Automated and faster library-free dia-PASEF analysis with a Spectronaut integrated workflow in ProteoScape

Javier Lopez[1], Sven Brehmer[1], Grzegorz Skoraczynski[2], Jose Abuin[1], Tejas Gandhi[2], Oliver Bernhardt[2], Lukas Reiter[2], Dennis Trede[1], Jonathan Krieger[3] and Tharan Srikumar[3]

[1]Bruker Daltonik GmbH, Fahrenheitstraße 4, 28359 Bremen, Germany [2]Biognosys AG, Wagistrasse 21, 8952 Schlieren (Zurich), Switzerland , [3]Bruker Ltd., 2800 Highpoint Dr. Suite#206, Milton, ON L9T 6P4, Canada

Introduction:

Bruker ProteoScape (BPS), formerly PaSER, has been transforming into a comprehensive proteomics data analysis platform that can integrate third-party tools while utilizing the concept of data streaming to realize fully customizable real-time processing workflows including on-the fly decision making based on the data generated. Recent advances in technology have decidedly allowed the field to migrate towards fast LC gradients, allowing for increasingly larger sample cohorts in studies in both data-dependent acquisition (DDA) and data-independent acquisition (DIA) modes. "Library-free" analysis of DIA data has also grown to be an increasingly alternative to generating project specific spectral library from DDA acquisitions. A significant downside of these "library-free" workflows is longer processing times due to the increased search space, especially when PTMs are considered.

Here we've integrated a workflow including a Spectronaut (SN) module in Bruker ProteoScape (BPS) to take advantage of the synergistic capabilities of the two software. We show that BPS is a natural platform for faster library free analysis of dia-PASEF data using Spectronaut's directDIA+ workflow by:

1. Allowing data to be processed at a single injection level, before combining at the project level.
2. Providing same results, including FDR control, for both single injection and project level analysis.
3. Providing an intuitive user interface for the entire workflow.

We utilize this new workflow on two datasets, including a phosphorylation enriched dataset and a multi-species dataset with know ratios to help evaluate precision and accuracy of the workflow and can show more than 40% savings in time-to-results vs analysis fully after the acquisition. BPS can provide quick visualization of both injection level data and project level data, while detailed peptide\spectrum level analysis can be explored in a standalone Spectronaut version.

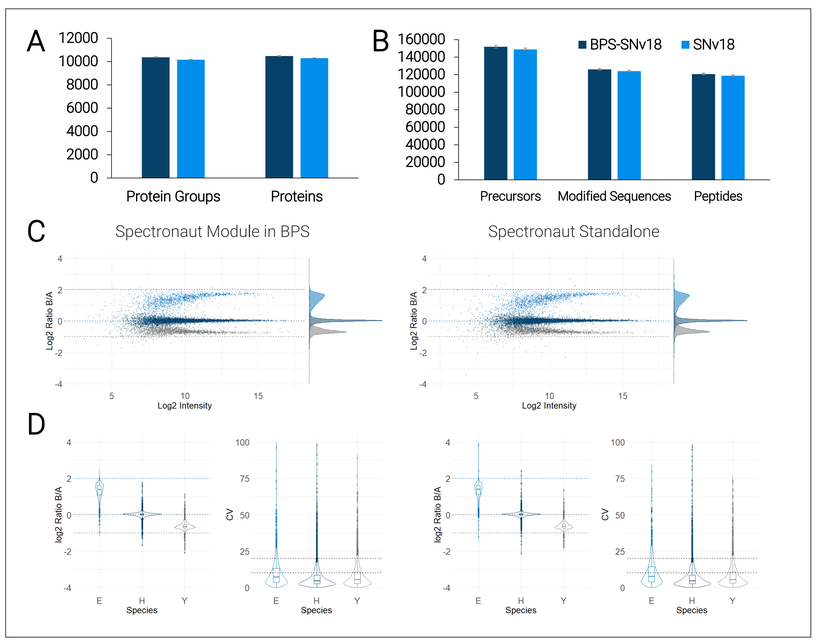

Fig. 2: The Spectronaut module in BPS and standalone version produce comparable results with regards to identification as well as quantitation accuracy and precision. Average number of identifications across all 10 replicates by Protein and Protein Groups (A) and by Precursor, Modified Sequences and Peptides are similar. (C) shows intensity vs ratio distribution by species for a dataset with 5 replicates. (D) The expected and measure ratios for all protein groups and the CV by species are all comparable.

Results:

A multi-species dataset, consisting of Human, Yeast and Ecoli mixed at ratio of 1:1;1:2;4:1, respectively between two samples was acquired in 5 replicate injections and processed using the BPS integrated Spectronaut directDIA+ workflow as well as in the standalone. The traditional processing pipeline, requiring users to manual trigger file transfer between acquisition and processing workstations, was 20% slower than then BPS pipeline (see Figure 1). Furthermore, users are able to evaluate the results of each individual run, on a run-by-run basis, allowing re-injection or re-analysis if the data is suboptimal. Both pipelines produced comparable results with regards to identification, quantitation and precision (see Figure 2). The BPS pipeline could be further optimized to increase system parallelism or via improvements to processing times in the Spectronaut module.

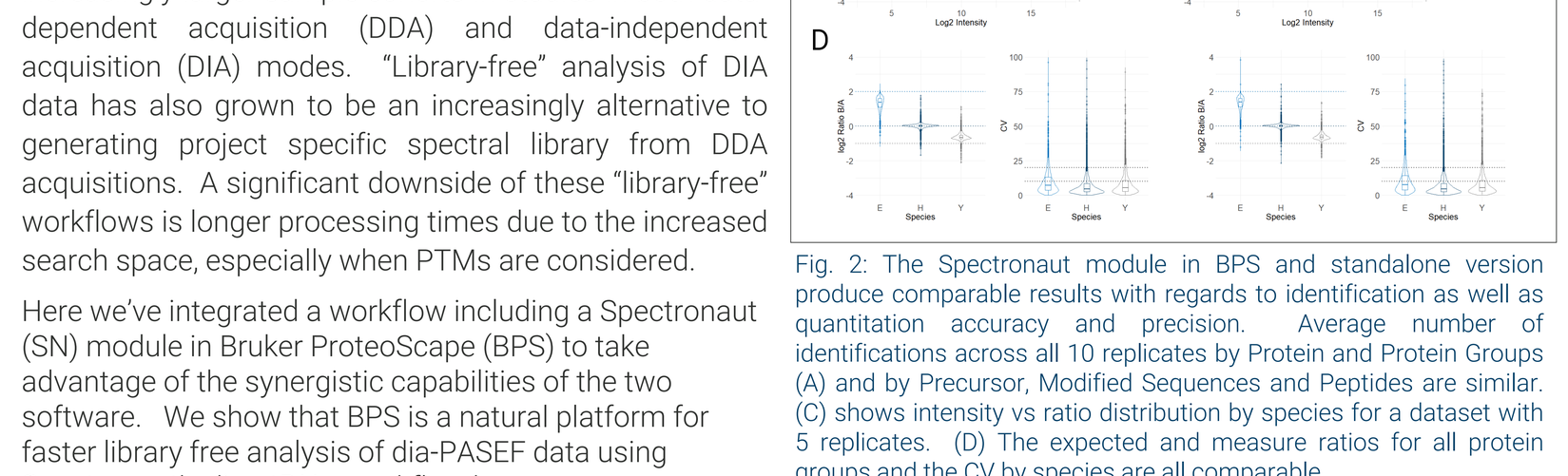

Fig. 1: BPS allows for automated dia-PASEF data processing. In this example with 2 samples and 5 replicates, with a 45min LC gradient and 15min LC overhead, users can get fully processed data 2.5h faster or ~ 40% faster. With traditional processing pipelines, users need to wait until all files have been acquired before transfer and subsequent processing. With BPS processing, run-by-run analysis is performed during the acquisition queue, so users are able to evaluate the data quality for each injection. At the end of the acquisition queue, users can trigger a multi-sample combine workflow, providing a project-wide view for further analysis.

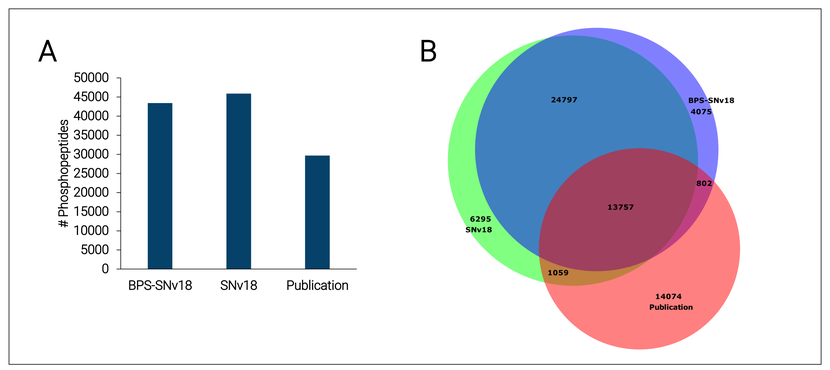

Fig. 3: A phosphorylation enriched dataset (Skowronek et al., 2022) was re-processed both in standalone Spectronaut as well as the BPS integrated versions. In both cases the identification of phosphopeptides were comparable (A) with >90% overlap (B).

We also utilized the directDIA+ PTM workflow for the re-analysis of the 21min gradient 'optimal' dia-PASEF phosphorylation enriched dataset from (Skowronek et al., 2022; PXD034128). BPS processing pipeline and standalone versions showed similar identifications rates and >90% overlap in the phosphopeptides identified with a PTM localization filter of 0.75 (see Figure 3). We also compared the phosphopeptides identified in the publication, which was processed using Spectronaut v16 and deeply fractionated phosphopeptides spectral library. Here overlap was poor and could indicate an area of further improvement for library-free workflows.

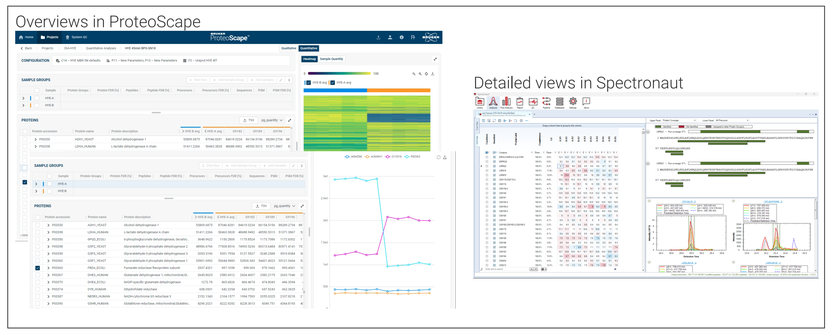

Fig. 4: The redesigned graphical user interface of BPS. Data\Results can be explored in BPS or the project file (*.sne) can be exported to be further explored in a standalone Spectronaut version.

Conclusion:

- ProteoScape has workflows integrating academic algorithms, such as DIA-NN, or 3rd party algorithms, such as Novor, and now Spectronaut greatly reducing time-to-results for large projects while simultaneously providing detailed results at the run-by-run level.

Technology