Poster Reprint

ASMS 2025
Poster number ThP 367

Evaluating Data Analysis Techniques for LC-IM-MS Data: Preprocessing, Untargeted Feature Finding, and DIA Fragmentation Alignment

Sarah M. Stow[1], Andrea Harrison[2], Bryson Gibbons[2], Olivier Chevallier[1], David A. Weil[1], Ruwan T. Kurulugama[1], Aivett Bilbao[2]

[1]Agilent Technologies, Inc., Santa Clara, CA;

[2]Pacific Northwest National Laboratory, Richland, WA

Introduction

The extra dimension ion mobility adds to traditional LC-MS analysis provides better separation of complex samples at rapid speeds and reduces chimeric spectra from DIA experiments through mobility separated fragmentation. The data rich files that result from IM experiments require a different set of algorithms to detect the features present. Four-dimensional (RT, DT, m/z, and intensity) feature finding algorithms provide important information from the precursor level and must consider both the LC separation as well as the IM for DIA data to extract the mobility aligned fragmentation patterns. Here we evaluate four-dimensional feature finding algorithms (vendor and third party) as well as methods for aligning precursor and fragment ions across various chemical species including lipid, proteomics and PFAS data.

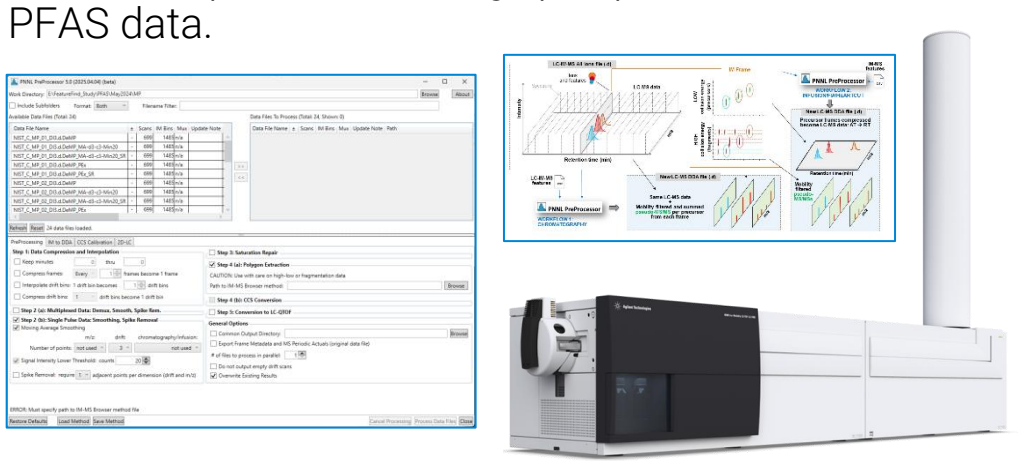

Figure 1. PNNL PreProcessor UI with IM to DDA diagram and 6560 IM-QTOF.

The data sets analyzed in this study exercise different elements of untargeted feature finding. The examples shown in Figure 2 highlight different analyte chemical classes (or isotope models), charge states (-1, +1, +2, and +3), and both positive and negative polarities.

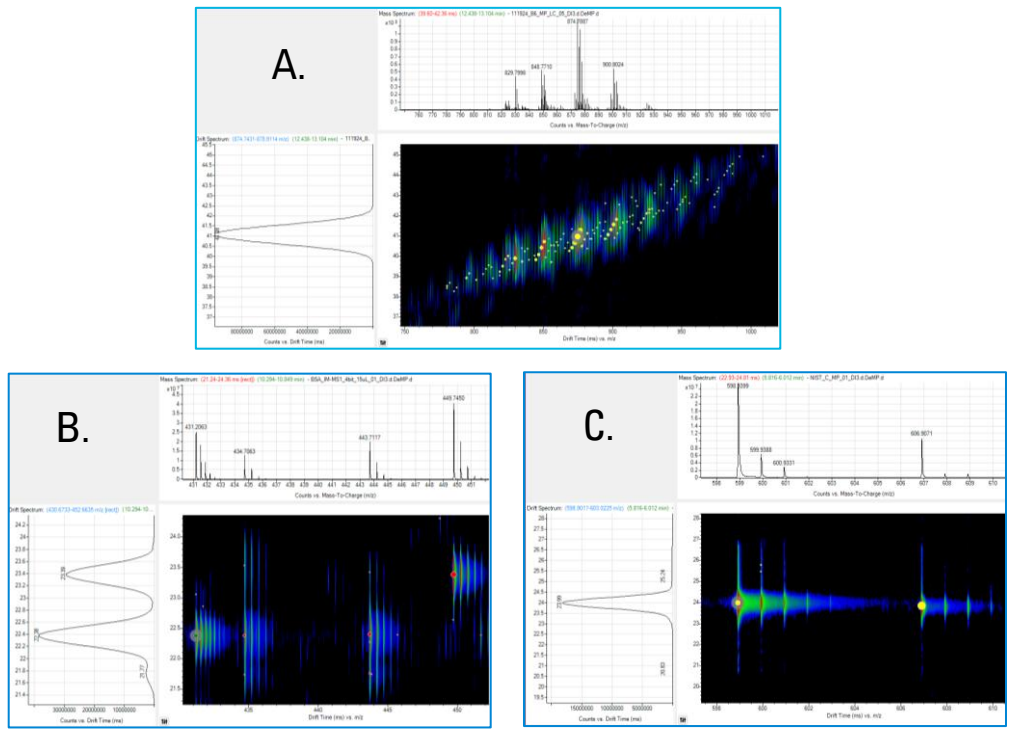

Figure 2. Example Features in each of the data sets evaluated: A. Lipids, B. Peptides, and C. PFAS.

Experimental

Data files were acquired on a 6560 Ion Mobility LC/Q-TOF (Agilent Technologies, Santa Clara, CA). A commercial LC (1290 series, Agilent Technologies, Santa Clara, CA) was used for LC separations prior to analysis with IM-MS. A lipid extract of the NIST SRM 1950, a BSA tryptic digest, and a PFAS sample were analyzed with both single pulse and multiplexed acquisition as well as both MS1 and mobility aligned fragmentation (MAF) DIA approaches. These data files have been processed with both Agilent and third-party software tools to evaluate preprocessing, feature finding, and isotope grouping techniques. New features in the PNNL PreProcessor[1,2] shown in Figure 3. streamline IM data analysis workflows for 6560 IM-QTOF data.

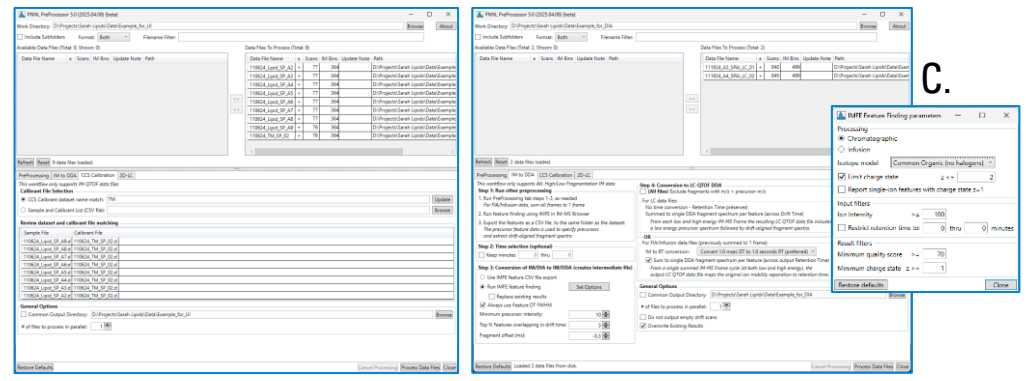

Figure 3. New features in the PNNL PreProcessor help to streamline IM data workflows. A. CCS Calibration tab, B. IM-to-DDA tab calls IMFE from within the PNNL PreProcessor. C. A new window contains IMFE parameters as in IM-MS Browser.

Three new features have been added to the PNNL PreProcessor for version 5.0:

- CCS Calibration tab (Figure 3A) allows tune mix CCS calibration to occur in the PNNL PreProcessor via dataset name matching or using a predefined CSV vs. requiring the user to manually use IM-MS Browser prior to the PNNL PreProcessor. This functionality is also available via command line for further workflow automation.
- The IM-to-DDA tab (Figure 3B) can now run feature finding as part of the processing. This combines two steps in the previous workflow – first, running IMFE in IM-MS Browser to generate a CSV feature list and then second, running the preprocessor to generate the DDA 3D file – into one step for the user. A parameters window (Figure 3C) mimics the one in IM-MS Browser.
- Polygon extraction is now supported for batch processing as shown in Figure 1. Removing unwanted background signal from data files speeds up downstream data analysis processing.

IM to DDA Workflow 1 and 2 Comparison for TG Lipids

The comparison of Workflow 1 and 2 for the IM to DDA conversion is illustrated below using TG lipids as an example. Workflow 2 is depicted in Figure 4. where the TG region of the LC separation is divided into 5 distinct summed data files. Feature finding with Agilent's IMFE (Ion Mobility Feature Extraction) is performed on each data file and then they are each converted to DDA format.

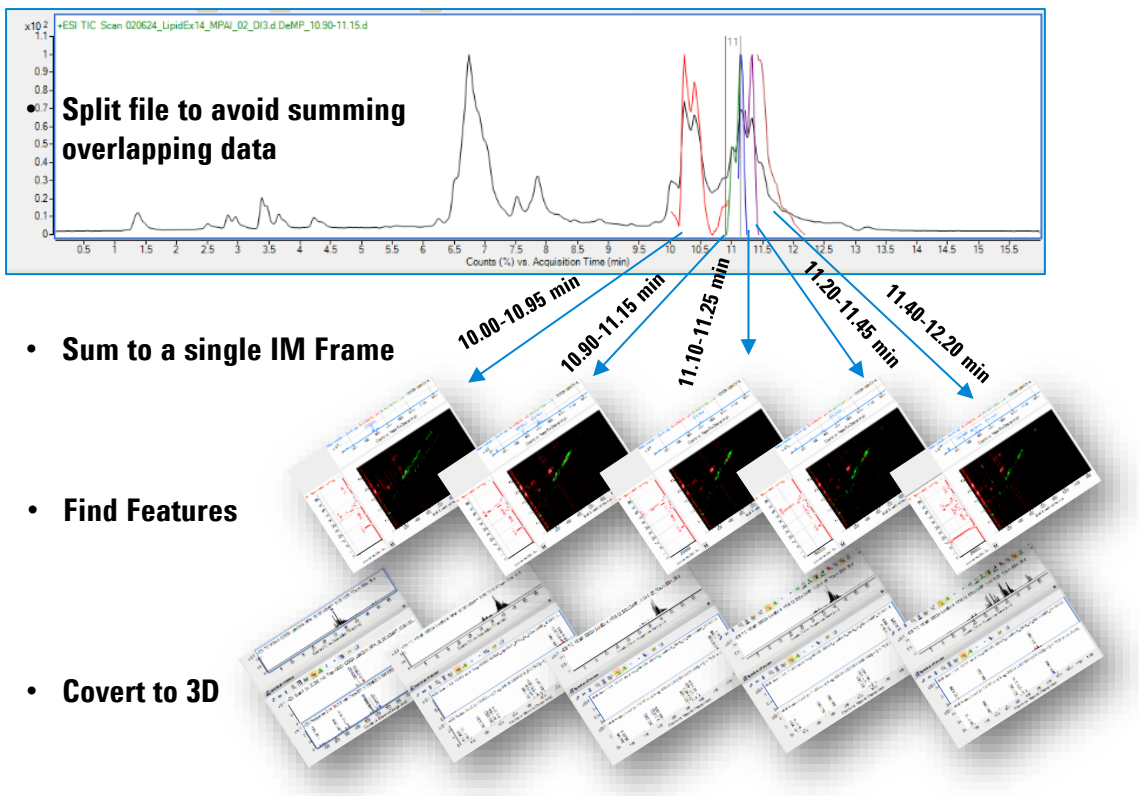

Figure 4. Example of Workflow 2 for IM to DDA Conversion for TG lipids.

The number of TG identifications is shown in Figure 5. When processed as 4D IM data files in MS DIAL[3], the number of TG identifications increases progressively from single pulse to demultiplexed and then to high-resolution demultiplex, as expected. However, when converted to DDA format using Workflow 1, HRdm does not lead to an increase in TG identifications, as all isomers generated by HRdm are collapsed back onto each other during conversion. In contrast, Workflow 2, depicted above, combined with manual data interpretation to combine results across the data files, successfully increases the number of TG lipids as expected with HRdm.

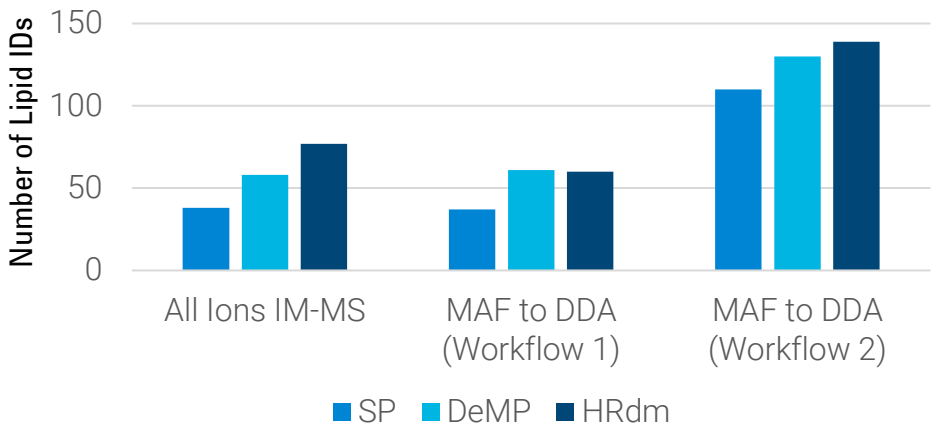

Figure 5. Number of TG lipid identifications for different processing approaches.

BSA Digest Processed with IM to DDA Workflow to Evaluate Streamlined Conversion and IMFE for Peptides

A BSA digest sample was processed via the IM to DDA workflow and sequence coverage was evaluated in MassHunter BioConfirm 12.1. In Figure 6. sequence coverage is shown for the full data file in teal (65.01%) and then for each individual charge state - +1 in orange (57.46%), +2 in blue (30.36%), and +3 in pink (10.46%). IMFE's accuracy in detecting charged species is critical, as missing a precursor in 4D data prevents fragmentation spectrum generation. For uncovered sequence regions, future work will assess whether chimeric interference from AIF or IMFE's failure to detect a feature caused a lack in sequence coverage.

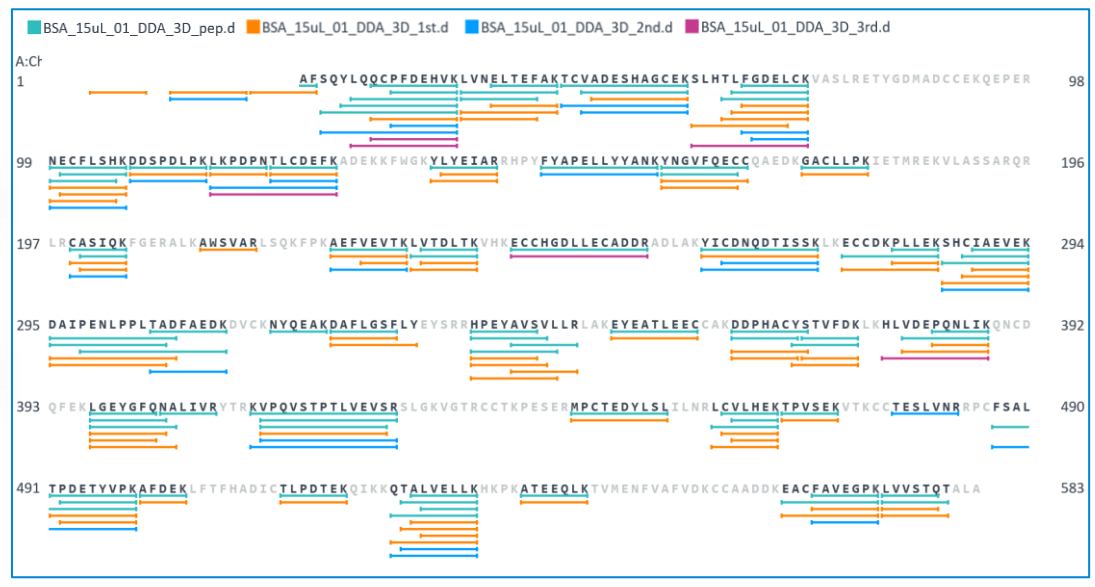

Figure 6. Sequence coverage for BSA digest sample shown according to precursor charge state from IMFE.

In Figure 7, the BSA digest data file is processed using two fragment ion extraction methods. The full width at half max of the precursor ion (teal) yields 63.12% coverage, while the full drift time width (purple) results in 57.8% coverage. Future work will determine if peak tailing in the drift dimension causes larger extraction windows which in turn cause more chimeric interferences in the fragmentation spectra result in lower sequence coverage for the drift time width approach.

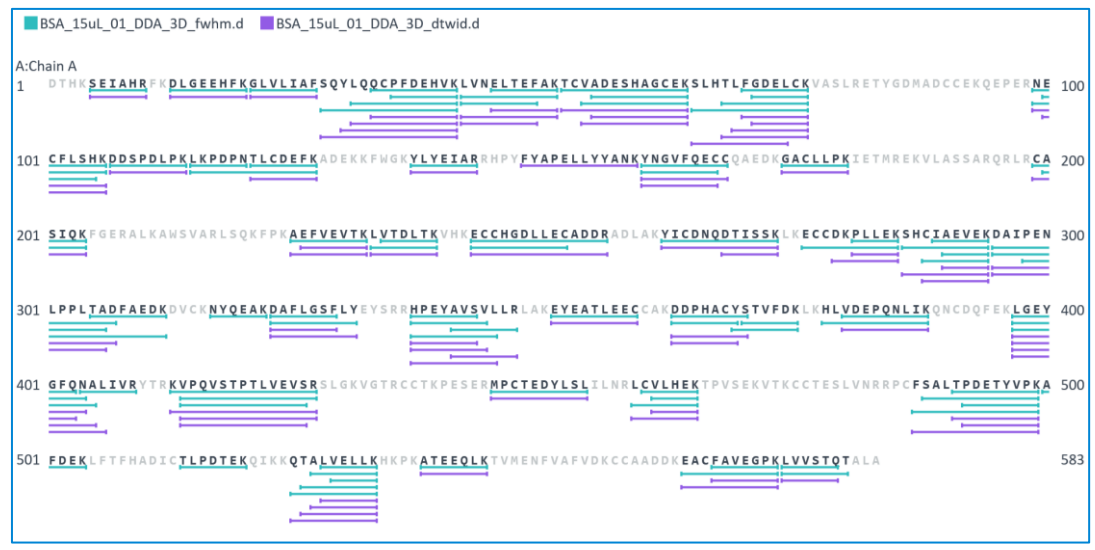

Figure 7. Sequence coverage for BSA digest sample shown for two extraction methods.

Results and Discussion

Polygon Batch Extraction in PNNL PreProcessor Improves Untargeted Feature Finding Time

IM-MS Browser allows exporting polygon regions as filtered IM or QTOF data files, but only one at a time. The PNNL PreProcessor now enables batch processing using the IM-MS Browser-defined polygon, as shown in Figure 8. In this example, PFAS compounds elute in a lower drift space than the complex sample background. Exporting the PFAS polygon improves untargeted feature finding across triplicate files, reducing processing time from 61 to 26 minutes (14 for polygon extraction, 12 for feature finding) on a 3.8 GHz, 6-core, 12-logical-processor system.

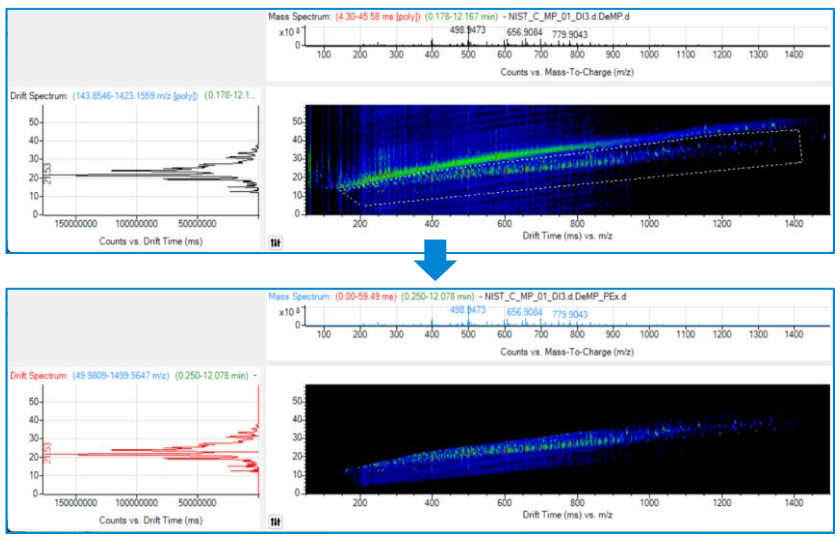

Figure 8. Drift Time (ms) vs. *m/z* plot for a PFAS sample in complex matrix before and after polygon extraction.

To assess IMFE, a targeted list of 45 PFAS compounds was used, with both original and preprocessed data files each missing one (different) compound. Q-scores for single replicates did fall below the single-file threshold of 70. Example of missed or low q-score feature from IMFE are discussed in more detail in the next section.

Feature finding results for Specific PFAS compounds

Perfluorotetradecanoic acid, which has two isomers separable by IM was not found in one of the triplicate data files when processed as the original demultiplexed file as shown in Figure 9. When IMFE was run allowing for single ion features this ion was found. Applying 3 point smoothing in the drift and chromatographic dimensions with the PNNL PreProcessor provided a better peak shape for the second isotope and allowed it be found when single ion features were turned off.

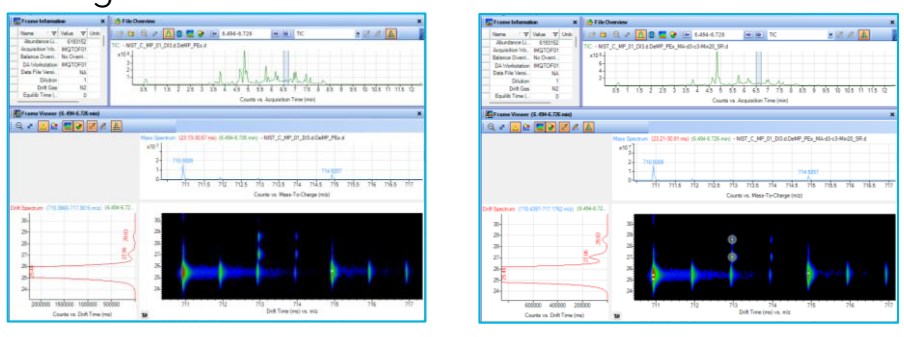

Figure 9. PFAS example where 3 point smoothing in drift and chromatographic dimensions improves IMFE performance.

Feature finding results for PFAS compounds cont.

Perfluorohexanesulfomide (RT 4.6) and Perfluorooctanesulfonic acid (RT 4.8) are present at very high concentrations in the sample set. When processed as the original data file, IMFE failed to identify either of these features in one of the replicates as shown in Figure 10. When the PNNL PreProcessor is used to apply smoothing in the drift and chromatographic regions is applied as well as saturation repair then both PFAS compound are found in the data file.

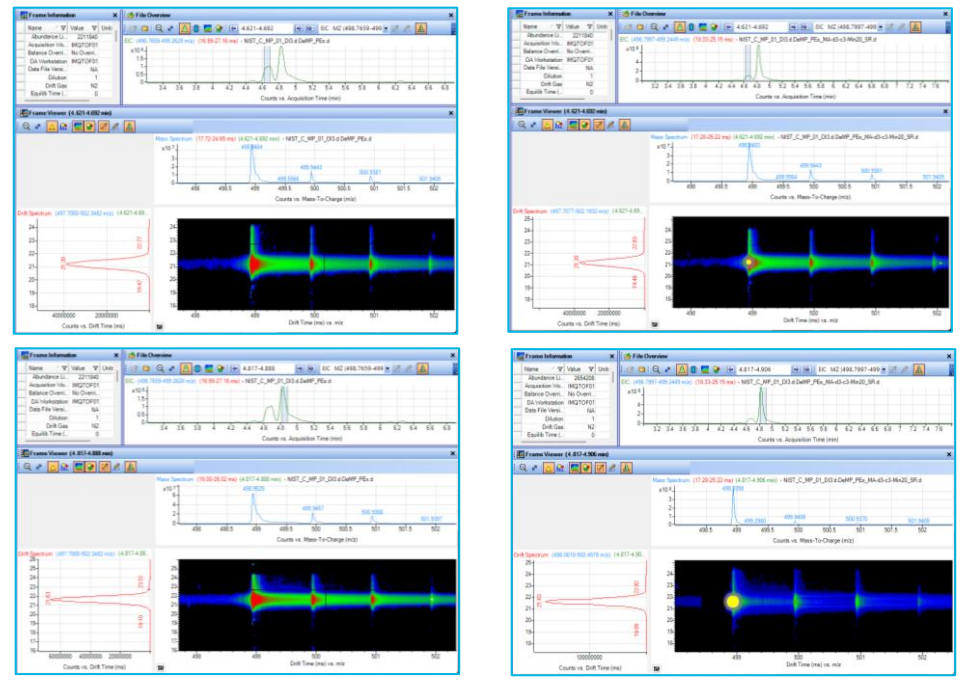

Figure 10. PFAS example where saturation repair improves IMFE performance.

Conclusions

- New features introduced in PNNL PreProcessor 5.0 provide streamlined data processing for LC-IM-MS workflows
- Workflow 2 for IM to DDA conversion must be used to report separate fragmentation spectra for IM isomers
- Initial investigations suggest FWHM provides tighter, more robust windows for fragmentation spectra extraction for IM to DDA
- Understanding why IMFE misses a feature can guide which preprocessing techniques may improve performance

References

[1] Bilbao, A. et. al. Journal of Proteome Research 2022, 21 (3), 798-807.

[2] Stow, S. M. et. al. Journal of the American Society for Mass Spectrometry 2024, 35 (8), 1991-2001.

[2] Tsugawa, H. et. al. Nature Methods 2015, 12 (6), 523-526.

https://www.agilent.com/en/promotions/asms

DE-006598

Published in USA, May 15, 2025

Agilent

Trusted Answers