TIMSrescore: timsTOF-optimized PSM rescoring boosts identification rates for immunopeptidomics and phosphoproteomics

Arthur Declercq[1], Jonathan Krieger[2], George Rosenberger[3], Ralf Gabriels[1], Robbin Bouwmeester[1], Tharan Srikumar[2], Sven Degroeve[1], Dennis Trede[3], Lennart Martens[1]

[1]VIB – Ugent Center for Medical Biotechnology, Gent, Belgium. [2]Bruker Ltd., Milton, Canada [3]Bruker Daltonik GmbH, Fahrenheitstraße 4, Bremen, Germany

Introduction:

The immunoproteome, the subset of protein involved in immune response, are key components of various disease mechanisms and recent progress in MS instrumentation and analysis have gained considerable attention for this special application. However, because immunopeptides are non-enzymatically cleaved, the necessary search space dramatically increases, resulting in lower peptide identification rates. It has been demonstrated that the introduction of orthogonal scores comparing predicted vs observed peptide properties can increase both sensitivity and specificity in these scenarios. Here, we introduce TIMSrescore, a timsTOF-optimized algorithm based on MS2Rescore (tims^{2}rescore), that provides these benefits for the timsTOF family of instruments.

Methods

The peptide fragmentation (MS2PIP) and collisional cross section (IM2Deep) predictors employed in TIMSrescore have been trained using a diverse set of timsTOF dda-PASEF PSMs. The updated models are available as part MS2Rescore (https://github.com/compomics/ms2rescore) and can be utilized independently from Bruker ProteoScape (BPS). The complete workflow in BPS is illustrated in Fig 1.

For evaluation we used 3 datasets. 1) Replicate injections of HeLa protein lysate digested with elastase. 2) Replicate phospho-enriched data from mouse cell lines with or without treatment with LPA provided by the laboratory of Prof. Stanely Stevens Jr. from a pilot project. The project was eventually analyzed by dia-PASEF and published https://doi.org/10.3389/fonc.2023.1048419. 3) The RCC tumor samples from Hoenisch et al., 2023. were also reprocessed with our pipeline. Representing triplicate measurements of HLA class I and II enriched samples.

Hoenisch Gravel, N. *et al.* TOFIMS mass spectrometry-based immunopeptidomics refines tumor antigen identification. *Nat Commun* 14, 7472 (2023).

Declercq, A. et al. MS2Rescore: Data-Driven Rescoring Dramatically Boosts Immunopeptide Identification Rates. Molecular & Cellular Proteomics 21, (2022).

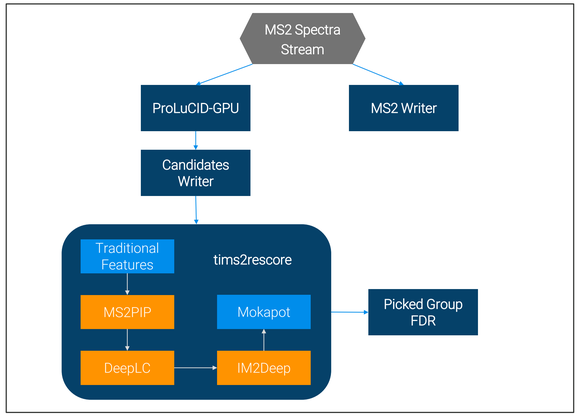
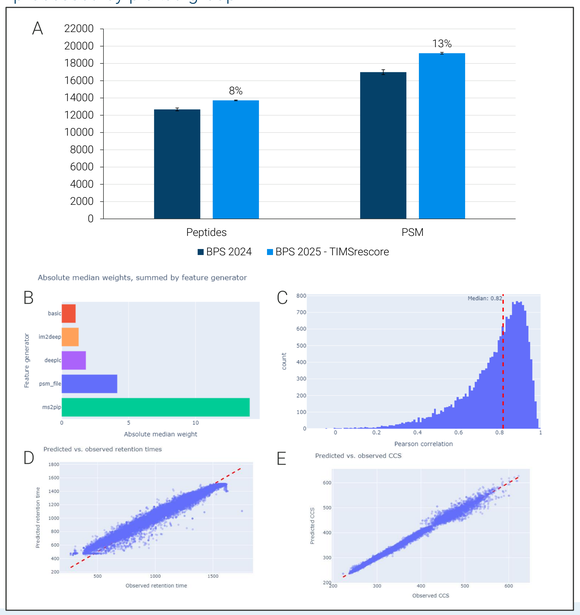


Fig. 1: The TIMSrescore workflow. As MS2 spectra are acquired from any timsTOF series instrument, data is streamed to ProteoScape. User defined search parameters are used by ProLuCID-GPU to generate a list of candidates for each spectra. Once acquisition ends, the candidates are written to parquet file and passed to the tims^{2}rescore module. tims^{2}rescore module adds many additional vectors to the traditional features provided by ProLuCID, including comparisons with predicted fragment ion intensities (via MS2PIP), retention time (via DeepLC) and CCS (via IM2Deep). This aggregated feature map is sent to Mokapot for PSM and peptide validation. The validated PSM list is then processed by picked group FDR.



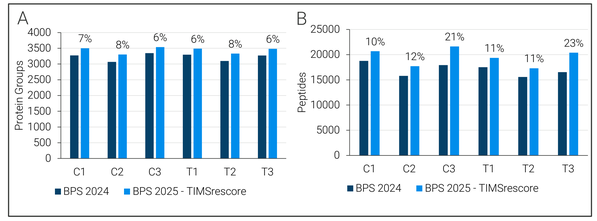


Fig. 3: Biological triplicates of control and treated phospho-enriched samples processed with and without TIMSrescore. TIMSrescore was able to increase the number of confidently identified (A) protein groups by ~7% (on average) and (B) peptides by ~15% (on average).

Results

Optimizing the MS2rescore feature generators for timsTOF data was critical for the improvements provided by the workflow. Details of the optimization will be provided in a manuscript that is in preparation.

In the HeLa elastase dataset, TIMSrescore increased peptide identification by 8% and PSMs by 13%. For the phosphoproteomics dataset, TIMSrescore increased recovery of peptides 10-23% and protein groups by 6-8%. For the immunoproteomic dataset, a considerable increase of 21% could be observed, indicating that rescoring is particularly impactful for large search spaces.

For all the datasets, the timsTOF optimizations of the fragmentation predictors were critical. For example, in the MHC-I dataset it led to an improvement of the median Pearson correlation to 0.88 from 0.53 (standard MS2PIP HCD model). IM2Deep was created based on principles of DeepLC, allowing for the CCS prediction for modified peptides, even if the modifications wasn't observed in the training data. The retention time predictions and ion mobility predictions had varying levels of (positive) contribution depending on the dataset.

Fig. 2: TIMSrescore workflow in BPS 2025 increases confidently identified peptides by >8% and PSM by 13%. 5ng of HeLa cell lysates digested with elastase were analyzed with 22min active LC gradient on timsTOF Ultra in triplicate. (A) The average number of peptides and PSMs are shown with and without rescoring. Feature usage in TIMSrescore. (B) Absolute median weights summed by feature generator. Each feature generator contributes to the overall rescoring process of separating target and decoy PSMs. (C) MS2PIP distribution of Pearson coefficients for all target PSMs (D) Deep-LC, and (E) IM2Deep model performance showing predicted vs observed RT and CCS.

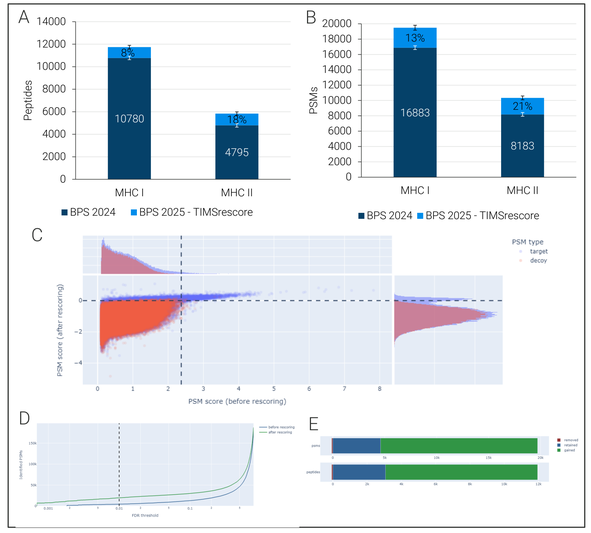


Fig 4: TIMSrescore for Immunopeptidomics. Peptide and PSM identifications with and without rescoring for both MHC-I (n=3) and MHC-II datasets (n=3). TIMSrescore increased confident MHC-I peptides by 8% and MHC-II peptides by 18%. Similarly, TIMSrescore increased MHC-I PSMs by 13% and MHC-II PSMs by 21% versus the standard workflow available in BPS for dda-PASEF analysis. (C) Scatterplot of Score Comparison for the MHC-I dataset. Target (blue) and decoy (red) PSMs before rescoring are shown on the x-axis and after rescoring are shown on the y-axis. The upper left quadrant are the PSMs only identified after rescoring. (D) False Discovery Rate Comparison. This shows the number of identified target PSMs in function of the FDR threshold. (E)Identification overlap showing PSMs, and peptides removed, retained and gained by the rescoring engine.

Conclusion

- TIMSrescore represents a timsTOF-optimized rescoring approach that can improve recovery of peptide identifications
- TIMSrescore increased identifications in all tested datasets with an average increase of 6-20%.
- TIMSrescore makes use of all available dimensions of timsTOF data including fragmentation, retention time, and the TIMS dimension

Technology