Real-time de novo sequencing of peptide antigens using Bruker ProteoScape™ for 'Run & Done' 4D-immunopeptidomics

Kirti Pandey[1], Rui Zhang[2], Qixin Liu[2], Mingjie Xie[2], Dennis Trede[3], Tharan Srikumar[4], Jonathan R Krieger[4], Bin Ma[2], George Rosenberger[5], Anthony W. Purcell[1]

[1]Monash University, Clayton, Australia, [2]Rapid Novor Inc., Kitchener, ON, [3]Bruker Dalton's GmbH & Co KG, Bremen, Germany, [4]Bruker Ltd., Milton, ON, [5]Bruker Switzerland AG, Fällanden, Switzerland

Introduction:

Bruker ProteoScape (BPS) was implemented on timsTOF instrumentation to enable real-time database interrogation of proteomics data. The provision of immediate peptide assignments vastly improves the speed of analysis and allows assessment of instrument performance, sample management and sample orientated decisions to be made immediately. Immunopeptidomics is the study of naturally processed peptides bound in complex with HLA molecules, that are presented on the cell surface to the immune system. There has been a resurgence of interest in immunopeptidomics fueled by improvements in instrument sensitivity and speed coupled to bioinformatic workflows that frequently rely on de novo sequencing to improve peptide-spectrum-matches. Here we implement a real-time de novo search algorithm to interrogate data acquired on a timsTOF Pro 2 instrument for immediate assessment of immunopeptidomics samples.

Methods

Novor was trained on a variety of timsTOF acquired data, where ground truth is taken from ProLuCID-GPU database search results filtered to 1% PSM FDR. These data included experiments with fixed collision energy measurements of deeply fractionated, GluC, Pepsin, Elastase, Chymotrypsin and Trypsin digested K562 lysates (Promega) across the range of 1-100eV. Training Novor on non-tryptic digests allowed it to learn a generalized model, particularly suited for sequencing of non-enzymatically digested peptides (Figure 1). On a BPS workstation, the MS/MS stream is processed by BPS Novor (during acquisition), and results are written to an output stream as well as persisted to disk.

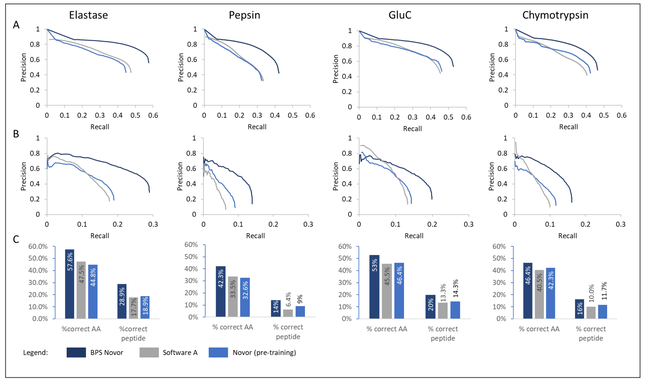

Fig. 1: Amino acid (A) and peptide (B) precision recall graphs for Human lysate digested with one of 4 enzymes, Elastase, Pepsin, GluC or Chymotrypsin. The percent of correct amino acids and peptides assigned by each algorithm is also shown (C).

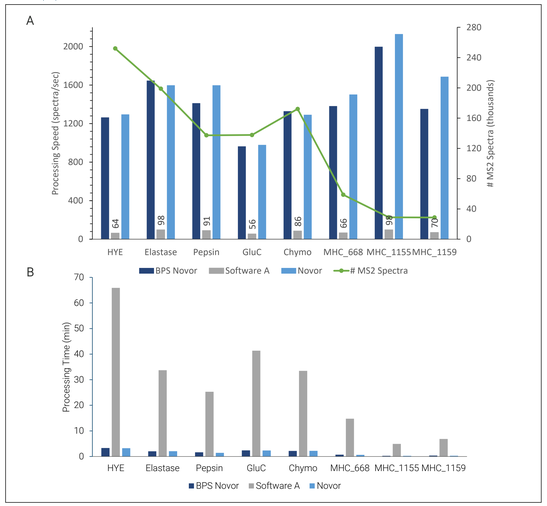

Fig. 2: Processing speed (A) and processing time (B) for 8 datasets that were benchmarked with BPS Novor, Software A and classical Novor. The datasets varied in number of spectra from >240,000 to 38,000.

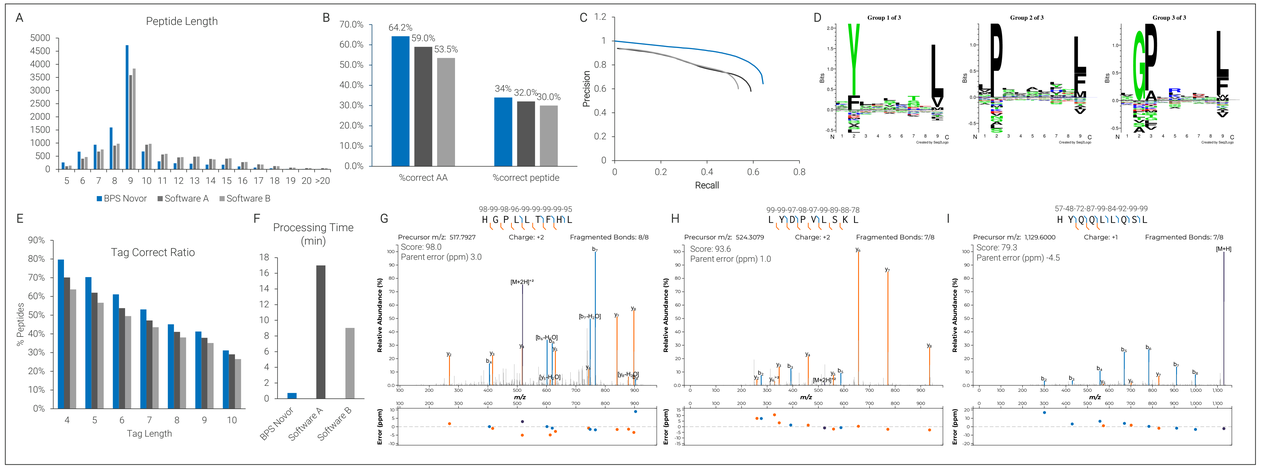

Fig. 3: Immunopeptidomics data derived from mouse CT26 colon cancer cell line. Data were processed in BPS Novor as well as two commercially available software. A) Peptide length distribution of all spectra with de novo score >70. B) Correct Amino Acid assignments and correct peptide assignments compared to ProLuCID database search results. C) Precision-recall graphs of the three software. D) Gibbs clustering of all 9-mer peptides identified with a de novo score > 70 by BPS Novor. E) Tag length vs correct tag assignments? F) Processing time in minutes for this dataset. (G-I) Examples of annotated spectra identified using the BPS Novor algorithm. Amino acid confidence shown above each residue.

Results

The timsTOF platform is capable of >150Hz scan speeds. To realize on-the fly real-time de novo sequencing, the algorithm needs to be capable of processing speed greater than the scan speed. We first evaluated the processing speed of Novor across 8 datasets, where each dataset consisted of >137,000 MS/MS spectra (Figure 2). Average processing time ranged from 86-199 seconds. This translated to an average processing speed of 1338±226 spectra/second. We utilized the BPS Novor module to produce de novo peptide sequences from real time acquired MS/MS spectra. We further show that training of the algorithm with timsTOF data improves the recall of the analysis at 75% precision on amino acid level to reach > 70% (vs 45% achieved with standard Novor that was not trained on timsTOF data). The implementation of this trained algorithm on the Bruker ProteoScape platform thus allows rapid decisions to be made on sample management and further data analysis. These studies pave the way for rapid interrogation of immunopeptidomics samples (Figure 3), where accurate and fast data are required to guide clinical decisions.

Conclusion

- A fast, accurate and precise peptide de novo sequencing algorithm has been integrated into Bruker ProteoScape, providing Run & Done capabilities to additional 4D-Proteomics applications.
- BPS Novor does not show a noticeable bias for digestion specificity or species and is 20x faster than competing products.
- Combined with PASEF technology on the timsTOF platform, BPS Novor provides enhanced sensitivity for real-time de novo sequencing for a variety of applications including immunopeptidomics.

Technology