Protein-grade FTMS Isotopic Simulator to guide the experiment design and data analysis in top-down proteomics

Konstantin O. Nagornov,[1] Anton N. Kozhinov,[1] Natalia Gasilova,[2] Laure Menin[2] and Yury O. Tsybin[1]

[1]Spectroswiss, EPFL Innovation Park, 1015 Lausanne, Switzerland; [2]Ecole Polytechnique Federale de Lausanne (EPFL), 1015 Lausanne Switzerland.

Introduction

Protein analysis by FTMS via middle-up, middle-down, intact, and top-down approaches may result in quantification errors or structural misinterpretation due to FT artifacts, *e.g.*, introduced by deconvolution or deisotoping. Simulating FTMS data may improve interpretation of complex protein experimental data by revealing these artifacts. The **FTMS Isotopic Simulator** is a software tool previously introduced by us to accurately simulate FTMS isotopic envelopes and mass spectra for selected compounds, Figure 1. Here, we extend its applicability to simulating mass spectra of proteins and protein mixtures, including complex product ion patterns in top-down data. It is now possible to simulate data for monoclonal antibodies and larger proteins.

Workflow

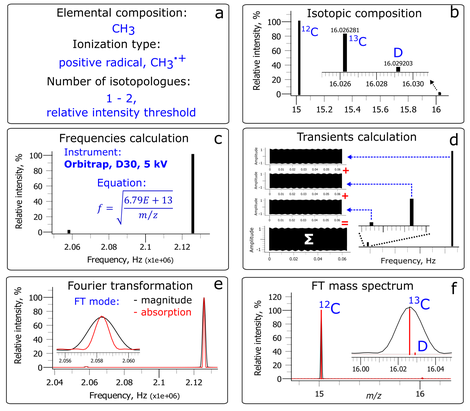

Figure 1. The user can define data processing parameters at several levels:
(i) Define molecule(s): elemental composition, amino acid sequence, charge carrier, modifications, isotopic enrichment/depletion, numbers of charge states and isotopologues;
(ii) Define experiment: FTMS instrument selection (any Orbitrap™ and ICR instruments), resolution target (at a given *m/z*), transient sampling rate, add noise, set intensity thresholds;
(iii) Define data processing: absorption or magnitude FT, apodization function, number of zero-fills, full or reduced profile mode, centroid mode, visualize transients and frequency spectra;
(iv) Use mass spectra: set ratios of intensities, visualize isotopic envelopes, plot multiple envelopes, extract high quality images, extract data points (mass spectra) for further analysis; and
(v) Results output: search experimental data with the simulated profiles/centroids, rank search results by the similarity score (*m/z*, intensity), save project for future use, create report as a PDF file.

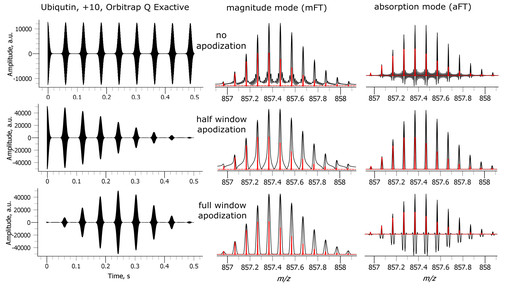

Figure 2. Simulated FTMS transients and isotopic distributions for analysis of 10+ ubiquitin (human) for Q Exactive™ Orbitrap™ instrument settings. Data are shown in their unapodized (top row) or apodized forms using half (middle row) or full (bottom row) windows based on the Kaiser function. Transients were processed using magnitude mode FT (mFT, middle column) and absorption mode FT (aFT, right column).

The First Beat Limit in FTMS of Proteins

Processing of a transient of rationally-defined length using an appropriate apodization window may maximize performance of protein analysis for measuring the average protein mass, Figure 3. In turn, this appropriate transient length could be suggested by accurate simulation of FTMS data.

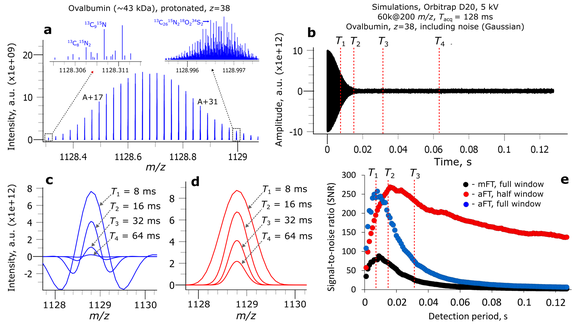

Figure 3. Panel (a) shows theoretical isotopic pattern of 38+ ovalbumin (~43 kDa). Expanded views show isotopic fine structures for A+10 and A+36 isotopologues. Panel (b) shows the corresponding transient of 128 ms length as simulated for Orbitrap™ Fusion™ instrument. Transient was mixed with noise signal (Gaussian distribution). The final transient at variable length from 1 ms up to 128 ms with a step of 1 ms was apodized using the Kaiser function and FT processed: panels (c, e) shown in blue - aFT, full window; panels (d, e) shown in red - aFT, half window; panel (e) shown in black - mFT, full window. Panel (e) SNR value was calculated as peak intensity divided by six standard deviations of noise.

The Second Beat Artifacts in FTMS of Proteins

FTMS artifacts may be expected to be present in the corresponding mass spectra when transients are at around the two isotopic beats duration, as confirmed and exemplified by Figures 4-6.

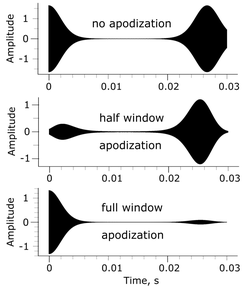

Figure 4. Transient of 30 ms length of 5+ insulin protein was simulated for Q Exactive™ HF FTMS instrument settings and shown as (top) unapodized, or apodized by application of (middle) half and (bottom) full windows of the Kaiser function. The corresponding mass spectra are shown in Figure 5.

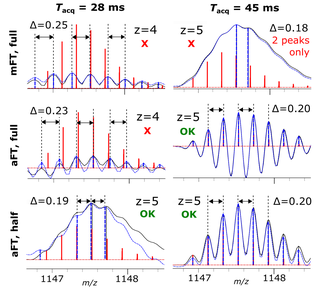

Figure 5. Expanded views into (black, solid line) simulated and (blue, dashed line) experimental profile mass spectra showing isotopic distributions of 5+ insulin. Peak centroids are shown for experimental peaks (blue, dashed line) and for simulated peaks (black, solid lines). Theoretical isotopic pattern is shown with red solid lines. Experimental data acquired using a Q Exactive™ HF FTMS, simulated data obtained using the corresponding settings. Comparison of the simulated vs. experimental results is the same FT parameters.

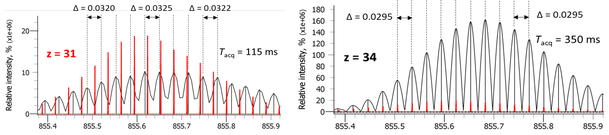

Figure 6. Artifacts in the analysis of protein isotopic distributions. Expanded views into the simulated profile mass spectra of 34+ carbonic anhydrase (CAH). The corresponding transients of different length were simulated for Q Exactive™ HF settings, apodized using the full window and mFT processed. Black line - profile mass spectra, red bars - theoretical isotopic patterns.

FTMS Artifacts in Protein Mixture Analysis

The computational capabilities of the FTMS Isotopic Simulator described here allow to generate and visualize broadband mass spectra with wide (user-defined) charge state distributions of proteins, Figures 7-8. These results show that reducing transient length 2-fold significantly changes the relative abundances of proteins in a mixture.

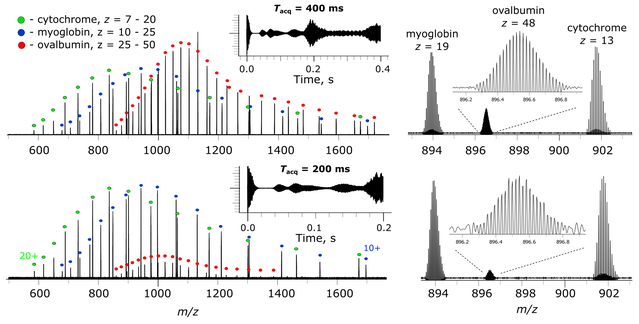

Figure 7. Analysis of relative intensity errors for a protein mixture. Simulated profile mass spectra (in absorption mode FT, full window) are shown for a protein mixture of cytochrome C (horse, ~12 kDa), myoglobin (bovine, ~17 kDa) and ovalbumin (chicken,~43 kDa). (Left column) broadband mass spectra show charge state distribution and (right column) expanded views show the corresponding isotopic distributions of proteins. The data were simulated with transient length of (top row) 400 ms and (bottom row) 200 ms for Q Exactive™ HF settings. Negative values of the mass spectra (full-window aFT) are truncated.

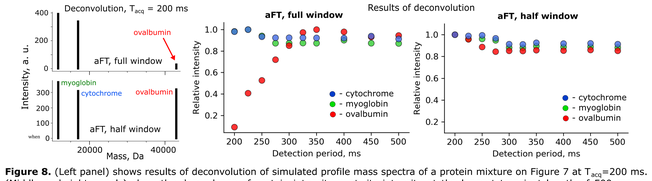

Figure 8. (Left panel) shows results of deconvolution of simulated profile mass spectra of a protein mixture on Figure 7 at T_{acq}=200 ms. (Middle and right panels) show the dependence of protein intensity w.r.t. its intensity at the longest transient length of 500 ms on detection period after deconvolution. Transient of variable length was apodized using (center panel) full and (right panel) the half window Kaiser function and aFT processed. All charge state patterns of all proteins were isotopically resolved for each transient length. The Hardklor software (University of Washington) was employed for data deconvolution.

Artifacts in Overlapping Proteoforms Analysis

In modern protein analysis, it is a common practice to detect unresolved proteoforms. Increasing the distance between the two IFS clusters of overlapping proteoforms starts to show a deviation between FT processing performed following half or full apodization window, Figure 9. When the half window is employed for FT processing (apodization), the discrepancy between the expected and observed abundances from the unresolved interfering proteoforms is not as substantial in comparison to the full window. The approach of processing FTMS data using different apodization windows can be applied to validate if a measured isotopic pattern can be potentially an overlap of several proteoforms.

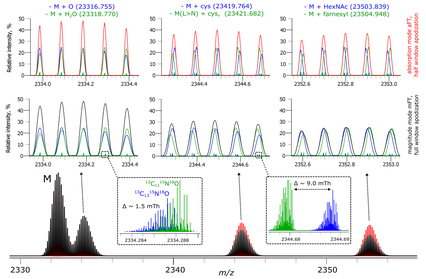

Figure 9. Simulated broadband profile mass spectrum and its expanded views are shown for native 10+ RoboArixtra protein (~23 kDa), M, and its proteoforms:
M+O – oxidation, +15.99491 Da;
M+H_2O – hydration, +18.01056 Da;
M+Cys – cysteinylation, +119.0041 Da;
M(L>N)+Cys - double leucine to asparagine substitution and cysteinylation, +120.92183 Da;
M+HexNAc – N-acetylhexosamine, +203.07937 Da;
M+farnesyl – farnesylation, +204.18780 Da.

The data was simulated with transient length of 1.8 s for a 12 T FT-ICR instrument. Simulated transient was apodized using the Kaiser function and FT processed: black color on middle panels - mFT, full window; red color on top panels - aFT, half window.

FTMS Artifacts in Top-Down Protein Analysis

Top-down analysis performance is sensitive to deconvolution and FT parameters such as apodization window type. Even a slight variation of a transient length may cause a significant variation in the number of identified product ions, and, thus, in the total sequence coverage, Figures 10 and 11. The number of deconvolved features remains almost the same at all transient lengths considered here, as isotopic envelopes of product ions were isotopically resolved. The reasons for mis-assignment of product ions in mass spectra generated using the full apodization window are the FT processing artifacts, *e.g.*, assignment of a wrong charge state to an isotopic distribution.

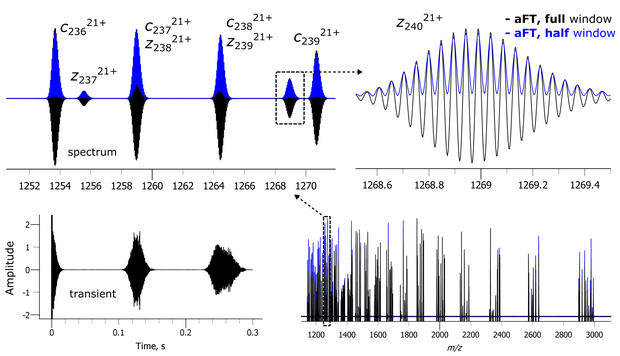

Figure 10. Top-down protein analysis example. A transient of selected *c*- and *z*-ions (total 194 ions) of N-terminal acetylated carbonic anhydrase, CAH (~29 kDa) was simulated for Orbitrap™ Fusion™ instrument settings. Wherein, mass and charge state of these ions were chosen to generate transient with two isotopic beats at 128 ms length. Simulated transient was aFT processed using the (black) full or (blue) half windows of the Kaiser function. Bottom right and top panels show the broadband profile mass spectrum and its expanded views.

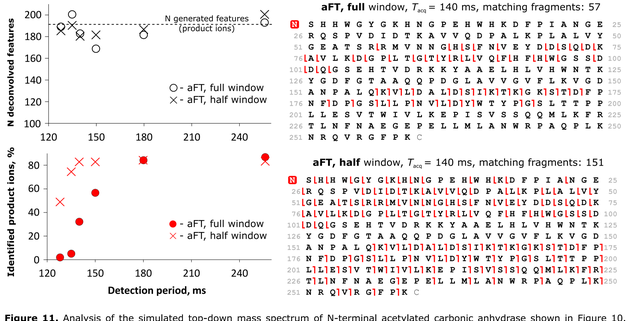

Figure 11. Analysis of the simulated top-down mass spectrum of N-terminal acetylated carbonic anhydrase shown in Figure 10. Transient simulated for Orbitrap™ Fusion™ instrument settings, followed by apodization of this transient at variable lengths using the full or half windows of the Kaiser function. Left panel shows the dependence of (top) total number of deconvolved features and (bottom) percentage of identified CAH product ions from total amount of simulated ones as a function of a detection period for data apodized using (circles) full and (crosses) half windows. (Right column) sequence coverage maps show *c*- and *z*-type ions identified in mass spectra acquired and processed for the following settings: top – Tacq = 140 ms, aFT, full window; bottom – T_{acq} = 140 ms, aFT, half window.

Conclusions

- The FTMS Isotopic Simulator efficiently simulates realistic FTMS data, e.g., top-down data, via transient generation;
- We use this tool here to showcase the potential artifacts in FTMS and FTMS/MS of proteins and protein mixtures;
- The FTMS Isotopic Simulator can be used to design experiments and support data analysis in FTMS;
- Simulated isotopic profiles and centroids can be used to search experimental data and rank results;
- The accurately simulated results can reveal FT processing-related artifacts, which may be lost in typical calculators;
- The didactic nature of the FTMS Isotopic Simulator can be found useful in training the FTMS scientists;
- It can find utility to benchmark and tune deisotoping, deconvolution and searching tools, and for quality control;
- The FTMS Isotopic Simulator is available for evaluation from the authors.

We acknowledge financial support from European Horizon 2020 RIA program (grant agreement #829157, TopSpec).