Benefits of direct feature extraction via correlation of experimental and accurately simulated mass spectra for glycoproteomics

Konstantin O. Nagornov,[1] Anton N. Kozhinov[1], Sergey Vakhrushev[2] and Yury O. Tsybin[1]

[1]Spectroswiss, EPFL Innovation Park, 1015 Lausanne, Switzerland; [2]University of Copenhagen, Copenhagen, Denmark

Introduction

- FTMS instruments, such as Orbitraps, often generate mass spectral datasets of very high complexity that require advanced data processing and analysis software tools;
- Considering the specific nature of the FTMS data, namely the time-domain transient processing and isotopic envelopes, enhances the feature extraction performance;
- Previously, we reported on developing the **FTMS Simulator** – a software tool to accurately simulate FTMS data via modeling the time-domain transients [1, 2];
- Here, we report on the feature extraction workflows for small molecule and protein analysis based on the data simulated with the FTMS Simulator.

Workflow

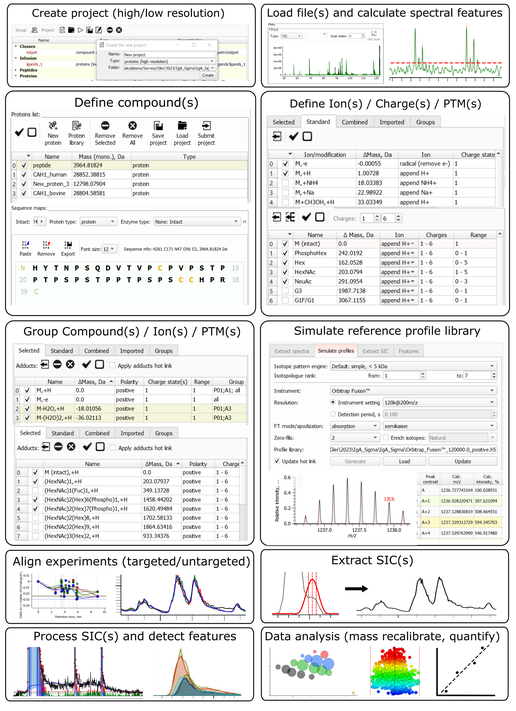

Figure 1. The project based MS and MS/MS workflow for data annotation and quantification via feature extraction from LC/MS or GC/MS experiments using transient mediated simulations.
- calculate mass spectral features: baseline correction, noise thresholding, three-point interpolation peak picking;
- simulate reference profile library: automatic instrument selection, resolution settings; simulation of isotopic envelope profile and centroids via transient calculation for compounds of interest;
- extract SIC(s): group compounds of interest (single compounds, charge state grouping, compound classes); isotopic group selection (A, A+1, etc.); mass tolerance based SIC extraction.
- feature detection: baseline correction, smoothing and noise thresholding of SIC(s), detect feature (SIC peak) based on matching experimental and simulated data; filter detected features: minimum points per peak, number of adducts, etc.
- data analysis: re-calibration of mass spectra; mass accuracy plots; dynamic global results matrix overview; hits map, quantification: absolute, semi, standard addition.

Small molecules (protein ligand interactions)

Small molecules analysis in protein ligand interactions, where large number (>20k) of reference compounds is searched in the large (up to 1 k) number of files, may be automated, accelerated and improved using the targeted deconvolution method, precalculated profile library and automatic detection parameters optimization, Figures 2-3.

Figure 2. Affinity LC-MS based ligand screening was performed using Orbitrap™ Q Exactive™. ~13'000 ligands were sampled and analysed using two (standard vs sample) 96-deep-well plates (132 ligands per single well). Compounds identification and similarity filtering in the samples were performed by extracting selecting ion current (SIC) chromatograms using two highest isotopologues. The binding index was calculated as a ratio of under the curve areas between standard and sample SIC features.

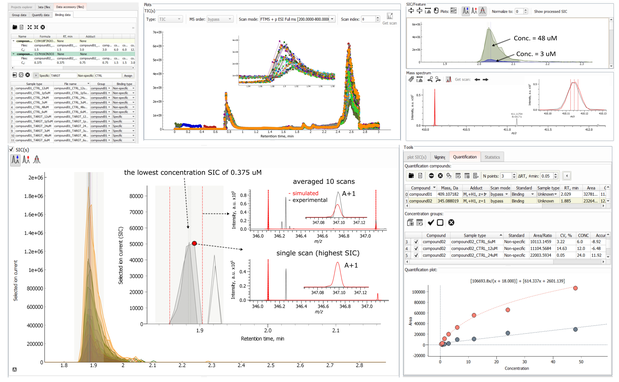

Figure 3. Mass spectrometry-based ligand binding assays experiment was performed using Orbitrap Q Exactive. 12 target ligands were sampled and analysed using two (non-specific, control vs specific, target) 96-deep-well plates; 8 different concentrations (8 wells) per ligand. The specific and non-specific dependencies of the under the curve area of SIC features on concentrations were used for calculation of ligand binding affinity. SICs were extracted and similarity filtered using two highest isotopologues.

References

1. Nagornov K.O., Kozhinov A.N., Tsybin Y.O. Transient-Mediated Simulations of FTMS Isotopic Distributions and Mass Spectra to Guide Experiment Design and Data Analysis, *JASMS*, 31, 9, 2020.
2. Nagornov K.O., Kozhinov A.N., Gasilova N., Menin L., Tsybin Y.O. Characterization of the Time-Domain Isotopic Beat Patterns of Monoclonal Antibodies in Fourier Transform Mass Spectrometry, *JASMS*, 33, 7, 2022.

Glycosylation analysis in proteomics

Targeted and semi-targeted PTM analysis (glycosylation, phosphorylation, etc.), may be also improved using the targeted deconvolution method via simulations and similarity filtering of grouped isotopic envelopes of different charge states. Wherein, the processing of filtered multiple unique MS and MS/MS scans per feature increases the confidence of identifications, Figures 4-6.

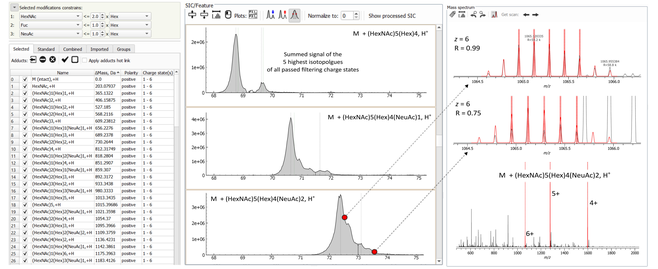

Figure 4. Targeted LC-MS data bottom-up analysis of IgA (Sigma) sample. The trypsin-generated data were acquired on Orbitrap™ Fusion™ instrument. Targeted heavily glycosylated peptide ~5 kDa was searched via reference PTM library of 256 glycoforms. SIC chromatograms were extracted using multiple charge states isotopic envelops and filtered via profile similarity between experimental and simulated data.

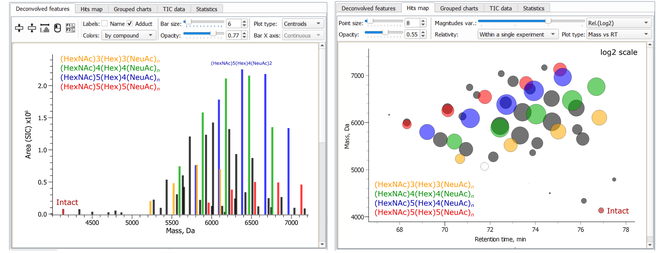

Figure 5. Targeted MS only bottom-up results analysis and visualization. (Left) the deconvolved mass spectrum and (right) bubble plot are shown for identified glycoforms of a target peptide from Figure 4. The glycoforms are grouped and colored by the corresponding number of HexNAc and Hex. The retention time filtering was additionally applied to increase confidence of identifications.

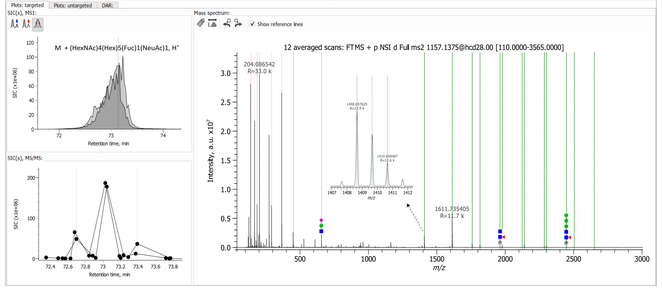

Figure 6. Targeted HCD bottom-up analysis of a glycoprotein on Orbitrap™ Q Exactive HF™. The glycoforms were identified based on the similarity filtering in both (top left) MS and (bottom left) MS/MS scans. Wherein, the only oxonium ions and characteristic *Y* fragments were used to calculate total similarity score in MS/MS spectra. To increase the identification confidence using better statistics, multiple MS/MS scans within the corresponding RT window were analyzed combinatorially.

Protein analysis (isotopic mass)

Protein analysis at high and low resolution may be improved using the combination of a standard untargeted deconvolution (Thrash, FlashDeconv, Hardklor, UniDec, etc.) and targeted deconvolution method via simulations, Figures 7-8.

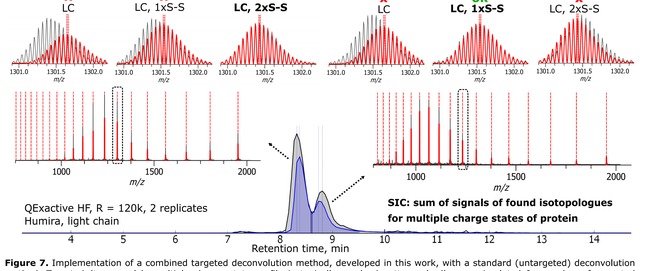

Figure 7. Implementation of a combined targeted deconvolution method, developed in this work, with a standard (untargeted) deconvolution method. Targeted (top panels): multiple charge state profile isotopically resolved patterns (red) were simulated for number of expected subunits of IdeS digested Humira and Orbitrap Q Exactive HF instrument. SICs were calculated for the simulated reference patterns and features were detected according to the "Feature Detection" algorithm (see the "Workflow" section). Untargeted: standard high resolution deconvolution was performed using the HardKlor algorithm, integrated into the software tool. The deconvolved spectra as output of both algorithms were analyzed together.

Protein analysis (average mass)

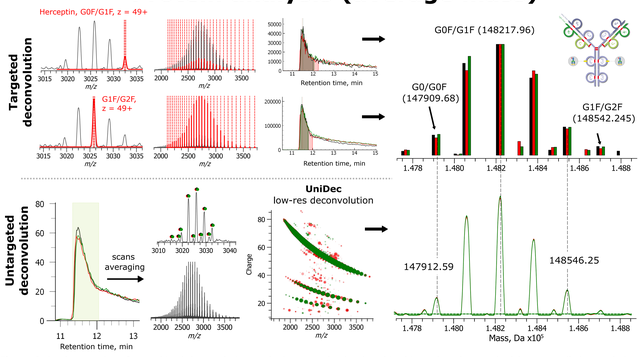

Figure 8. Implementation of a combined deconvolution method: targeted - developed in this work and a standard (untargeted) deconvolution method. Targeted (top panels): broadband charge state profile patterns (red) were simulated for a number of expected proteoforms of Herceptin and Orbitrap Q Exactive HF instrument. SICs were calculated for the simulated reference patterns and features were detected according to the "Feature Detection" algorithm (see the "Workflow" section). Untargeted (bottom panels): standard low resolution deconvolution was performed using the UniDec algorithm (DOI: 10.1021/acs.analchem.5b00140), integrated into the software tool. (Right panel) the deconvolved spectra as output of both algorithms were analyzed together.

Conclusions

- FTMS specific software tool was developed for targeted feature extraction from LC/GC-MS data for molecules of any size. The tool is supported with a graphical user interface, demonstrated excellent analytical specificity, high sensitivity, and quantitative precision when applied to metabolomics, complex mixture and protein analysis;
- Accurate simulation of FTMS (profile) mass spectra for small and large molecules support accurate, rapid, and targeted deconvolution based on feature extraction;
- Features are confirmed by similarity scoring between the isotopic envelopes of the experimental and simulated data. The isotopic envelopes can be matched individually or grouped. The groups can contain charge state distributions, proteoform clusters (e.g., glycosylation profiles of mAbs), or classes of compounds in complex mixture analysis.

We acknowledge Fernando Almeida from OMass Therapeutics for the provided data and discussions.

Conflict of interest: Nagornov K.O., Kozhinov A.N. and Tsybin Y.O. are employees of Spectroswiss.

Poster presented at the 71st ASMS Conference, Houston, TX, USA, June 3-8, 2023