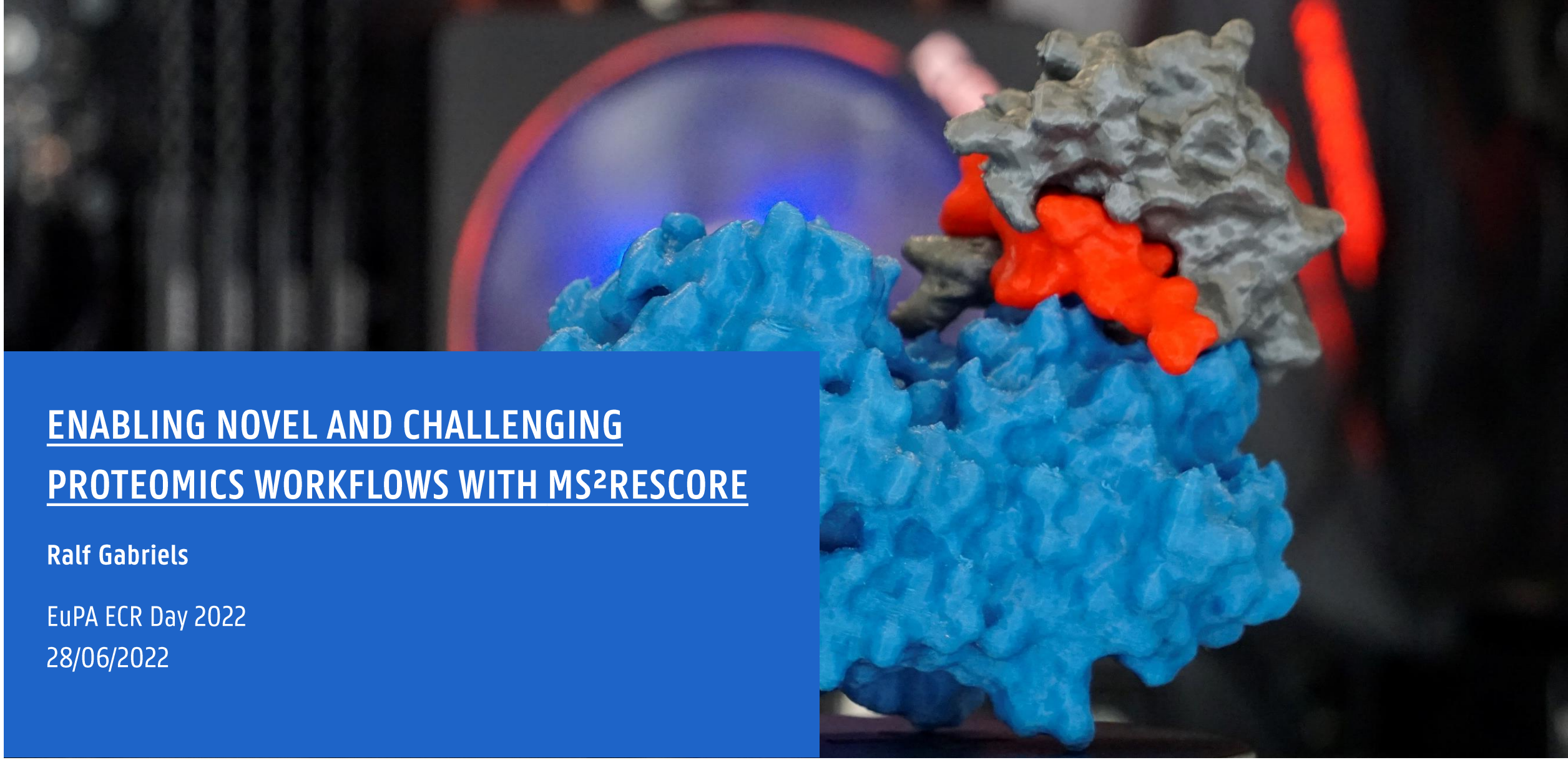


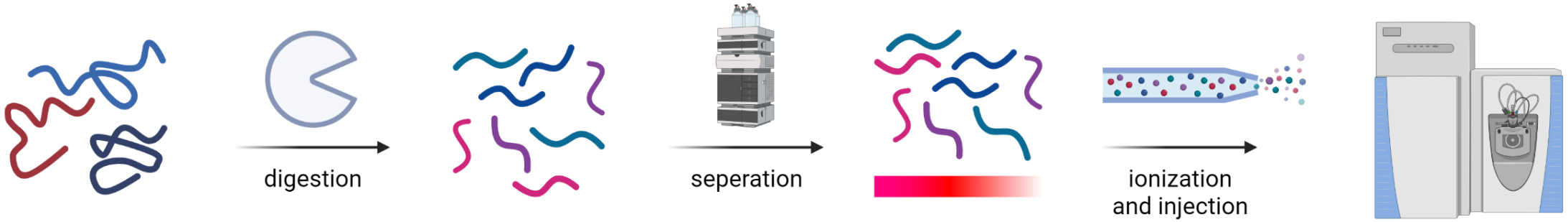
# ENABLING NOVEL AND CHALLENGING PROTEOMICS WORKFLOWS WITH MS²RESCORE

**Ralf Gabriels**

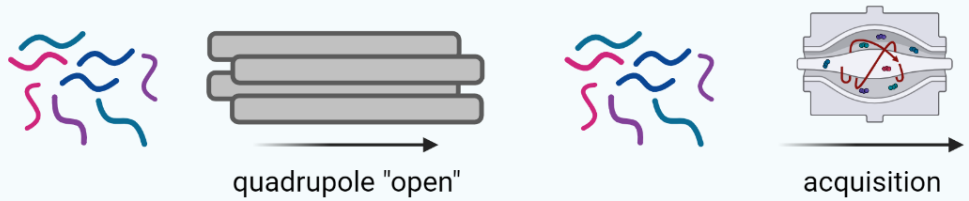
EuPA ECR Day 2022
28/06/2022

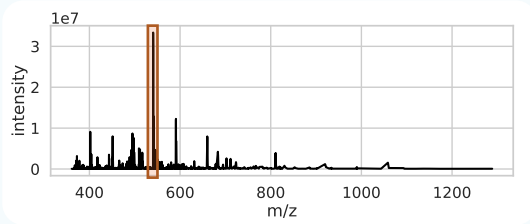
GHENT UNIVERSITY
FACULTY OF MEDICINE AND HEALTH SCIENCES
VIB-UGENT CENTER FOR MEDICAL BIOTECHNOLOGY
CompOmics
fwo

# Mass spectrometry-based proteomics recap

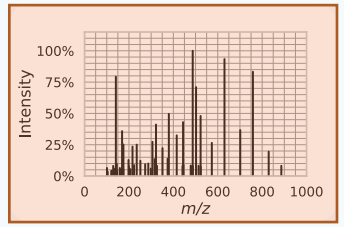



**MS1 survey scan:**
Find precursors (= full peptides)



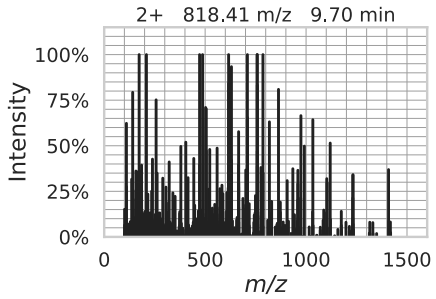



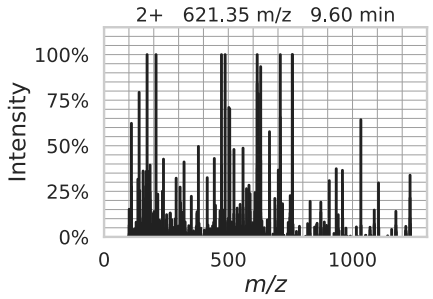

**MS2 scan:**
Get fragmentation spectrum for peptide at selected m/z window

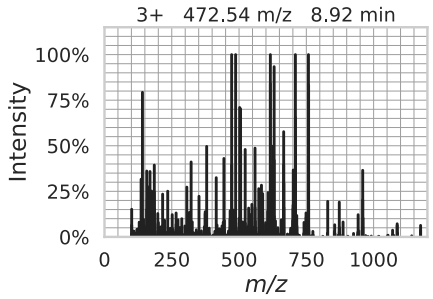


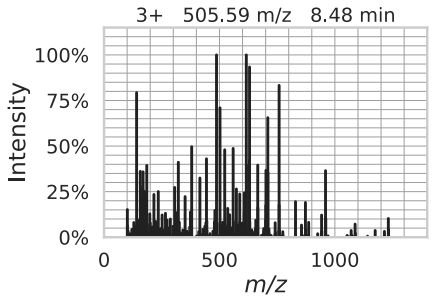

Figures created in BioRender.com

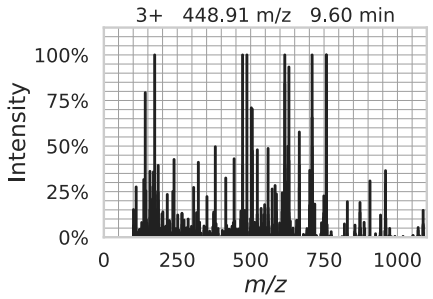
# How do we link MS2 spectra back to the original proteins?

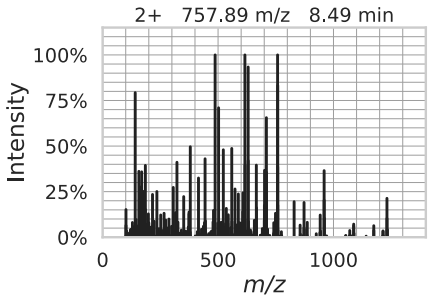


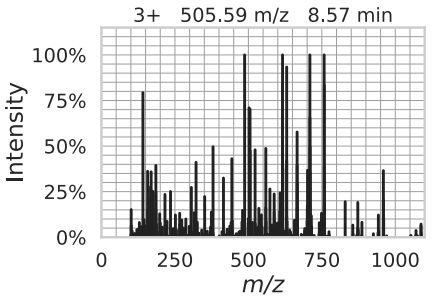



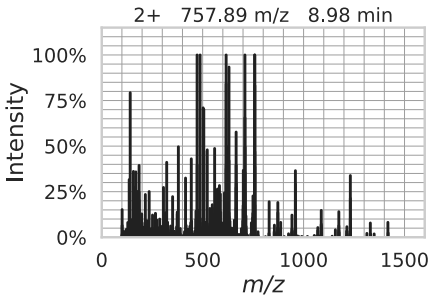



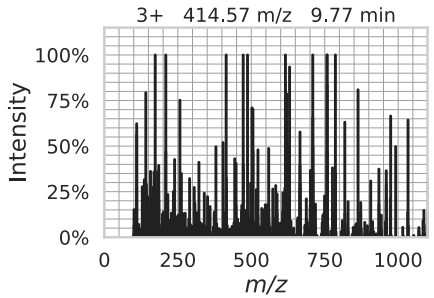



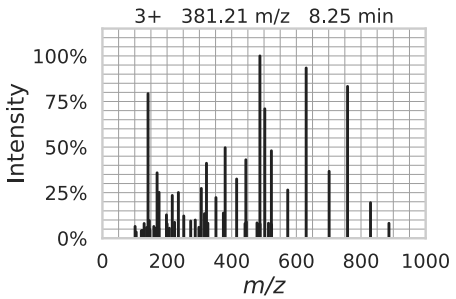



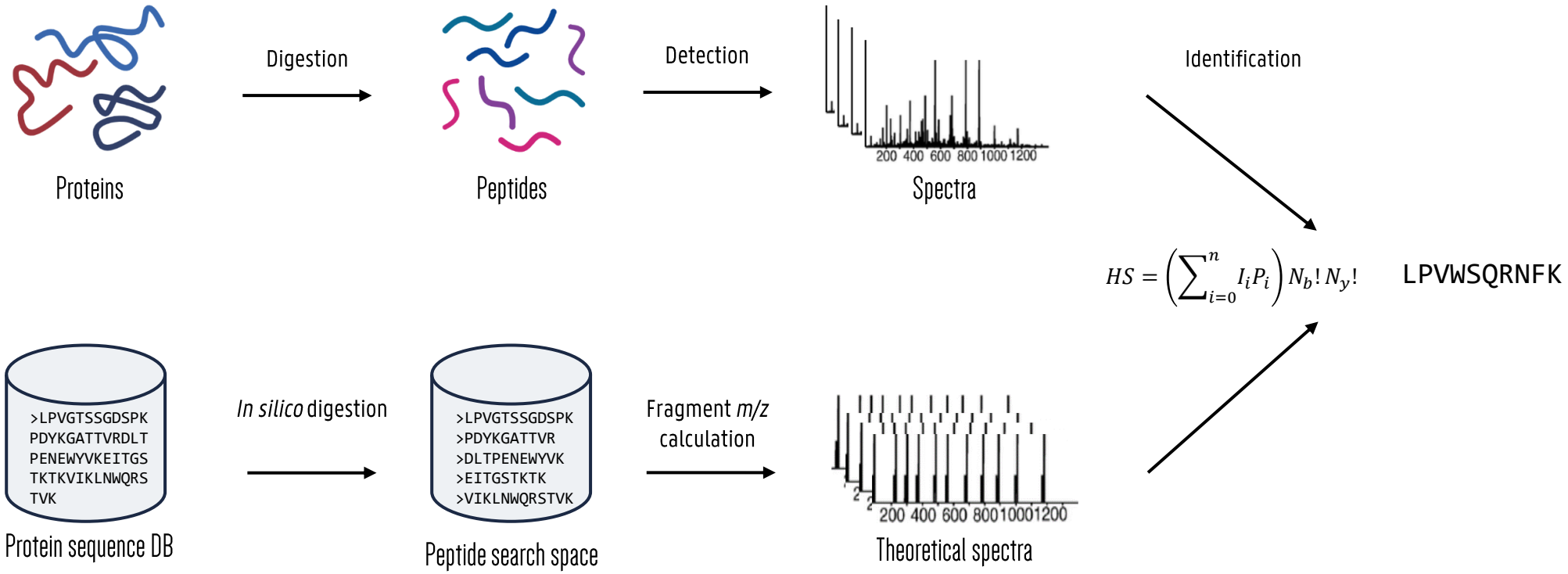



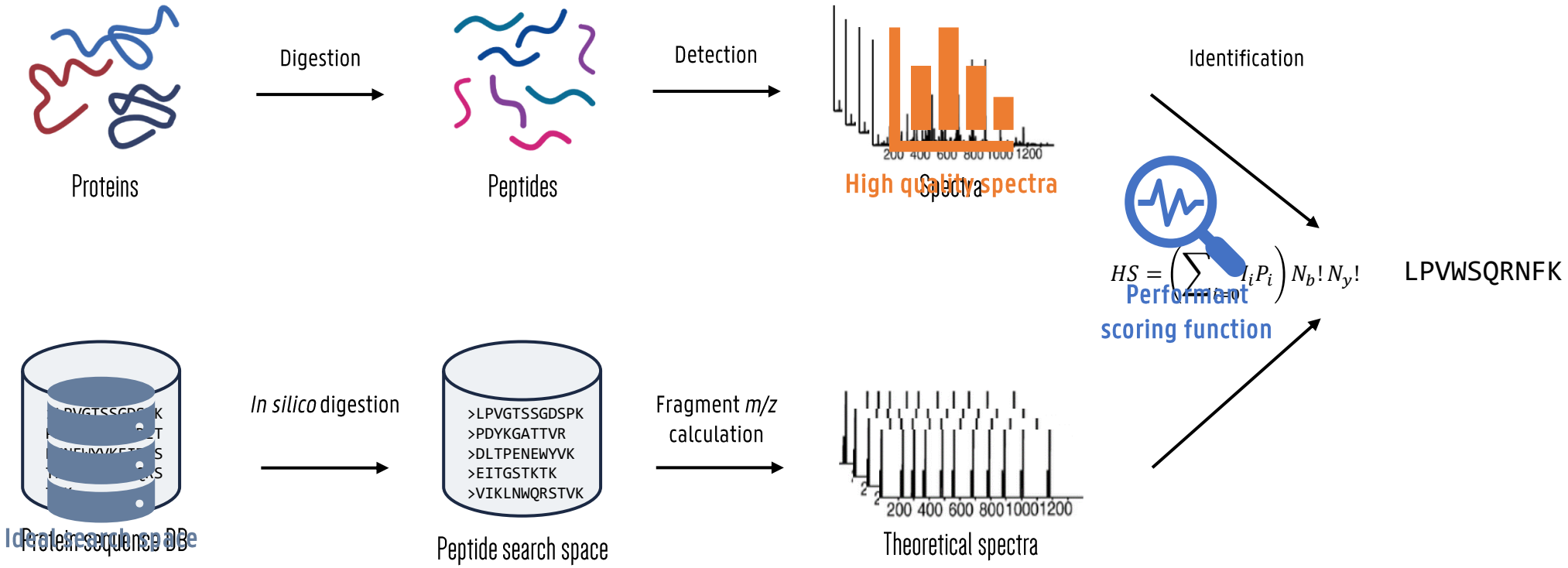



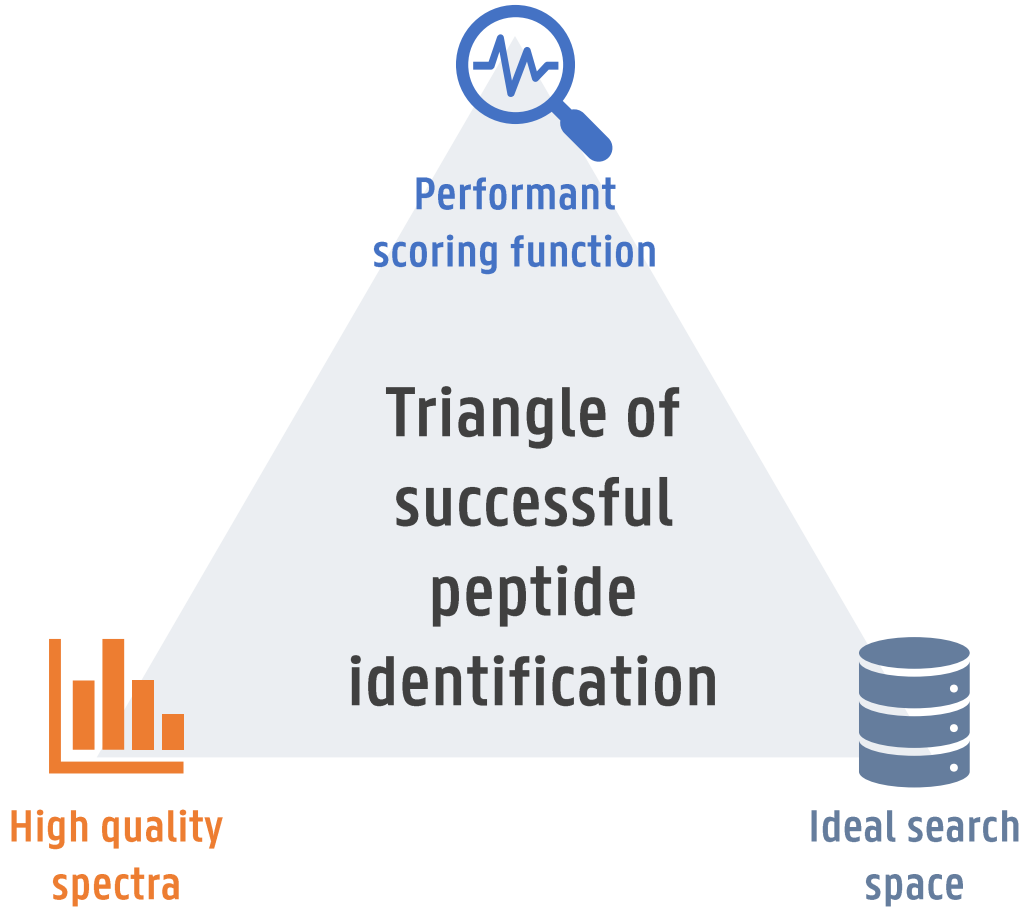
# Proteomics search engines replicate the LC-MS steps *in silico*



# Key elements to successful peptide spectrum identification?



# Key elements to successful peptide spectrum identification?



# Key elements to successful peptide spectrum identification

Triangle of successful peptide identification

DIA
Low abundance peptides
...

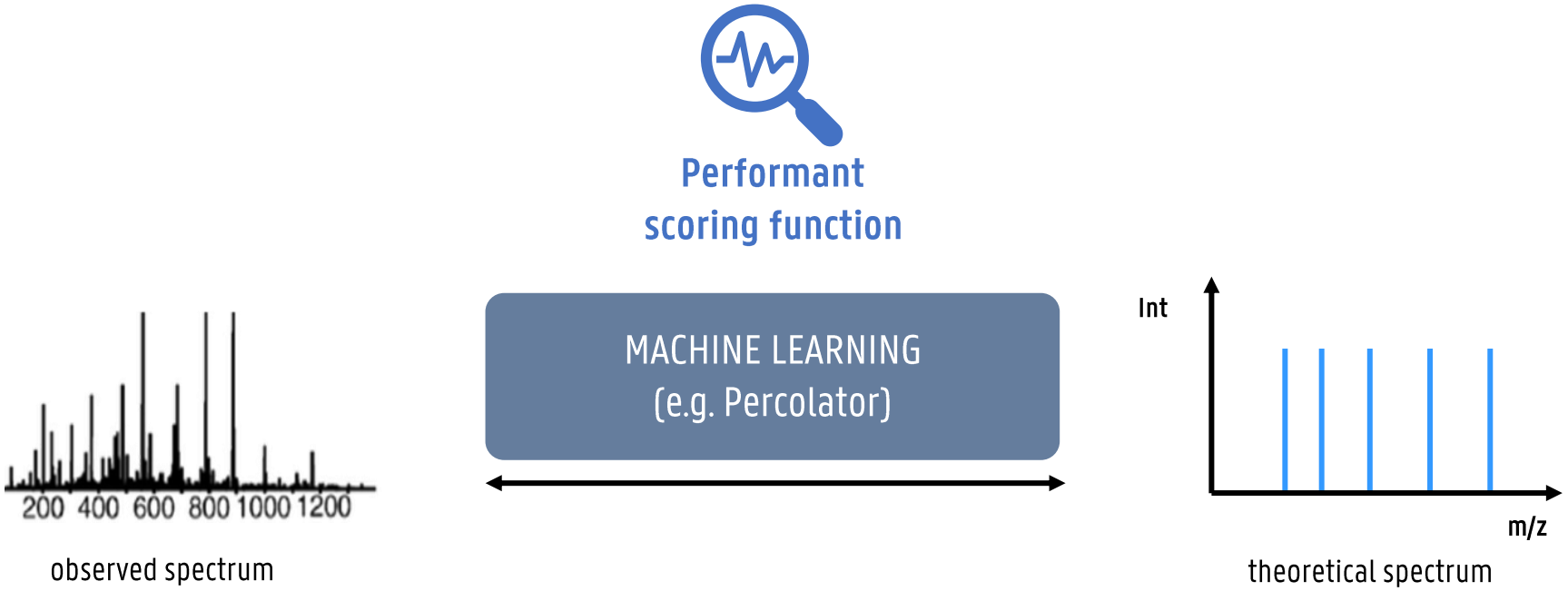
High quality spectra

Ideal search space

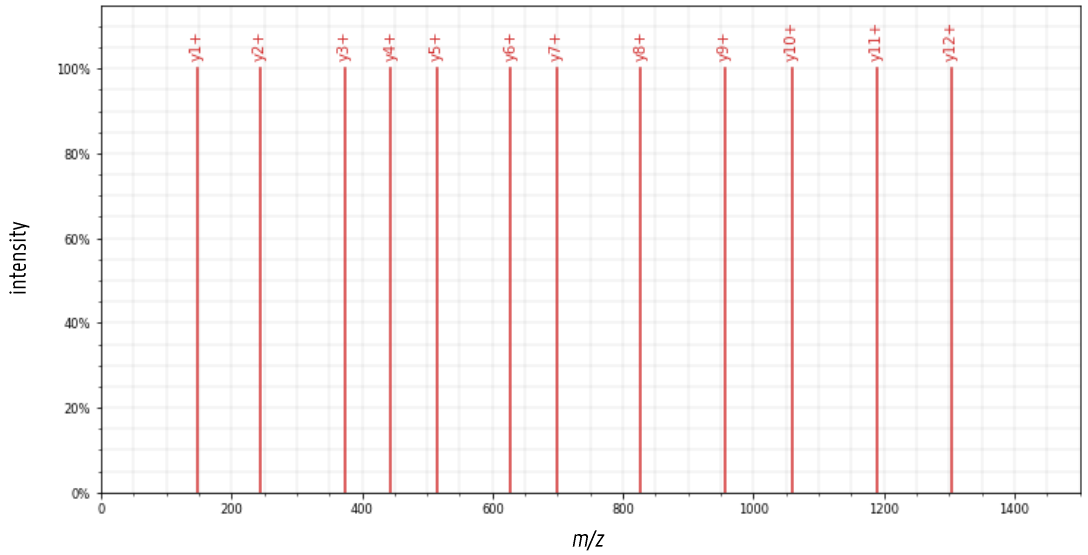
Proteogenomics
Metaproteomics
Immunopeptidomics
...

novel (but challenging) proteomics workflows

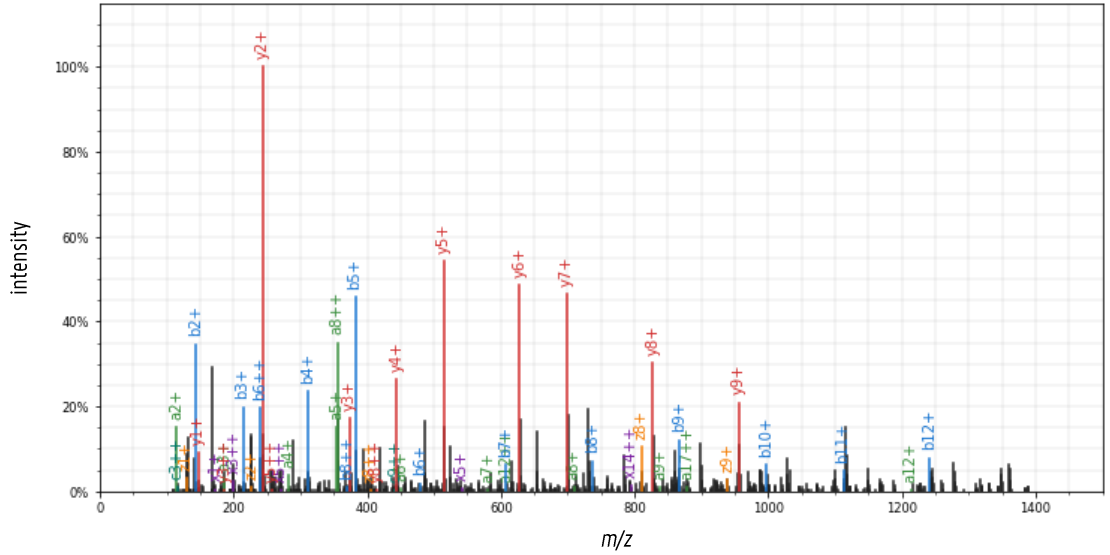
# Machine learning can generate a dynamically optimized scoring function, tailored to each data set



# MS²PIP brings a more realistic view on peptide fragmentation to PSM scoring

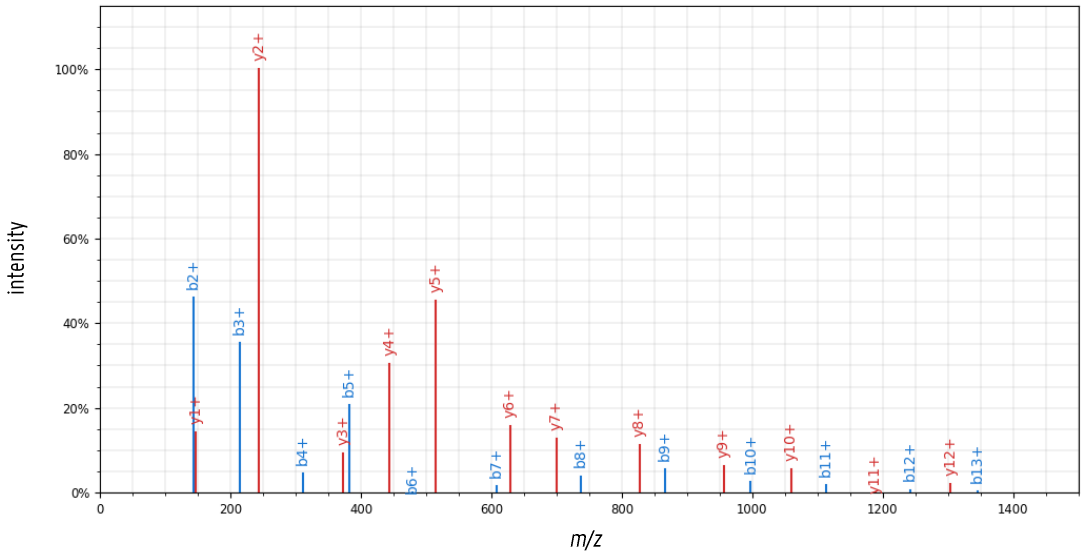



theoretical spectrum

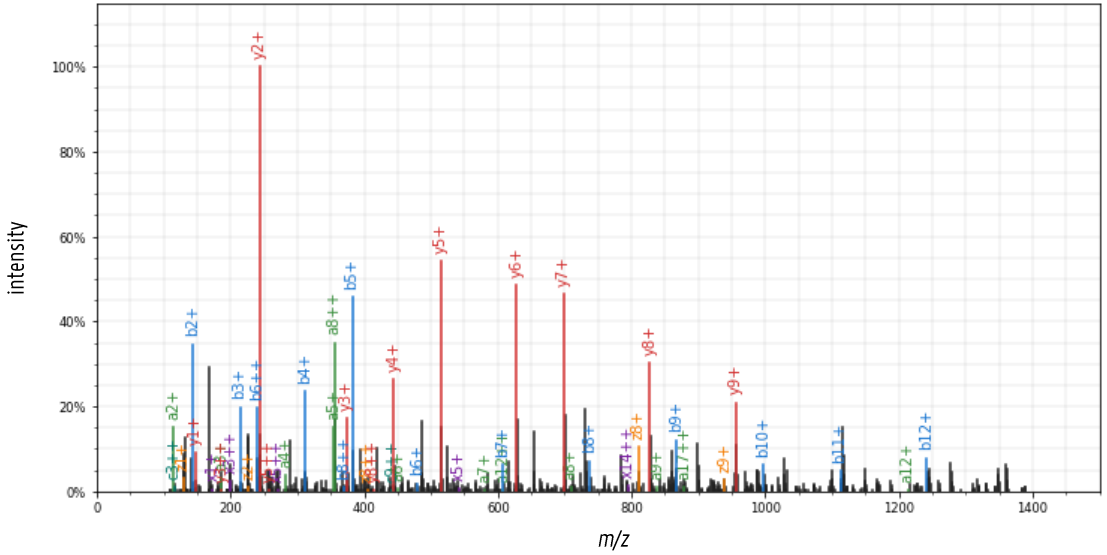



observed spectrum

# MS²PIP brings a more realistic view on peptide fragmentation to PSM scoring

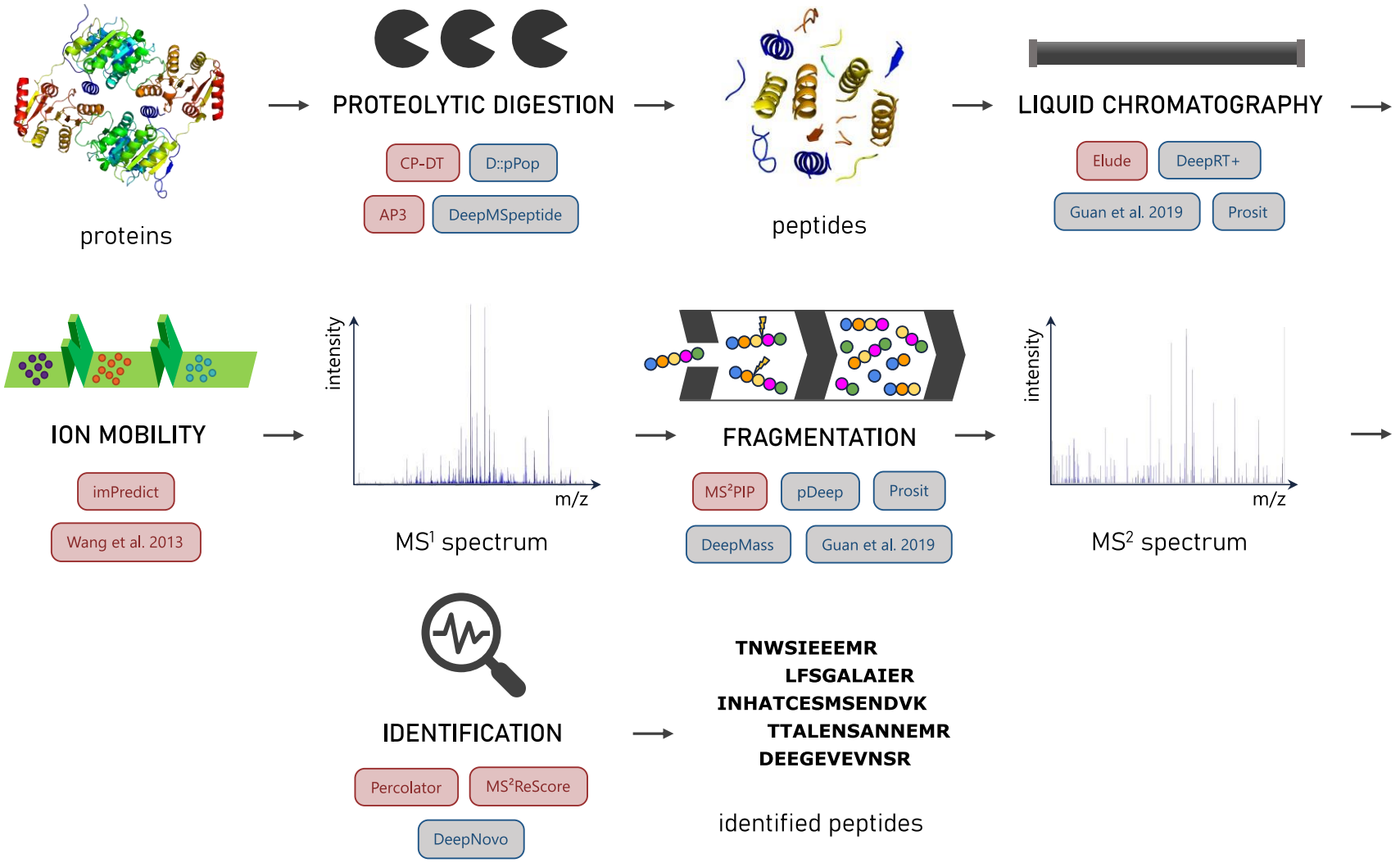



MS²PIP-predicted spectrum



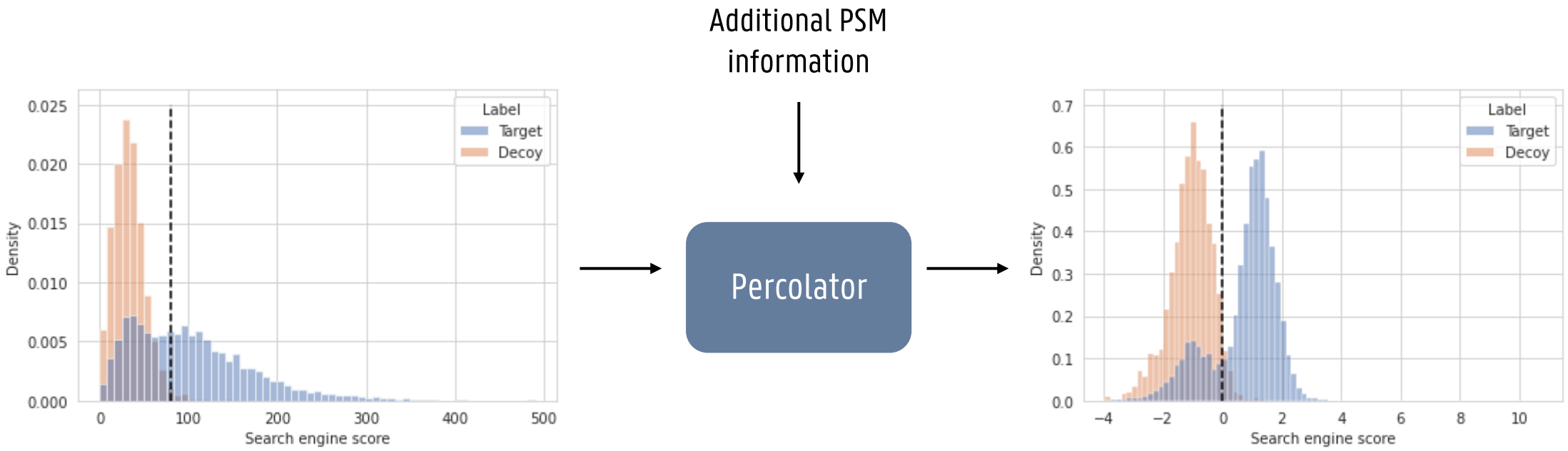

observed spectrum

# Nearly the whole LC-MS workflow can be modeled

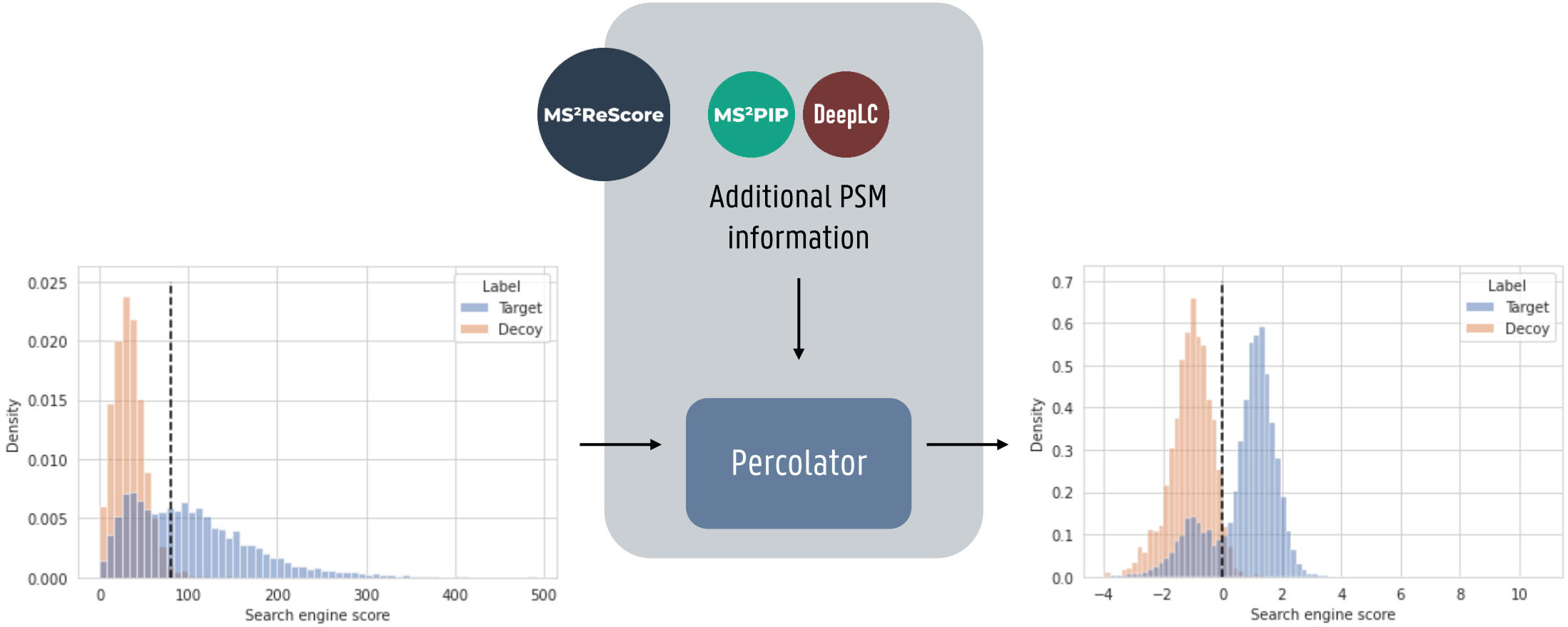



Ralf Gabriels* & Robbin Bouwmeester* et al. (2020) Proteomics. doi:10.1002/pmic.201900351

# Machine learning-based information can be passed to Percolator for improved rescoring



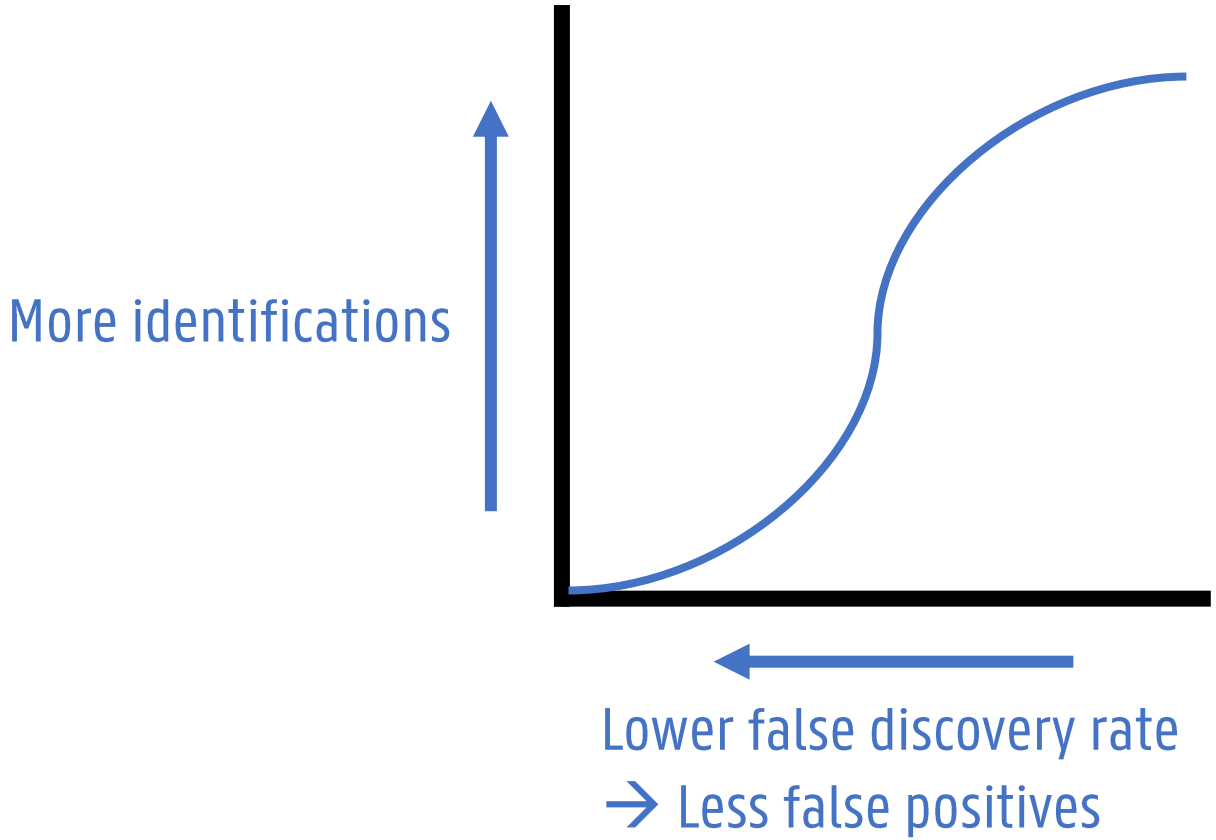
# Machine learning-based information can be passed to Percolator for improved rescoring



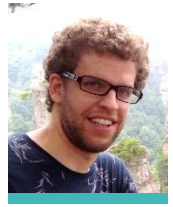


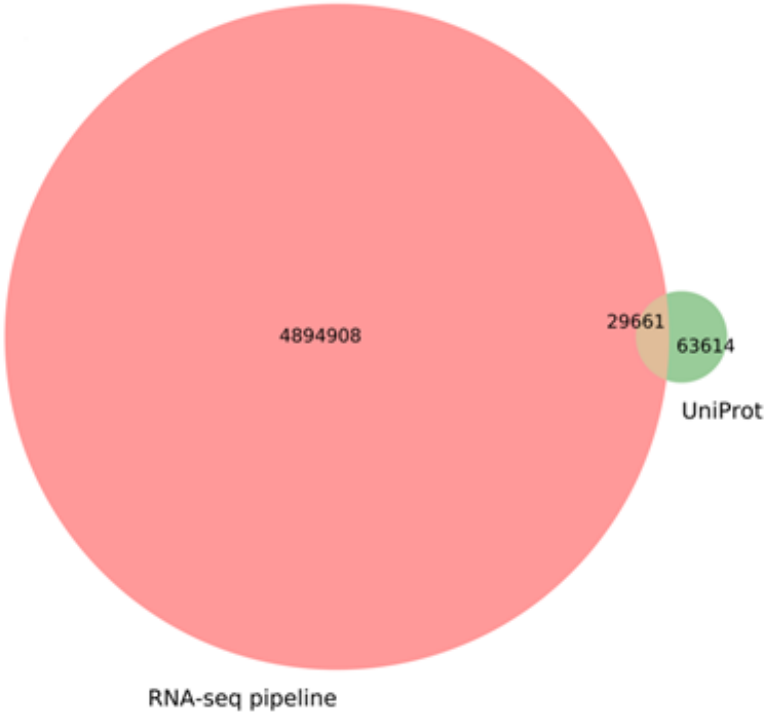
Rescoring with MS²PIP: Ana S C. Silva et al. (2019) Bioinformatics. doi:10.1093/bioinformatics/btz383
DeepLC: Robbin Bouwmeester et al. (2021) Nature Methods. doi:10.1038/s41592-021-01301-5
MS²Rescore: Arthur Declercq, [...] Ralf Gabriels (2021) bioRxiv. doi:10.1101/2021.11.02.466886

# Intermezzo: Explaining FDR/identification rate plots



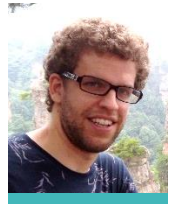
# MS²Rescore in proteogenomics: Higher ID rate, at a 10-fold lower FDR threshold

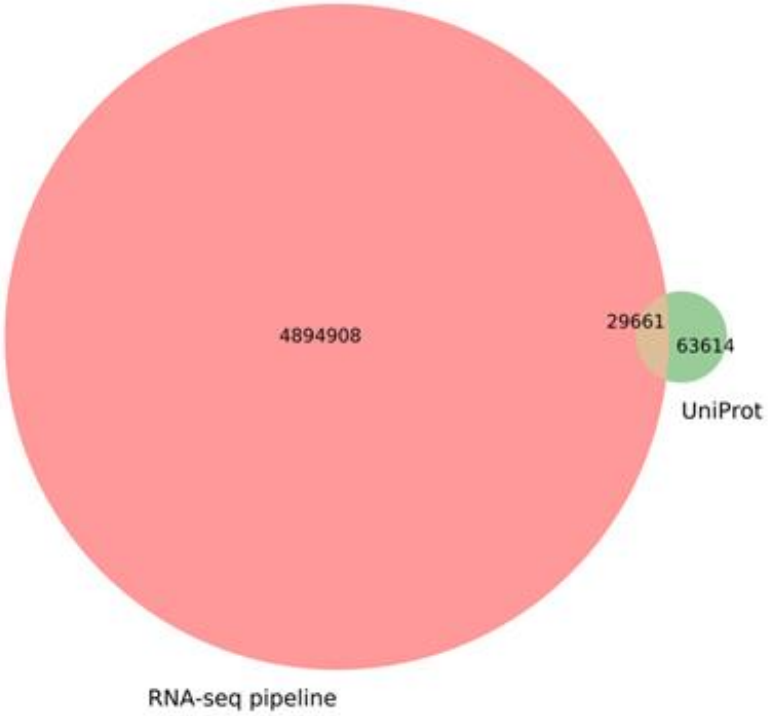


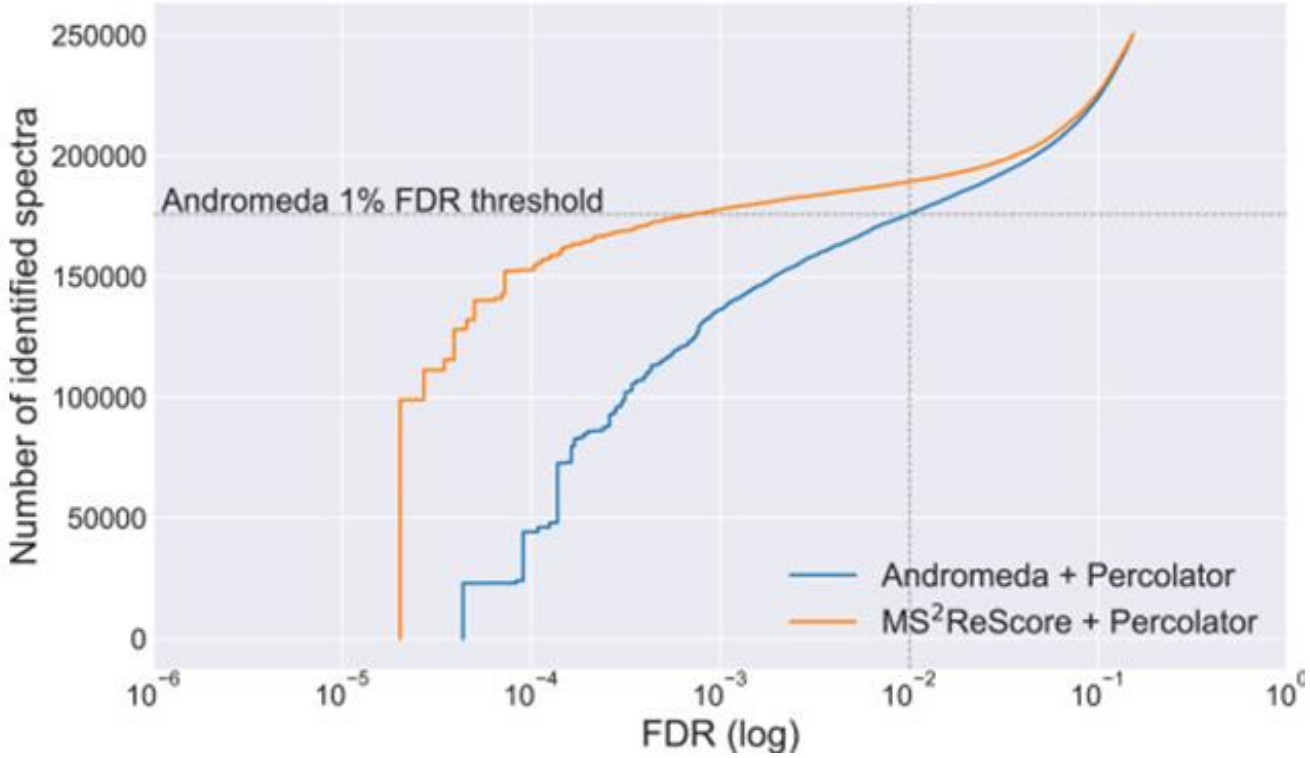



Nanopore RNA-seq:
- Extended search space
- Human HCT116 cells
- HCD orbitrap acquisition

# MS²Rescore in proteogenomics: Higher ID rate, at a 10-fold lower FDR threshold

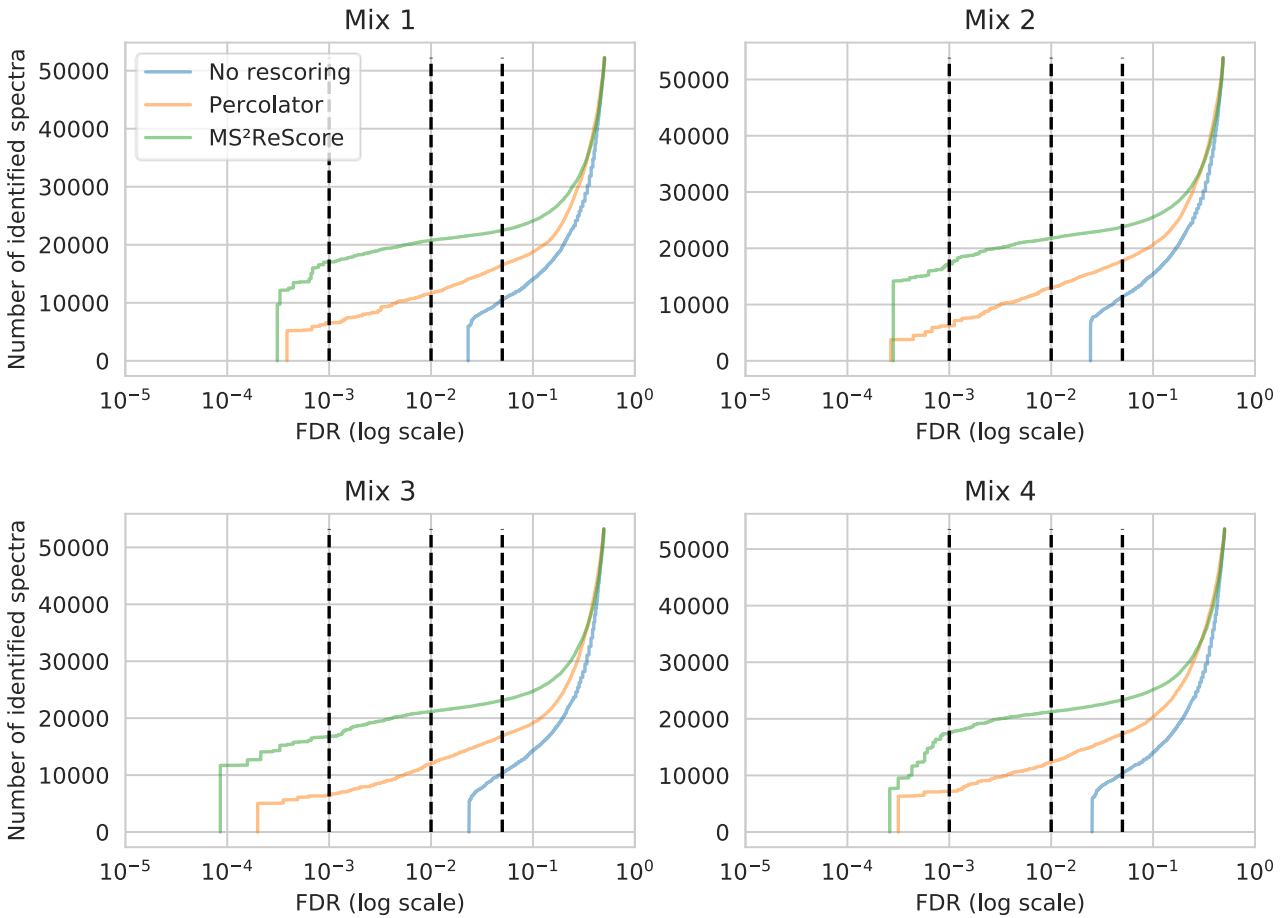





Steven Verbruggen (2021) MCP. doi:10.1016/j.mcpro.2021.100076

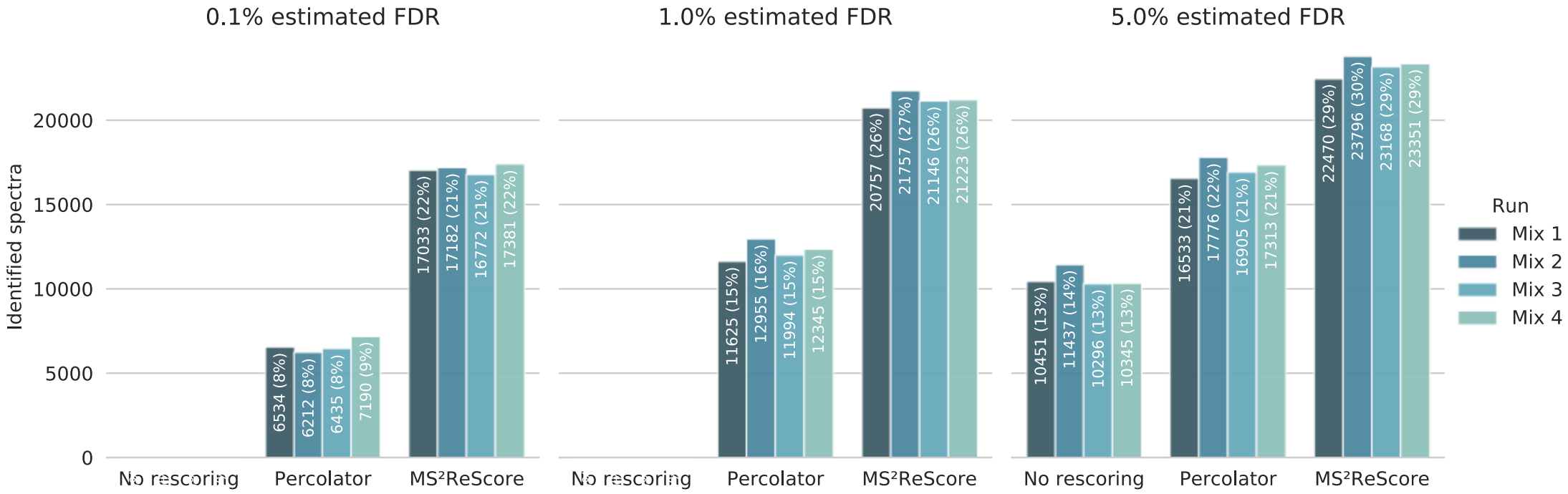
# MS²Rescore in metaproteomics: From 0 to 20 000 identified spectra



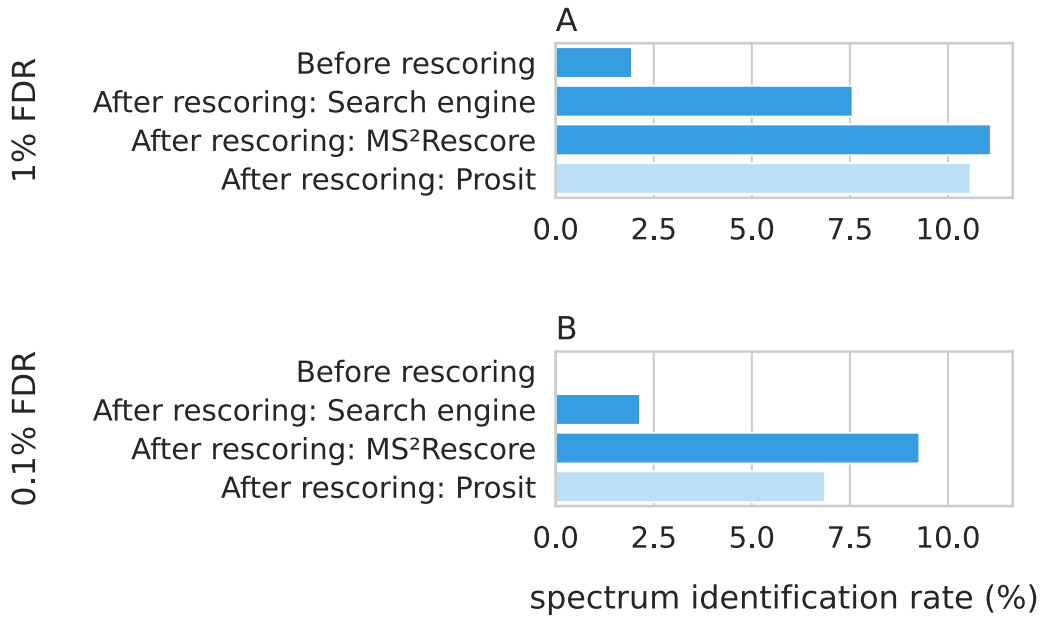

iPRG 2020 study:

- 4 microbial mixes, with unknown sequences
- Extremely large search space
- CID ion trap acquisition

# MS²Rescore in metaproteomics: From 0 to 20 000 identified spectra

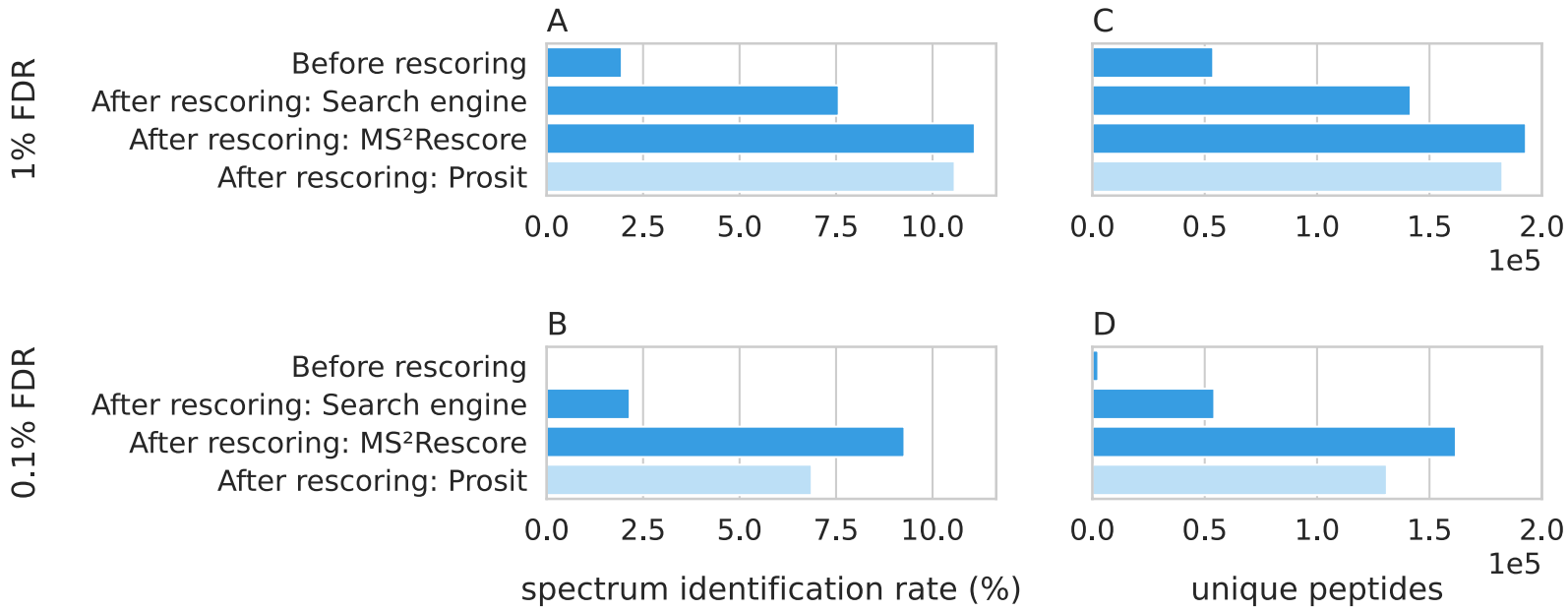


# MS²Rescore in immunopeptidomics: +46% spectrum identification rate

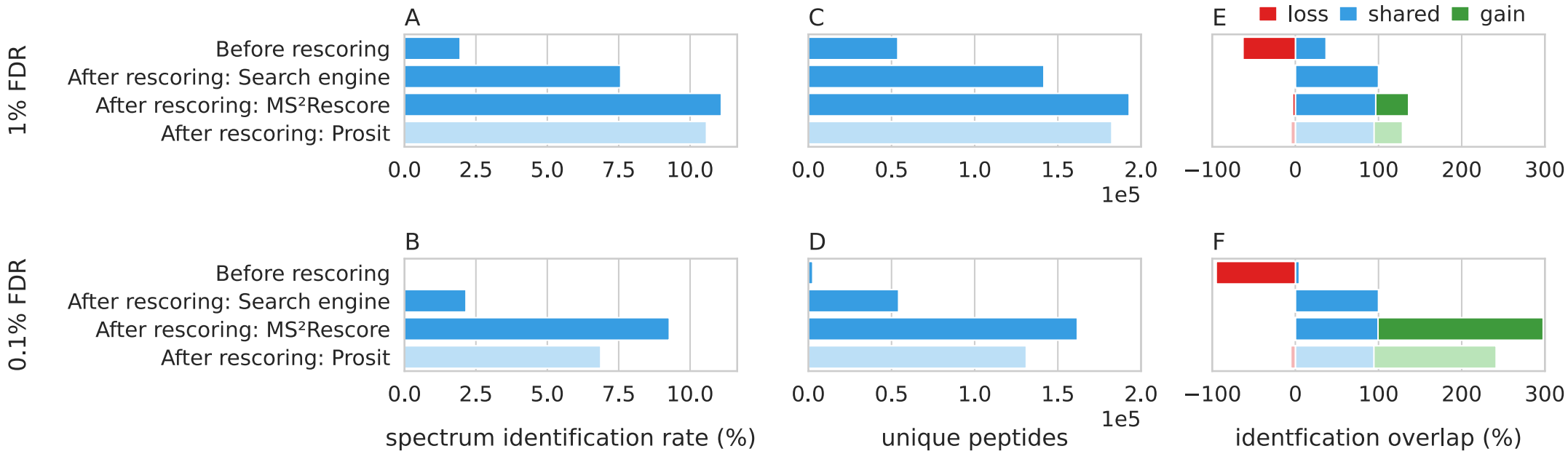



Immunopeptidomics:
- non-specific cleavage
- peptides with low ionization and fragmentation efficiency

# MS²Rescore in immunopeptidomics: +36% unique identified peptides

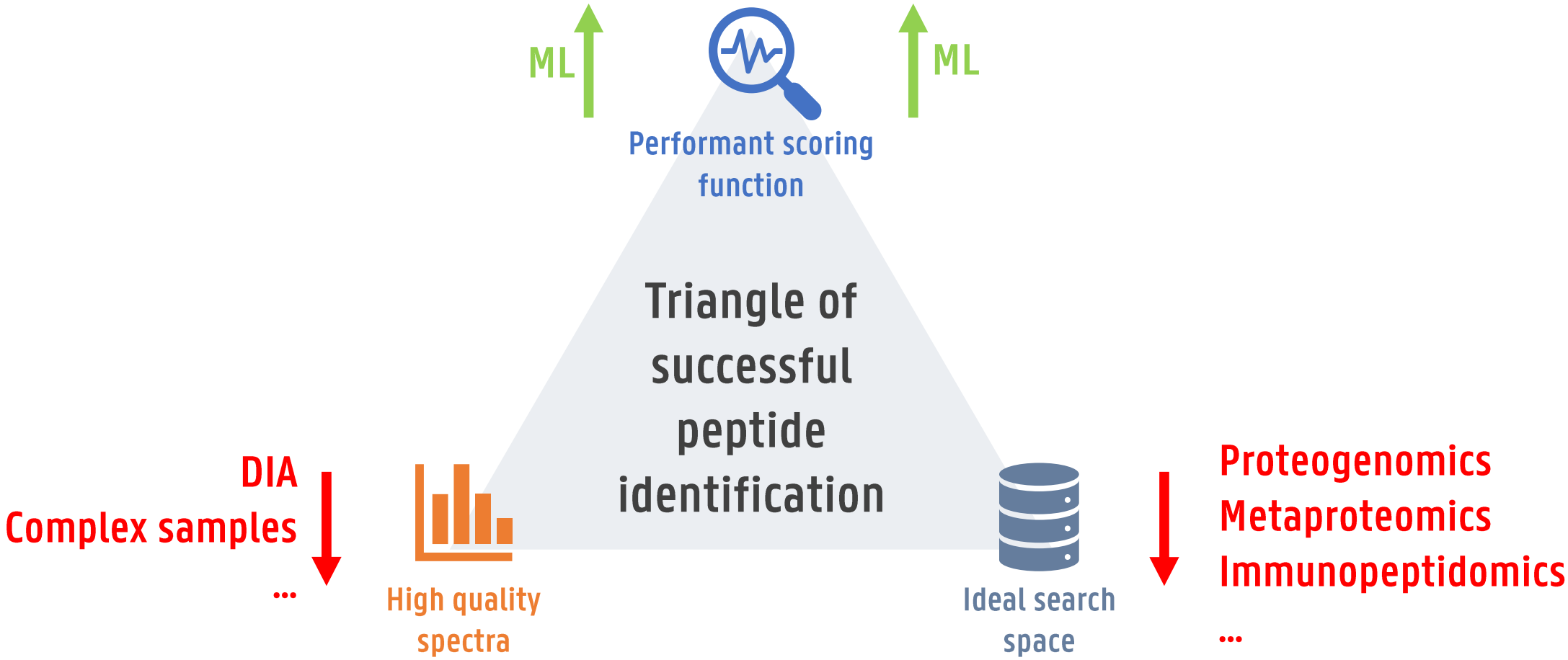


# MS²Rescore in immunopeptidomics: +36% unique identified peptides

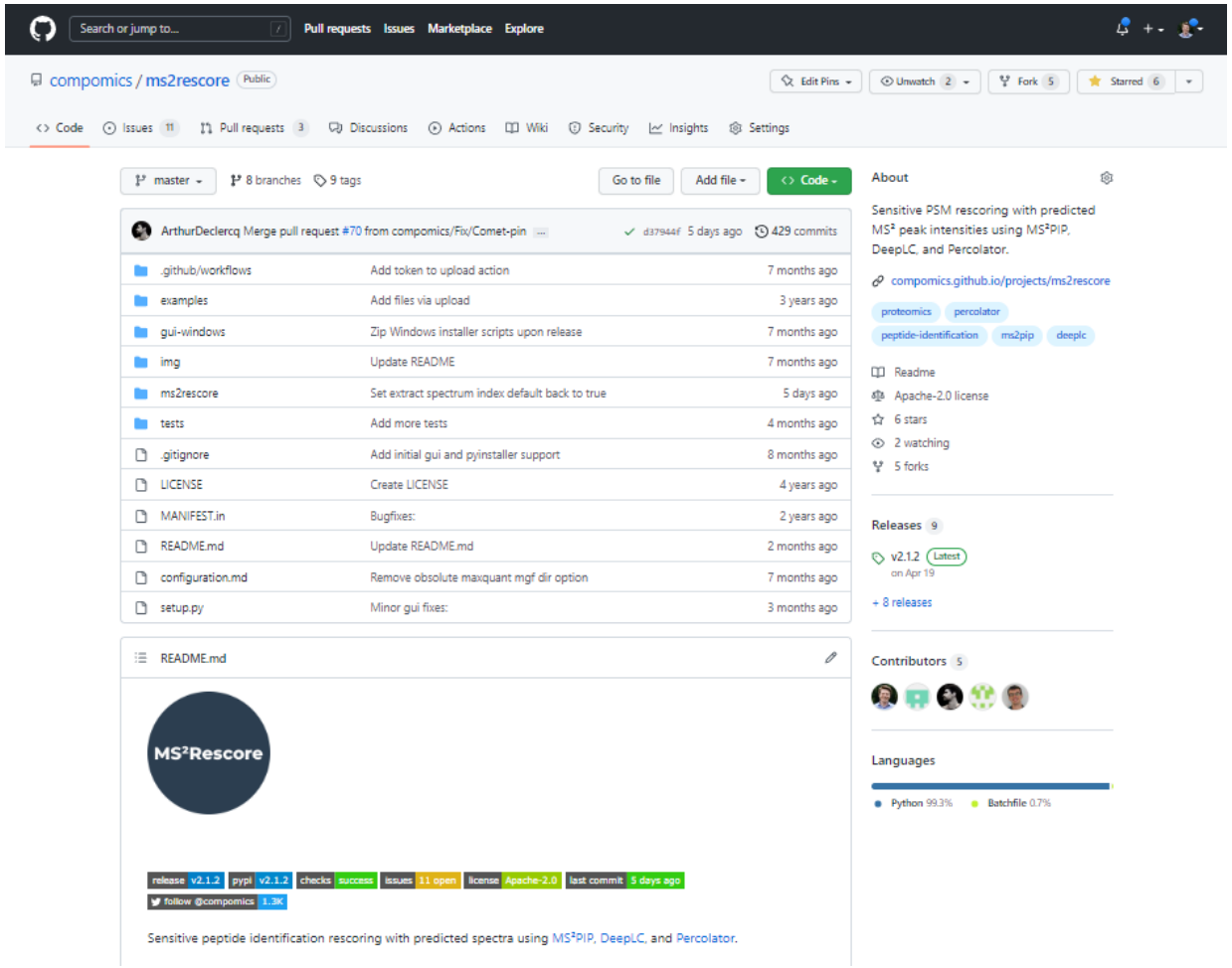


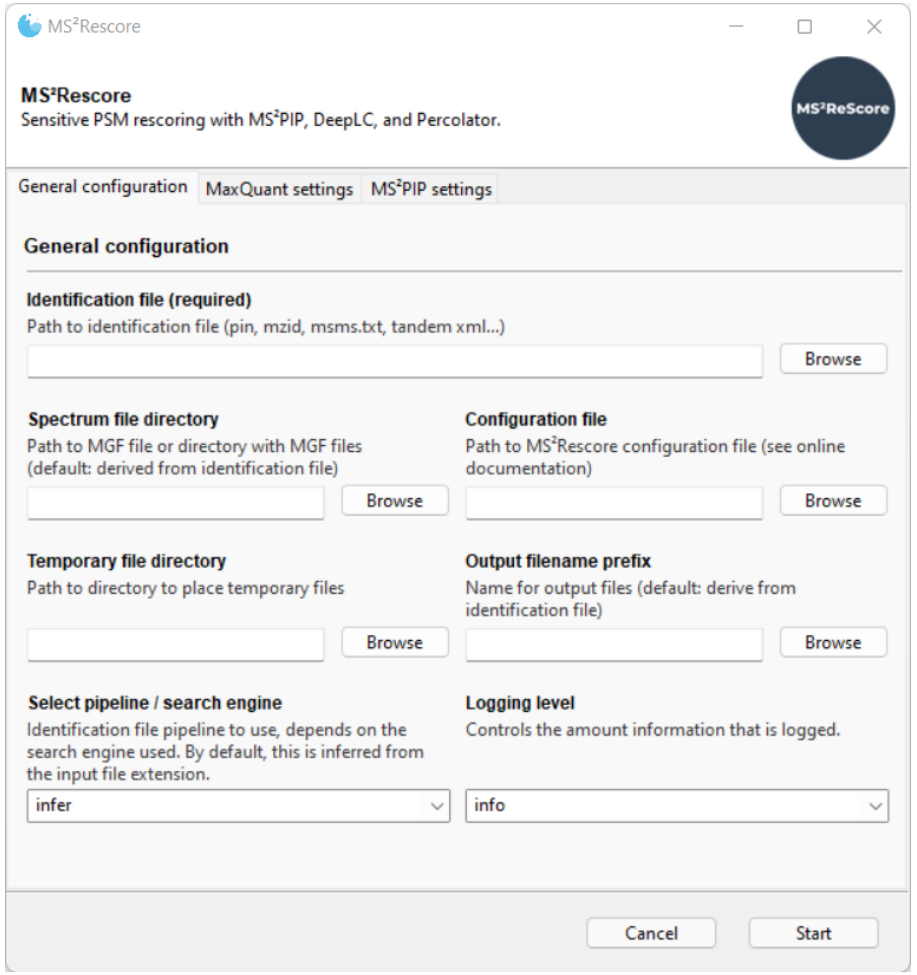
# MS²Rescore in immunopeptidomics: Recovery of low-abundance peptides

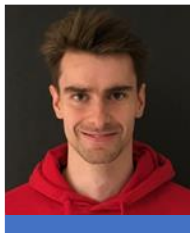
# In challenging experimental workflows, machine learning can rescue peptide identification rates



# MS²Rescore is freely available on GitHub as a Python package, or as an easy-to-install GUI



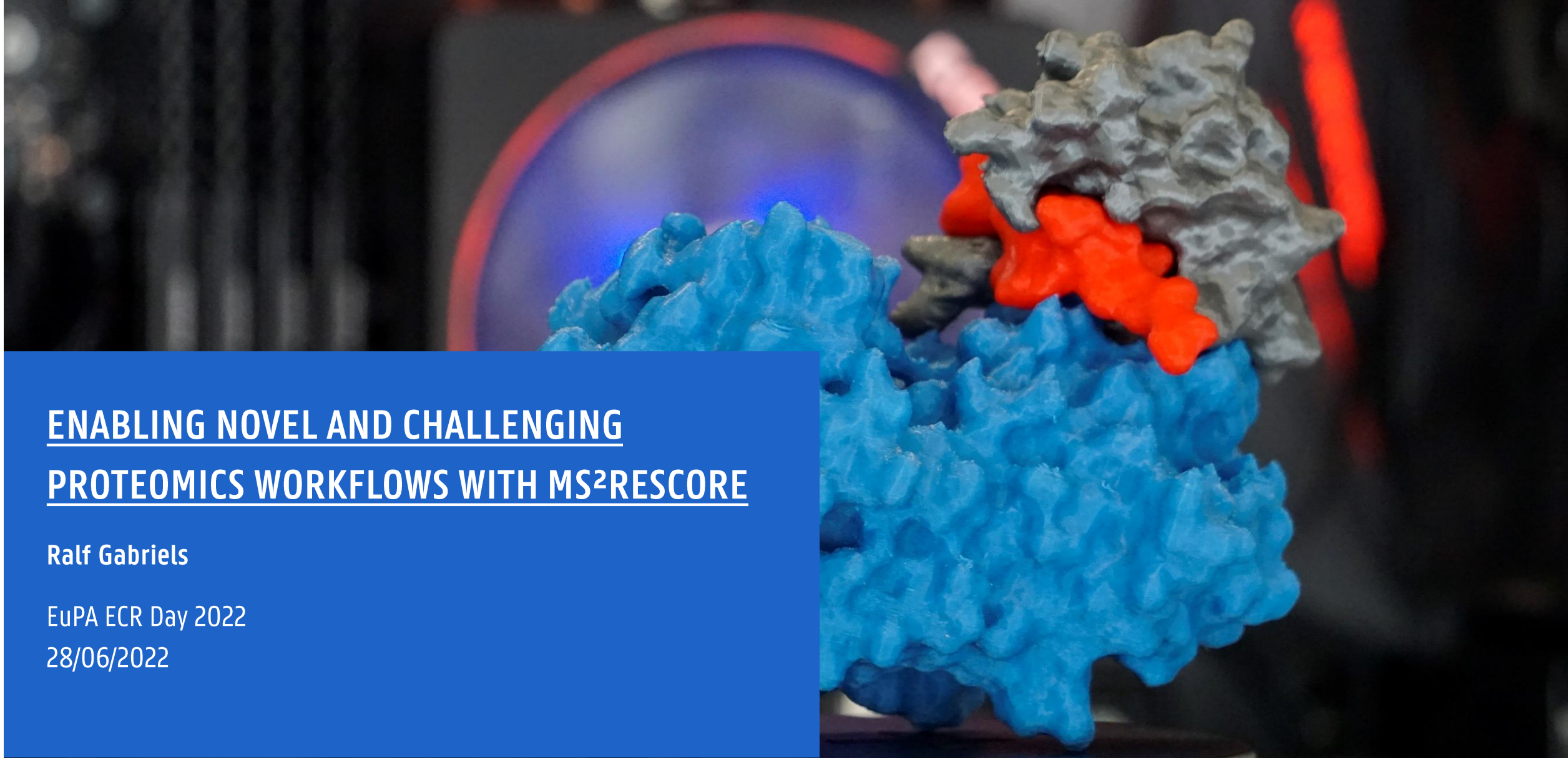





@compomics

www.compomics.com

@RalfGabriels



# ENABLING NOVEL AND CHALLENGING PROTEOMICS WORKFLOWS WITH MS²RESCORE

**Ralf Gabriels**

EuPA ECR Day 2022
28/06/2022

GHENT UNIVERSITY
FACULTY OF MEDICINE AND HEALTH SCIENCES
VIB-UGENT CENTER FOR MEDICAL BIOTECHNOLOGY
CompOmics
fwo