SCIENCE MEETS LIFE

MS^2 peak intensity prediction for specific PTMs, fragmentation techniques and instruments

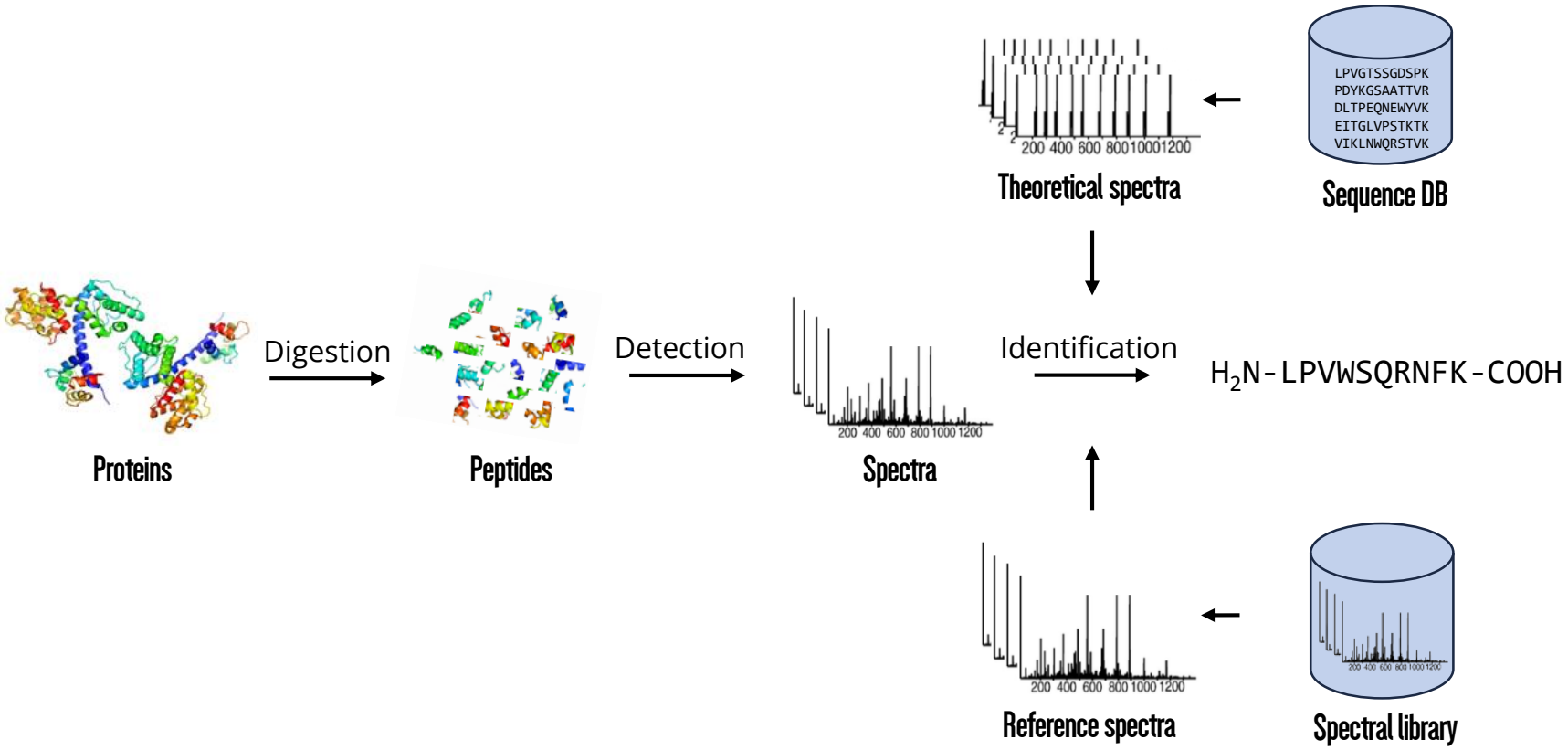
Ralf Gabriels

VIB-UGENT CENTER FOR MEDICAL BIOTECHNOLOGY
GHENT UNIVERSITY
CompOmics
fwo

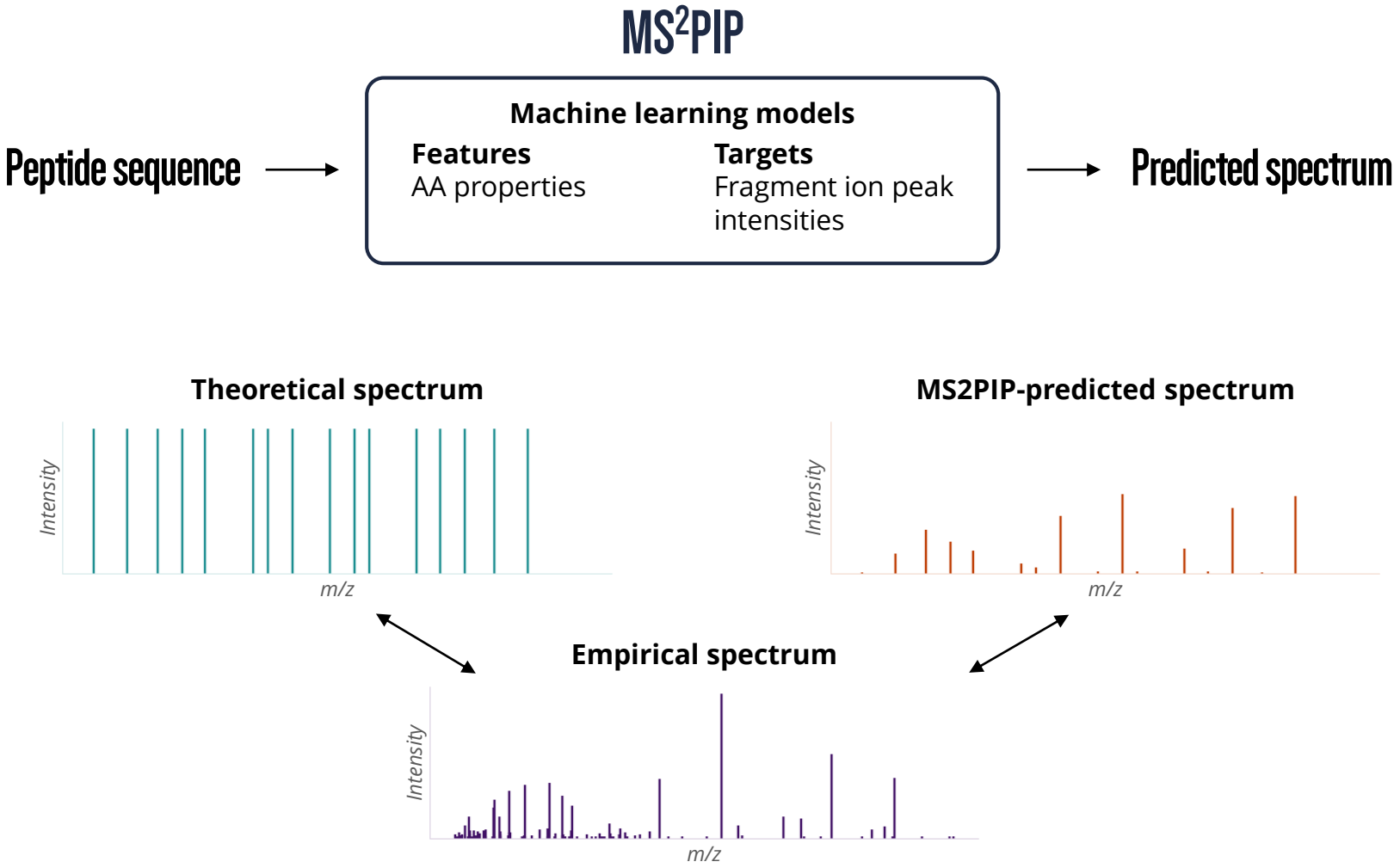
26/09/2018

f-TALES - Big Data in Life Sciences and Biomedicine

MS² peptide spectra are identified using sequence databases or spectral libraries



MS²PIP combines the best of both worlds



We are all trained to recognize people

Image adapted from www.freepik.com/free-vector/casual-characters_774641.htm

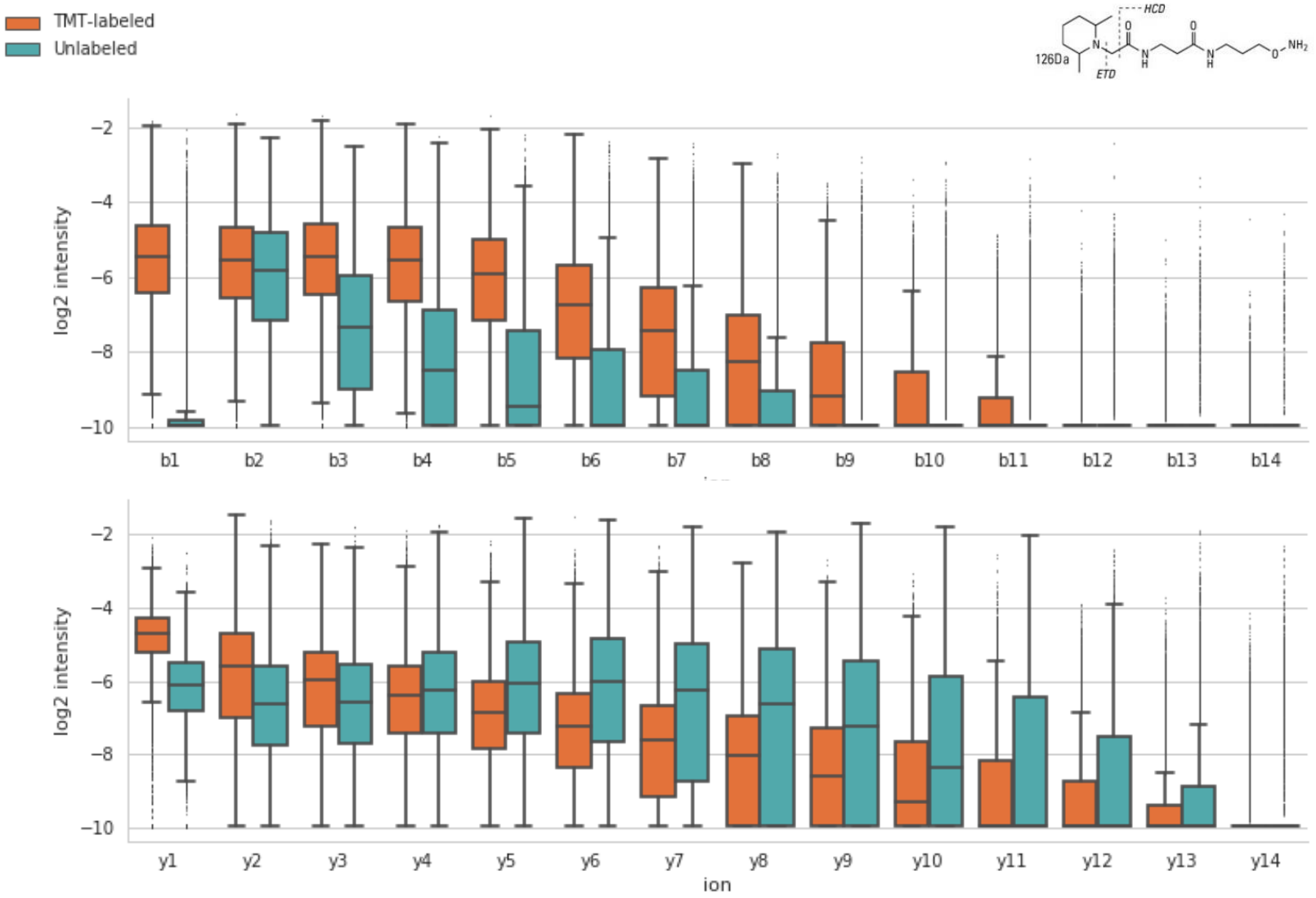
However, recognizing penguins is not as easy

Image adapted from www.sciencing.com

Unless you are a trained zookeeper

Image adapted from www.kisspng.com/png-rico-kowalski-skipper-penguin-madagascar-penguins-819595

TMT-spectra are the proverbial penguins

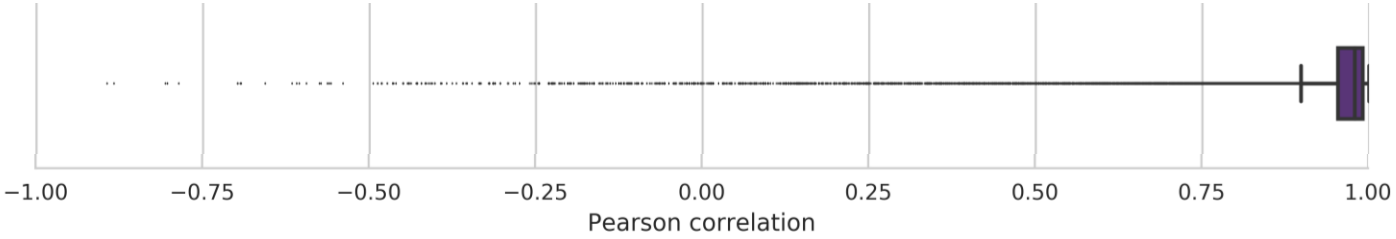


TMT structure adapted from www.thermofisher.com/order/catalog/product/90401

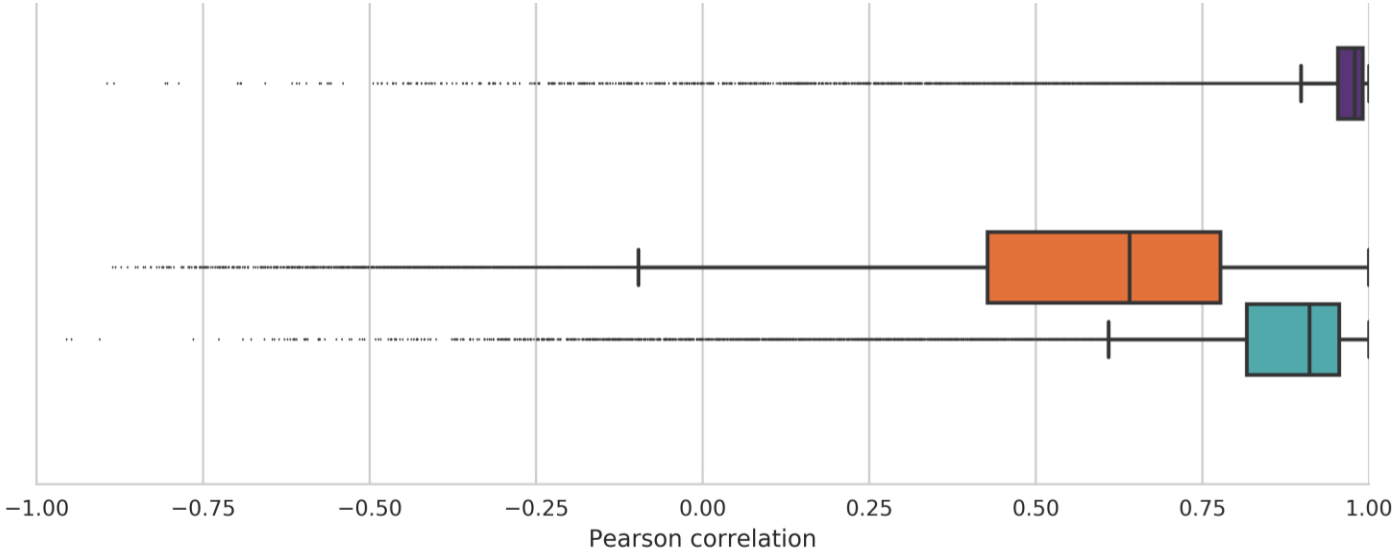
The solution is to retrain MS²PIP on the appropriate MS² spectra

Dataset	Source	Reference	# Peptides
Train data	TMT Spectral library (90%)	Shen J. et al. Journal of Proteome Research (2018)	492 083
External unseen TMT data	TMT experiment	Kris Gevaert Lab, VIB-UGent	45 015
Unlabeled reference dataset	Unlabeled experiment	Chick J.M. et al. Nature Biotechnology (2015)	14 256

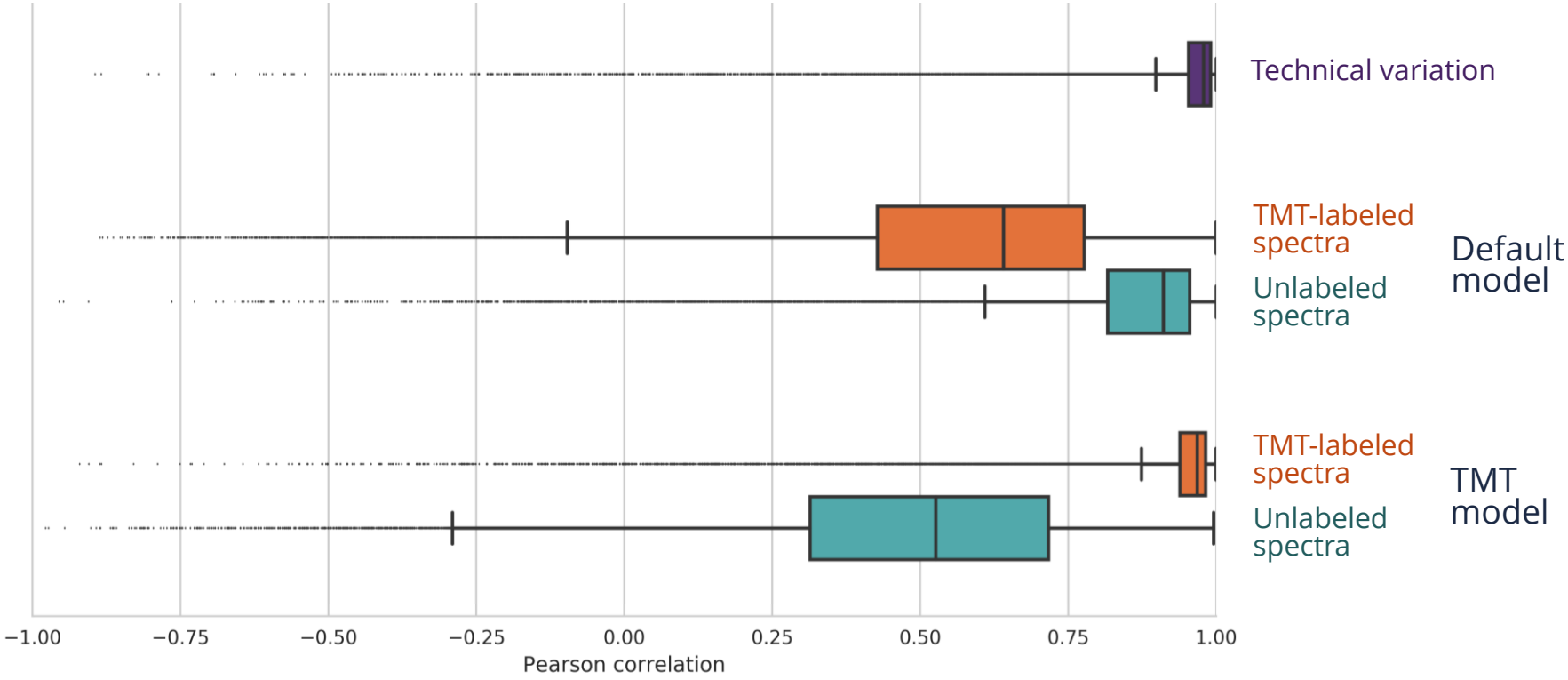
TMT model performance is only limited by the technical variation



TMT model performance is only limited by the technical variation

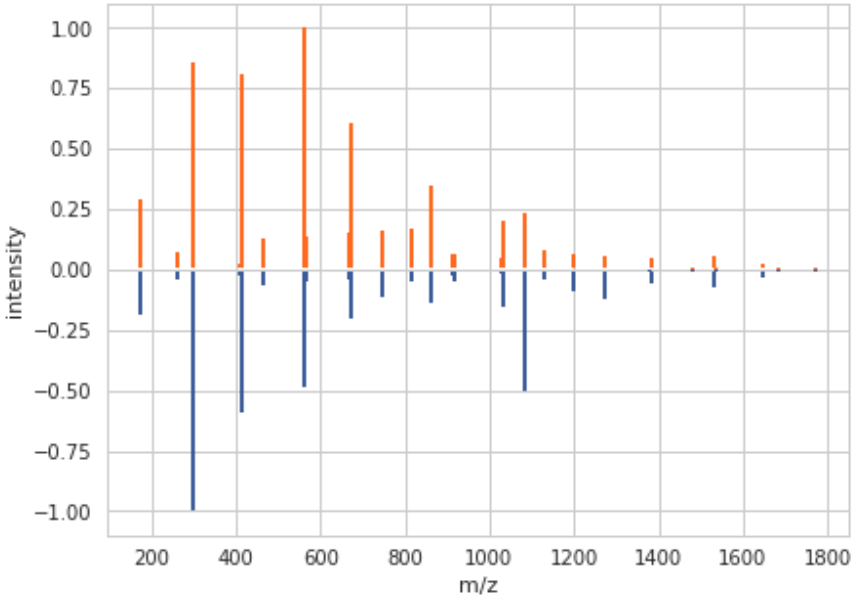


TMT model performance is only limited by the technical variation



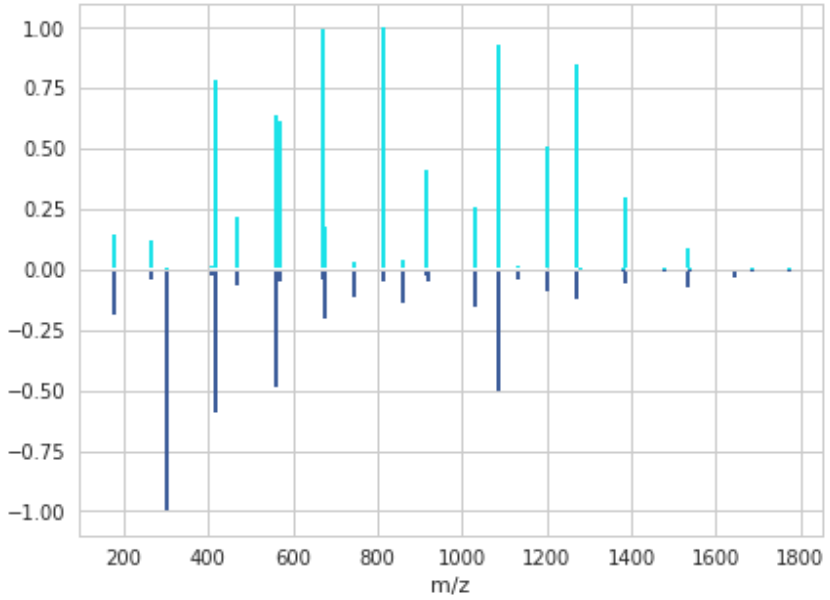
Some example spectra

TMT model prediction



Empirical TMT spectrum

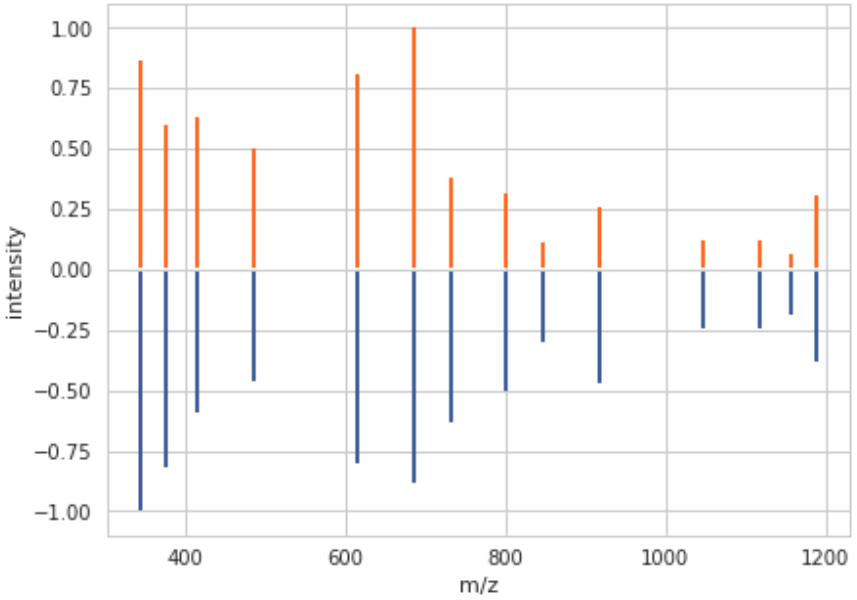
Default MS²PIP prediction



Empirical TMT spectrum

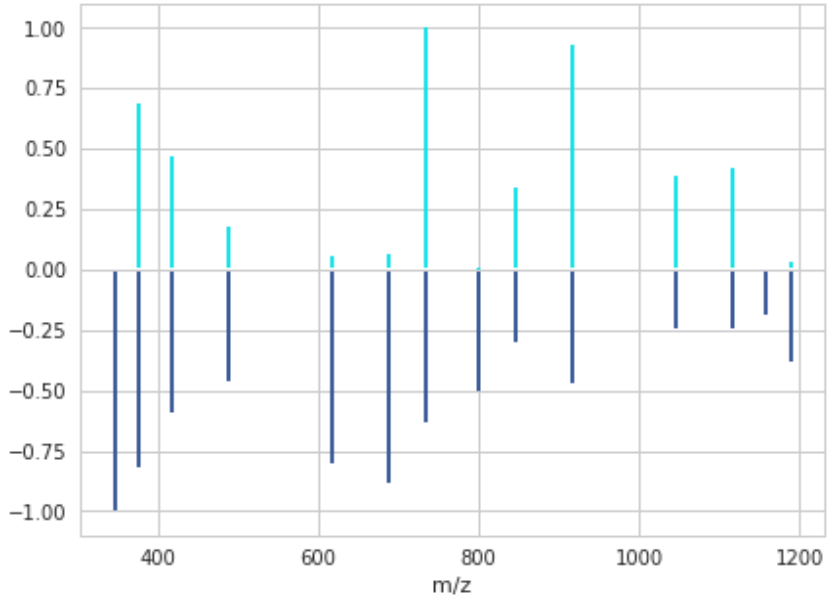
Some example spectra

TMT model prediction



Empirical TMT spectrum

Default MS²PIP prediction



Empirical TMT spectrum

Many other specific models can be trained

TTOF5600 instruments

ET(hc)D fragmentation

Phosphorylation

Non-tryptic (immuno-) peptides

All models are (or will be) available on the user-friendly MS²PIP Server

SCIENCE MEETS LIFE

VIB-UGENT CENTER FOR MEDICAL BIOTECHNOLOGY

GHENT UNIVERSITY

Comp omics

fwo

www.compomics.com
@CompOmics

iomics.ugent.be/ms2pip

MS²PIP SERVER

MS² Peak Intensity Prediction

MS²PIP is a tool to predict MS² signal peak intensities from peptide sequences. It employs the XGBoost machine learning algorithm and is written in Python.

You can install MS²PIP on your machine by following our extended install instructions found on the MS²PIP GitHub repository. For a more user friendly experience, we created this web server. Below, you can easily upload a list of peptide sequences, after which the corresponding predicted MS² spectra can be downloaded in a CSV or MGF file format.

More advanced users can also access MS²PIP Server through our RESTful API. Swagger-generated documentation can be found here and an example Python script to access the API can be found here.

If you use MS²PIP for your research, please cite the following papers:

- Degroeve, S., Maddelein, D., & Martens, L. (2015). MS²PIP prediction server: compute and visualize MS2 peak intensity predictions for CID and HCD fragmentation. *Nucleic Acids Research*, *43*(W1), W326–W330.

Ralf Gabriels

Ralf.Gabriels@VIB-UGent.be

@CompOmics

www.compomics.com