BRUKER

diaPASEF: label-free quantification of highly complex proteomes

The timsTOF Pro with diaPASEF provides reproducible and accurate qualitative and quantitative results in complex proteomics samples – making it perfectly suited for data-independent acquisition approaches.

Abstract

Data-independent acquisition (DIA) promises reproducible and accurate protein quantification across large sample cohorts. 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 diaPASEF [1]. In this application note, we apply diaPASEF for label-free quantification of a set of samples consisting of a mixture of three proteomes (digests of HeLa, Yeast, *E. coli*) each to investigate its potential for the in-depth analysis of complex proteomics samples. In brief, data processing with Spectronaut (Biognosys) revealed more than 8,000 proteins that could be reliably identified and quantified at a false discovery rate of 1% in the three proteome samples. Fragment ion-based quantification was highly reproducible,

Authors: Stephanie Kaspar-Schoenefeld[1], Kristina Marx[1], Tejas Gandhi[2], Lukas Reiter[2], Florian Meier[3], Andreas-David Brunner[3], Max Frank[4], Annie Ha[4], Markus Lubeck[1], Oliver Raether[1], Ute Distler[5], Stefan Tenzer[5], Ruedi Aebersold[6,7], Ben Collins[6], Hannes Rost[4], Matthias Mann[3,8];
[1]Bruker Daltonik GmbH, Bremen, Germany; [2]Biognosys, Schlieren, Switzerland; [3]Max Planck Institute of Biochemistry, Martinsried, Germany; [4]Donnelly Centre for Cellular and Biomol. Research, Toronto, Canada; [5]University of Mainz, Mainz, Germany; [6]ETH Zurich, Zurich, Switzerland; [7]University of Zurich, Zurich, Switzerland; [8]NNF Center for Protein Research, Copenhagen, Denmark.
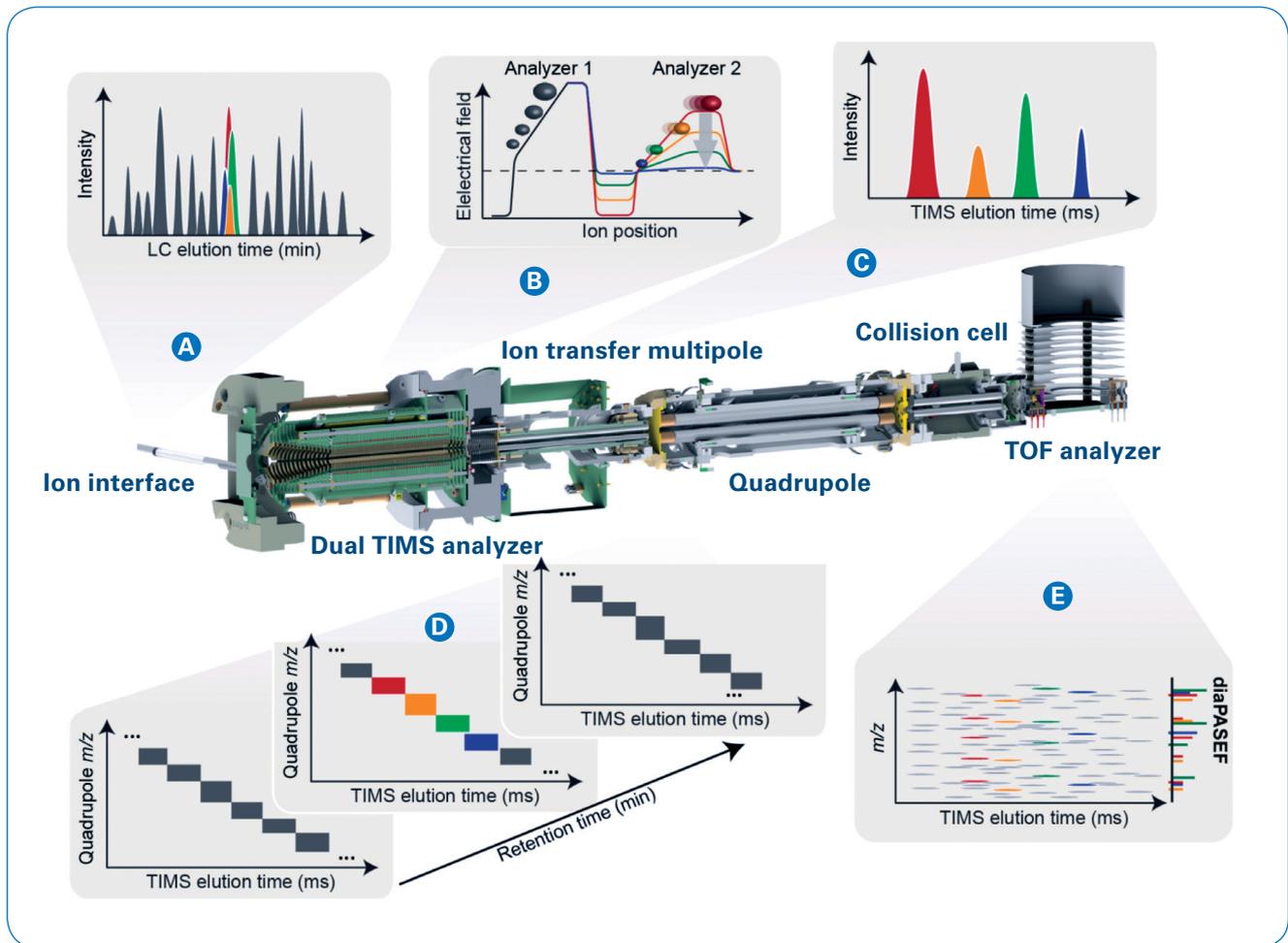
Figure 1: Overview diaPASEF workflow. (A) Peptides eluting from the chromatographic column are ionized and enter the mass spectrometer through a glass capillary. (B) In the dual TIMS analyzer, the first TIMS section traps and stores ion packets, and the second resolves them by ion mobility. (C) Ion mobility separated ions are released sequentially from the second TIMS analyzer as function of decreasing electrical field strength. ((D), (E)) For diaPASEF 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.

with a median coefficient of variation (CV) below 10%, and accurate with experimental ratios close to the theoretical ones.

Introduction

Mass spectrometry (MS)-based quantitative proteomics has become a powerful technology to elucidate the dynamics of proteomes, and a wide range of approaches have been developed to answer the panel of challenges set by biological sample analysis. Among those, Data-dependent acquisition (DDA) -based approaches allow for the best

sensitivity/selectivity combination, as the MS system physically isolates a precursor ion before fragmenting it to generate highly specific fragment ion spectra. However, while analyzing sample cohorts, and as a result of the high complexity of a majority of 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" (MBR) approaches [2].

The timsTOF Pro in combination with PASEF (Parallel Accumulation Serial Fragmentation) is setting new

standards of uncompromised speed, sensitivity and resolution for shotgun proteomics. Especially, the > 100Hz MS/MS speed and the accurate measurement of Collisional Cross Sections (CCS), adding an additional 4th dimension to perform MBR, partially compensate for the stochastic effect in DDA approaches.

Data-independent acquisition (DIA) approaches have been developed to overcome the "missing value problem". They promise reproducible and accurate protein quantification across large sample cohorts by using wide selection windows, rather than

selecting individual peptides in DDA, to insure that all precursor ions are fragmented in every sample. The mass spectrometer typically cycles through many isolation windows covering the *m/z* range of interest.

As the high reproducibility of the CCS values of the timsTOF Pro makes library-based approaches very attractive, we have worked with groups at the ETH Institute for Molecular and Systems Biology (IMSB) in Zurich, the Max Planck Institute in Martinsried, and the University of Toronto to transfer the PASEF principle to DIA resulting in a new acquisition method, called diaPASEF [1]. We previously showed the identification of more than 7,000 proteins in a single, 2 h gradient from a 200 ng human cell lysate peptide sample measured with diaPASEF. Remarkably, more than 3,000 proteins

were identified from only 10 ng peptide sample loaded on column [3]. Here we explore the potential of diaPASEF for label-free quantitative proteomics by analyzing a set of complex three proteome samples, which were mixed in well-defined ratios.

Methods

The samples containing tryptic digests of HeLa, Yeast and *Escherichia coli* (*E. coli*) were prepared according to [4]. To generate the hybrid proteome sample, tryptic peptides were combined in the following ratios: Sample HYE-A was composed of 65% w/w human, 30% w/w yeast, and 5% *E. coli*. Sample HYE-B was composed of 65% w/w human, 15% w/w yeast, and 20% *E. coli* proteins.

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 C18 column (25 cm x 75 μm i.d., 1.6 μm, IonOpticks, Australia) using a 100-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 at 50°C. For diaPASEF, 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 diaPASEF 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.
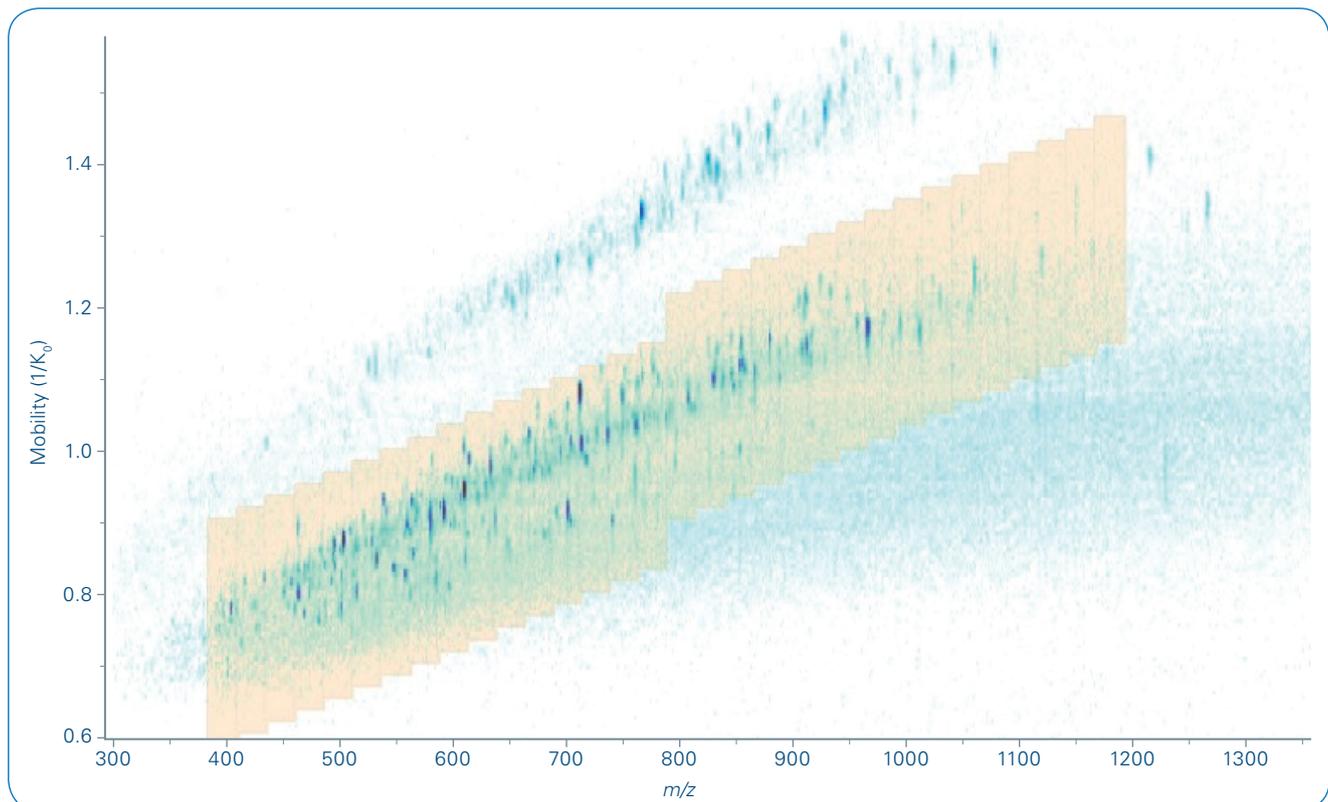

Figure 2: Method scheme for diaPASEF. The applied method consists of two windows in each 100 ms diaPASEF scan. Sixteen of these scans cover the diagonal scan line for doubly and triply charged peptides in the m/z – ion mobility plane with narrow 25 m/z precursor windows.

For targeted data extraction, we used a project-specific library generated from 5 high-pH reverse-phase peptide fractions per organism (HeLa, *E. coli*, Yeast) acquired with PASEF technology. MS data were collected over an m/z range of 100 to 1,700. During each MS/MS data collection, each TIMS cycle was 1.1 seconds and included 1 MS and 10 PASEF MS/MS scans, with an average of 12 precursors selected for each PASEF MS/MS scan, resulting in an MS/MS data acquisition rate of 109 Hz.

Raw files were processed using a new version of Spectronaut 13 (v13.6) which adds beta support for processing such data. The ion mobility enhanced library was generated using Spectronaut's Pulsar database search engine with 1% false discovery rate (FDR) control at PSM, peptide and protein level. For the subsequent targeted analysis of diaPASEF data, Spectronaut used 1% FDR at peptide and protein level.

Results and Discussion

We investigated the potential of diaPASEF for label-free quantification of complex proteomics mixtures. diaPASEF combines the well-known PASEF method [3,5], which exploits trapped ion mobility spectrometry to achieve high duty cycle, efficient ion usage and improved peptide identification rates and reproducibility of identification across multiple samples with DIA (Figure 1). By coupling DIA isolation windows to the precise ion mobility elution of the corresponding ions, diaPASEF allows multiplexing of DIA windows in a single 100 ms precursor ion mobility separation.

To evaluate the quantitative capabilities of diaPASEF, we compared two samples both consisting of tryptic digests of human, yeast
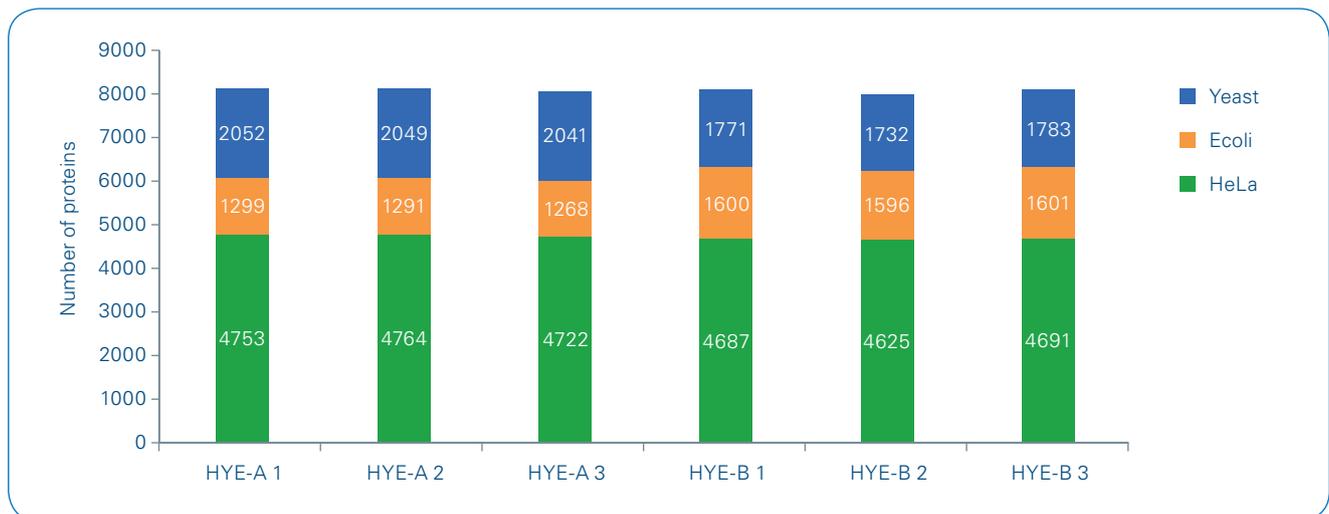

Figure 3: Number of identified proteins at an FDR of 1% for the two samples (HYE-A, HYE-B) identified with at least 2 peptides per protein.
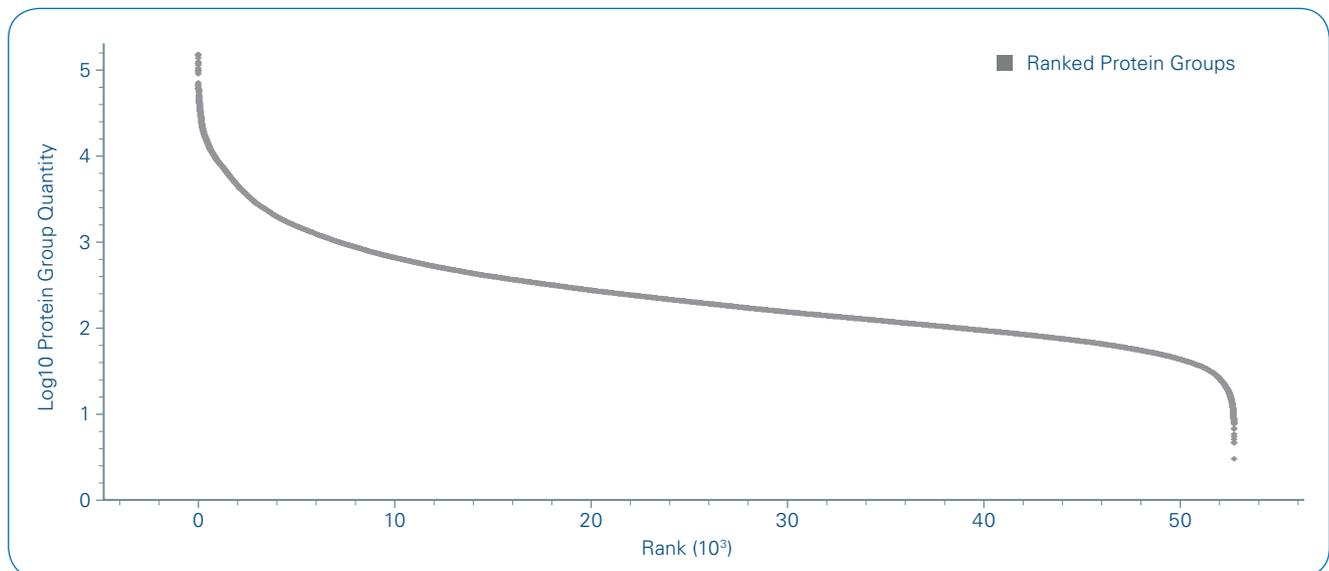

Figure 4: Dynamic range of identified proteins covering about 5 orders of magnitude. Proteins are ranked by their log10 protein group quantity.
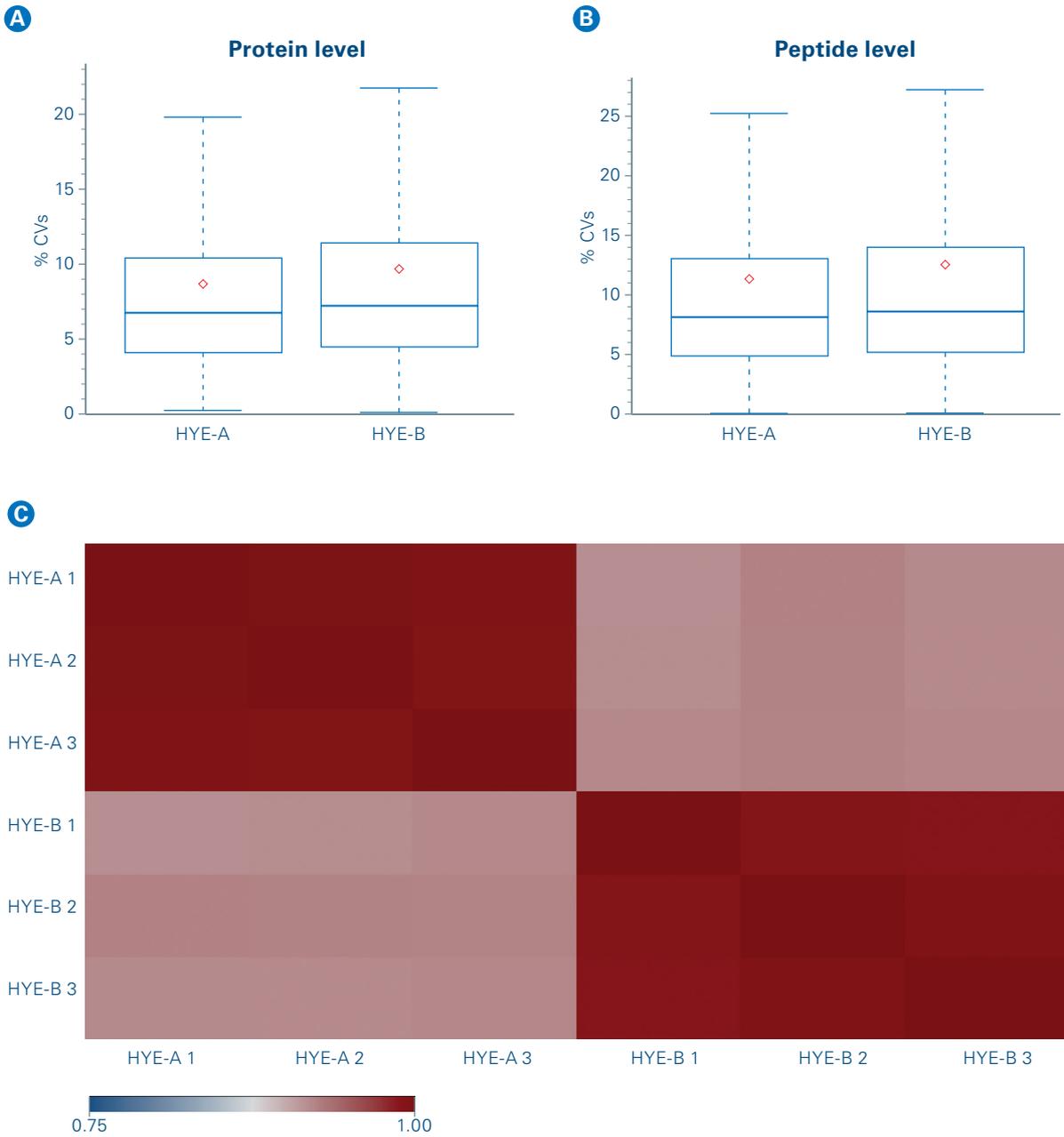
Figure 5: Reproducibility of diaPASEF measurements. Boxplots show the CV distribution in % per condition on protein group level Ⓐ and peptide level Ⓑ revealing the excellent reproducibility of diaPASEF with median CVs of 6.7% and 7.2% on protein group level and 8.1% and 8.6% on peptide level for sample HYE-A and sample HYE-B, respectively. Ⓒ Sample correlation matrix of precursor quantities between the two samples, HYE-A and HYE-B, together with the technical replicates.

and *E. coli* cell lysates. The three proteomes were mixed in defined ratios (1:1 for human, 1:4 for *E. coli* and 2:1 for yeast) with human proteins representing the background proteome.

With a 100 min LC gradient and a very complex sample consisting of three proteomes, we reasoned that a diaPASEF method with a somewhat lower duty cycle, but higher precursor selectivity should be beneficial. We devised a method with two windows in each 100 ms diaPASEF 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 precursor isolation windows

(Figure 2). Each diaPASEF 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 diaPASEF scan). Approximately 6.25% of the doubly and triply charged ions are used for MS/MS in every scan. In comparison, a standard DIA experiment with 32 windows would use only half of the ions in each MS/MS scan.

To identify and quantify peptides from the diaPASEF data, a software capable of analyzing the four-dimensional data is required. Here we applied Spectronaut software (version 13.6, Biognosys), which has been specifically developed for DIA-based quantitative proteomics. The processing includes fully auto-

mated in-run calibration, automatic interference correction, targeted analysis and automatic quality control.

The Spectronaut workflow generates spectral libraries using the integrated Pulsar database search engine directly from the fractionated DDA PASEF runs, which are used for targeted data extraction. The experiment specific spectral library from the analysis of 5 fractions per organism (HeLa, *E. coli*, Yeast), in two replicates each, consists of 13,013 protein groups (6,790 HeLa; 2,423 *E. coli*; and 3,800 Yeast proteins) and 136,682 unique peptide sequences (71,637 HeLa; 29,933 *E. coli*; and 35,112 Yeast peptides).
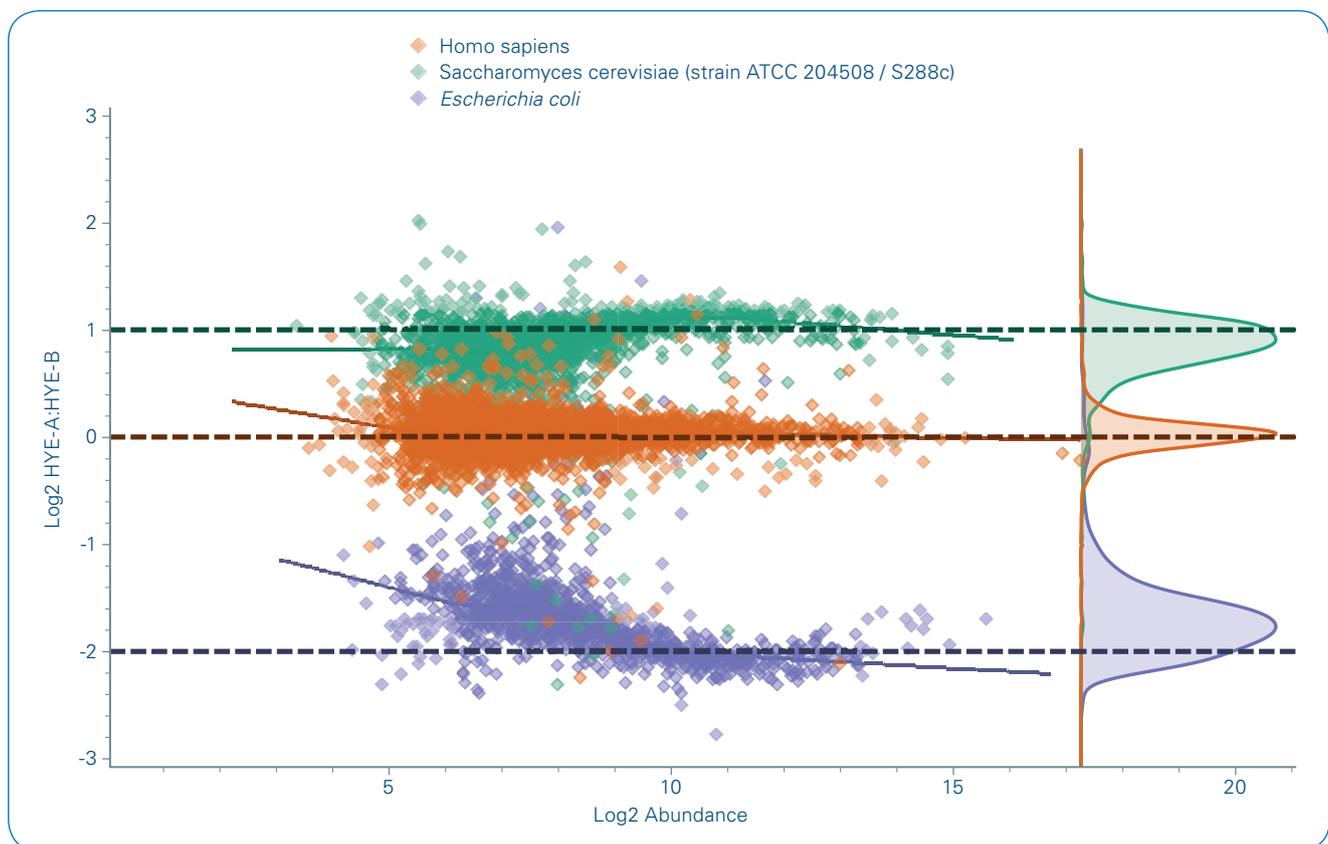

Figure 6: Normalized protein level quantification results. Log-transformed ratios (log2(HYE-A:HYE-B) of proteins (human proteins in orange, Yeast proteins in green, E. coli *proteins in purple) were plotted over the log-transformed intensity of sample HYE-B. Dashed colored lines represent the expected (log2(HYE-A:HYE-B) values for human, yeast and* E. coli *proteins.*

Using this comprehensive library for targeted data extraction, we were able to identify and quantify on average 8,080 and 8,029 proteins for sample HYE-A and HYE-B, respectively when excluding single-hit proteins and at a 1% false discovery rate (Figure 3). Whereas the average number of identified HeLa proteins is comparable for sample HYE-A and sample HYE-B (4,746 and 4,668), the number of *E. coli* (HYE-A: 1,286 / HYE-B: 1,599) and Yeast (HYE-A: 2,047 / HYE-B: 1,762) proteins differed most likely due to the different ratios present in the samples. On peptide level on average 81,410 unique sequences could be identified and quantified in total, among them 54,091 HeLa, 12,806 *E. coli*, and 14,513 Yeast peptides. In general, diaPASEF reveals highly reproducible number of identifications. For the described data set we were able to achieve 95.9% data completeness on protein group level, representing very few missing values. The identified proteins covered a dynamic range of around 5 orders of magnitude (Figure 4).

Reproducibility of peptide and protein quantification determines the confidence in applied proteomics approaches. The great advantage of DIA approaches in general and very specific of diaPASEF lies not only in the very reproducible identification of 1,000s of proteins but also in a high quantitative consistency and accuracy. The median coefficient of variation for the replicate runs was well below 10% for peptide and protein level (Figure 5), illustrating the excellent reproducibility of the timsTOF Pro's MSMS level quantification. A very high sample correlation with average correlation values of 0.99 is observed between the technical replicates of sample HYE-A and HYE-B. Correlation between the two samples was lower with average correlation values of 0.92.

The experimental design enabled the evaluation of the quantitative accuracy of diaPASEF in a complex proteomics mixture with well-known theoretical regulation ratios. Background human proteins were spiked in equal amounts resulting in a theoretical ratio of 1:1. We found the human proteins to be centered at a log2 ratio of sample HYE-A versus HYE-B of 0 across the complete dynamic range (Figure 6). For yeast and *E. coli* proteins we achieved a very good global accuracy on the timsTOF Pro with regulation ratios measured close to the theoretical ones (Median ratio sample HYE-A/HYE-B for yeast: 1.9 (expected value: 2.0) and for *E. coli*: 0.3 (expected value: 0.25).

Conclusions

- Biognosys' DIA processing software Spectronaut now has beta support for diaPASEF data.

- More than 8,000 proteins could be reproducibly identified in single 100 min LC runs from 200 ng sample injected taking advantage of the new features of the timsTOF Pro including high ion usage efficiency from the TIMS device.

- diaPASEF delivers extremely reproducible identification and quantification information over a concentration of 5 orders of magnitude 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/applications/life-sciences/proteomics

References

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

[2] Cox J et al (2014) Mol Cell Proteomics, doi: 10.1074/mcp.M113.031591

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

[4] Navarro P et al (2016) Nat Biotechnology, doi: 10.1038/nbt.3685

[5] Meier F et al (2015) J Proteome Res., doi: 10.1021/acs.jproteome.5b00932