Digging deeper into phosphoproteomes through AI-driven deconvolution of chimeric spectra

Florian Seefried, Siegfried Gessulat, Michael Graber, Vishal Sukumar, Samia Ben Fredj, Patroklos Samaras, Markus Schneider, Tobias Schmidt, Daniel P. Zolg, & Martin Frejno
[1]MSAID GmbH, Germany

+ Introduction

1 CHIMERYS™: An intelligent search algorithm for DDA data

Prediction-driven identification of peptides from experimental data

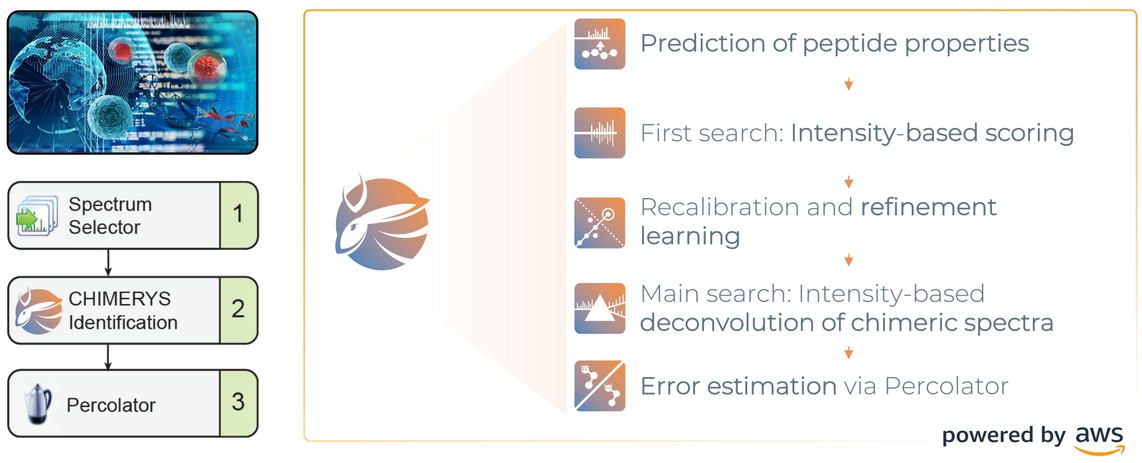

Figure 1 – CHIMERYS workflow in Thermo Scientific™ Proteome Discoverer™ 3.x software

2 CHIMERYS: tackling chimeric peptide tandem mass spectra

"One peptide per spectrum" rarely exists, even in DDA measurements

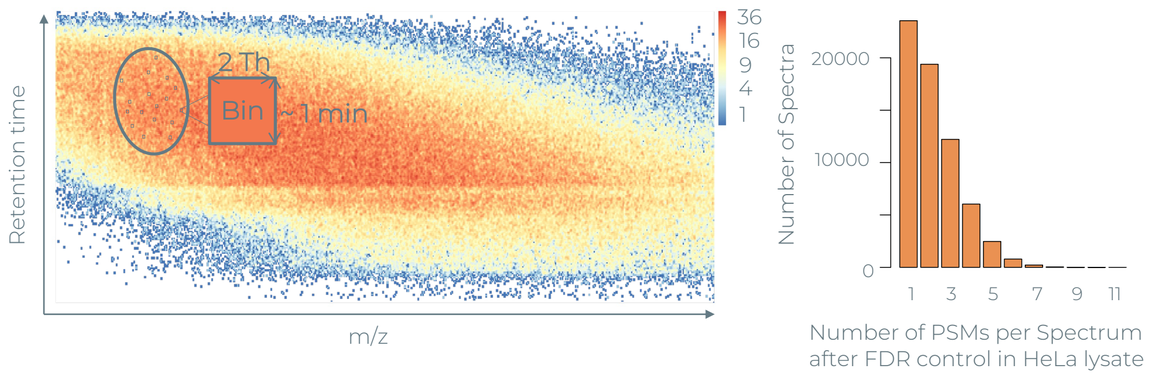

Figure 2 – a) Visualization of a human spectral library[1] **b)** Number of PSMs per spectrum of unmodified peptides in a complex HeLa cell lysate measured for 1h on a Orbitrap™ Q Exactive HFX mass spectrometer

+ Results

3 Extending INFERYS™ predictions to phosphorylation

Adding phosphorylated peptides to the training data to obtain an all-in-one model

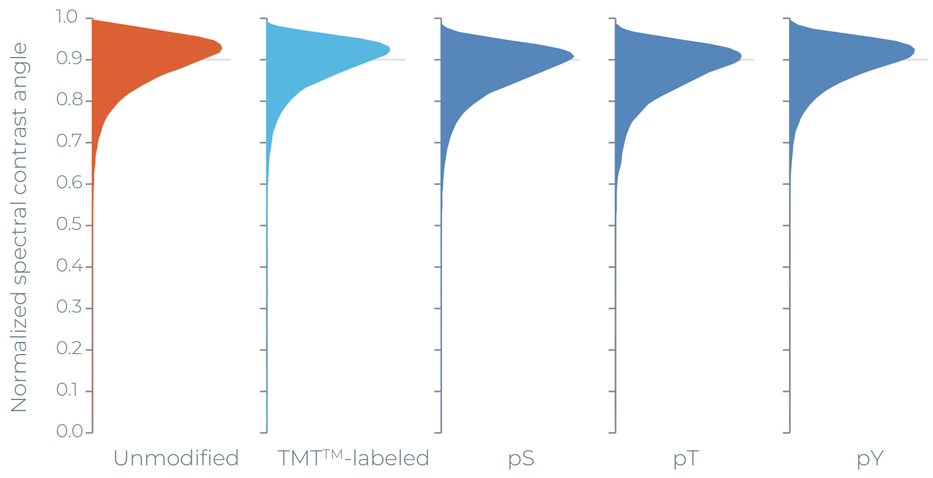

Figure 3 – Bean plots of spectral similarity between prediction and hold-out dataset

4 Retention time prediction via transfer learning

Transfer learning for accurate adaption of retention time prediction to measurement

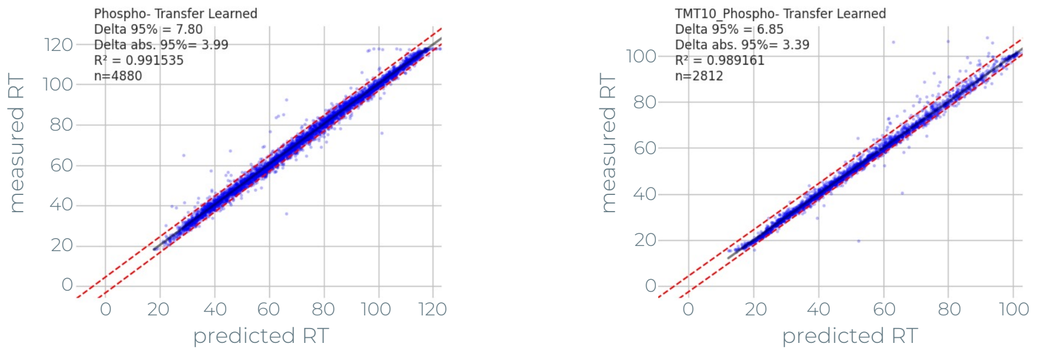

Figure 4 – a) Label-free phospho dataset **b)** TMT10-labeled phospho dataset

5 Application of CHIMERYS 2.0-beta to phosphoproteomics

CHIMERYS proportionally increases identifications for pSTY phosphorylated peptides

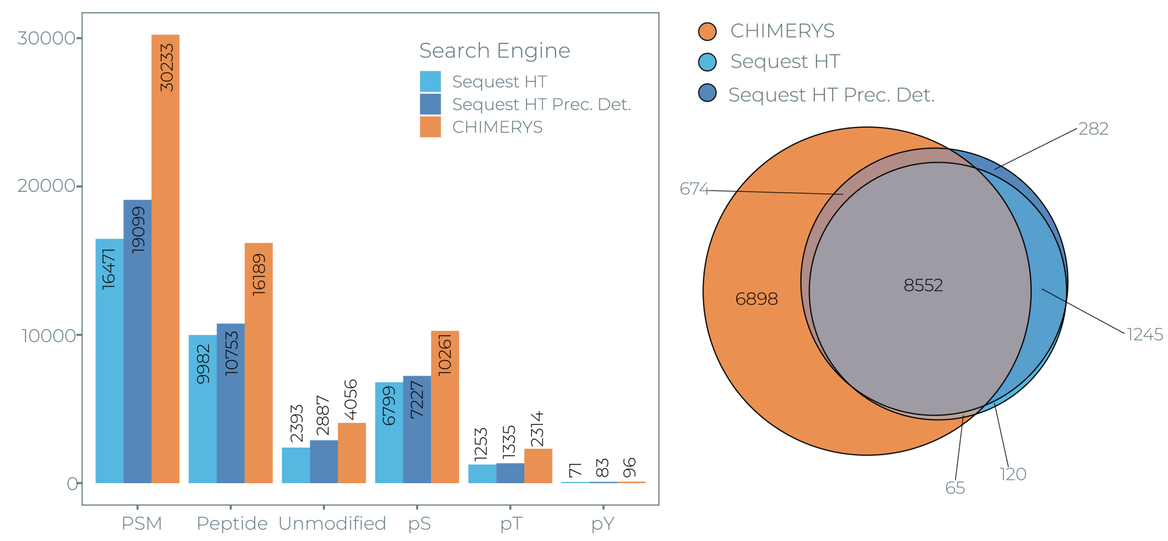

Figure 5 – a) Comparison of identifications from different algorithms on a phospho-enriched one-shot sample measured for 1h on an Orbitrap Q Exactive HF[2] **b)** Venn diagram of phosphorylated peptide sequences

6 High spectral similarity for phosphorylated peptide species

Spectral angle distributions of predicted vs. identified spectra is slightly downshifted from unmodified analytes; extrapolation to multiply phosphorylated peptides feasible

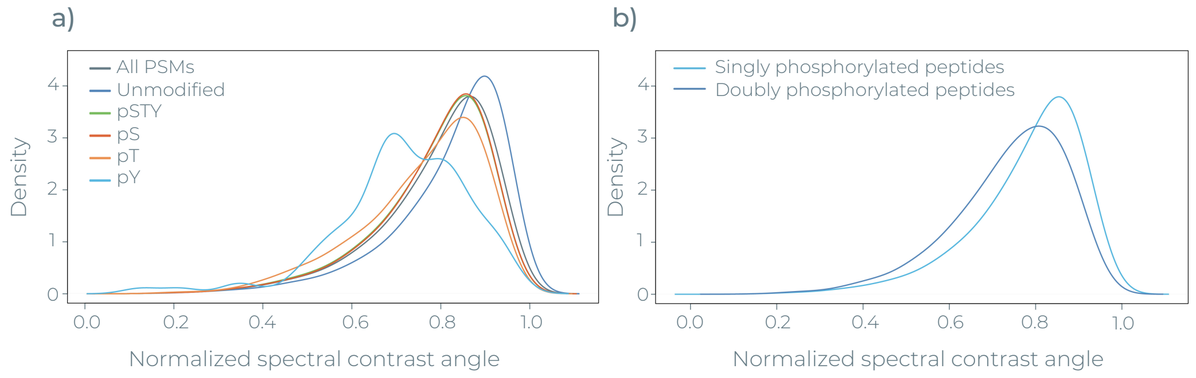

Figure 6 – a) Similarity by modified residue **b)** Similarity by number of phospho sites

7 Sequence motifs of exclusive identifications corroborate IDs

Proline-directed phosphorylation of serine residues detected in all peptide subsets

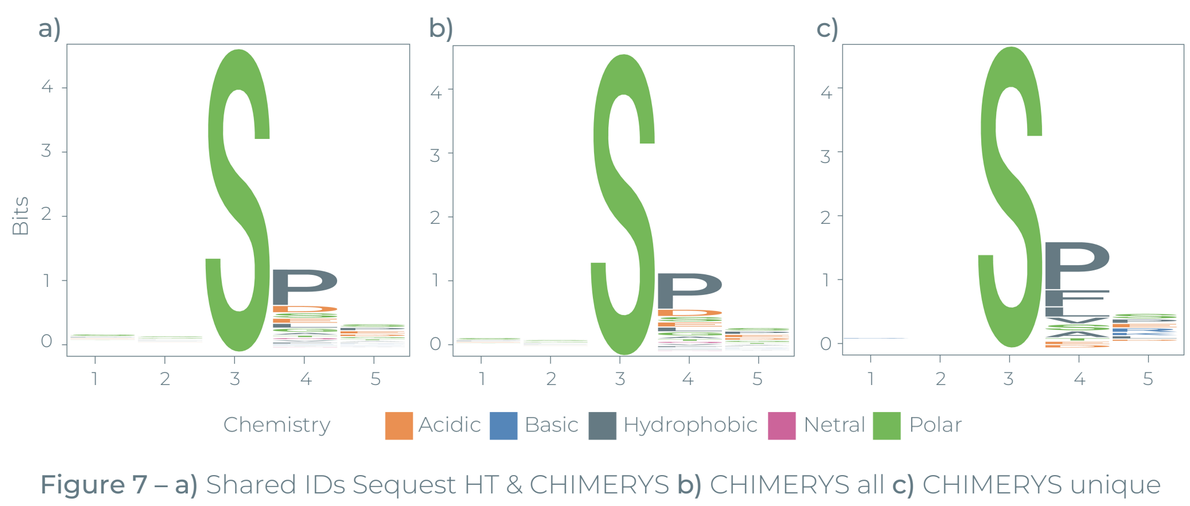

Figure 7 – a) Shared IDs Sequest HT & CHIMERYS **b)** CHIMERYS all **c)** CHIMERYS unique

+ Conclusions

- CHIMERYS is an intelligent search algorithm tackling chimeric spectra in DDA data
- A development version of the machine learning model INFERYS was extended to accurately predict properties of phosphorylated peptides
- A development version of CHIMERYS identifies more phosphorylated PSMs, peptides and individual sites than Sequest HT
- Spectral similarity and logo plot similarity corroborate the validity of identifications
- Additional efforts will tackle PTM localization and test new INFERYS models

+ Related Content

Poster PP02.91 (Tuesday, December 6th | AM Session 11:45 AM - 1:00 PM):

A unifying, spectrum-centric approach for the analysis of peptide tandem mass spectra

MSAID BOOTH #211 | https://www.msaid.de/conferences/hupo2022

References
[1] Zhu, T et. al, 2020. DPHL: A DIA Pan-human Protein Mass Spectrometry Library for Robust Biomarker Discovery
[2] Frejno, M. et al, 2020, Proteome activity landscapes of tumor cell lines determine drug responses