Application note | 001826

Environmental

A comprehensive software workflow for non-targeted analysis of per- and polyfluoroalkyl substances (PFAS) by high-resolution mass spectrometry (HRMS)

Authors

Juan M. Sanchez and Ralf Tautenhahn

Thermo Fisher Scientific, San Jose, CA

Keywords

PFAS, liquid chromatography, untargeted, environmental, data reduction, Compound Discoverer software, mass spectrometry, high resolution accurate mass (HRAM) spectrometry, non-targeted analysis, per- and polyfluoroalkyl substances (PFAS), PFOS, PFOA, GenX, PFCs, Orbitrap Exploris mass spectrometers

Goal

Provide an overview of the new untargeted PFAS analysis workflow capabilities within Thermo Scientific™ Compound Discoverer™ software

Introduction

The ubiquity and toxicity of a highly stable group of small molecules collectively known as per- and polyfluoroalkyl substances (PFAS) recently garnered concerns among health and environmental regulatory agencies globally.[1] Regulatory monitoring of PFAS has traditionally focused on the development of targeted quantitative methods by LC-MS/MS. These methods are limited in scope due to the lack of available certified reference standards. Over 9,000 known PFAS (with more PFAS being actively discovered) dictate the need for a comprehensive non-targeted analysis of PFAS by high-resolution accurate mass (HRAM).

Numerous individual techniques effective at discriminating PFAS in complex matrices by using intrinsic attributes such as signature product ions, progressive retention times tied to chain length, and CF_2-specific Kendrick mass defect are well documented in the literature.[2-4] Additionally, fluorine's physicochemical attributes, such as a characteristic negative mass defect and the formation of homologous series containing predictable CF_2 patterns resulting from industrial PFAS synthesis techniques, may be exploited to

thermo scientific

simplify the detection and annotation of novel PFAS. Here we present a fusion of the most prominent untargeted PFAS analysis techniques leveraged within a single workflow using Compound Discoverer software as a turnkey solution.

Experimental

Essential elements of a singular, comprehensive PFAS workflow

The Compound Discoverer software PFAS workflow (Figure 1) is a pre-assembled combination of customizable interconnected nodes with parameters optimized for the analysis of PFAS. It can be applied to high resolution accurate mass spectrometry (HRAM) data that has been acquired from a variety of matrix types, such as simple water, complex municipal waste leachate, and biological tissues. Full MS dd-MS^2 data acquisition as well as the availability of at least one blank file and three replicates per sample are recommended. The workflow leverages formula prediction based on HRAM and spectral best fit. Formula prediction is constrained by a maximum of 50 fluorine atoms to provide optimal coverage for the observable chemical space where PFAS reside.

Spectra from authentic PFAS standards are searchable via the Thermo Scientific™ mzCloud™ spectral library as well as the manually curated FluoroMatch Suite database[5,6] of over 700 compound agnostic PFAS signature product ions. The lack of authentic standard availability, sparse coverage in spectral libraries, and limitations with negative mode *in silico* fragmentation are circumvented by this manually curated negative mode signature product ion database, enabling MS^2 matching of PFAS absent from spectral libraries. Accurate mass is also leveraged via searches against the EPA's DSSTOX database via ChemSpider™, a manually curated mass list of 40 noble PFAS compound classes, and an extensive mass list of known and theoretical PFAS. In addition to this, general background subtraction as well as peak quality filters accounting for peak shape and frequency in replicates can be used. For matrices where no blank is available, a suitable sample containing relatively low levels of PFAS may be employed with appropriate modification to the parameters in the Mark Background Compounds node. Analytical approaches compiled from the literature[7] including mass defect filtering thresholds specific to fluorine containing compounds, chemical transformations, and Kendrick mass defect (MD) for the identification of homologous series are also built in. Onboard visualization tools encompassing Kendrick MD plots, molecular networks, and an orthogonal discrimination approach independent of fragmentation, provide in-depth data interrogation enabling the identification of unknown targets for fragmentation in follow-up experiments.

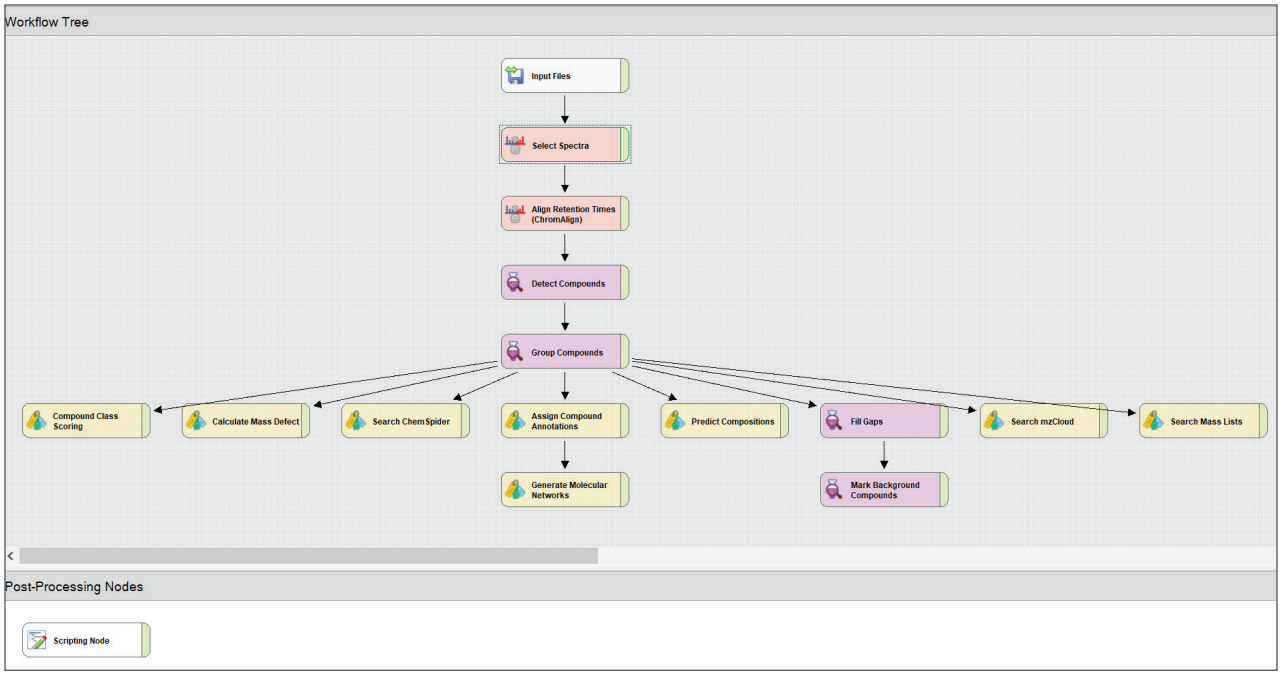


Figure 1. Compound Discoverer workflow tree. This workflow illustrates the nodes used for analysis of PFAS-containing samples as well as their connectivity. The compound class scoring node permits queries against the fine signature fragment and FluoroMatch Suite databases. The Calculate Mass Defect node carries out standard and CF_2 Kendrick mass defect calculations. The Search Mass Lists node enables searching of PFAS mass lists. Assign Compound Annotations assigns the hierarchy for identity assignment from available sources. Predict Compositions applies established formula assignment rules for formula prediction. Search mzCloud enables spectral library searches. Search ChemSpider enables searches within a specified mass tolerance range or with a certain elemental composition. Generate Molecular Networks enables class-based clustering of PFAS based on fragmentation similarity as well as chemical transformations.

Results and discussion

Results and data reduction techniques

Post-processing, insights may be gained from the results view containing overlayed visualizations of chromatograms as well as mass spectra (Figure 2). Step-by-step data processing instructions are available here. Results encompassing retention time, peak areas, formula, and other information pertinent to analysis via LC-MS/MS are displayed in the compounds table and its sub-tables within this view. The number of entries in the Compounds table depends on upstream parameters selected in the workflow nodes such as the peak quality filter, as well as the complexity of the matrix. The coupling of complex matrices such as municipal waste leachate with a low peak intensity threshold is required to avoid the loss of low abundance PFAS, resulting in tens of thousands of entries and making data interpretation challenging. Intelligent experimental design leveraging grouped replicates and blanks enables the use of peak quality filters and blank subtraction to reduce the number of entries. Additional contributing factors for intelligent experimental design include the use of pooled QC samples and internal standards.

Further reduction may be achieved by using additional result filters, such as those listed in Figure 3A, capitalizing on the intrinsic properties of PFAS as well as fragmentation. A mass defect filter leveraging pre-calculated values from the Compounds table was applied to retain PFAS based on optimized mass defect ranges established in the literature. Class coverage was used to ensure that at least three fragments from the experimental data matched with the FluoroMatch Suite database of over 700 manually curated PFAS-related, but not exclusive, product ions serving as a coarse filter. Following this coarse filter, fine filters with lower thresholds may be more confidently applied to retain only compounds matching either the mzCloud library or fine signature fragment database. The fine signature fragment database contains a more limited selection of mostly exclusive PFAS product ions providing specificity. Lastly, the assigned formulas are constrained to contain more than two fluorine atoms, eliminating all non-PFAS compounds from this initial pass. Custom tags as well as the checked compounds column are used to assign confidence to these annotations and mark them for visualization downstream.

Compounds lacking fragmentation data, as well as those whose top formula assignment lacked the required number of fluorine atoms but contained an alternate predicted composition, ChemSpider match, or mass list match fulfilling this requirement, may be re-examined using a fragmentation independent discrimination method adapted from the literature. Briefly, the orthogonal discrimination tool relies on a scripting node to estimate the number of carbons independently from Compound Discoverer software's formula assignment. This calculation[7] uses the measured A0 (first isotopic peak) and A1 (corresponding monoisotopic peak) distribution as inputs to approximate the number of carbons. In turn, this enables the creation of new m/C (molecular mass divided by number of carbon atoms) and md/C (mass defect divided by number of carbon atoms) ratios. When

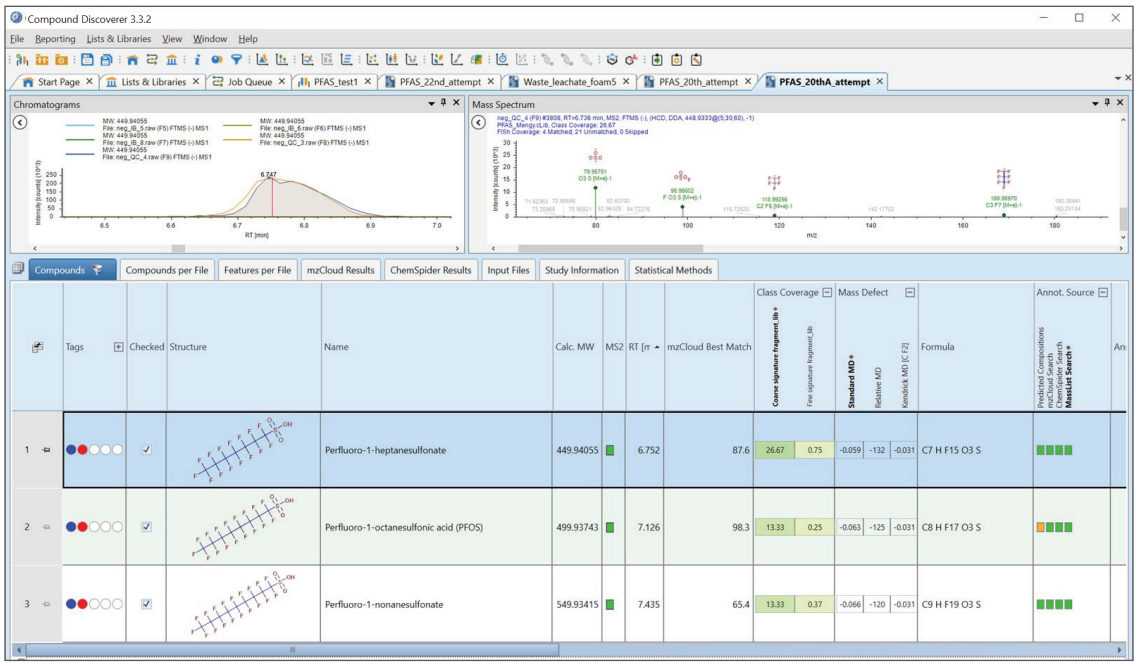
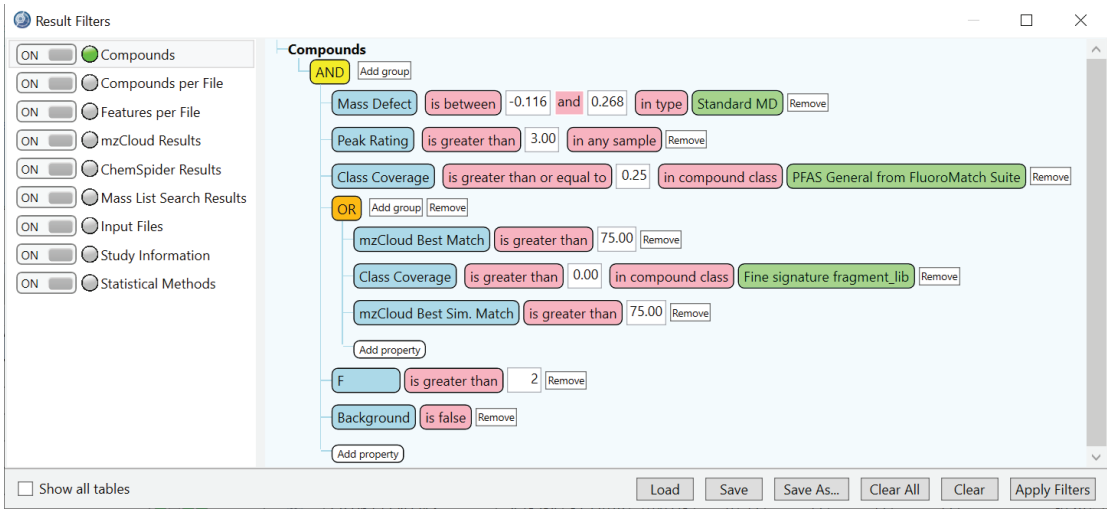


Figure 2. Results view. The Results view provides a way to interact with the analysis data and organize information into a table with columns that contain pertinent information from an LC-MS/MS analysis, such as retention time, *m/z*, fragmentation library scores, and formulas among other items. This view also shows visualizations for overlaid chromatograms as well as mass spectrum information.

these ratios are plotted via Compound Discoverer software's onboard visualization tools, PFAS containing molecules cluster on the bottom right quadrant as shown in Figure 4. This tool is not suitable for compounds where no A1 is detected as they plot to the origin despite some targets being putatively identified as PFAS via the previous fragmentation-based filtering scheme as well as known retention times. PFAS where multiple carbons are substituted by oxygen or other atoms in the structure's backbone also fall outside the region of interest. These compounds tend to appear higher to the upper left than normal PFAS with unsubstituted CF_2 chains. To optimize the power of this tool, edge cases that broaden the region of interest are omitted. The clustering seen here also validates the original findings since the compounds that survived fragment-based filtering and received check marks cluster only on the bottom right and are displayed as light blue circles. Checked compounds appearing outside this window should be scrutinized and the reason for their location understood.

Filter 3B describes a filtering schema focusing only on the region where most of the traditional PFAS reside. This filter also accounts for all other potential formulas containing more than two fluorine atoms to prevent formula misassignments and discover previously missed targets. Knowledge of retention time trends for homologous series, branched isomers, manual assessment of key product ions such as SO_3- in perfluoro sulfonic acids, and general analytical chemistry knowledge is applied here to conserve only targets of interest and assign them a checked status. Data reduction is achieved in a water sample from 373 compounds with no filters to 28 and 60 compounds utilizing the fragment-based filtering and fragmentation independent orthogonal discrimination filtering approaches, respectively (Figure 5). Compounds remaining in the plot after using the fragment-based filtering approach (Figure 5B) are retained when the more permissive fragmentation independent orthogonal discrimination filter (Figure 5C) is applied. This data reduction approach narrowed the compound entries in a complex matrix such as municipal waste leachate from 14,000 to less than 100 facilitating further onboard visualization and data interpretation.

A



B

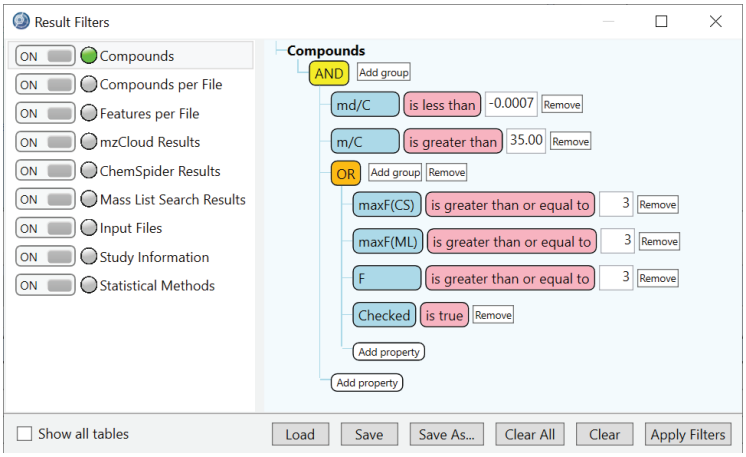


Figure 3. Result filters. Combinations of result filters with logical gates are displayed for data reduction. Figure 3A shows the fragment-based filtering approach leveraging mzCloud and compound class libraries. Figure 3B shows filters amenable to a fragmentation independent orthogonal QC approach.

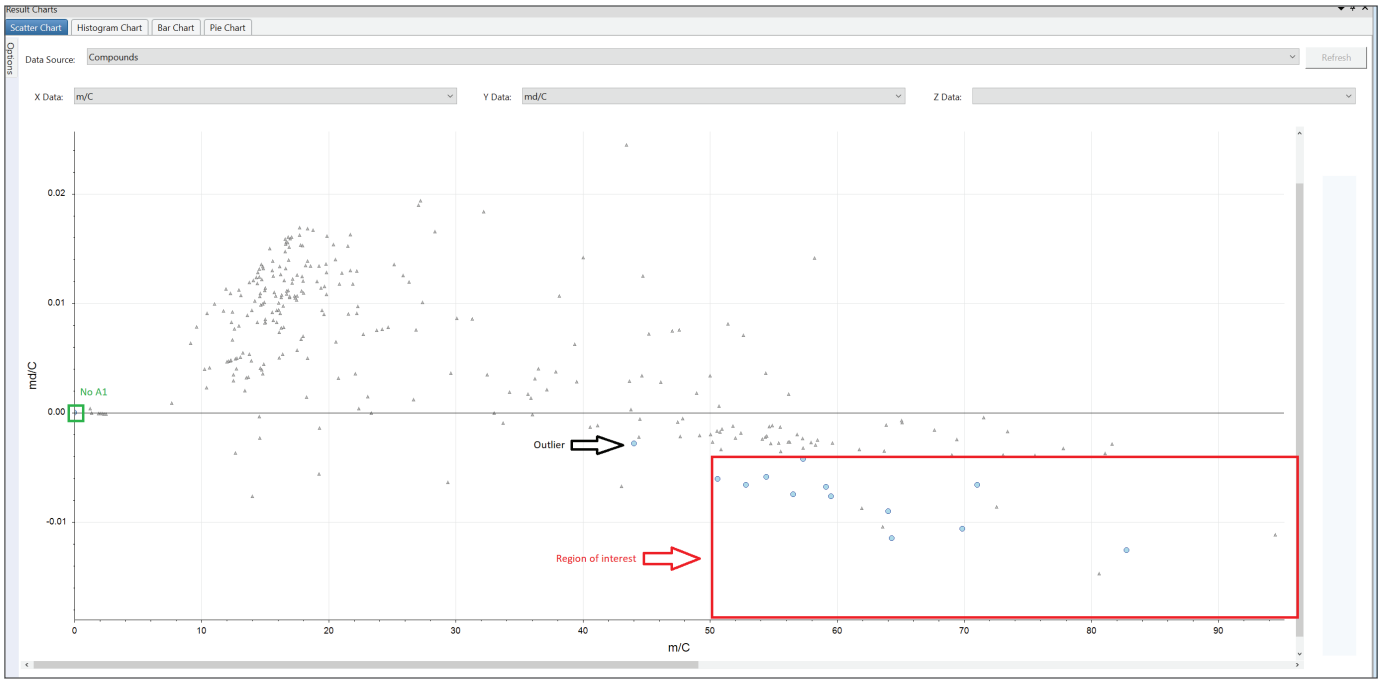


Figure 4. Orthogonal PFAS discrimination. Visualization of the orthogonal PFAS discrimination approach. The region of interest is outlined by a red rectangle. Compounds incompatible with this approach, which lack an A1, are outlined by green square located at the origin. An outlier with multiple carbon atom substitutions by oxygen within the main PFAS chain is denoted by a black arrow.

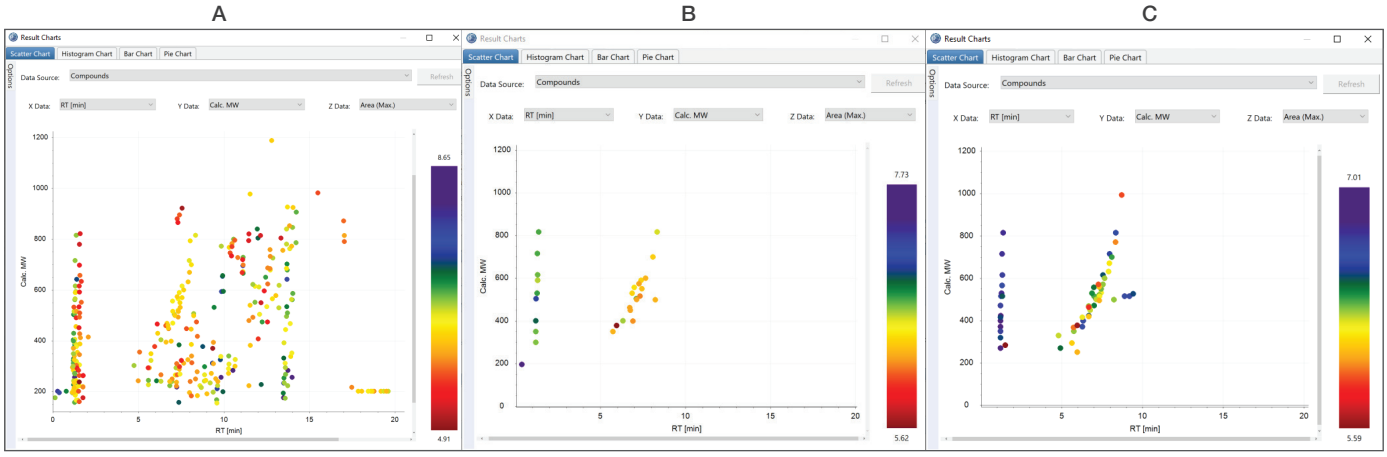


Figure 5. Multistep filtering visualization. Figure 5A shows no filters applied, 373 compounds displayed. Figure 5B shows fragment-based filtering approach, 28 compounds retained. Figure 5C shows fragmentation independent orthogonal discrimination filter applied, 60 compounds retained. The Z axis is log-transformed.

Workflow performance Orbitrap Exploris 120 mass spectrometer

To assess the performance of this workflow, data was obtained from water samples spiked with 13 PFAS standards at EPA relevant concentrations using a Thermo Scientific™ Orbitrap Exploris™ 120 mass spectrometer. A Thermo Scientific™ Vanquish™ Core Binary UHPLC system with PFAS Retrofit Kit was used for chromatographic separation. The data was processed using the PFAS Compound Discoverer software workflow. For this matrix, a sensitivity of 91% was calculated using published formulas.[8] Selectivity could not be calculated in a similar manner due to the lack of blank data.

Visualizing meaningful compounds onboard using Compound Discoverer software

Beyond the detection of PFAS, onboard visualization capabilities including volcano plots, principal component analysis, Kendrick MD plots, and molecular networks enable the transition from discovery to insight. The identification of homologous series is an important aspect of PFAS analysis, providing increased confidence in assigned identifications while also informing on the target's provenance. Compound Discoverer software uses integrated Result Charts to plot all data contained within the multiple tables visible in the results. Data from the Compounds

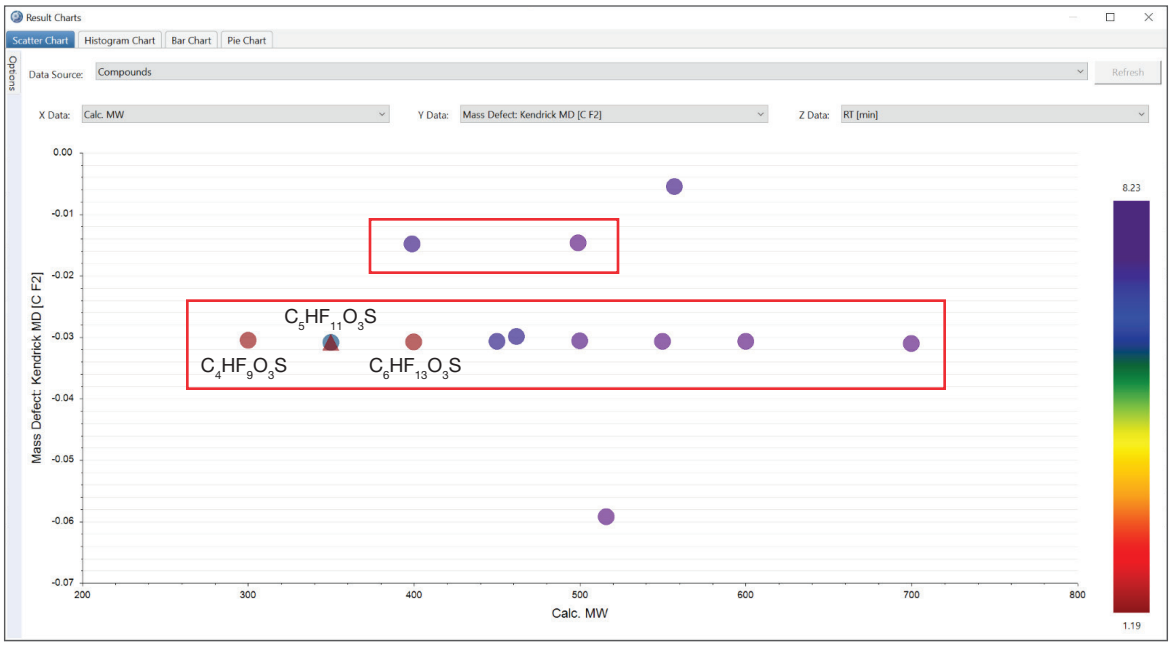


Figure 6. Kendrick mass defect. CF_2 Kendrick mass defect is visualized as a built-in result chart. Part of a homologous series is labeled, from left to right: perfluorobutanesulfonic acid, $C_4HF_9O_3S$; perfluoropentanesulfonic acid, $C_5HF_{11}O_3S$; and perfluorohexanesulfonic acid, $C_6HF_{13}O_3S$.

table are plotted in three dimensions using the Kendrick MD of approximately 50 Da or exactly one CF_2 to elucidate the presence of homologous PFAS series (Figure 6). Homologous PFAS series will share the same Kendrick MD but differ in molecular weight by 50 Da and have increasing RT based on PFAS chain length. The retention time is color coded as a third dimension to provide a simple verification of this trend.

Two homologous series are identified at Kendrick MD [CF_2] of -0.03 and -0.015. On the longer series, there is one overlapping PFAS with alternate branching not following the retention time trend with a molecular weight of 349.9471 Da corresponding to perfluoropentanesulfonic acid. This overlap is resolved by plotting either of the overlapping compounds as a triangle. The signature homologous series patterns are observable in Figure 6. To the left of this PFAS compound containing a five-carbon chain, perfluorobutanesulfonic acid containing a four-carbon chain may be found with a loss of 50 Da and to the right perfluorohexanesulfonic acid containing a six-carbon chain with a gain of 50 Da. This visualization is linked to the Compounds table, enabling the selection of groups of compounds and assignment of checked status for additional investigation.

A few other putatively identified PFAS not belonging to either series are also displayed here. Their relationship may be examined by forming a molecular network. The molecular networking node includes a pre-selected CF_2 chemical transformation to simplify the grouping of homologous series. This node accounts for similarities between MS^2 spectra and is capable of clustering PFAS based on class. Class-based clustering for perfluorosulfonic acids and perflurosulfonamide is shown (Figure 7A). The cluster of perflurosulfonic acids, retention times, MS^2 spectra match scores, as well as their respective PFAS chain shortening transformation is showcased by the link between perfluorononane sulfonate and perfluroctane sulfonate (Figure 7B). The increasing retention time for these two PFAS from 7.126 to 7.435 minutes for the 8 and 9 carbon chains, respectively, further validates Compound Discoverer software's findings.

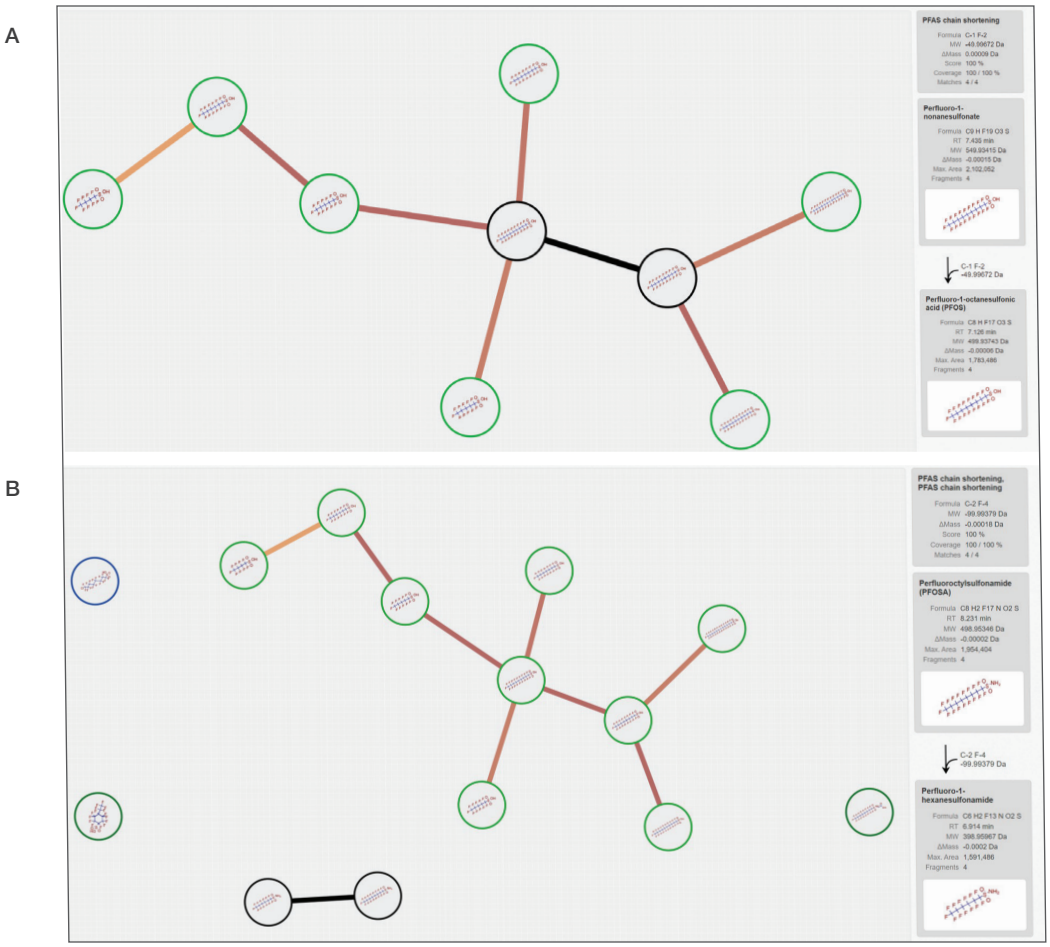


Figure 7. Molecular networking. Figure 7A shows molecular networks for homologous series clustering by class. Figure 7B shows a cluster containing several linked perfluoro sulfonic acid homologues.

Conclusion

Compound Discoverer software is a powerful platform providing a comprehensive turnkey solution for the untargeted analysis of PFAS in complex matrices. Access to the mzCloud spectral library to provide similarity searches—as well as the potential to leverage in silico fragmentation in positive mode and matching against a manually curated compound class library of PFAS signature product ions in negative mode—provides unparalleled capabilities. The incorporation of analysis techniques and best practices from the literature, compilation of PFAS databases, a custom scripting node (available here) for orthogonal discrimination, and a myriad of onboard visualization tools enables a simplified approach for analyzing this concerning class of small molecules. When challenged with analyzing PFAS, the untargeted PFAS workflow available in Compound Discoverer software version 3.3 SP2 can provide labs with an integrated solution to achieve meaningful insights.

References

1. Liu, Y.; D'Agostino, L.A.; Qu, G.; Jiang, G.; Martin, J.W. High-Resolution Mass Spectrometry (HRMS) Methods for Nontarget Discovery and Characterization of Poly- and per-Fluoroalkyl Substances (Pfass) in Environmental and Human Samples." TrAC Trends in Analytical Chemistry 2019, 121, 115420. https://doi.org/10.1016/j.trac.2019.02.021
2. Getzinger, G.J.; Higgins, C.P.; Ferguson, P.L. Structure Database and in Silico Spectral Library for Comprehensive Suspect Screening of per- and Polyfluoroalkyl Substances (Pfass) in Environmental Media by High-Resolution Mass Spectrometry. Analytical Chemistry 2021, 93(5), 2820–27. https://doi.org/10.1021/acs.analchem.0c04109
3. Gonzalez de Vega, R.; Cameron, A.; Clases, D.; Dodgen, T.M.; Doble, P.A.; Bishop, D.P. Simultaneous Targeted and Non-Targeted Analysis of per- and Polyfluoroalkyl Substances in Environmental Samples by Liquid Chromatography-Ion Mobility-Quadrupole Time of Flight-Mass Spectrometry and Mass Defect Analysis. Journal of Chromatography A 2021,1653, 462423. https://doi.org/10.1016/j.chroma.2021.462423
4. Jia, S.; Marques Dos Santos, M.; Li, C.; Snyder, S.A. Recent Advances in Mass Spectrometry Analytical Techniques for per- and Polyfluoroalkyl Substances (PFAS). Analytical and Bioanalytical Chemistry 2022, 414(9), 2795–2807. https://doi.org/10.1007/s00216-022-03905-y
5. Koelmel, J.P.; Paige, M.K.; Aristizabal-Henao, J.J.; Robey, N.M.; Nason, S.L.; Stelben, P.J.; Li, Y. et al. Toward Comprehensive per- and Polyfluoroalkyl Substances Annotation Using FluoroMatch Software and Intelligent High-Resolution Tandem Mass Spectrometry Acquisition. Analytical Chemistry 2020, 92(16), 11186–94. https://doi.org/10.1021/acs.analchem.0c01591
6. Koelmel, J.P.; Stelben, P.; McDonough, C.A.; Dukes, D.A.; Aristizabal-Henao, J.J.; Nason, S.L.; Li, Y. et al. Fluoromatch 2.0—Making Automated and Comprehensive Non-Targeted PFAS Annotation a Reality. Analytical and Bioanalytical Chemistry 2021, 414(3), 1201–15. https://doi.org/10.1007/s00216-021-03392-7
7. Kaufmann, A.; Butcher, P.; Maden, K.; Walker, S.; Widmer, M. Simplifying Nontargeted Analysis of PFAS in Complex Food Matrixes. Journal of AOAC International 2022, 105(5), 1280–87. https://doi.org/10.1093/jaoacint/qsac071
8. Jacob, P.; Wang, R.; Ching, C.; Helbling, D.E. Evaluation, Optimization, and Application of Three Independent Suspect Screening Workflows for the Characterization of Pfass in Water. Environmental Science: Processes & Impacts 2021, 23(10), 1554–65. https://doi.org/10.1039/d1em00286d

Acknowledgment

Thermo Fisher Scientific wishes to thank the Higgins lab at the Colorado School of Mines for sharing their data.

Learn more at **thermofisher.com/compounddiscoverer**