

An AI-driven leap forward in peptide identification through the deconvolution of chimeric spectra



M. Frejino¹, D.P. Zolg¹, T. Schmidt¹, S. Gessulat¹, M. Graber¹, F. Seefried¹, M. Rathke-Kuhnert¹, S. Ben Fredj¹, P. Samaras¹, K. Fritzscheier², F. Berg², W. Nasir², D. Horn³, B. Delanghe², C. Henrich², B. Kuster⁴, M. Wilhelm⁴
¹MSAID GmbH, Germany, ²Thermo Fisher Scientific (Bremen) GmbH, Germany, ³Thermo Fisher Scientific, USA, ⁴Technical University of Munich, Germany

+ Introduction

1 CHIMERYS: tackling chimeric peptide tandem mass spectra

"One peptide per spectrum" rarely exists, either by chance (DDA) or by design (DIA)

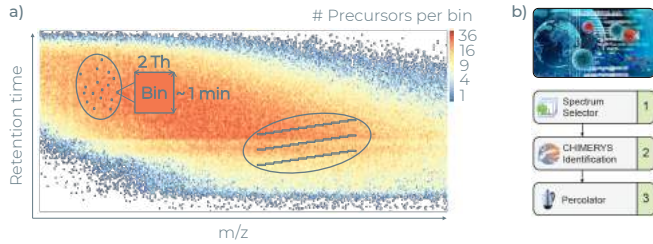


Figure 1 – a) Visualization of a human spectral library b) CHIMERYS workflow in PD 3.0

+ Results

2 Deconvolution of chimeric spectra doubles peptide identifications

CHIMERYS IDs concur with various workflows and drastically increase PSMs

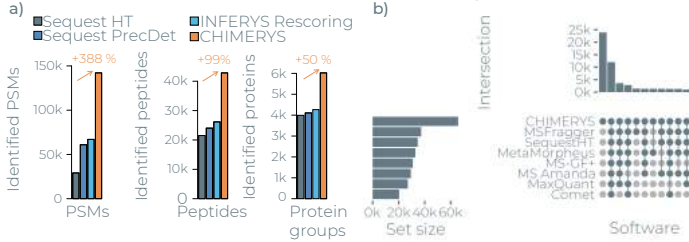


Figure 2 – a) CHIMERYS vs Sequest variants b) CHIMERYS vs open-source software

3 CHIMERYS excels both for deep-proteomes and fast gradients

Fractionated samples & high-throughput applications profit from ML-aided scoring

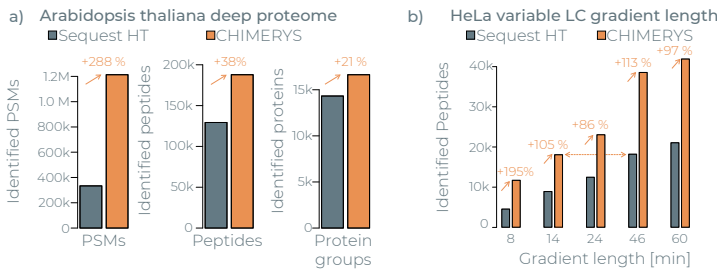


Figure 3 – a) Deeply fractionated dataset b) Improvement of IDs over gradient length

4 Entrapment experiments with wide-window DDA data

1x human & 9x human shuffled fasta: accurate FDR estimation in wide-window DDA

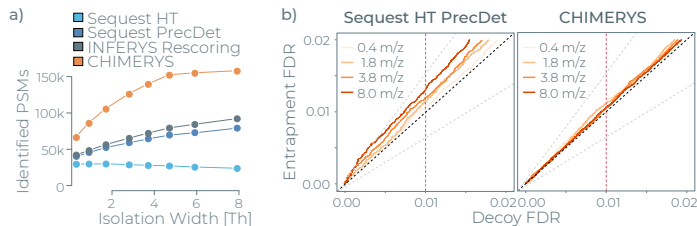


Figure 4 – a) Increasing isolation windows b) Entrapment on wide-window DDA data

References
 Zhu, T et al, 2020 | Van Puyvelde, B. et al, 2022

CHIMERYS®, INFERYS® and MSAID® are registered trademarks of MSAID GmbH. Thermo Scientific™ Proteome Discoverer™ software are trademarks of Thermo Fisher Scientific Inc. Sequest™ is a trademark of the University of Washington

5 Comparison to peptide-centric quantification on DIA data

Different scoring approaches show correlation of quantification

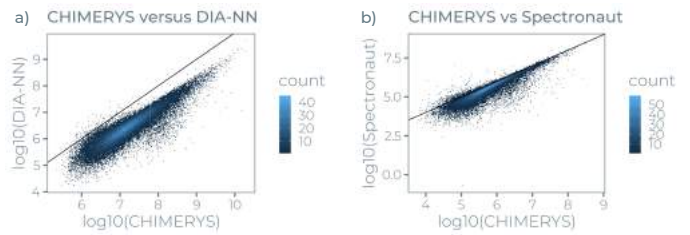


Figure 5 – a) Comparison to DIA-NN b) Comparison to Spectronaut 16

6 High overlap of IDs and accurate quantification of DIA data

Different scoring approaches show overlap of IDs & CHIMERYS quantifies accurately

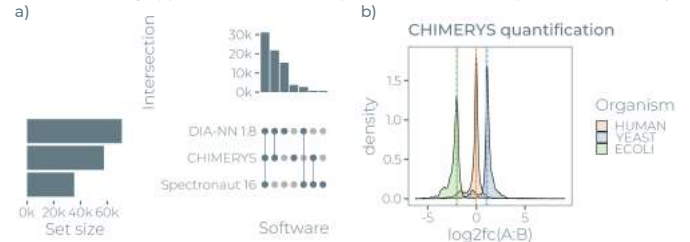


Figure 6 – a) CHIMERYS vs other DIA software b) Recovery of known ratios

7 Aggregated, coefficient-based quantification matches Skyline

Automated scoring is correlated to expert-performed, manual quantification of XICs

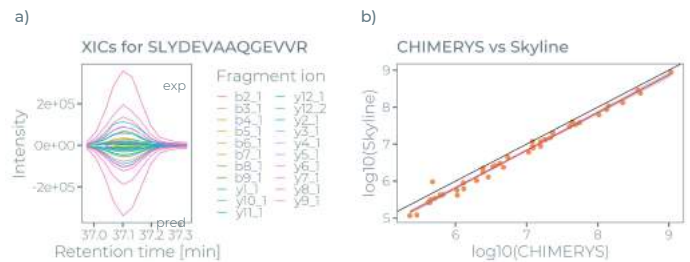


Figure 7 – a) Mirror XICs validate IDs b) High correlation over 4.5 orders of magnitude

+ Conclusions

- + Deep-learning based predictions enable systematic spectrum-centric data analysis
- + A novel deconvolution algorithm yields striking results for chimeric DDA spectra
- + Deconvolution concept can be translated from wide-window DDA to DIA and PRM
- + Coefficient-based quantification matches current MS2-based quantification
- + We predict a bright future for ML-based applications in proteomics

+ Related Content

Talk: An AI-driven leap forward in peptide identification through deconvolution of chimeric spectra – Monday, August 15th – Afternoon session

<https://www.msaid.de/conferences/asbmb2022>