Comparative evaluation of a new processing pipeline for PASEF Label-Free Quantification analysis.

FP 375

Patrick Emery[1], Manuel Chapelle[2], Pierre-Olivier Schmit[2] and Aurélie Même[2]

1 Matrix Science ltd, London, UK.
2 Bruker France S.A., 67160 Wissembourg, France

Introduction

Parallel Accumulation Serial Fragmentation (PASEF) data acquisition strategies have changed the way proteomics data are recorded in many different way: on top of the combined speed and sensitivity increase, the additional separation of target ions in the ion mobility dimension as well as the systematic measurement of their collisional cross-section (CCS) has dramatically increased the data files information content. The recently introduced 4D-proteomics approaches are making an extensive use of the ion mobility separation and of the CCS information to increase the identification reliability, the data completeness, and the quantitation accuracy. Here we have evaluated the performances of a newly introduced processing pipeline and compared it to the established MaxQuant and Peaks Studio platforms.

Methods

Sample A and B have been created by mixing tryptic digests from human cell line, yeast and E.Coli in two different ratios (A : 65%-15%-20% and B : 65%-30%-5%), dried down and resuspended in 20µl (94,9% H2O, 5% CH3CN, 0,1% TFA). 500ng of each sample have been injected as quadruplicates on a 25cm Aurora nanocolumn (IonOptiks) with a 60 min gradient using a nanoElute nano-LC system coupled to a timsTOF PRO mass spectrometer (Bruker) operated in PASEF acquisition mode. Data have been processed for identification and label-free quantification using MaxQuant 2.0.1.0 (Cox Lab), Peaks X the (BSI) or Mascot Distiller 2.8 (Matrix Science) and the same mixed proteome database. For all search engines, protein false discovery rates have been adjusted to 1% and a minimum of one unique peptide was requested for identification. We report the identification and quantitation performances for each of these processing pipelines.

Processing Pipeline	Identified protein groups				Quantified protein groups			
	Total	Human	Yeast	E.Coli	Total	Human	Yeast	E.Coli
MaxQuant 1.6.17.0	8489	4718	2462	1309	7768	4365	2190	1213
MaxQuant 2.0.3.0	8808	4884	2561	1363	8598	4842	2488	1268
Peaks XPro	9102	5082	2634	1386	7140	4447	1873	820
Mascot Distiller 2.8	10575	5723	3313	1539	8906	4947	2585	1374
Mascot Distiller 2.8 2 peptides	8812	4872	2539	1400	5535	3245	1299	991

Table 1: ID and quant results for protein groups (60 min gradient).
The same dataset (acquired in 2019) has been processed with different software pipelines on the same workstation. 2 stringency thresholds were tested for Mascot Distiller (1 or 2 unique peptides required for quantitation).

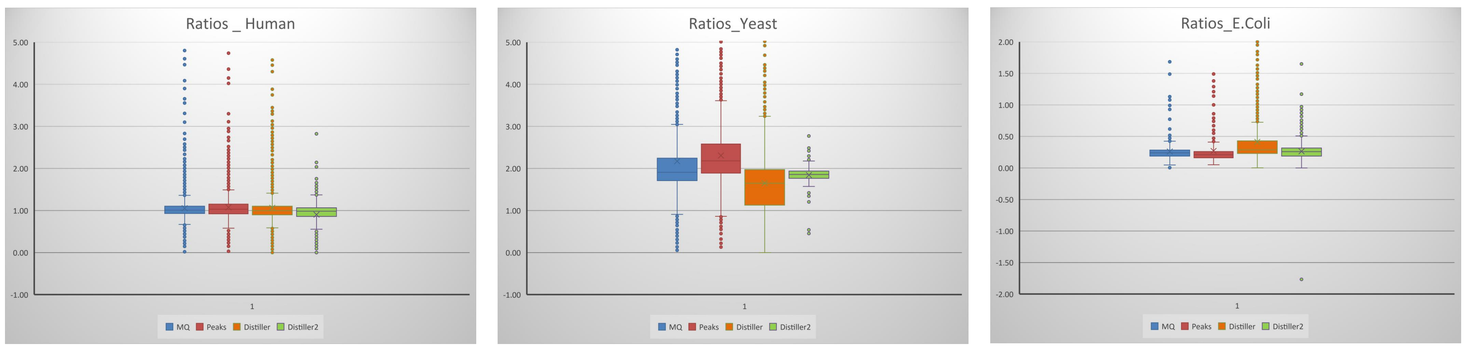

Fig. 1: Ratio accuracy (protein groups – 60 min gradient).
Display of A:B ratio for all quantified protein groups, the average values of Group A being normalized to 1 for each protein group.

Results & discussion

The three pipeline all allowed to identify more than 8800 protein groups using a 1% protein FDR threshold. The quantified/identified ratios were 97.6%, 78,4% and 84,2% for MaxQuant, Peaks and Mascot Distiller, respectively. If Mascot Distiller reported 20% more ID's than MaxQuant, the observed gap is down to 3.6% when considering the quantified protein groups (Table 1). Interestingly, the proportion of Ecoli proteins quantified is higher for MaxQuant and Mascot Distiller, whereas Peaks reports a higher proportion of human quantified human proteins. However, if all pipelines deliver similar ratio accuracy for the Human proteins, it seems that the ratios calculated by Mascot Distiller for The Yeast and E.Coli proteins are slightly compressed (on top of having a broader distribution – Fig.1). Requesting a to perform the quantitation on 2 peptides for Mascot Distiller allows to correct this bias and narrows the intra-specie ratio distribution but does also result in loosing 38% of the quantified proteins. We now need to further adjust the stringency settings to retrieve some of the missing proteins while preserving the ratio accuracy. We also focused on some of the Human protein outliers : the example displayed in Fig.2 is representative of most of the cases we have investigated. In these cases, the ratio determined by the pipeline is not representative of the one that be determined by performing a manual integration of the same peptides, underlining that the compound finder algorithms can further be improved.

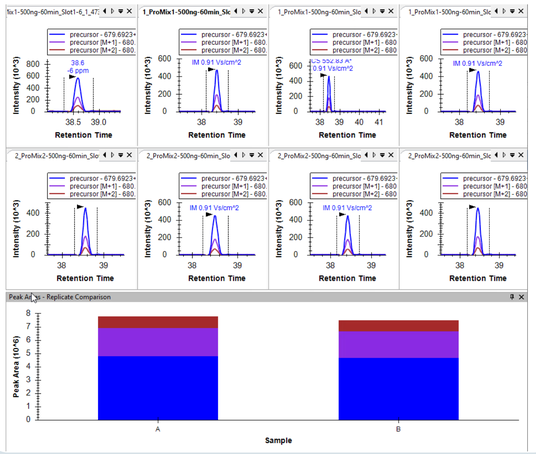

Fig. 2: Highlighting remaining ratio extraction issues
Example of the Human Sp Q9ULJ6 protein. The theoretical A:B ratio is 1:1, the ratio given by MaxQuant is 1:2.2, based on the triply charged form of the QHLQNPANFHNAATELLDWCGDPR peptide. The figure display the Skyline integration for this peptide trace for all 8 injection as well as the concatenated group intensities. The ratio detected from Skyline is 1:1.

Conclusions

- The ion mobility dimension is fully supported in the three processing pipelines that have been evaluated
- In this example, more than 8000 protein groups could be accurately quantified from 60 minutes gradient runs, the 2021 processing pipelines showing slightly improved results compared to their 2020 counterparts
- The results can be further improved, on the one hand by using last generation acquisition methods (an old dataset has been used to compare to results obtained with the older versions of the software) and on the other hand by continuing to improve the data extraction algorithms.

timsTOF PRO