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

Martin. Frejno[1], Daniel P. Zolg[1], Tobias Schmidt[1], Siegfried Gessulat[1], Michael Graber[1], Florian Seefried[1], Magnus Rathke-Kuhnert[1], Samia Ben Fredj[1], Patroklos Samaras[1], Kai Fritzemeier[2], Frank Berg[2], Waqas Nasir[2], David Horn[3], Bernard Delanghe[2], Christoph Henrich[2], Bernhard Kuster[4], Mathias Wilhelm[4]
*[1]MSAID GmbH, Germany, [2]Thermo Fisher Scientific (Bremen) GmbH, Germany, [3]Thermo Fisher Scientific, USA, [4]Technical 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)

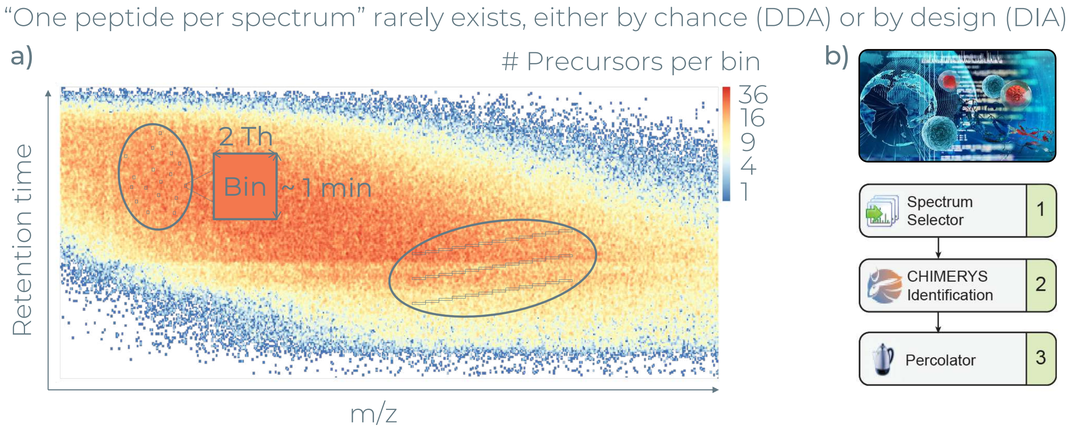


**Figure 1 – a)** Visualization of a human spectral library[1] **b)** CHIMERYS workflow in PD 3.0

## + Results

### 2 Deconvolution of chimeric spectra doubles peptide identifications

CHIMERYS IDs concur with various workflows and drastically increase PSMs

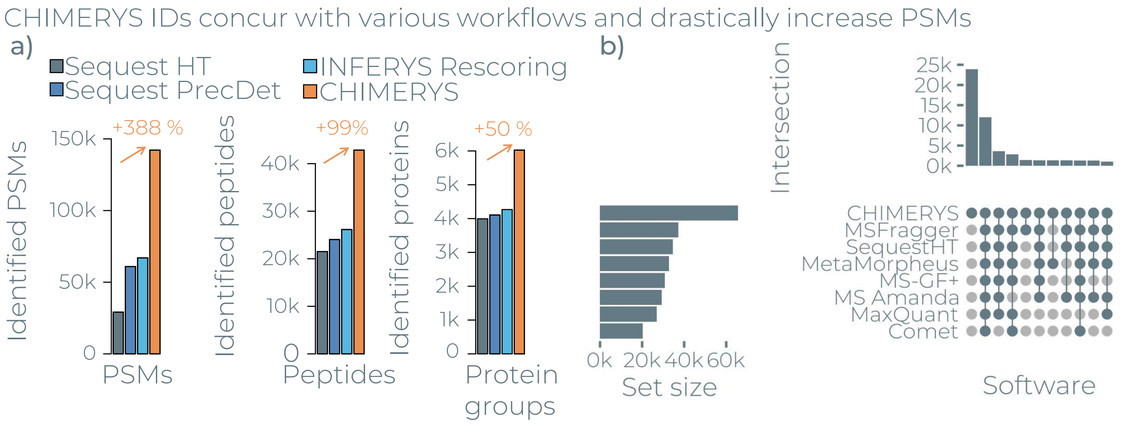


**Figure 2 – a)** CHIMERYS vs Sequest variants **b)** CHIMERYS vs open-source software

### 3 CHIMERYS excels both for deep-proteomes and fast gradients

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

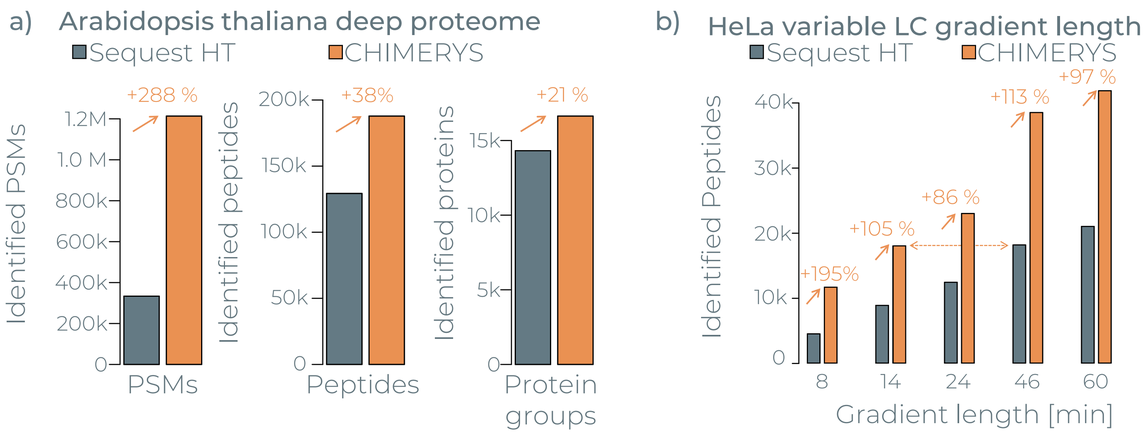


**Figure 3 – a)** Deeply fractionated dataset **b)** Improvement of IDs over gradient length

### 4 Entrapment experiments with wide-window DDA data

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

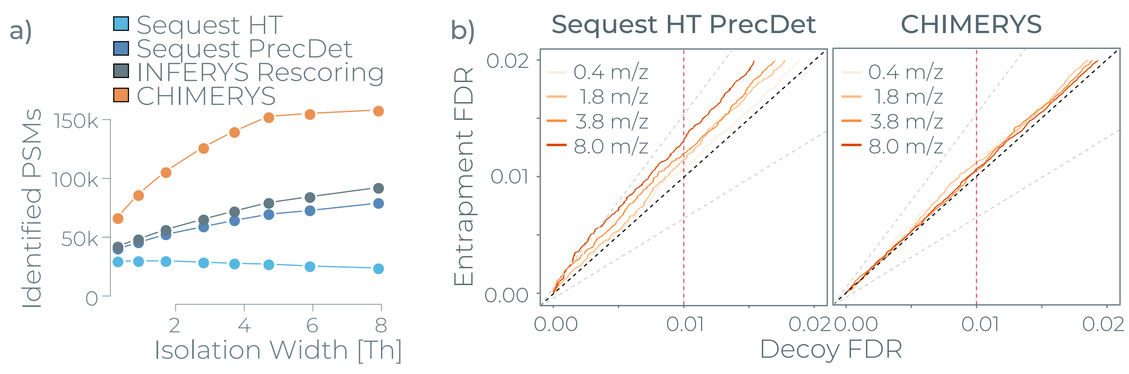


**Figure 4 – a)** Increasing isolation windows **b)** Entrapment on wide-window DDA data

### 5 Comparison to peptide-centric quantification on DIA data[2]

Different scoring approaches show correlation of quantification

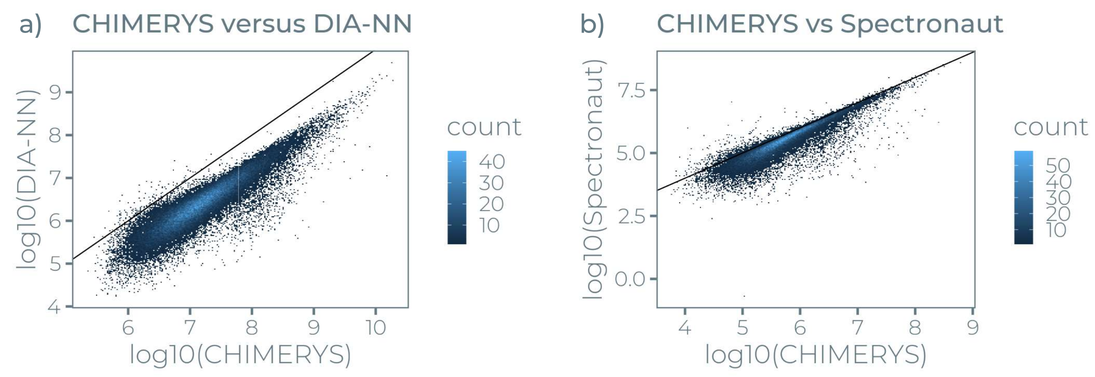


**Figure 5 – a)** Comparison to DIA-NN **b)** Comparison to Spectronaut 16

### 6 High overlap of IDs and accurate quantification of DIA data

Different scoring approaches show overlap of IDs & CHIMERYS quantifies accurately

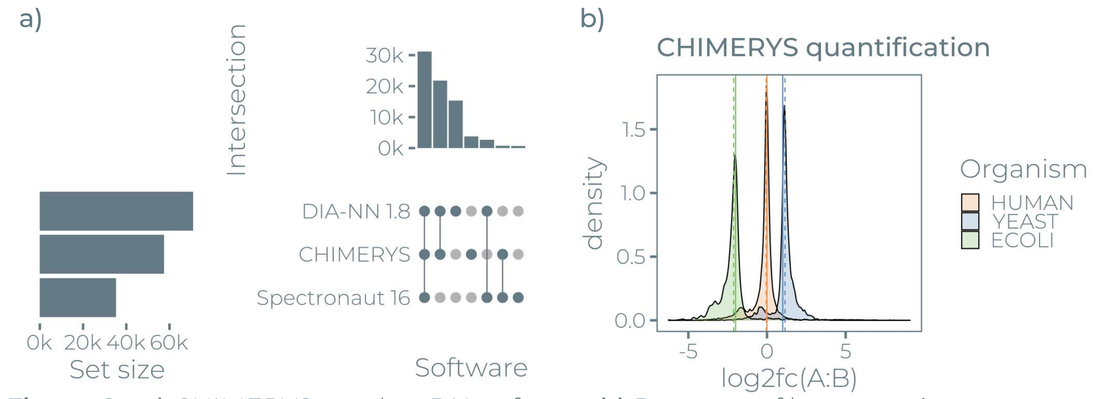


**Figure 6 – a)** CHIMERYS vs other DIA software **b)** Recovery of known ratios

### 7 Aggregated, coefficient-based quantification matches Skyline

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

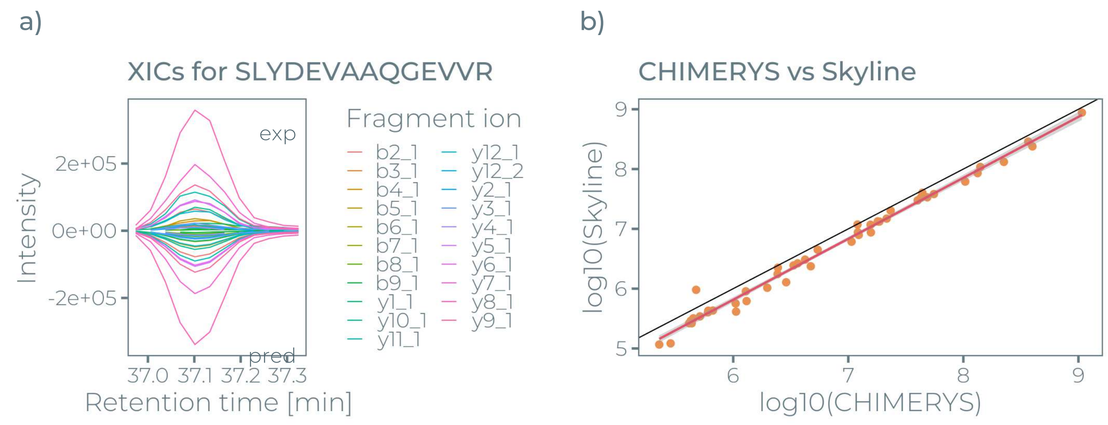


**Figure 3 – a)** Mirror XICs validate IDs **b)** High correlation over 4.5 orders of magnitude

## + 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

**Poster IM-PA-020:** An end-to-end machine learning workflow for MS-based proteomics (Abstract #536)
Poster Session A – Monday 29th and Tuesday 30th August 2022

https://www.msaid.de/conferences/imsc2022

References
[1] Zhu, T et. al, 2020. DPHL: A DIA Pan-human Protein Mass Spectrometry Library for Robust Biomarker Discovery
[2] Van Puyvelde, B. et al, 2022. A comprehensive LFQ benchmark dataset on modern day acquisition strategies in proteomics