Deep proteome coverage at scale combined with reproducible quantitation on the timsTOF HT

Stephanie Kaspar-Schoenefeld[1], Andreas Schmidt[1], Markus Lubeck[1], Pierre-Olivier Schmit[2], Torsten Mueller[1] and Narayanaganesh Balasubramanian[3]

[2] Bruker Daltonics GmbH & Co. KG, Bremen, Germany, [3] Bruker France S.A., Wissembourg, France, [3]Bruker Daltonics, Billerica, MA, USA

Introduction

In the realm of proteomics, accurate and precise relative protein quantitation is the key to unravel the complex secrets of biological processes. Data independent acquisition (DIA) offers the advantage of comprehensive proteome coverage and reliable quantitation of proteins, while reducing missing information leading to more robust results. Dia-PASEF is an advanced variant of DIA, capitalizing on the additional dimension of separation unlocked on the timsTOF platform by trapped ion mobility separation (TIMS). Demonstrating these capabilities resulted in optimizing a dia-PASEF acquisition method to analyze both single proteome samples and complex hybrid proteome mixtures for benchmarking.

Methods

Tryptic digests of HeLa, yeast and E.coli were used to evaluate the performance of dia-PASEF for short gradients on the timsTOF HT. Digests were either loaded directly onto the column or combined in defined ratios, resulting in complex benchmark sample sets. Samples were loaded directly on a 15 cm C18 column (75 µm inner diameter, 1.6 µm particle size, Aurora, IonOpticks) using a nanoElute 2 nano HPLC (Bruker) coupled to a timsTOF HT mass spectrometer (Bruker) via a CaptiveSpray 2 ionization source (Bruker) using a 15-min acetonitrile (ACN) gradient. For the dia-PASEF acquisition, a window placement scheme optimized via the py-diAID tool was used. Data were processed in Spectronaut (v18, Biognosys) using library-free mode (directDIA+™). False discovery rate (FDR) was controlled at 1% for peptide and protein group level.

Results

High proteome coverage from 15-minute gradients

We investigated the performance of the timsTOF HT mass spectrometer for full proteome analysis using short gradients. First, we analyzed yeast, a eukaryotic organism, which is known to express around 4500 proteins. With a very short turnaround of just 15 minutes active LC gradient, we were able to cover nearly the complete yeast proteome by identifying on average 4408 protein groups (Fig. 1 A) with 50,450 peptides. When analyzing a human protein digest representing a higher complex proteome, we identified nearly 105,000 stripped peptide sequences from close to 8000 protein groups (Fig. 1 B). In both cases 95% of the protein groups were identified and quantified with CV values below 20% from triplicate injections, underlining the excellent reproducibility of the presented setup.

Accurate and reproducible label-free quantitation from complex proteome mixtures

For evaluation of the presented setup for quantitative proteomics we used samples mixed in defined ratios. From the three-model proteome mixture (HeLa, yeast, E.coli) we were able to identify on average 12,893 protein groups from 173,851 peptide precursors using library-free data processing (Fig. 2 A). On average 10 peptides per protein have identified and quantified. Obtaining a high sequence coverage is advantageous as it enables the comprehensive mapping of proteoform diversity, going beyond simple identification numbers. The great advantage of DIA approaches in general and specifically of dia-PASEF lies not only in the very reproducible identification of 1000s of proteins but also in a high quantitative consistency and accuracy. In our case, the combination of high protein coverage with enough data points per peak contributes positively to reach an even better quantitative accuracy. As a result, the median coefficient of variation (CV) for the replicate runs was at around 5% for all sample sets on protein group level (Fig. 2 B).

The chosen experimental design enabled the evaluation of the quantitative accuracy of dia-PASEF in complex proteomics mixtures with pre-determined theoretical ratios. Background human proteins were spiked in equal amounts resulting in a theoretical ratio of 1:1. We found the relative quantitation of the corresponding human proteins to be centered at the expected ratio across the complete dynamic range for the three different sample sets measured (Fig. 2 C), median log2 ratios of –0.01, -0.02, -0.05). The ratios obtained for yeast and E. coli were close to the expected ratios with low levels of standard deviation demonstrating the ability of the timsTOF HT to differentiate small and high changes in abundance making it perfectly suited for investigation of proteome-wide changes in protein abundance in different biological contexts.

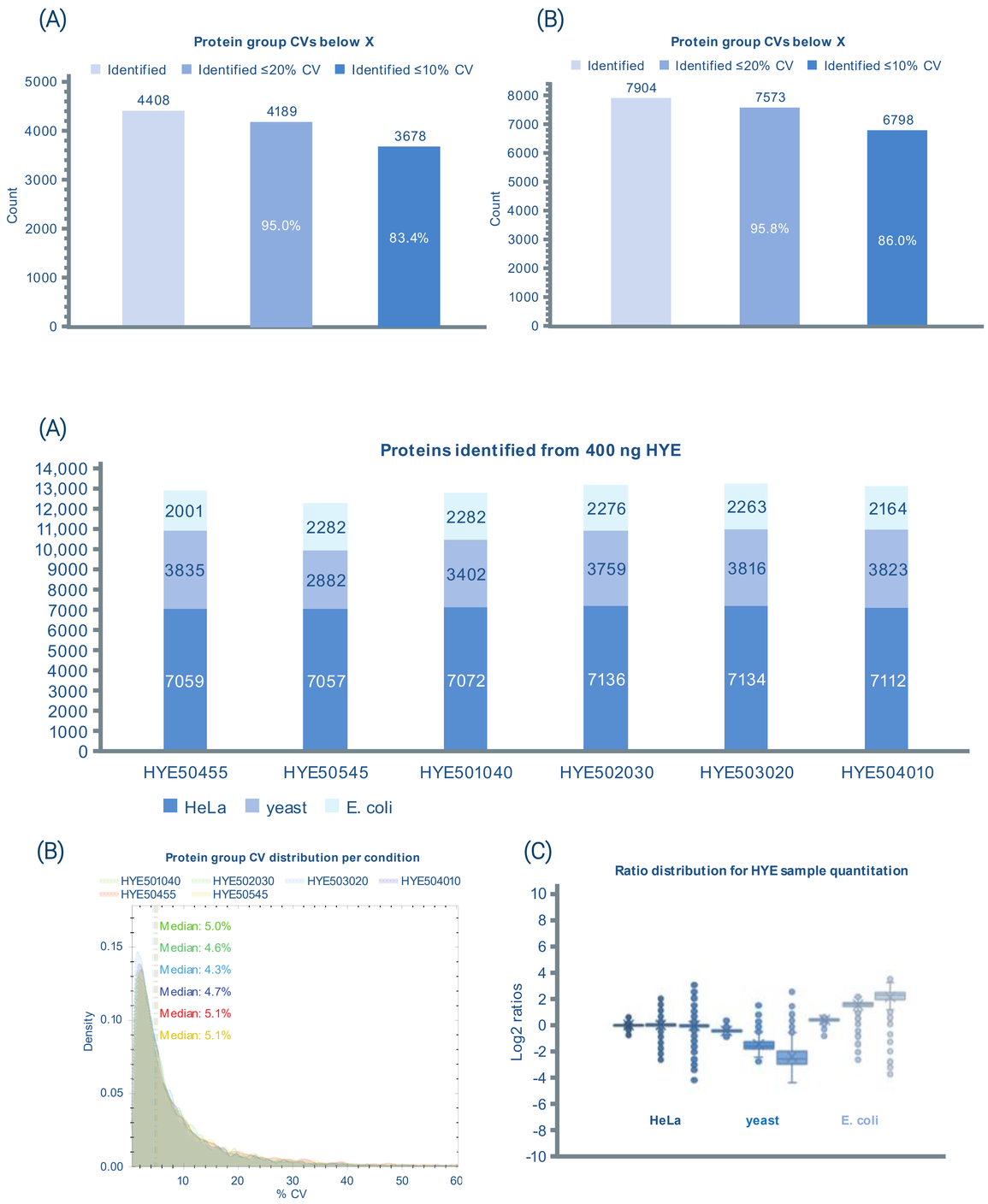

Fig. 1: Reproducible and in-depth protein identification from yeast and HeL protein digest using library-free data processing (directDIA+, Spectronaut v18). . A: Number of protein groups identified and quantified below 10 and 20% coefficient of variation from triplicate injections of yeast (A) and human protein digest (B) using dia-PASEF on a 15-minute gradient.

Fig. 2: In-depth protein identification and reproducible quantitation using dia-PASEF from complex proteomics samples. A: Average number of identified protein groups from the mixed proteome samples fore each of the three species from triplicate injections. B: Coefficient of variation (CV) values of protein groups from triplicate injections. C: Box and Whisker plot of obtained log2 ratios for all three species (HeLa, yeast and E.coli) for the different quantitative comparisons.

Conclusion

- dia-PASEF on the timsTOF HT enables high proteome coverage and accurate quantitation in short gradients of 15 minutes.
- Nearly full yeast proteome coverage can be achieved with on average 4408 protein groups identified. Remarkably, 95% of those protein groups were identified and quantified with coefficient of variation values of below 20%.
- Analysis of complex mixed proteomes results in identification of nearly 13,000 protein groups using library-free data processing.
- Accurate and reproducible quantitation can be achieved using dia-PASEF and short gradients in combination with complex sample types. This makes the setup perfectly suited to run high-throughput biological studies.

timsTOF HT