The GlycoPaSER prototype as a real-time N-glycopeptide identification tool based on the PaSER parallel computing platform

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]

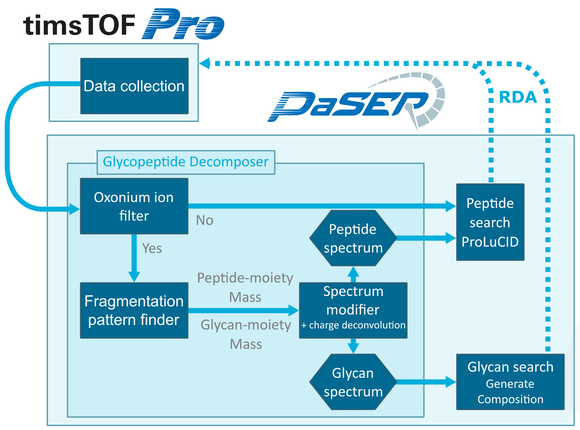
[1] Translational Metabolic Laboratory, RadboudUMC, Nijmegen (NL); [2] Bruker Daltonics GmbH, Bremen (DE); [3] Bruker Ltd, Milton (CA); [4] Bruker S.R.O., Brno-City (CZ); [5] Department of Neurology, RadboudUMC, Nijmegen (NL)

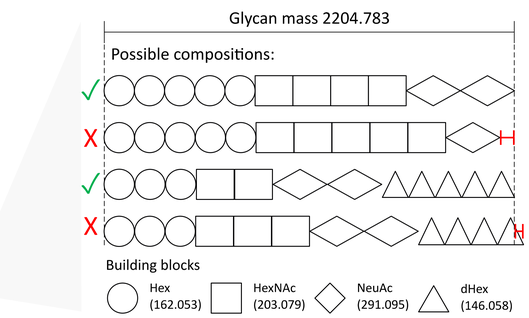
Introduction

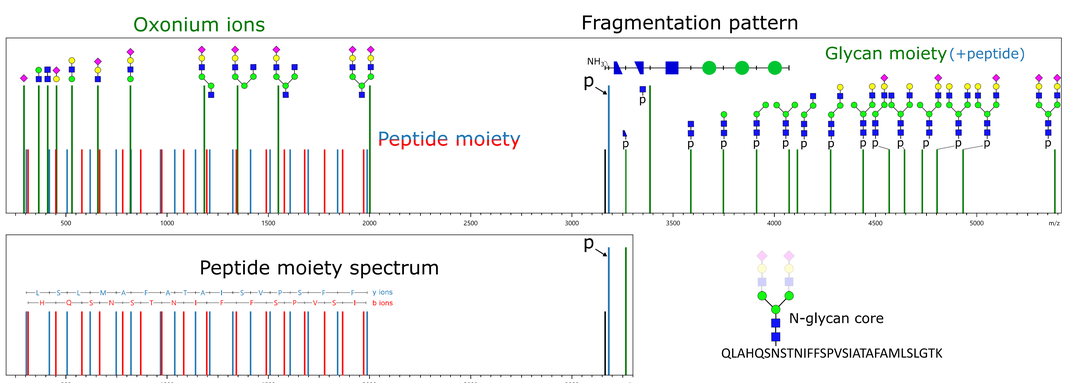
Glycosylation of proteins has strong implications on protein function and has high potential for serving as biomarkers in clinical applications. Holistic glycoproteomics in blood offers unique possibilities for functional diagnoses of different human diseases by providing site-specific data for glycosylation of hundreds of proteins in a single measurement. We develop GlycoPaSER, software modules for the PaSER (Parallel Search Engine in Real-time) computational platform, which 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, results dependent acquisition (RDA), 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 to a decomposer module which first determines if a spectrum was derived from a glycopeptide based on oxonium ions. Next, the glycopeptide decomposer 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.

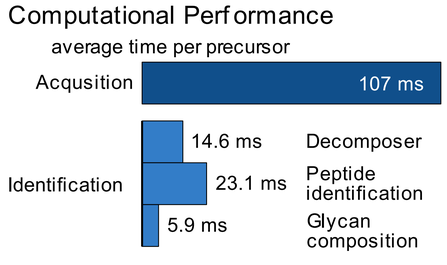


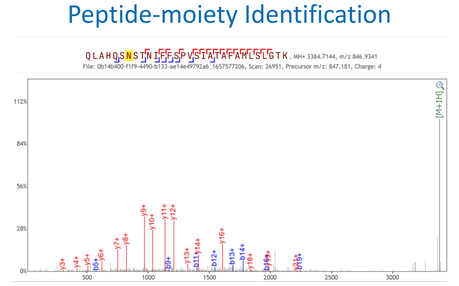


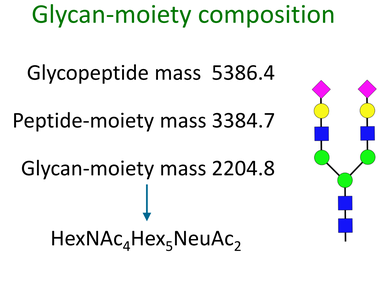


Results

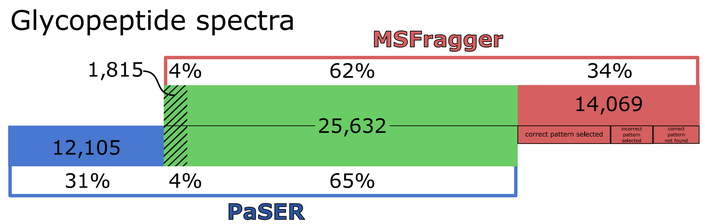
Comparing the glycopeptide PASEF acquisition time to the GlycoPaSER identification time shows that the PaSER modules are quick enough to perform the identification in real-time, during data acquisition. Therefore, right after the end of the LC-MS/MS measurement the glycopeptide identification results are available for subsequent downstream analysis. These glycopeptide results include the identification of the peptide moiety and potential glycan composition(s) based on the glycan moiety mass.

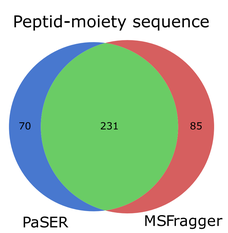


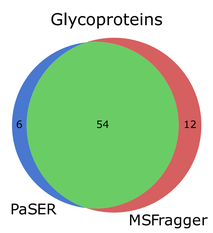




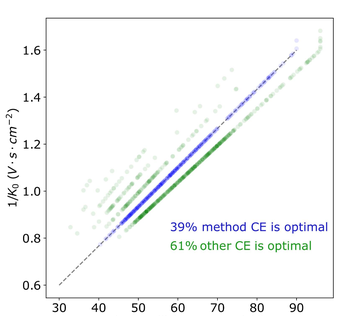
The identifications from a real-time search on PaSER were compared to results of an MSFragger search results performed on the same data. The amount of unique spectra identified by each of the tools probably stems from the different approaches taken for glycopeptide identification. Nonetheless, when both tools identify the same spectrum, 93% of the identifications are the same.



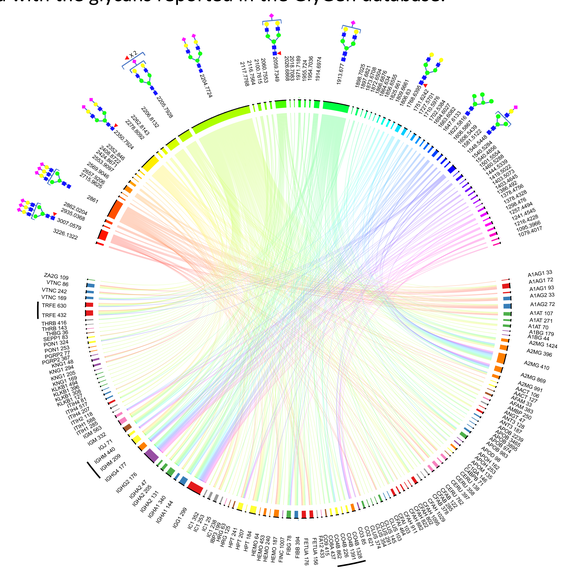




GlycoPaSER enables results dependent acquisition (RDA) which can improve the data quality by adjusting acquisition parameters in real time. For example, our optimized collision energy for glycoproteomics is optimal for only 39% of the fragmented spectra.



The intricate relationship between the glycosylation sites and the types of glycans that are found on them are visualized by the chord diagram. We can observe that the most frequently detected glycan mass correspond to complex di- and tri-antennary glycans which are the dominant glycans of the plasma N-glycome. Furthermore, some key plasma glycoproteins were identified with the glycans reported in the GlyGen database.



Conclusions

- Provided proof-of-concept for N-glycopeptide identification in real-time.
- The identification is quick enough to run in real time (up to 30 Hz).
- The performance is comparable to MSFragger.
- The glycopeptide identifications are in line with the literature.
- Performance improvement and new features are under development.
- GlycoPaSER enables results dependant acquisition (RDA) for glycopeptides.

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.