achieved for complex HeLa digest, with 7600 protein groups identified at 90 min gradient time and almost 6400 using a 30 min gradient. These identification rates, combined with the robust and reproducible quantification provided by dia-PASEF, brings the field of proteomics even closer to achieving good coverage of the proteome with reasonable throughput.

Monitoring drivers of disease

A key challenge in moving proteomics into the clinical space is the analysis and interpretation of data resulting from high-throughput shotgun proteomics workflows. Primary data analysis (protein identification and quantitation) can be solved with powerful computing technology, but automatic interpretation of proteomics data to establish 'disease vs. healthy' is challenging. Such capabilities require advanced artificial intelligence (AI) and deep learning algorithms, which is the focus of current research efforts to achieve true diagnostic proteomics.

For example, researchers at the Röst Lab at The University of Toronto are using dia-PASEF to monitor the human proteome throughout a person's lifetime, and looking at how machine learning could be used to analyse the data.

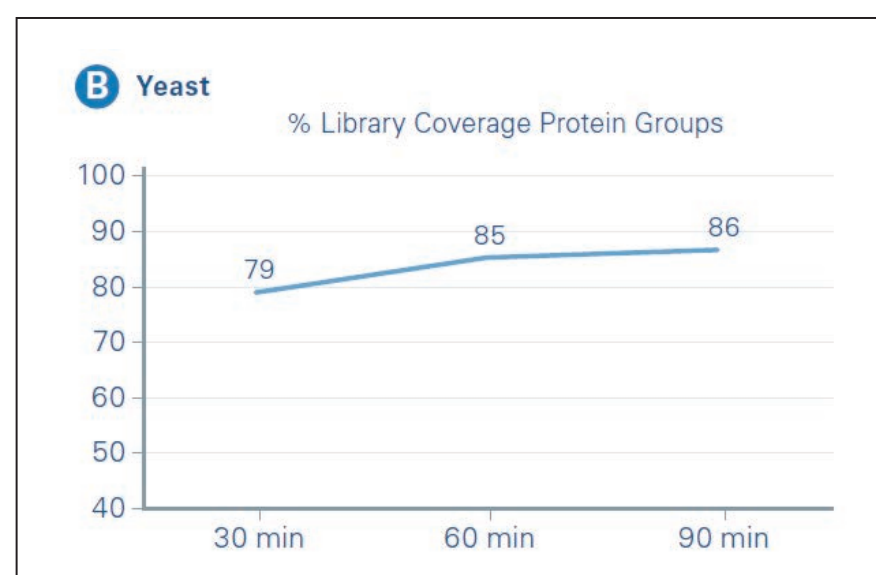
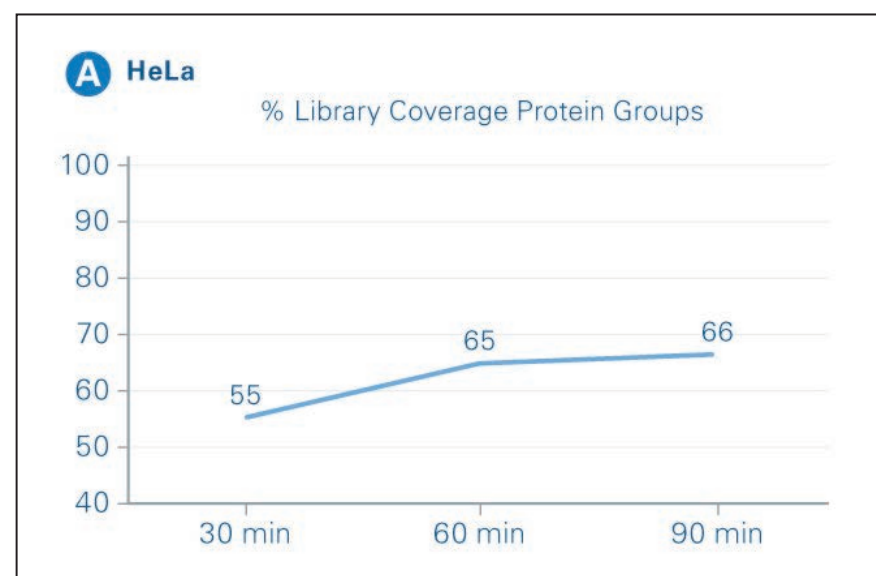


Figure 4. Library coverage for (A) HeLa and (B) Yeast samples on protein group level. Shown are the results for three different gradient lengths (30, 60, 90 min).

By analysing large patient cohorts in this way, deviations from a healthy state can be monitored and the analyses associated with that deviation can be quantified. This approach aims to improve understanding of how biological systems work, how they react to these deviations, and how signals are processed. By applying this process to individual patients and patient cohorts, the molecular profile of a single patient can be traced over long periods of time. This can be used to understand how their molecular profile changes over time, in response to environmental and lifestyle changes, ultimately enabling the observation of transitioning from a healthy profile to a disease profile, and making a diagnosis in the early stages of disease.

Studying patient cohorts longitudinally is important for comparing an individual's molecular profile over time, rather than a population average, to progress towards a precision medicine approach. Machine learning has been introduced to proteomics and metabolomics studies relatively recently, to alleviate the more time-consuming steps involved in DIA, such as the generation of the spectral library.

Early experiments from the Röst Lab show that it could be possible to predict these spectral libraries, which would allow the group to mine DIA data more efficiently and directly, without relying on prior experimental evidence and measurements that are usually completed using DDA.

Continuing the 4D proteomics revolution

Due to the 'missing value problem' facing traditional DDA approaches, proteomics researchers have turned to DIA as an alternative approach that promises reproducible and accurate protein quantification across large sample cohorts by eliminating the selection of individual peptide ions for identification. Combining this with the inherent efficiency of PASEF, to create the new acquisition mode dia-PASEF, moves the concept of identifying thousands of proteins from minimal sample volumes into reality. Knowing that observed differences are due to biological variation, rather than variation between peptides analysed in one sample versus another, will allow clinical researchers to accurately monitor the progression of disease. The recent application of deep learning to proteomics is now allowing

researchers to mine dia-PASEF data more efficiently, making the most of the method's boosted throughput, efficient ion utilisation, rapid peptide identification and quantification, and unparalleled sensitivity. These advanced technologies are set to continue the evolution of MS-based proteomics.

For more information on dia-PASEF, please visit:

<https://www.bruker.com/products/mass-spectrometry-and-separations/lc-ms/o-tof/timstof-pro.html>.

References

1. Wormwood KL, Deng L, Hamid AM, DeBord D and Maxon L (2019) *The Potential for Ion Mobility in Pharmaceutical and Clinical Analyses*. In: Woods A., Darie C. (eds) *Advancements of Mass Spectrometry in Biomedical Research. Advances in Experimental Medicine and Biology*, vol 1140. Springer, Cham.
2. Beck S (2016) *Novel quadrupole time-of-flight mass spectrometry for shotgun proteomics*, PhD Thesis, Faculty of Chemistry and Pharmacy, Ludwig-Maximilians-University Munich.
3. Angel TE, Aryal UK, Hengel SM, Baker ES, Kelly RT, Robinson EW and Smith RD (2012) *Mass spectrometry-based proteomics: existing capabilities and future directions*. *Chem Soc Rev*, 41(10):3912–3928.
4. Kaspar-Schoenefeld S, Meier F, Brunner AD, Frank M, Ha A, Lubeck M, Raether O, Collins B, Aebersold R, Rost H and Mann M, "Parallel accumulation - serial fragmentation combined with data-independent acquisition (diaPASEF)." *AppNote LC-MS 157*, https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/ApplicationNotes/1869385_LCMS-157_diaPASEF_ebook.pdf
5. Kaspar-Schoenefeld S, Marx K, Gandhi T, Reiter L, Meier F, Brunner AD, Frank M, Ha A, Lubeck M, Raether O, Distler U, Tenzer S, Aebersold R, Collins B, Rost H and Mann M. "diaPASEF: label-free quantification of highly complex proteomes." *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
6. Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA, Bache N, Hoerning O, Cox J, Räther O and Mann M. *Online parallel accumulation – serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer*. *Mol Cell Proteomics*. 2018; mcp.TIR118.000900. © the American Society for Biochemistry and Molecular Biology.
7. Meier F, Brunner AD, Frank M, Ha A, Bludau I, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Bache N, Aebersold R, Collins BC, Röst HL, Mann M. *diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition*. *Nat Methods*. 2020 Dec;17(12):1229-1236. doi: 10.1038/s41592-020-00998-0.



Read, Share and Comment on this Article, visit: www.labmate-online.com/article