Towards real-time glycopeptide identification on the timsTOF Pro - PaSER platform: virtual precursor enabled peptide-moiety identification

Gad Armony[1], Sven Brehmer[2], Lennard Pfennig[2], Sebastian Wehner[2], Fokje Zijlstra[1], Tharan Srikumar[3], Dennis Trede[2] Gary Kruppa[4], Dirk Lefeber[1,5], Alain van Gool[1], Hans Wessels[1]

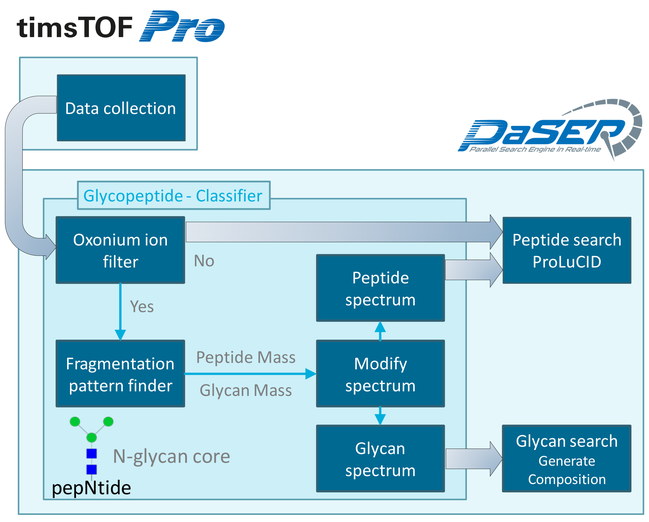
1 Translational Metabolic Laboratory, RadboudUMC, Nijmegen (NL); 2 Bruker Daltonics GmbH, Bremen (DE); 3 Bruker Ltd, Milton (CA); 4 Bruker S.R.O., Brno-City (CZ); 5 Department of Neurology, RadboudUMC, Nijmegen (NL)

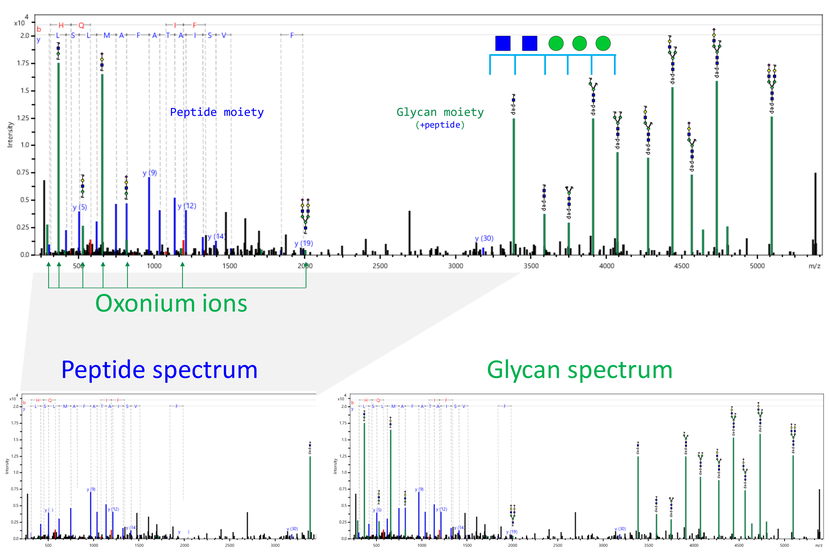
Introduction

Glycosylation of proteins has strong implications on protein function and has high potential for serving a biomarkers in clinical applications. Holistic glycoproteomics in blood offers unique possibilities for functional diagnoses of different human diseases by providing site-specific data for glycosylation of hundreds of proteins in a single measurement. We develop software modules for the PaSER (Parallel Search Engine in Real-time) computational platform to efficiently handle data generated by PASEF-DDA on timsTOF Pro instruments. These modules enable glycoproteomics in clinical environments by performing (semi) real-time data processing, data analysis, on-the-fly acquisition parameter adjustment, reporting, and data management.

Methods

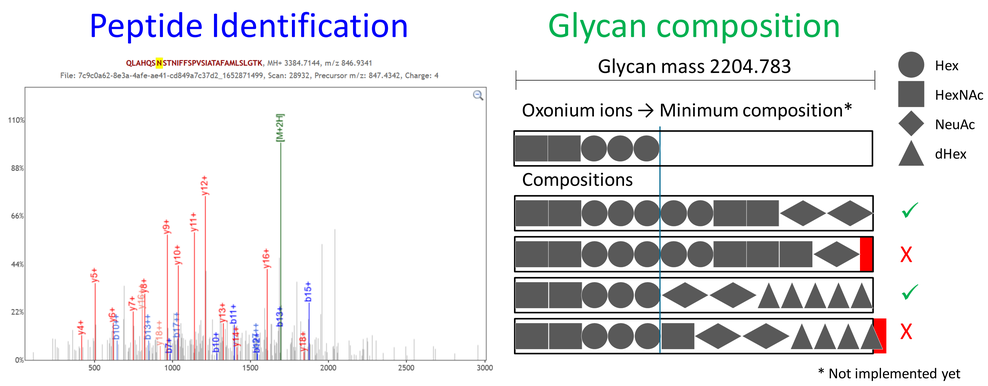
To enable real-time glycopeptide identification capabilities on PaSER, we split the glycopeptide identification into two distinct processes: peptide moiety identification and glycan moiety identification. The fragmentation spectra are streamed (Kafka) to a classifier module which first determines if a spectrum was derived from a glycopeptide based on oxonium ions. Next, the classifier uses the fragmentation pattern of the constant N-glycan core structure to determine the mass of the peptide moiety part. For peptide identification, the spectrum is processed and submitted to the database search engine in PaSER (ProLuCID). As a first step towards full glycan characterization, the glycan mass is used to calculate possible compositions using oxonium ions as constraints.



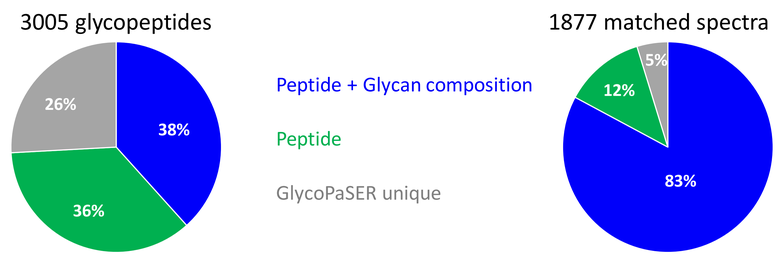


Results

With the classifier implemented in the PaSER platform, glycopeptide characterization from fragmentation spectra is performed in real-time during data acquisition. Therefore, right after the end of the LC-MS/MS measurement the glycopeptide identification results are available for subsequent downstream analysis. Currently, these glycopeptide search results include the identification of the peptide moiety and potential glycan composition(s) based on the glycan moiety mass and oxonium ion information.



The identifications from the real-time search are compared to MSFragger search results performed on the same data. On the left, overlap for all glycopeptides - unique peptide sequence and glycan composition (mass). On the right, overlap between identifications of the same spectra.



Conclusions

- Delivered proof-of-concept for N-glycopeptide identification in real-time.
- Processing time per spectrum fits in PASEF duty cycle (<25 msec).
- Current focus is on improving the performance of peptide moiety identification by ProLuCID and on improving glycan moiety characterization using glycan fragmentation data.

Contact information
Gad Armony, RadboudUMC, The Netherlands
E: gad.armony@radboudumc.nl

This research was part of the Netherlands X-omics Initiative and partially funded by NWO (project 184.034.019), supported by ZonMw Medium Investment Grant (40-00506-98-9001), Health-Holland TKI-LSH PPS grant LSHM21032 "EnFORCE" and EUROGLYCAN-omics (ERARE18-117), under the frame of E-Rare-3, the ERA-Net for Research on Rare Diseases.