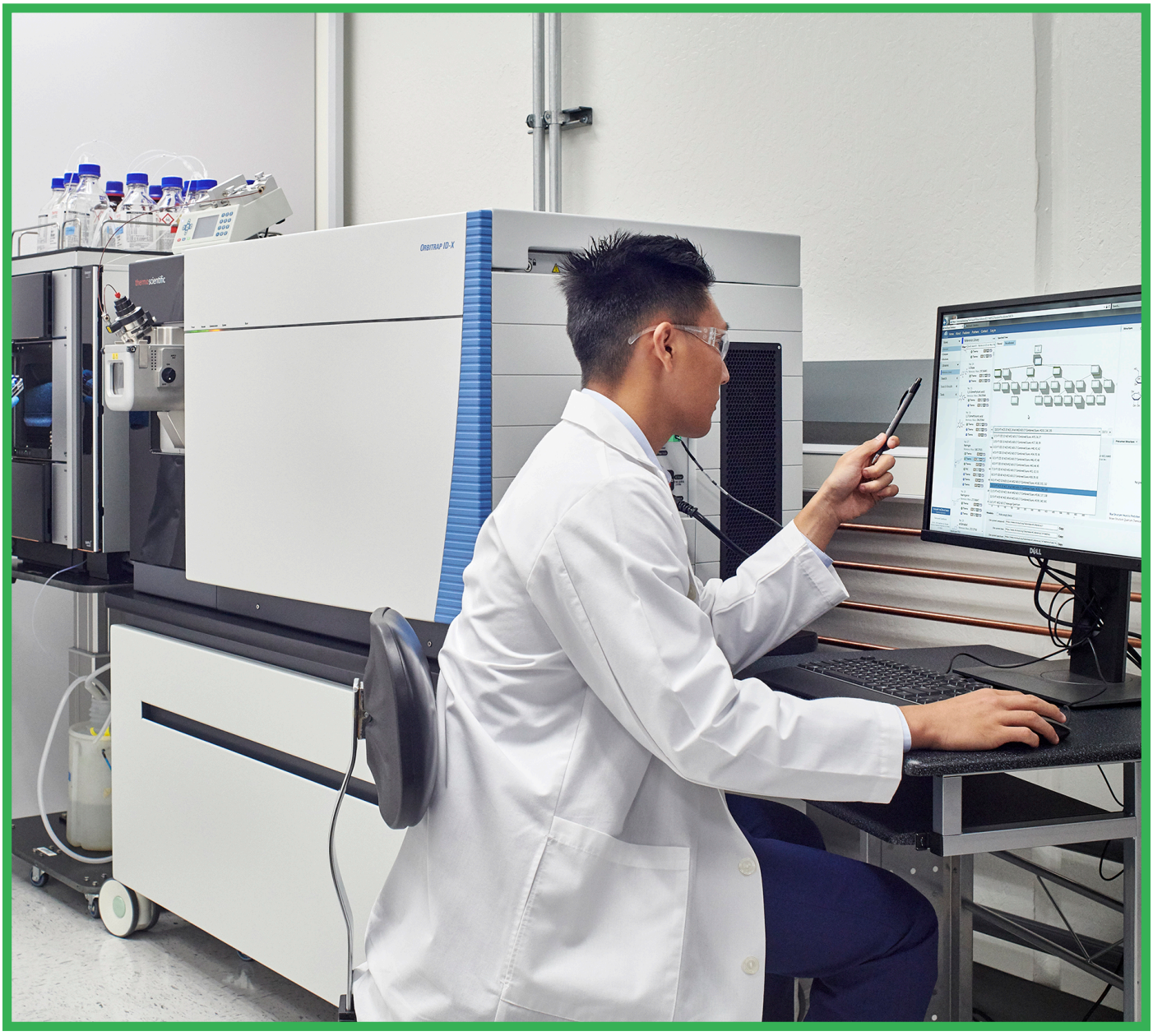
The ultimate workflow for small molecule discovery

Transform your small molecule characterization and identification studies, regardless of sample type



BROUGHT TO YOU BY INDEPENDENT SCIENCE PUBLISHER

SelectScience®

IN PARTNERSHIP WITH

ThermoFisher SCIENTIFIC

Introduction

Discovery-based small molecule research has re-emerged as a key sector due to the creation and expansion of many routine applications. Targeted screening and quantitative workflows may only profile a fraction of the sample-specific compounds, potentially missing critical factors, whether this means limited coverage of metabolites and degradants, outstanding unknown or unexpected pesticides, or the misidentification or characterization of natural products that could potentially provide health benefits.

As a result, there is a drive to expand knowledge bases on representative samples by performing global identification and characterization. The introduction of more sensitive and selective mass spectrometers enables researchers not only to identify unknown compounds of interest in complex samples, but also to utilize multistage mass spectrometry (MS^n) data for structural characterization.

Contents

- Untargeted metabolomics
- Lipid annotation and profiling
- Natural products
- Pesticide screening
- Top tips
- Featured products
- Additional resources

However, simply increasing the mass spectral performance or MS^n capabilities does not single-handedly solve the small molecule bottleneck. The primary challenge is discerning which of the measured precursor features are attributed to compounds of interest and which are attributed to the background matrix, creating a need for

the collection of high-quality MS and tandem mass spectra, as well as robust data processing to confidently assess experimental hypotheses.

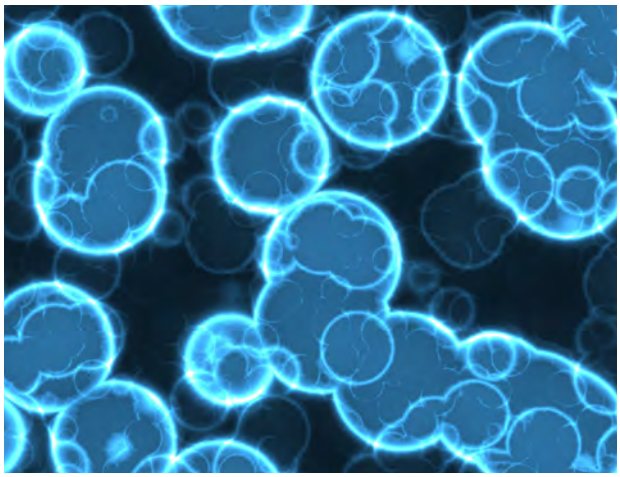
Additionally, each application may require different strategies based on the type of matrix, the expected number of compounds of interest, and dynamic range relative to the matrix. Therefore, the experimental tools contained in a workflow must address the broad scope of small molecule research, while maintaining a high level of effectiveness and efficiency. Above all, these workflows must be easy to set up and use for complex data acquisition strategies.

The Orbitrap ID-X™ Tribrid™ mass spectrometer from Thermo Fisher Scientific is designed to be the optimal solution for this rapidly developing research sector. The Tribrid™ architecture aims to extend the dynamic detection range, increase acquisition speed, provide unique screening capabilities, and maintain high mass measurement accuracy at all MS^n stages (MS^2 and higher MS^n stages), including low- (CID) or high-energy collision-induced dissociation (HCD). The ID-X workflow utilizes AcquireX intelligent MS^n data acquisition methods to manage comprehensive and efficient data acquisition. In addition, a combination of the Thermo Scientific™ Compound Discoverer™ software and Thermo Scientific™ Mass Frontier™ Spectral Interpretation software (and Lipid Search) promises unique data processing strategies that utilize this high-quality MS^n data for exhaustive characterization, identification and structural annotation, and relative quantitation and profiling.

This application-based eBook provides an overview of the experimental requirements and goals for a diverse subset of small molecule applications, including how the Orbitrap ID-X workflow is designed to adapt for maximum performance and efficiency, regardless of the sample type.

Here we highlight four different application areas:

- Untargeted metabolomics
- Lipid annotation and profiling
- Natural products
- Pesticide screening



Untargeted metabolomics

The study of untargeted metabolomics focuses on the comprehensive detection of endogenous metabolites to generate a metabolic profile for a given biological system. This type of study requires the detection and MS^n sampling of all metabolites covering a wide range of structures, molecular weights, and expression levels. Differential expression analysis is often performed to determine which metabolites are affected by the biological stimulus and identify key pathways with the aim of discovering potential biomarker panels or targeted methods.

The overall effectiveness of untargeted metabolomic studies is dictated by how accurately you can characterize metabolites. Difficulties arise due to the potential number of metabolites associated with the most complex biological matrices. Additionally, each metabolite may be ionized with a range of different adducts, which can dilute the overall signal, increase the complexity of the full-scan mass spectrum, and reduce the efficiency of data dependency. The more comprehensive the metabolite identification, the greater the potential insight into metabolic pathways and mechanisms that underlie health and disease states.

Application highlight:
In this application note, a comprehensive workflow for the AcquireX Deep Scan method of unknown metabolite detection and annotation is described. This application note highlights the critical role untargeted

metabolomics plays in understanding the molecular underpinnings of biological systems and demonstrates how unknown compound annotation can be accelerated with confidence, from spectra to structure, in untargeted metabolomics experiments.

Lipid annotation and profiling

Lipids play an important role in many biological functions and current research focuses on developing LC-MS-based global profiling methods for phenotyping composition and relative abundance based on external influencing factors such as diet, age, and health.

There are two primary challenges associated with lipid data acquisition:

1. Identifying sample-specific lipids relative to the background matrix
2. Acquiring tandem MS data containing diagnostic product ions to enable confident annotation

The structure of lipids may result in minor differences between compounds, from small changes in aliphatic chain length to degrees and positions of unsaturated C-C bonds. Therefore, even UHPLC separation can result in co-isolation of similar lipid structures. The co-elution of isobaric or isomeric lipids can inhibit detection of MS/MS-based structure-specific product ions required for annotation. In addition, higher-energy collisional dissociation (HCD) often results in very predictable fragmentation with very few product ions, further reducing detailed structural annotation.

An additional challenge for lipid studies is the lack of comprehensive software solutions available that merge identification, annotation, and relative quantitation to support the study size needed for confident biological analyses.

Application highlight

This application note describes a novel workflow that leverages the intelligent MS capabilities of the Orbitrap ID-X and AcquireX method to maximize lipid profiling and annotation. The incorporation of exclusion and inclusion lists shows increased LC-MS^n acquisition efficiency, providing the high-resolution (120,000 @ *m/z* 200) data needed to ensure separation between isobaric lipids and potential matrix ions. This profiling method also combines HCD and on-resonance CID at different MS^n stages to increase fragmentation and therefore increase annotation confidence.

Natural products

The study of natural products has seen recent growth due to the identification, verification, and validation of compounds that elicit biological effects similar to synthetic drugs. One group of plant metabolites, flavonoids, are thought to be particularly beneficial to human health as powerful antioxidants with anti-inflammatory and immune system benefits.

Flavonoids have chemically similar core structures that can be modified by many reactive groups at numerous carbon core sites, in addition to secondary

metabolic pathways that generate O- and C-glycosylation. This has led to the creation of over 10,000 reported flavonoids and makes unknown flavonoid annotation particularly difficult.

There are very few commercially available standards that can be used to generate reference spectra for unknown flavonoid identification and the structure and corresponding modifications of flavonoids can limit the production of diagnostic product ions that are necessary for confident structural annotation. This lack of well-annotated spectral libraries means researchers often perform manual assignment of fragmentation data. The requirement of substantial *a priori* knowledge significantly slows down and limits this research.

Application highlight

In this application note, a novel structure-based ion tree acquisition scheme is implemented to address the challenges of unknown flavonoid annotation. This intelligent MS^n method, as performed on the Orbitrap ID-X Tribrid mass spectrometer, uses real-time product ion spectral analysis to determine when to acquire MS/MS spectra using low- or high-energy dissociation and when to extend tandem mass spectral acquisition to higher MS^n orders, increasing the acquisition efficiency and enabling more unknown flavonoids to be detected and annotated.

Pesticide screening

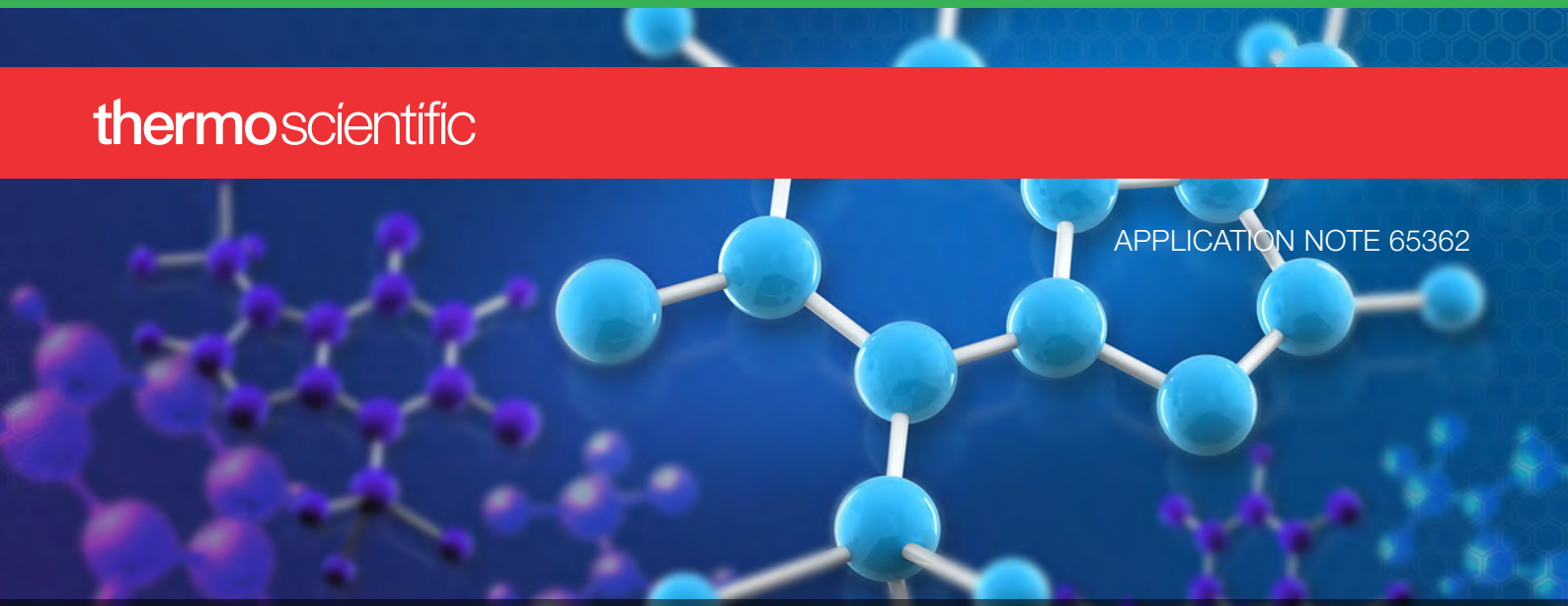
With the globalization of food sources and the growth of pesticide development and application, maximum accepted residue levels are regulated to protect both human health and the environment. These established regulations have created a need for the development of LC-MS-based analytical methods to confidently detect pesticides.

Established multiresidue pesticide screening methods must address a series of challenges. The number and diversity of structures requires LC-MS methods that can detect large numbers of pesticides per method, handle many different matrices in an efficient manner, and enable unknown detection and annotation. Methods must also be capable of detecting these pesticides at low levels (ppb to ppt). However, despite the use of extraction kits, resultant samples may still contain matrix compounds increasing the complexity. Therefore, newly created detection methods must be able to maintain the required sensitivity and selectivity to only measure ion signals attributed to the pesticide and avoid potential false positive/negatives.

Application highlight

This application note presents a multiresidue pesticide screening method, using non-targeted LC-MS acquisition and high-resolution accurate mass methods on an Orbitrap ID-X Tribrid mass spectrometer. A set of 250 pesticides were spiked into the same matrix at levels ranging from 0.01 to 200 ppb to demonstrate automated data acquisition, efficiency and reproducibility in detection, and the product ion spectral quality necessary to minimize false-positive and false-negative results.

thermo scientific



APPLICATION NOTE 65362

Accelerated unknown compound annotation with confidence: from spectra to structure in untargeted metabolomics experiments

Authors
Amanda Souza, Ioanna Ntai, and Ralf Tautenhahn,
Thermo Fisher Scientific

Keywords
Untargeted metabolomics, nontargeted metabolomics, discovery, metabolite profiling, unknown identification, compound annotation, qualitative analysis, elemental composition, database search, spectral library search, Compound Discoverer software, mzCloud spectral library, HighChem Fragmentation Library, high-resolution accurate-mass MS, Orbitrap ID-X Tribrid mass spectrometer, mzLogic algorithm

Goal

Use Thermo Scientific™ Compound Discoverer™ 3.0 software to process untargeted metabolomics for unknown compound annotation. Demonstrate the utility of analyzing multiple analytical measurements from mass spectral data to apply a consensus approach across different annotation sources for confident annotation assignments.

Introduction

Untargeted metabolomics aims to comprehensively detect endogenous metabolites to generate a metabolic profile of a given biological system. This unbiased approach is often without a priori knowledge of the molecular make-up of a sample. Untargeted metabolite profiling globally provides the potential association of a metabolite or set of metabolites to a biochemical phenotype that can be further associated with metabolic pathways and biological function. These associations are insightful in an array of research and discovery settings such as defining signatures of disease and mechanisms of cellular function in human health, defining metabolic response in plant and animals resulting from a changing environment, or evaluating food composition in a commercial setting. The application of untargeted metabolomics plays a critical role in understanding the molecular underpinnings in biological systems.

ThermoFisher SCIENTIFIC

One of the greatest challenges in untargeted metabolomics analysis by mass spectrometry (MS) is the identification of unknown compounds.[1-4] The gold standard in the metabolomics field is to confirm the identification of an unknown analyte by comparing the generated mass spectra against that of a purified reference standard. Yet, there are numerous instances where this is not possible. It could be that the reference standard is commercially unavailable, and therefore requires chemical synthesis. Another instance may require isolation and purification from a select biological source. The most challenging scenario is when the unknown compound is truly novel and has yet to be characterized by researchers. Despite these practical setbacks, modern mass spectrometers and data processing tools provide the means for confident annotation of unknown compounds.[5]

Specifically, untargeted metabolomics analysis using high-resolution accurate-mass (HRAM) tandem Orbitrap™ mass spectrometry generates multiple analytical measurements that when taken collectively via consensus evaluation, builds confidence in the compound annotation or compound class association (Table 1). These multiple measurements start with the acquisition of ultra-high resolution mass spectra to distinguish between molecular ions of closely related mass and detection of associated adduct ions, isotope pattern, and, of particular importance, isotope fine structure. This information, combined with accurate mass measurements, provides confident elemental composition prediction. The elemental formulae are then used for chemical structure database searching.

Next, fragmentation spectra provide an added level of knowledge about the unknown compound. Fragmentation spectra are used to search against spectral libraries. Fragment ions can also be compared against theoretical fragments of putative structures by *in silico* fragmentation prediction. Multi-stage fragmentation (MS^n) by ion trap instruments enables further structural characterization by establishing product ion relationships. Lastly, the novel mzLogic™ algorithm uses actual fragmentation spectra to prioritize putative chemical structure candidates. Collectively, this information facilitates the unknown annotation process by reducing the number of possible candidate compounds determined by the data and increases the confidence in annotation assignments.

The information-rich data generated by Thermo Scientific™ Orbitrap™ mass spectrometers must be efficiently extracted in a single data processing pipeline that streamlines the unknown compound annotation process and presents data in an intelligible format. Thermo Scientific Compound Discoverer 3.0 software expedites the annotation assignment process in untargeted metabolic profiling experiments from HRAM Orbitrap data using numerous annotation tools to fully integrate multiple analytical measurements. To start, the software predicts elemental composition using the power of HRAM Orbitrap data with isotope fine structure and incorporates fragmentation information. The elemental composition predicted from spectral data is then used to search against a chemical database for matching potential structures. The Compound Discoverer software is fully integrated with the ChemSpider™ chemical structure database providing access to over 250 data sources and

Table 1. Multiple analytical measurements for unknown compound annotation. Data types include MS^1, MS^2 and MS^n mass spectral information along with chromatographic separation.

Data type	Measured property	Information
Accurate mass	Monoisotopic mass	Elemental composition (mass tolerance)
Adducts	Accurate mass difference	Assignment of molecular species (M+H, M+NH_4, 2M+H)
Isotopic pattern	Isotope distribution	Constrain possible elemental formulas
Ultra-high resolution	Isotopic fine structure	Separate isobaric species Direct confirmation of isotopes: ^{13}C, ^{15}N, ^{34}S, ^{18}O, ^{2}H Confirm elemental formula (isotopes and ratios)
MS^2	Product ions Neutral losses	Sub-structures and their elemental composition Compound class (signature neutral loss or product ions)
MS^n	Product ion relationships	Mass spectral tree (precursor ion fingerprinting) Collision energy profile (breakdown curves)
LC-MS	Polarity Hydrophobicity	Retention time order—differentiate isomers Compared to literature/reference standards

68 million chemical structures representing a vast chemical space of endogenous and exogenous origin (http://www.chemspider.com/). Furthermore, corresponding MS^2 fragmentation spectra are searched against the mzCloud™ spectral library, which hosts an extensive array of HRAM MS^n fragmentation spectra (https://www.mzcloud.org/).

For compounds lacking direct spectral library matches, the mzLogic algorithm takes advantage of actual experimental fragmentation data to prioritize candidate compounds resulting from the ChemSpider database search. The mzLogic algorithm utilizes sub-structure annotations from the mzCloud library mapped back to the experimental fragment ions to rank the most probable chemical structures in the ChemSpider database results list. Using actual data to rank chemical structures allows users to focus on reasonable compound candidates. Additionally, the Fragment Ion Search (FISh) scoring algorithm incorporates *in silico* fragmentation of a proposed chemical structure to explain fragment ions structures based on literature-defined chemical reactions using the HighChem™ Fragmentation Library™.

Building cumulative evidence across different annotation sources for each individual compound enables consensus evaluation moving toward greater confidence, which in turn gives rise to greater certainty in the subsequent metabolic pathway analyses. If a purified standard is indeed available, the Compound Discoverer software considers chromatographic retention times and spectral matching using a customizable, in-house library to confirm compound identification. It should be noted that while the scope of this application note is to demonstrate unknown compound annotation, the Compound Discoverer software is a complete data analysis program for untargeted metabolomics experiments that operates by first applying a data reduction strategy to generate meaningful compounds that are experimentally related and subsequently providing statistical capabilities geared for differential analysis, visualization tools, and pathway analysis.[6,7]

Here we demonstrate the use of the Compound Discoverer 3.0 software for confident annotation of unknown compounds using the aforementioned annotation tools with an untargeted metabolomics analysis. The NIST Standard Reference Material (SRM) 1950, Metabolites in Frozen Human Plasma, was analyzed with a novel, automated acquisition strategy to generate more fragmentation spectra using the Thermo Scientific™ Orbitrap ID-X™ Tribrid™ mass spectrometer coupled to a Thermo Scientific™ Vanquish™ Horizon Ultra High Pressure Liquid Chromatograph (UHPLC) system.

Experimental conditions

Sample preparation

SRM 1950 was purchased from the National Institute of Standards and Technology (NIST). The plasma sample was prepared via protein precipitation with the addition of four volumes of 80% methanol. The sample was centrifuged, and the supernatant collected. The sample extract was evaporated to dryness, then reconstituted in water containing 0.1% formic acid, and subsequently transferred to a deactivated autosampler vial. The solvent blank was prepared using the reconstitution solution with direct transfer to a deactivated autosampler vial. A total of 2 µL was injected onto the stationary phase.

Instrument and method setup

The sample was analyzed using the Orbitrap ID-X Tribrid mass spectrometer (Table 2) coupled to a Vanquish Horizon UHPLC system. Data were acquired using Thermo Scientific™ Xcalibur™ 4.2 software and Thermo Scientific™ Standard Integration Software (SII) for Xcalibur 1.4. The chromatographic separation was obtained with a Thermo Scientific™ Hypersil GOLD™ column (1.9 µm, 150 × 2.1 mm). The column was eluted isocratically at a flow rate of 300 µL/min with 100% mobile phase A (0.1% formic acid in water) for 3 min followed by a linear gradient to 50% mobile phase B (0.1% formic acid in methanol) over 8 min, and then to 98% mobile phase B over 1 min. The column compartment temperature was held at 45 °C. Sample analysis was performed using the AcquireX Deep Scan setup (Figure 1), an automated, data-informed acquisition strategy for real-time determination of ion inclusion and exclusion with repeated sample interrogation. In brief, full scan MS was acquired first on the solvent blank and next on the plasma extract to respectively generate an ion exclusion and inclusion list. Data-dependent acquisition with the automatically generated lists was performed on the plasma extract. The plasma extract was repeatedly injected with subsequent exclusion and inclusion list updates to generate the greatest number of unique fragmentation spectra for sample-related precursor ions. Detailed instrument parameters are provided in Tables 3 and 4.

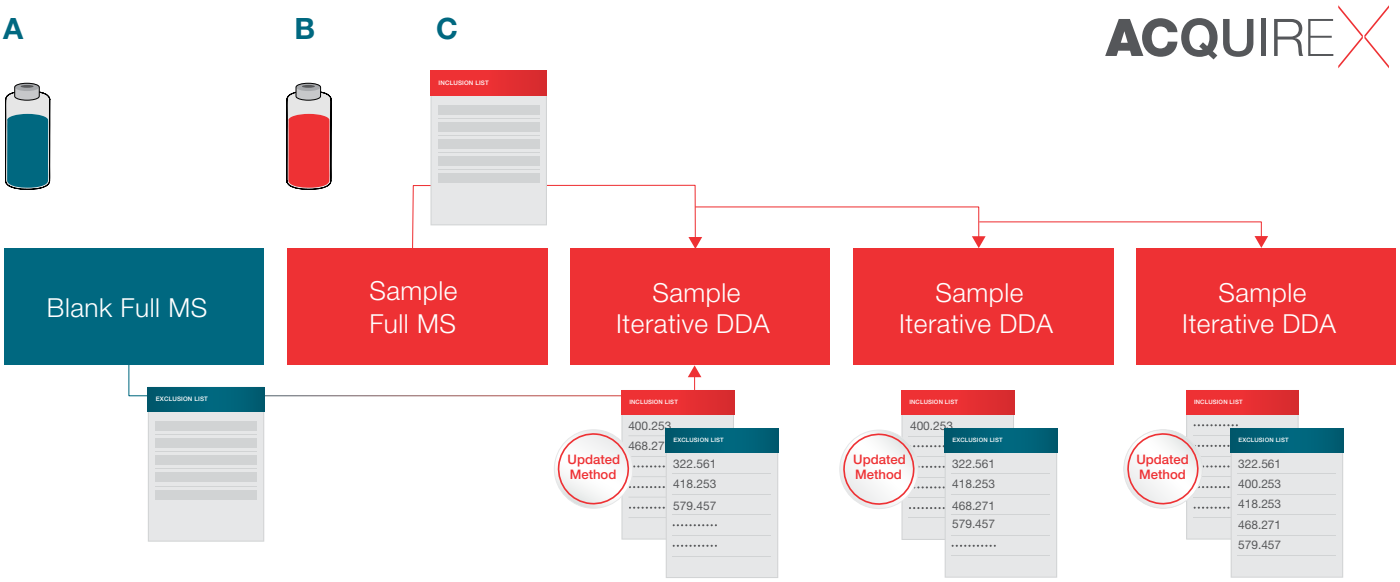


Figure 1. AcquireX Deep Scan acquisition strategy. A) This approach employs data driven intelligence by first injecting a solvent blank (or experimental blank) to generate a three-dimensional matrix of precursor ions (*m/z,* retention time, and intensity) assigned as background ions that are unrelated to the experimental sample. Background ions are automatically added to an ion exclusion list for use in the data-dependent method. B) Next, the experimental sample is injected to generate a three-dimensional matrix whereby valid chromatographic peaks are detected and assigned as candidate ions for fragmentation by automatic addition to the ion inclusion list. Data generated from these two injections are then automatically incorporated into the subsequent data-dependent acquisition method, which is then applied to the sample. C) Upon completion of the first data dependent MS^2 run, the acquisition process automatically moves ions already selected for fragmentation from the inclusion list to the exclusion list. The acquisition method is updated and subsequently applied to a repeat injection of sample. This iterative acquisition reduces redundant sampling of reoccurring precursor ions from injection to injection thereby generating more fragmentation spectra of unique compounds.

Table 2. Mass spectrometer conditions for data acquisition using full scan mode and AcquireX Deep Scan setup, an intelligent data-dependent acquisition strategy using multiple injections.

MS conditions	
Ion Source	Thermo Scientific™ OptaMax™ NG ion source
Ionization	ESI Positive Mode
Qualitative Acquisition	AcquireX Deep Scan
Source conditions	
Sheath Gas Flow Rate	40 Arbitrary Units (AU)
Auxiliary Gas Flow Rate	8 AU
Sweep Gas Flow Rate	1 AU
Spray Voltage	3500 V
Ion Transfer Tube Temperature	275 °C
RF Lens	35%
Vaporizer Temperature	320 °C

Table 3. Instrument method parameters for high-resolution full scan mode of the solvent blank and the NIST SRM 1950 human plasma to generate MS^1 quantitation.

Full Scan acquisition parameters	
Scan Range (*m/z*)	67–1000
Orbitrap Resolution	120,000 FWHM @ 200 *m/z*
AGC Target	$1e^5$
Max Injection Time	50 msec

Table 4. Instrument method parameters for the high-resolution data-dependent MS^2 fragmentation of the NIST SRM 1950 human plasma using the AcquireX Deep Scan setup for iterative injections and real-time dynamic updating of precursor ion exclusion and inclusion lists. Data generated with this method exclusively provides the fragmentation spectra used for compound annotation in the data analysis process.

Data-dependent Scan Mode Acquisition Parameters	
Full scan MS	
Scan Range (*m/z*)	67–1000
Orbitrap Resolution	120,000 FWHM @ 200 *m/z*
AGC Target	$1e^5$
Max Injection Time	50 msec
Data-dependent MS^2 Fragmentation	
Top-speed Mode	0.6 sec cycle time
Activation	HCD
dd-MS^2 Resolution	30,000 FWHM @ 200 *m/z*
AGC Target	5e4
Max Injection Time	54 msec
Quad Isolation Width	1.5 Daltons
Normalized Stepped Collision Energy	20, 35, 50%
Intensity Threshold	$2e^4$
Dynamic Exclusion (s)	2.5 sec

Data processing

Data were processed using the Compound Discoverer 3.0 software. To expedite the data processing setup, a pre-defined processing template was used. A modified version of the Metabolomics Max ID workflow template was employed to provide exhaustive compound annotation from multiple measurements following unknown peak detection. This template is useful for a limited number of raw data files where the peak intensity threshold is set to a pre-determined minimum low value, and detection filters are turned off to accommodate very low abundant compounds. Briefly, the workflow includes unbiased unknown compound detection, elemental composition prediction, database searching at the precursor level against the ChemSpider database, application of the mzLogic algorithm to rank putative candidates generated from the ChemSpider database, searching against a custom-built metabolomics database, as well as MS^2 spectral matching against the mzCloud spectral library. The Compound Discoverer 3.0 software is fully integrated with the ChemSpider database and the mzCloud spectral library for automated and expedited data processing. Ten data sources were selected via the ChemSpider database comprising both endogenous and exogenous entries: Aggregated Computational Toxicology Resource (ACToR), BioCyc, Drug Bank, EAWAG Biocatalysis/Biodegradation Database, Environmental Protection Agency (EPA) DSSTox, EPA ToxCast, Federal Drug Administration Unique Ingredient Identifier (UNII), FooDB, Human Metabolome Database, and the Kyoto Encyclopedia of Genes and Genomes™ (KEGG). Data were processed with and without mzLogic for comparison purposes. A more detailed illustration of the workflow and associated nodes is depicted in Figure 2.

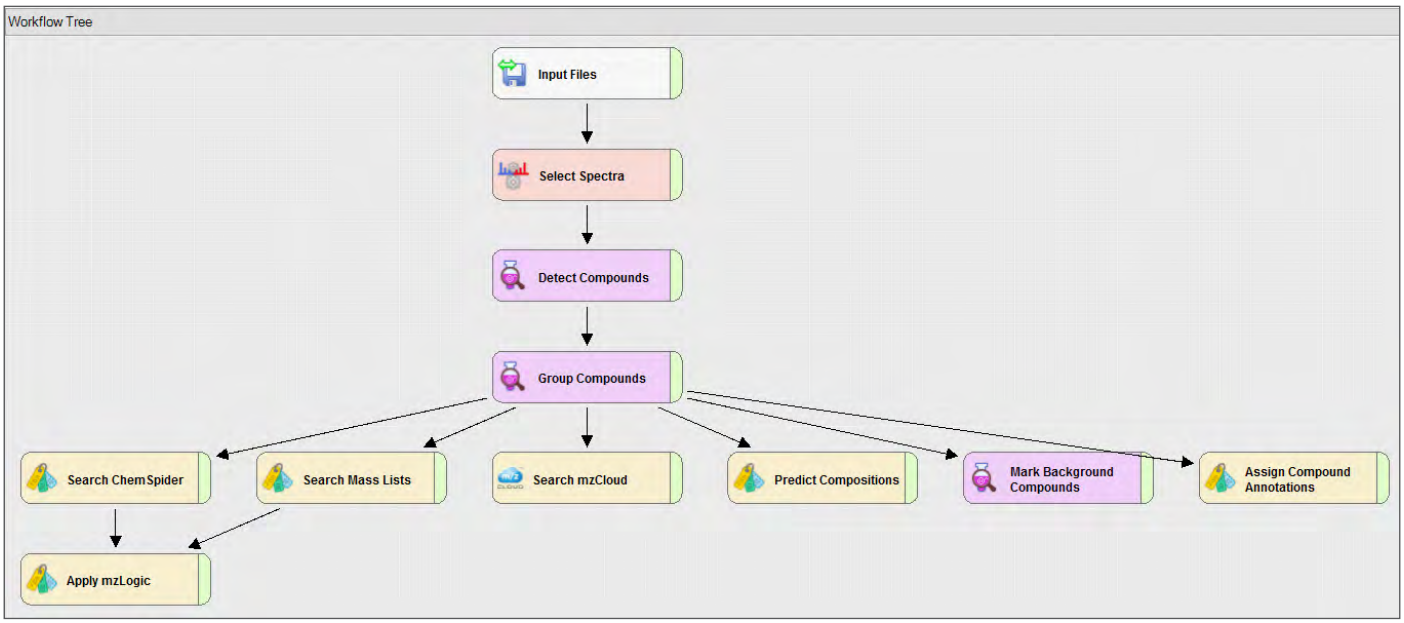


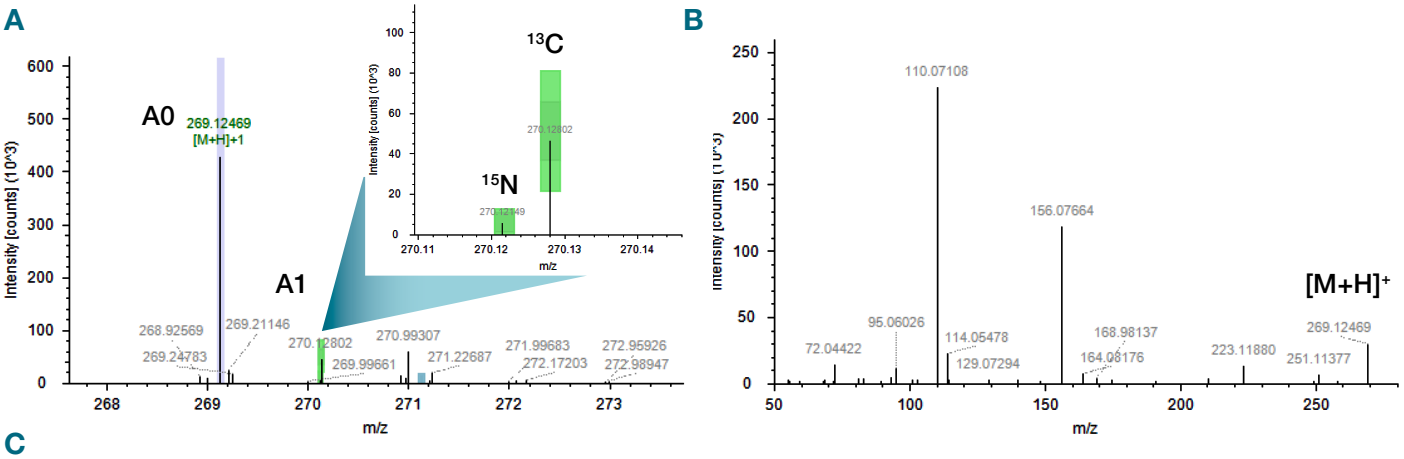
Figure 2. A workflow tree from the Compound Discoverer 3.0 software displaying select data processing nodes and the associated workflow connections. Included are preliminary data processing nodes like Input Files and Select Spectra nodes. Unknown peak detection is implemented via the Detect Compounds node. Information across multiple raw data files are integrated through the Group Compounds node. Numerous unknown compound annotation nodes are utilized for both MS^1 and MS^2 spectral data: Predict Compositions node, Search Mass List node, Search ChemSpider node, Apply mzLogic node, and the Search mzCloud node. The Assign Compound Annotations node prioritizes the annotation source. Lastly, the Mark Background Compounds node incorporates a solvent blank (experimental blank) to indicate compounds arising from the sample preparation.

Results and discussion

The role of HRAM in unknown compound annotation

High-resolution mass spectrometers increase the ability to distinguish between two molecules of closely related mass, which is highly advantageous for untargeted metabolomics analyses of complex matrices such as human plasma extracts. Additionally, information obtained from accurate mass lends the ability to apply narrow search criteria against theoretical values of exact mass and sufficiently high resolution allowing for the detection of isotope pattern and isotopic fine structure. HRAM obtained with Orbitrap mass spectrometers and isotopic fine structure increases confidence in elemental composition prediction, which is further aided with fragmentation data. Performing database searches using elemental formula is preferred over molecular weight or *m/z* due to increased specificity, which in turn reduces complexity. Taken together, these attributes are beneficial for the annotation of unknown compounds.

Figure 3A shows a narrow mass range of the full scan spectrum displayed in the Compound Discoverer 3.0 software indicating the monoisotopic peak and respective isotope pattern for the expected compound at *m/z* 269.1247. Of particular interest is the isotope fine structure achieved with high resolution to differentiate the nitrogen 15 isotope from the carbon 13 isotope. Mass and intensity tolerances are applied. In the example, the measurement reveals the presence of the nitrogen atom, which is applied to the elemental composition prediction. This confidently eliminates chemical formulas lacking nitrogen, thus, reducing the number of candidate formulas. The Compound Discoverer software additionally considers fragment ions to rank putative chemical formulas by confirming fragment ions that match a subset of the precursor ion's elemental composition. Figure 3B shows the MS^2 spectrum of the isolated precursor ion using an isolation window of 1.5 Daltons (Da). The presence of the $[M+H]^+$ ion confirms the precursor ion isolation. Figure 3C displays the predicted composition results for the expected compound, acetylcarnosine. Supporting data is provided for each possible candidate formula including delta mass, the number of matched isotopes, the number of matched fragment ions, spectral fit (Sfit) and coverage values. The elemental formula predicted for *m/z* 269.1247 is $C_{11}H_{16}N_4O_4$. High resolution, accurate mass, and fragmentation data collectively contribute increased confidence for the predicted chemical formula.



C

Structure Proposals	Compounds per File	Predicted Compositions	mzCloud Results	ChemSpider Results	Mass List Search Results

	Compound Match	Formula	Molecular Weight	ΔMass [Da]	ΔMass [ppm]	RDBE	H/C	Rank	# Matched Iso.	# Missed Iso.	# Matched Frag.	SFit [%]	Pattern Cov. [%]	MS Cov. [%]	MSMS Cov. [%]
1		C11 H16 N4 O4	268.11715	0.00026	0.98	6.0	1.5	1	3	0	13	89	98.51	100.00	99.00
2		C12 H12 N8	268.11849	-0.00107	-4.01	11.0	1.0	2	3	0	6	70	99.01	100.00	58.98
3		C13 H21 N2 P S	268.11631	0.00111	4.15	5.0	1.6	3	3	0	5	71	94.14	100.00	11.48
4		C11 H26 O P2 S	268.11796	-0.00054	-2.02	0.0	2.4	4	2	0	2	92	94.50	98.84	2.34
5		C5 H17 N8 O3 P	268.11612	0.00130	4.83	2.0	3.4	5	2	1	11	36	94.26	90.34	73.00
6		C4 H16 N10 O2 S	268.11784	-0.00042	-1.57	2.0	4.0	6	2	0	7	57	91.29	90.34	16.51

Figure 3. HRAM and fragmentation data to predict elemental composition. A) MS^1 survey scan showing color-coded isotope pattern fit for the detected compound of *m/z* 269.1247. The violet vertical box indicates the monoisotopic protonated molecule of the expected compound, acetylcarnosine, matching the centroid. Green boxes represent the theoretical isotope pattern with defined tolerances for mass and relative intensity. Inset is the region of the A1 isotope cluster displaying the ^{13}C and ^{15}N isotopes. B) MS^2 fragment ion spectrum for the isolated precursor ion of *m/z* 269.1247. C) Predicted composition results table for the selected compound displaying multiple candidates in rank order. Several variables are considered including delta mass, the number of matched isotopes, and the number of matched fragment ions.

Database search as a starting point

Chemical structure databases aid in generating putative candidate compounds. Public repositories containing known, well-defined chemical compounds can be searched against for annotation. Information found in these databases generally include chemical structure, related chemical characteristics, metadata like biological activity, associations, and may include fragmentation spectra either from actual data or *in silico* prediction of fragmentation. The Compound Discoverer software is fully integrated with the ChemSpider database for automatic searching specifically for precursor mass information. Two search modes can be utilized. The searches can be performed either by elemental formula or by mass. The advantage of searching by elemental formula over mass is that the elemental composition prediction incorporates several variables from the mass measurement, as described in the previous section for more confident formulas. A search by formula can reduce the number of possible matches when searching databases compared to mass alone. Nevertheless, database searching provides a good starting point for compound annotation via precursor matching.

Figure 4A shows a narrow mass range from the full scan spectrum in the Compound Discoverer 3.0 software for the expected compound phenylalanine. The molecular ion and associated isotopes for the amino acid are detected. Elemental composition is predicted using both full scan data and associated fragment ions (Figure 4B) resulting in the formula $C_9H_{11}NO_2$ (Figure 4C). Applying this formula to the ChemSpider database search against the selected data sources generates 419 possible candidate matches (Figure 4D). DL-phenylalanine was ranked highest based on the number of references in the ChemSpider database, which is over 13,000 (Figure 5). The ChemSpider results table in the Compound Discoverer 3.0 software displays an interactive ChemSpider ID for easy network access in addition to chemical structure, molecular formula, and delta mass. While a database search finds putative candidate compounds, caution should be used when this is the sole annotation source because the search is simply based on precursor information only. As shown here, over 400 possible annotations were generated for this single chemical formula. This results from the fact that one molecular formula can represent various possible chemical structures, such as structural isomers where there is the same number of atoms for each element but with different spatial arrangement.

Searching data sources containing endogenous and exogenous substances can be highly useful in an untargeted metabolomics experiment, particularly when analyzing samples obtained from "free-ranging" organisms such as humans and animals. The SRM 1950 human plasma sample was pooled from a collection of 100 fasted donors, both male and female.[8] A putative ChemSpider match for benzoylecgonine, a metabolite of the drug compound cocaine, was detected in the sample (Figure 6); illustrating the value of comprehensive small molecule analysis considering more than just endogenous metabolites.

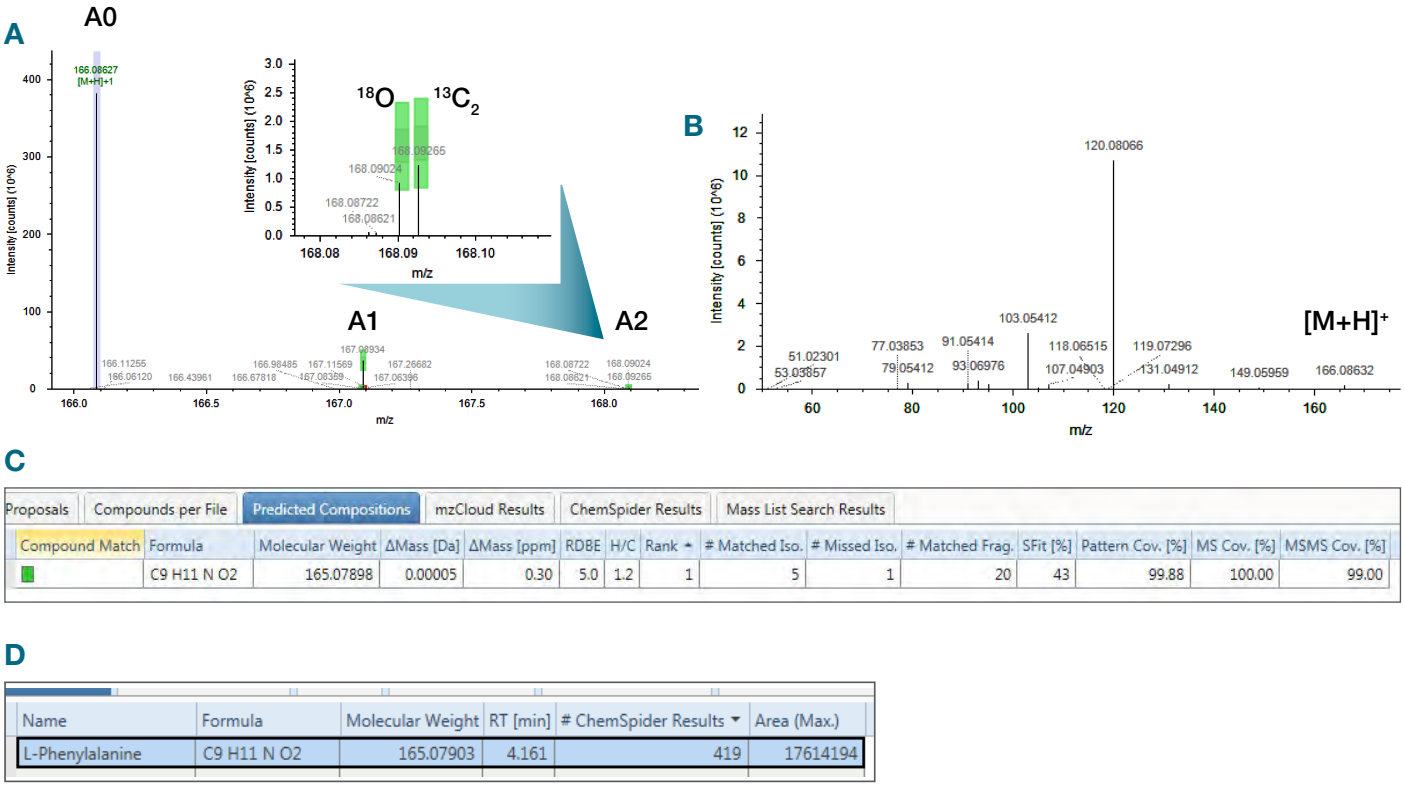


Figure 4. Elemental composition for database searching. A) MS^1 survey scan showing color-coded isotope pattern fit for the expected compound phenylalanine, *m/z* 166.0863. The inset showing the A2 isotope cluster containing the $^{13}C_2$ and ^{18}O isotopes. B) MS^2 fragment spectrum for the isolated precursor ion of *m/z* 166.0863. C) One candidate elemental formula for the expected compound and associated variables to support this: delta mass, the number of matched isotopes, and the number of matched fragment ions. D) Applying the elemental formula, $C_9H_{11}NO_2$ to the ChemSpider database search resulted in 419 candidate annotations.

roposals | Compounds per File | Predicted Compositions | mzCloud Results | ChemSpider Results | Mass List Search Results

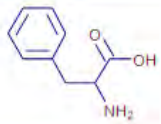
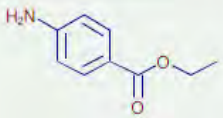
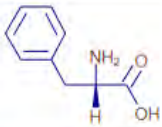
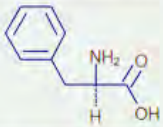
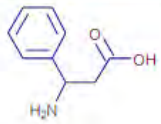
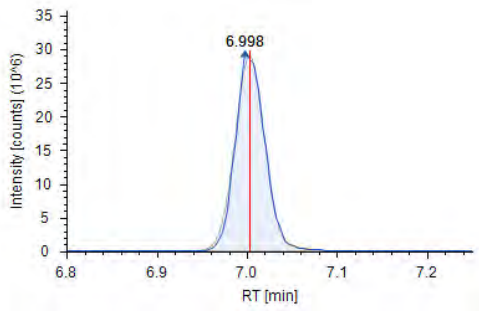
Compound Match	Structure	Name	Formula	Molecular Weight	ΔMass [Da]	ΔMass [ppm]	CSID	# References
		DL-Phenylalanine	C9 H11 N O2	165.07898	-0.00005	-0.29	969	13669
		Benzocaine	C9 H11 N O2	165.07898	-0.00005	-0.29	13854242	1589
		L-(-)-Phenylalanine	C9 H11 N O2	165.07898	-0.00005	-0.29	5910	942
		D-(+)-Phenylalanine	C9 H11 N O2	165.07898	-0.00005	-0.29	64639	408
		3 Amino-phenylpropionic acid	C9 H11 N O2	165.07898	-0.00005	-0.29	62403	285

Figure 5. Top five candidate annotations based on the ChemSpider database search for the expected compound phenylalanine, *m/z* 166.0863. Here, putative candidates are ranked based on the number of references found in the ChemSpider database. Using this logic, the phenylalanine isomers, benzocaine, and 3-aminophenylpropionic acid are ranked highest.

A



B

Name	Formula	Molecular Weight	RT [min]	# ChemSpider Results	Area (Max.)
Benzoylecgonine	C16 H19 N O4	289.13170	7.003	64	1148631

C

Proposals | Compounds per File | Predicted Compositions | mzCloud Results | ChemSpider Results | Mass List Search Results

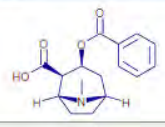
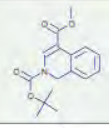
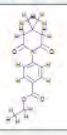
Compound Match	Structure	Name	Formula	Molecular Weight	ΔMass [Da]	ΔMass [ppm]	CSID	# References
		Benzoylecgonine	C16 H19 N O4	289.13141	-0.00029	-1.01	395095	75
		2-tert-Butyl 4-methyl isoquinoline-2,4(1H)-dicarboxylate	C16 H19 N O4	289.13141	-0.00029	-1.01	24534363	65
		ethyl 4-(4,4-dimethyl-2,6-dioxopiperidino)benzoate	C16 H19 N O4	289.13141	-0.00029	-1.01	2104375	59

Figure 6. Choosing data sources for endogenous and exogenous compounds. A) Chromatogram of the expected compound, benzoylecgonine. B) Searching the elemental formula, $C_{16}H_{19}NO_4$, against the ChemSpider database using data sources containing both endogenous and exogenous compounds resulted in a total of 64 candidate annotations. C) The top-ranked candidate hit is benzoylecgonine, a primary metabolite of the drug cocaine. Entries listed here are ranked based on the number of references for each compound in the ChemSpider database.

The power of fragmentation spectra in parallel with a library hosting an immense number of spectra

Coupling ion dissociation techniques to HRAM MS provides another measurement for compound characterization to build confidence toward unknown compound annotation. Isolation of the precursor ion population via a narrow isolation window improves the quality giving purer fragmentation spectra. For this reason, a data-dependent acquisition was chosen for the analysis and the high-quality fragmentation data was then used to search against a spectral library. Unlike a database, a spectral library is a collection of actual fragmentation spectra. Typically, each compound entry contains multiple spectra representing several different collision energies. The mzCloud spectral library is a highly curated, public library of endogenous and exogenous small molecules containing almost 3 million fragmentation spectra. Each compound entry in the library generally includes two collisional techniques: higher energy collision-induced dissociation (HCD) and ion trap resonant collision-induced dissociation (CID). For each dissociation technique, fragmentation spectra span a wide range of collision energies in iterations of 5 or 10%, thus generating a complete breakdown curve showing the reduction of precursor ion intensity with increasing collision energy while fragment ions increase concurrently. This systematic collection of spectra eliminates constraints in how data are acquired in terms of collision energy for subsequent spectral library matching. Furthermore, ion trap technology provides MS^n with repeated isolation and fragmentation of product ions beyond MS^2, such as MS^3, MS^4 to MS^n, producing spectral trees for each library entry. Finally, the mzCloud spectral library was generated using actual spectra collected from purified reference material. Each spectrum is recalibrated for exact mass and noise removed and is further structurally annotated making this a very high-quality spectral library. The mzCloud library can be searched by uploading individual fragmentation spectrum; however, the Compound Discoverer 3.0 software is fully integrated with the mzCloud library, enabling batch searching for a fully automated analysis. Searching against a high-quality, fully curated spectral library with ample fragmentation spectra from existing compounds provides reinforcing knowledge about the molecular makeup of an unknown compound.

Two search types were implemented: identity search and similarity search. An identity search matches both precursor ion and fragment ions, while a similarity search only searches for fragment ions, allowing for matches indicative of substructure. Figure 7 shows a mzCloud library match for the amino acid methionine of the chromatographic peak at 2.16 min (Figure 7A) and a monoisotopic peak at *m/z* 150.0584 (Figure 7B). The full scan spectrum and the isotope fine structure is in complete accord with the elemental formula $C_5H_{11}NO_2S$. Figure 7C displays the mirror plot for the MS^2 spectrum of the experimental data (top) and the matching spectrum found in the mzCloud library (bottom) where multiple ions overlap including the precursor ion and several product ions. This identity match further points toward a confident annotation of methionine. There are instances when the library may not contain an expected compound yet, the fragmentation data is still highly informative toward annotation, particularly when product ions match between the library and experimental data for sub-structure elucidation. This approach is commonly known as Precursor Ion Fingerprinting (PIF),[9] which is immensely insightful for determining the degree of similarity between unknown and candidate compounds. Figure 8 follows this evaluation process. A chromatographic peak at 1.20 min (Figure 8A) with a predicted chemical composition of $C_{12}H_{26}N_2O_5$ generated seven ChemSpider database matches (Figure 8B). Of the seven candidates, the top-ranked entry is 1,4-bis[bis (2-hydroxyethyl)amino]-2-butanone (Figure 8C) based on the molecular weight of the precursor ion. At the same time, taking the fragmentation data into account resulted in the mzCloud library similarity match to carnitine (Figure 8D) suggesting this compound is related to carnitine given that there are several overlapping product ions in addition to the presence of the precursor ion for carnitine as seen in the mirror plot (Figure 8E). Though the molecular weight of the expected compound is heavier by 117.0794 Da compared to carnitine, the fragmentation data indicates a carnitine-like compound rather than the top-ranked hit from the ChemSpider database. Fragmentation spectra are meaningful to further build confidence toward unknown compound annotation.

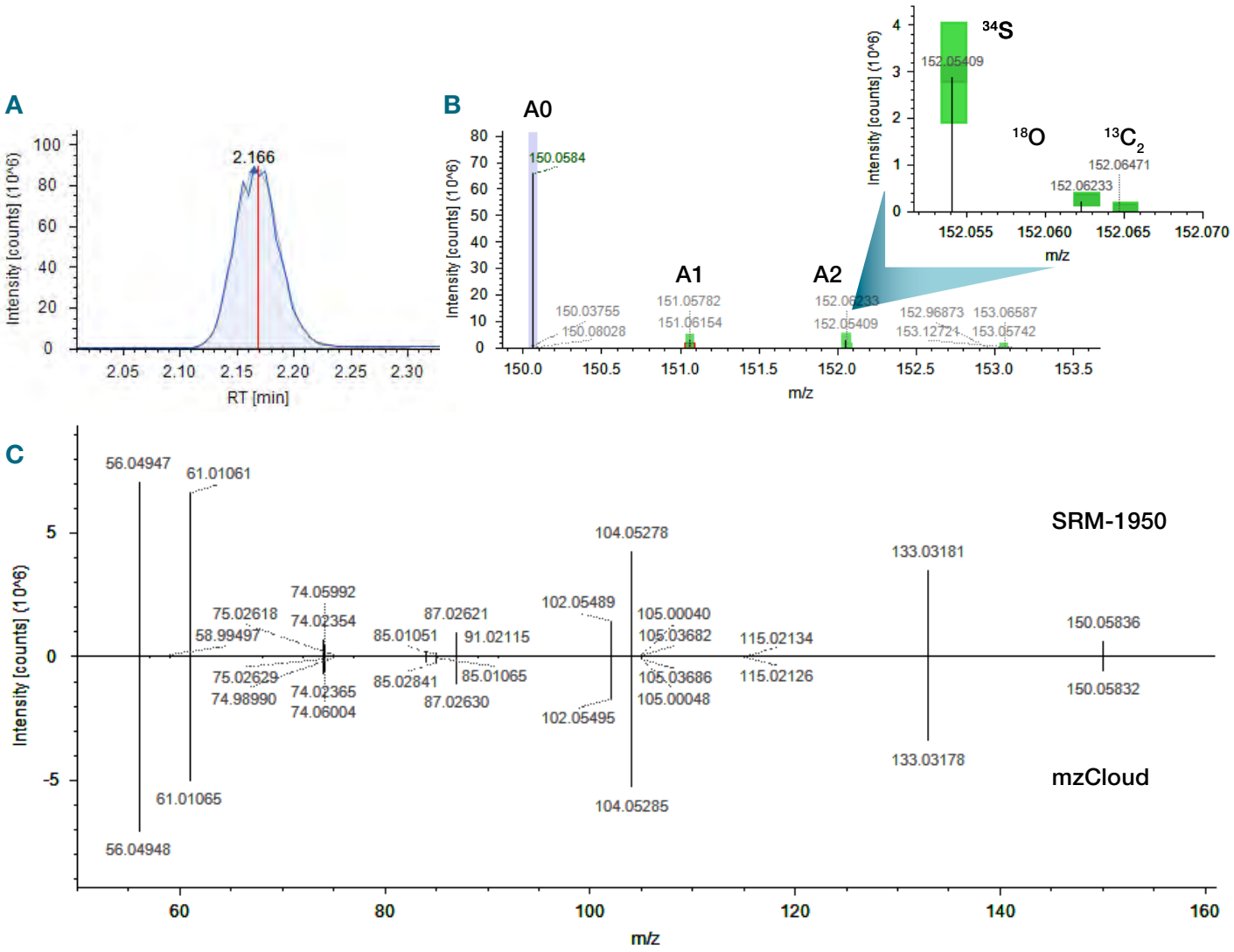


Figure 7. Identity match—MS² spectral search against the mzCloud spectral library. A) Chromatogram of the expected compound, methionine. B) MS¹ survey scan showing color-coded isotope pattern fit for the expected compound methionine, *m/z* 150.0584. The inset shows the A2 isotope cluster displaying the ^{34}S, $^{13}C_2$ and ^{18}O isotopes. C) Mirror plot of the MS² fragmentation spectrum from the SRM-1950 (top) and matched spectrum from the reference library in the mzCloud library (bottom).

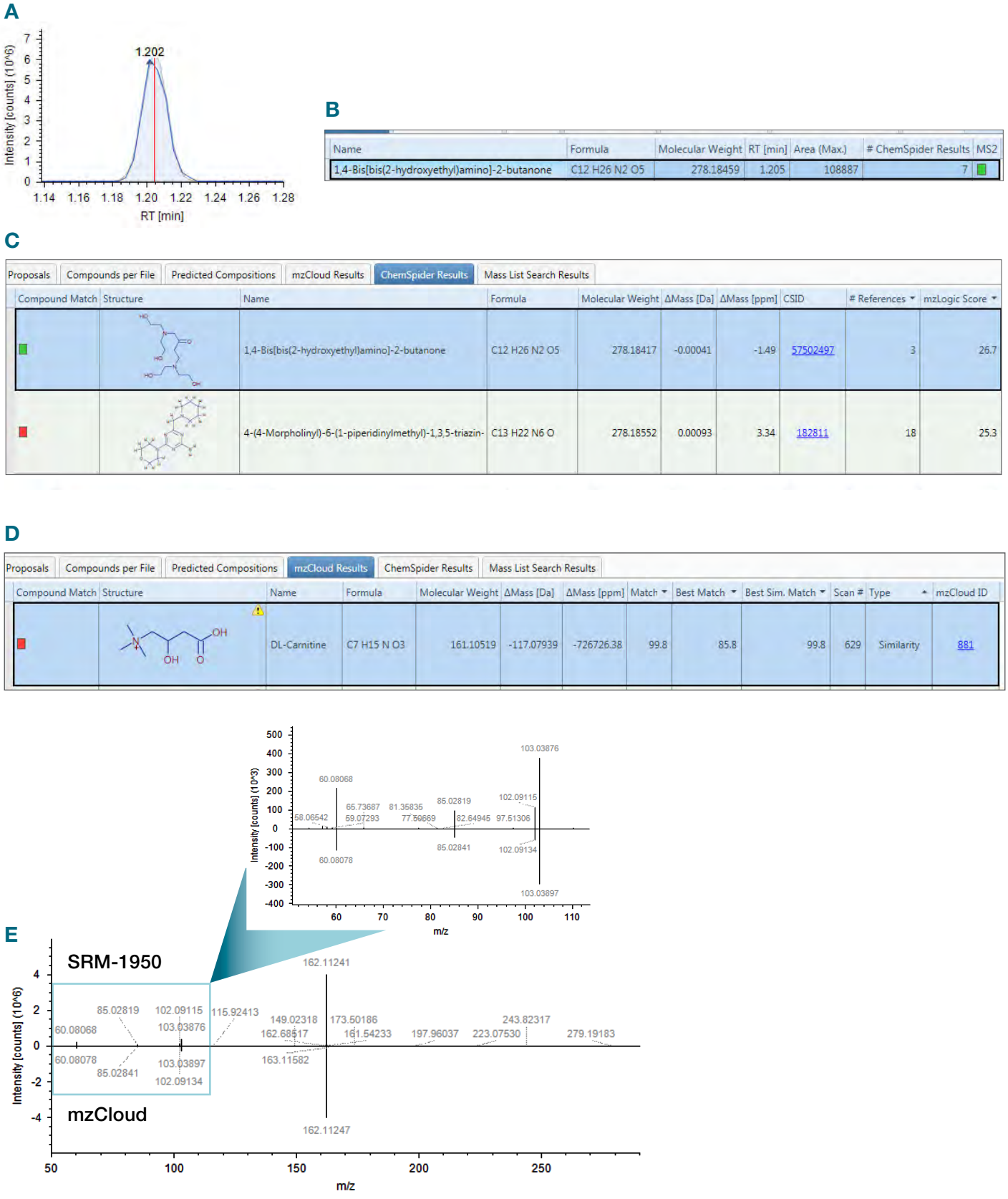


Figure 8. Similarity match—MS² spectral search against the mzCloud spectral library. A) Chromatogram of the expected compound, 1,4-bis[bis(2-hydroxyethyl)amino]-2-butanone. B) Searching the elemental formula, $C_{12}H_{26}N_2O_5$, against the ChemSpider database resulted in a total of 7 candidate annotations. C) The top ranked candidate compound shows three references, with a delta mass of 1.29 ppm. D) The mzCloud library search resulted in a similarity match to carnitine, suggesting that this compound is rather related to carnitine, and thus carnitine-like. The molecular weight of the expected compound is 278.1846 Daltons (Da) while the molecular weight of carnitine is 161.1052 Da, a delta mass of 117.0794 Da. The caution triangle indicates a naturally positively charged compound. E) Mirror plot of the MS² fragmentation spectrum from the SRM-1950 (top) and the matched spectrum from the reference library in mzCloud (bottom). Several overlapping product ions are displayed including the precursor ion and lower molecular weight ions like *m/z* 103.0397, 85.0284, and 60.0808 (zoomed area).

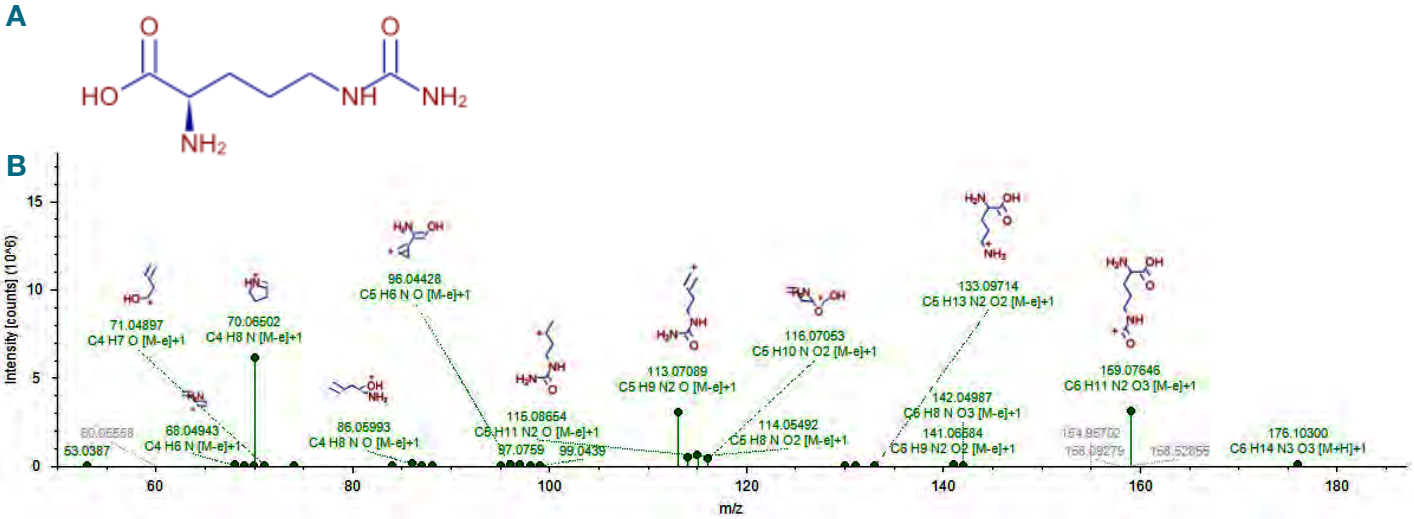


Figure 9. FISh scoring using *in silico* prediction. A) The chemical structure for the expected compound citrulline, molecular weight 175.0959 Da. B) MS^2 fragmentation spectrum for the isolated precursor ion of *m/z* 176.1030. The FISh coverage score is 77.1% where 27 product ions were successfully matched (green) and 8 ions remain unmatched. Structural annotations are displayed for all matched ions when sufficient computer monitor display space is allowed. The FISh coverage score is determined by the number of matched centroids divided by the number of used (matched and unmatched) centroids multiplied by 100.

Knowing where to break the chemical bond

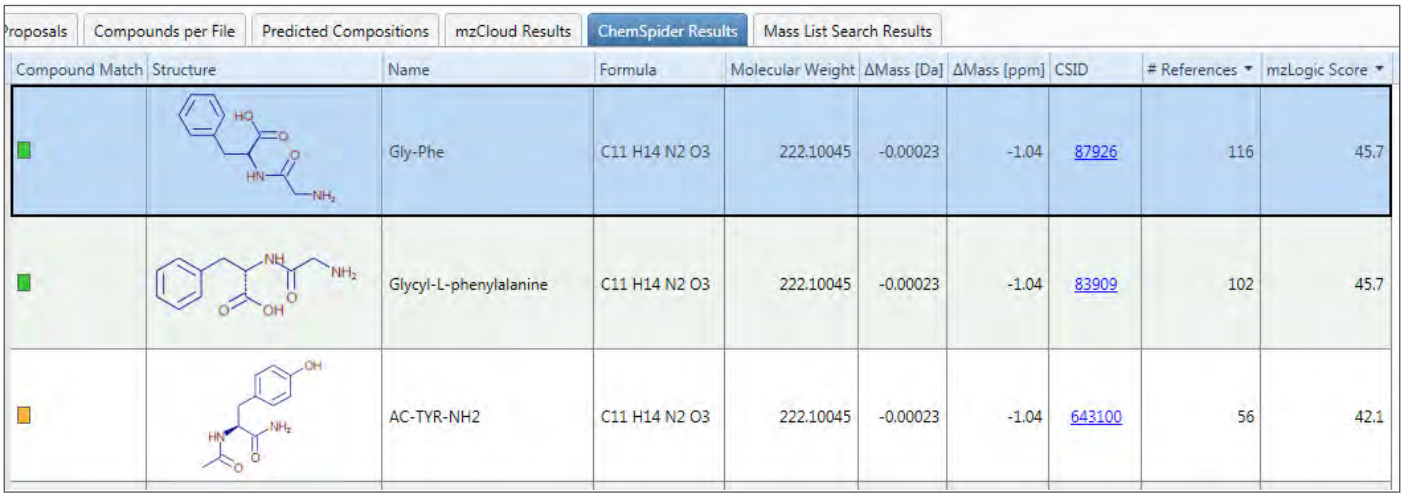
Another way to evaluate fragmentation data of an unknown compound is to apply an *in silico* fragmentation prediction algorithm to parent structures in order to generate potential ion fragments. Applying reaction logic, *in silico* fragmentation prediction in the Compound Discoverer 3.0 software is generated based on the HighChem Fragmentation Library, which consists of 31,901 fragmentation schemes and 136,169 decoded mechanisms taken from literature publications based on mass spectrometry. Erroneous fragmentation mechanisms are eliminated using this practical approach since reaction mechanisms are rigorously evaluated, both manually and automatically, producing a high-quality database. The *in silico* processing method in the Compound Discoverer 3.0 software is by the FISh scoring algorithm. Incorporating FISh capability in an untargeted metabolomics analysis aids to structurally explain product ions present in the fragmentation spectrum via a proposed chemical structure. Figure 9 demonstrates the applicability of the FISh processing method. An expected compound in the SRM 1950 extract is proposed as citrulline. Both the ChemSpider database and mzCloud library matching support this annotation. Taking this one step further, FISh was then applied to the chemical structure citrulline (Figure 9A) and associated with the fragmentation spectrum generated for the corresponding precursor mass of *m/z* 176.1030. Of the possible product ions assessed, more than 75% in the spectrum can be structurally related to citrulline (Figure 9B). While the selected example illustrates FISh applied to MS^2, this processing method can be applied to any level of MS^n data. The ability to define a product ion by structural annotation for a proposed compound contributes additional knowledge in evaluating annotations for unknown compounds.

The mzLogic algorithm—combining the database search with the mzCloud Spectral Library

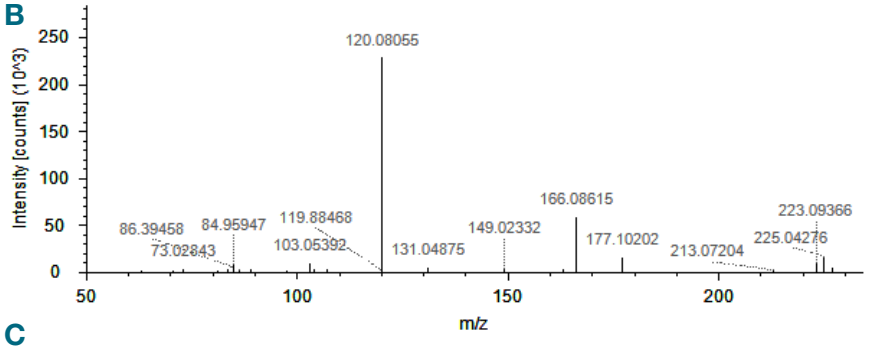
Real fragmentation data can also be used to rank order results from the ChemSpider database search. As previously descrww a search against the ChemSpider database may result in numerous putative candidate compounds in which the number of references prioritizes the results list. This is a plausible approach; however, it does not take into consideration any structural information. Leveraging sub-structural knowledge and the extensive spectral fragmentation information in the mzCloud spectral library, the new mzLogic algorithm can prioritize the list of candidate compounds from the ChemSpider database based on the unknown's fragmentation spectra. The mzLogic algorithm first obtains all possible parent chemical structures resulting from the ChemSpider database search. Next, the results from the similarity search against the mzCloud library are included to map structural annotations of matched fragment ions back to the experimental data. The partial structures are overlaid to the parent structures generated from the database search to obtain the best fit. The candidate structures that can best be explained with the maximum common sub-structure and the highest spectral match score are then prioritized. The mzLogic score is easily viewed in the Compound Discoverer 3.0 software in a column format within the ChemSpider database results table, or any other result table containing chemical structure, such as the Mass List Search results table.

Figure 10 shows the application of the mzLogic algorithm to rank order the ChemSpider database search results to generate the annotation of Gly-Phe. The compound was top ranked with a score of 45.7 (Figure 10A) out of a total of 185 candidates (not shown). The fragmentation spectrum (Figure 10B) was used to perform a similarity search against the mzCloud library to generate fragment structures of similar compounds. The structural annotations are used to determine the maximum common substructure to rank the ChemSpider database search results (Figure 10C). The mzLogic algorithm uses fragmentation information for structural explanation via the mzCloud library, thus providing another annotation tool contributing to confident annotations based on mass spectral fragmentation data.

A



B



C

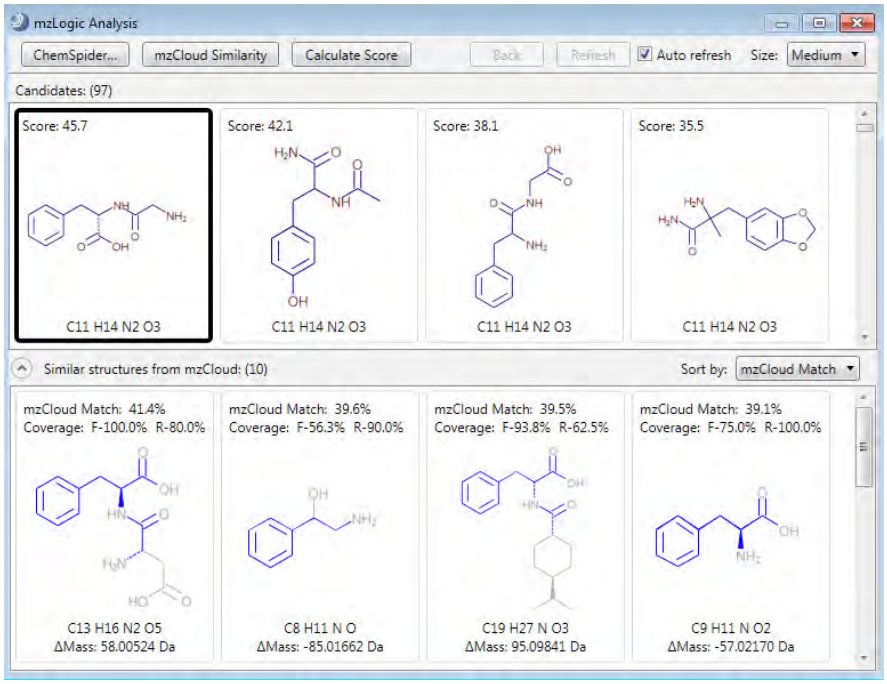


Figure 10. The mzLogic algorithm ranks database search results using fragment structures obtained via the mzCloud library. A) A ChemSpider database search resulted in 185 matched compounds (not shown) where the top match is the dipeptide Gly-Phe. With the mzLogic algorithm in place, putative candidates are first ranked based on mzLogic scoring followed by the number of references in the ChemSpider database. B) Fragmentation data were acquired for this compound, but spectral library search did not result in an identity match. C) Results from the mzLogic analysis. The top panel shows the top ranked structural candidates based on the mzLogic score. The bottom panel shows compound structures from the mzCloud library where structural similarity overlapping with the selected candidate compound is highlighted in blue. The black box shows Gly-Phe selected with a score of 45.7 and the corresponding structural similarity colored in blue.

Consensus for increased annotation confidence and pathway certainty

Confidence in unknown compound annotation increases when multiple analytical measurements are collectively evaluated.[10,11] Several annotation tools were used for the analysis of the SRM 1950 human plasma extract. In the Assign Compound Annotation processing node, users can define the preference order of the annotation source for the assigned annotation. For this analysis, annotation assignments were based in the following order of data sources starting with the most preferred: mzCloud spectral library search, predicted elemental composition, mass list search, and the ChemSpider database search. Accordingly, complete matching across all sources for a given annotation using maximum information (full scan spectra, fragmentation spectra, and chemical structure) increases the confidence in the annotation. The consensus approach is easily viewed in the Compound Discoverer 3.0 software via a color-coded status indicator (Figure 11) illustrating the value of achieving agreement across multiple annotation sources. While not all compounds detected in the analysis may have full matches to all sources, there is still confirmatory information contributing to compound annotation as with partial matching or chemical relatedness (i.e., similarity search). Confidence in an unknown compound annotation is essential in untargeted metabolomics analysis. Using multiple analytical measurements to gain greater compound knowledge increases confidence in compound annotation. Confident compound annotations facilitate reliable biochemical pathway analyses.

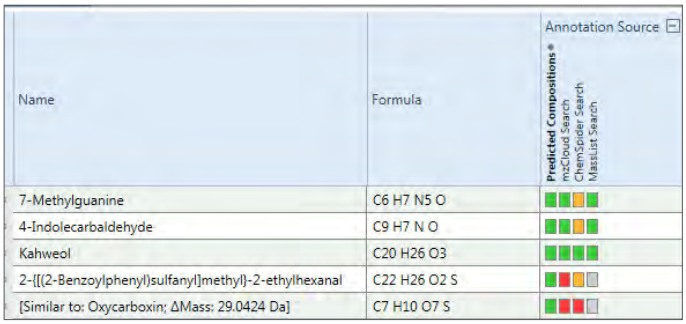


Figure 11. Application of the consensus approach for confident compound annotation using results by several annotation sources generated from multiple analytical measurements. Shown are five compounds listed in the Main Table results. The annotation source results are displayed in column format for each compound via colorized rectangles representing match status. The color green indicates a full match while the color orange is a partial match. The color red represents a mismatched mass in the case of the mzCloud library search (similarity match) and represents no match in the case of ChemSpider database search. The third compound listed is Kahweol, a diterpenoid molecule found in the beans of Coffea arabica. This compound matched all four sources used in the analysis: elemental composition prediction, mzCloud library search, ChemSpider database search, and the mass list search. Other compounds listed in the table, 7-methylguanine and 4-indolecaraldehyde, matched three sources, while the last two entries resulted in the mzCloud library similarity search only.

Conclusion

The SRM 1950 human plasma extract was analyzed using the Orbitrap ID-X Tribrid MS and the Compound Discoverer 3.0 software. HRAM data were generated for MS^1 and MS^2. Fragmentation spectra were acquired using the novel AcquireX Deep Scan acquisition strategy. The data-driven, intelligent acquisition approach prioritized precursor ion selection for subsequent fragmentation using ion inclusion and exclusion lists via preliminary full scan analysis of the blank and matrix sample. The unknown compounds were annotated using multiple annotation sources in the processing workflow to increase confidence in the assignment of unknown compound annotations. Predicted elemental composition from the high-quality mass spectra obtained using the HRAM Orbitrap instrument was used to search the ChemSpider database. The database provides access to several relevant metabolomics databases such as HMDB, BioCyc, KEGG, Yeast Metabolome Database in addition to chemically relevant sources of synthetic origin like ACToR, DrugBank, and EAWAG. Structure candidates produced from the ChemSpider database search were ranked using the mzLogic algorithm that takes into account experimental fragment ions. The fragmentation spectra were searched against the

thermo scientific

mzCloud library using both identity and similarity search. For further confirmation, *in silico* fragmentation of proposed chemical structures using the FISh algorithm was used to structurally explain fragment ions.

Results generated from this experiment demonstrate the utility of analyzing multiple analytical measurements from mass spectral data against different data sources. Both endogenous and exogenous compounds were annotated in the human plasma extract. Expected endogenous metabolites like acetylcarnosine and phenylalanine were annotated by the predicted elemental composition and ChemSpider database search. Incorporating databases of synthetic origin allowed for the annotation of benzoylecgonine, a primary metabolite of the drug cocaine. The mzCloud library search generated an identity match to methionine while the similarity search detected carnitine-like compounds. *In silico* fragmentation from the FISh algorithm structurally explained more than 75% of the fragment ions for citrulline. The mzLogic algorithm enabled the Gly-Phe annotation to be prioritized out of 185 candidate compounds. In some instances, the compound annotation was the same across different sources indicating agreement. In other instances lacking an identity match, a partial match associated compound class. Employing different annotation tools in the Compound Discoverer 3.0 software to analyze the SRM 1950 extract for multiple analytical measurements built confidence in the annotation assignments of the unknown compounds.

References

1. Patti, G., Dunn, W., Creek, D., and Sumner, L. (2018). The journey from features to compound identification in metabolomics when will we get there? *The Scientist*, eBook #65354.
2. Dunn, W.B., Erban, A., Weber, R.J.M., Creek, D.J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J., and Viant, M.R. (2013). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, 9 (Suppl1), 44–66.
3. Loulou Peisl, B.Y., Schymanski, E.L., and Wilmes, P. (2018). Dark matter in host-microbiome metabolomics: tackling the unknowns–a review. *Analytica Chimica Acta*, 1037, 13–27.
4. Brown, M., Dunn, W.B., Dobson, P., Patel, Y., Winder, C.L., Francis-McIntryre, S., Begley, P., Carroll, K., Broadhurst, D., Tseng, A., Swainston, N., Spasic, I., Goodacre, R., and Kell, D.B. (2009). Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. Analyst, 134, 1322–1332.
5. Peake, D. (2018). Review paper 65356: High-resolution compound identification in metabolomics: a review of current practices.
6. Stratton, T., and Tautenhahn, R. (2018). White paper 65210: Compounding insights for small molecule research Thermo Scientific Compound Discoverer software.
7. Souza, A., and Tautenhahn, R. (2018). Technical note 65204: Features or compounds? A data reduction for untargeted metabolomics to generate meaningful data.
8. Phinney, K.W. et al. (2013). Development of a Standard Reference Material for Metabolic Research. *Analytical Chemistry*, 85, 11732–11738.
9. Sheldon, M.T., Mistrik, R., and Croley, T.R. (2009). Determination of ion structures in structurally related compounds using precursor ion fingerprinting. Journal of the American Society of Mass Spectrometry, 20, 370–376.
10. Rochat, B. (2017). Proposed confidence scale and ID score in the identification of known-unknown compounds using high resolution MS data. Journal of the American Society of Mass Spectrometry, 28 (4), 709–723.
11. Blaženović, I., Kind, T., Ji, J., and Fiehn, O. (2018). Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites, 8, 31.

Find out more at
www.thermofisher.com/compounddiscoverer

ThermoFisher SCIENTIFIC

thermoscientific

APPLICATION NOTE 72942

Increased confidence of insect lipidome annotation from high-resolution Orbitrap LC/MS^n analysis and LipidSearch software

Authors

David Peake[1], Reiko Kiyonami[1], Yasuto Yokoi[2], Yelena Aldefinskaya[3], Jeff Gilbert[3], and Daniel Gachotte[3]

[1]Thermo Fisher Scientific, San Jose, California, USA

[2]Mitsui Knowledge Industry Co., LTD, Tokyo, Japan

[3]CORTEVA agriscience, Indianapolis, Indiana, USA

Keywords

Lipidomics, untargeted lipidomics, lipid annotation, structural characterization, LipidSearch, MS^n, Orbitrap ID-X, AcquireX, data dependent, LC/MS, phosphatidylcholine, triglyceride, insect, larvae, western corn rootworm

Goal

- Demonstrate the utility of LC/MS and data-dependent MS^n acquisition for simultaneous lipid profiling and more complete lipid annotation using the Thermo Scientific™ Orbitrap™ ID-X™ Tribrid™ mass spectrometer and Thermo Scientific™ LipidSearch™ software.
- Show that the use of the Thermo Scientific™ AcquireX intelligent acquisition method combined with neutral loss and product ion scan filters increases the overall quality and number of lipid annotations.
- Investigate the utility of combining HCD MS^2 and CID MS^n for structural determination of phosphatidylcholine and triglyceride species in lipid extracts from insect larvae.

Introduction

The application of lipidomics to phenotypical analysis of a wide range of plant tissues[1] and insects is becoming a more important aspect of agricultural research[2]. Insect lipids are highly structurally diverse species that perform many important functions including storage of metabolic energy, contributing to the structure of membranes, protection against dehydration and pathogens, and circulating energy molecules.[3] In addition, essential lipids such as sterols (precursors to hormones) and polyunsaturated fatty acids

ThermoFisher SCIENTIFIC

are only available from the diet. Understanding how lipid molecular species change in response to both diet and age is critical to define nutrient requirements and fitness relationships. However, conducting detailed studies of insect lipid composition has been technically challenging due to the complex nature of insect lipid extracts and the lack of software for automated lipid annotation.

We evaluated the robustness of a new high-resolution LC/MS^n approach to perform non-targeted lipid profiling experiments. The western corn rootworm (WCR) larvae (*Diabrotica virgifera virgifera*) was chosen due to its economic impact, estimated at 2 billion US dollars associated with its control and corn production loss in the Americas.[4,5] In this study, the effects of larval growth and diet were investigated using a Thermo Scientific™ Orbitrap™-based mass spectrometer and LipidSearch software. New software algorithms were introduced specifically to reduce false positives, improve quantitation using labeled internal standards, and automate processing of LC/MS^n structural data obtained by higher-energy collisional dissociation (HCD) and linear ion trap collisional induced dissociation (CID) fragmentation methods.

The analytical challenges of identifying isomeric/isobaric lipids using an infusion workflow was recently reviewed.[6] The choice of the mass spectrometric method, mass resolution used for analysis, and the specific structural information obtained with the optimized MS/MS conditions determines the level of annotation that is reported for every lipid species. A fundamental issue in lipid identification is that lipids dissociate in a very predictable way to give relatively few product ions, and thus structural details are often incomplete. Obtaining more complete lipid structural information can be addressed in several different ways, including the following:

- Selective derivatization to shift the mass of isomeric species or to change fragmentation
- Separation by chromatographic or other means such as differential mobility (FAIMS)
- CID MS^n analysis to obtain a series of selective transitions revealing structure
- Alternative dissociation techniques such as UV photodissociation (UVPD)

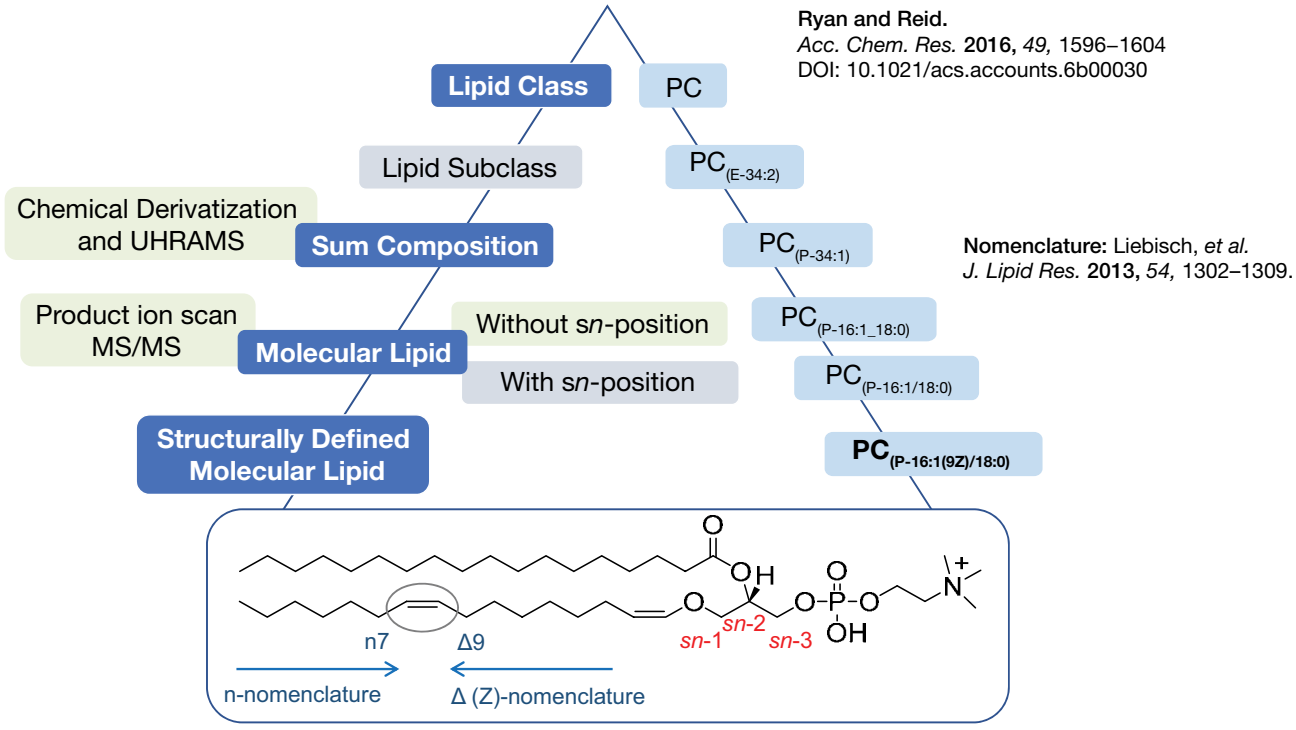
Here, we use a reversed-phase UHPLC separation and lipid class-based LC/MS^n analysis in combination with the AcquireX data acquisition strategy[7] to provide deeper annotation and higher confidence in lipid annotation for isomeric phosphatidylcholine (PC), diglyceride (DG), and triglyceride (TG) lipids.

As shown in Figure 1, ultra-high resolution (more than 100,000 resolution) MS data provides lipid annotation at the sum composition (elemental formula) level. The addition of MS/MS data provides fatty acyl information and allows annotation at the molecular lipid level. However, MS^2 information is often incomplete or ambiguous due to the presence of isomeric mixtures. CID MS^n information is essential when it is not possible to determine the fatty acyl composition from a mixture of isomers, such as in the analysis of triacylglycerol lipids. Determination of structurally defined molecular lipid species (Figure 1) requires additional structural details (e.g., fatty acyl position on the glycerol backbone, double bond location, *cis* vs. *trans*) from chemical reaction/derivatization or MS^n in combination with alternative activation methods such as UVPD (ultra-violet photodissociation).

Experimental methods

Sample preparation

Fresh frozen insect larvae (approximately 100 mg) were weighed and softened in hot isopropanol (IPA) for 20 minutes to inhibit lipases.[8] The samples were then homogenized using a motorized glass rod and re-extracted for an additional 10 minutes. All organic solvent contained 0.01% of butylated hydroxytoluene. After the addition of chloroform and water (1:0.4), the samples were agitated vigorously for 30 minutes at room temperature. For tissue containing a significant amount of neutral lipids, a second extraction with chloroform-methanol (2:1) was performed with 30 minutes agitation at room temperature. The organic phases were pooled and washed with 1 M potassium chloride followed by water, and the organic layer was then filtered (PTFE, 0.45 µm) and evaporated to dryness under nitrogen prior to reconstitution in isopropanol-methanol-chloroform (0.45:0.45:0.1) containing SPLASH® LIPIDOMIX® standards (Avanti Polar Lipids, Inc) at a 1:10 dilution. In this study, four different larvae populations (natural and artificial diet, Instar 1 and Instar 3 developmental stages) were extracted and analyzed in triplicate as shown in Figure 2.



Adapted from: Ekroos, K. In Lipidomics, Ch 1. 2012 Wiley-VCH Verlag GmbH & Co

Figure 1. Hierarchal scheme of lipid classification

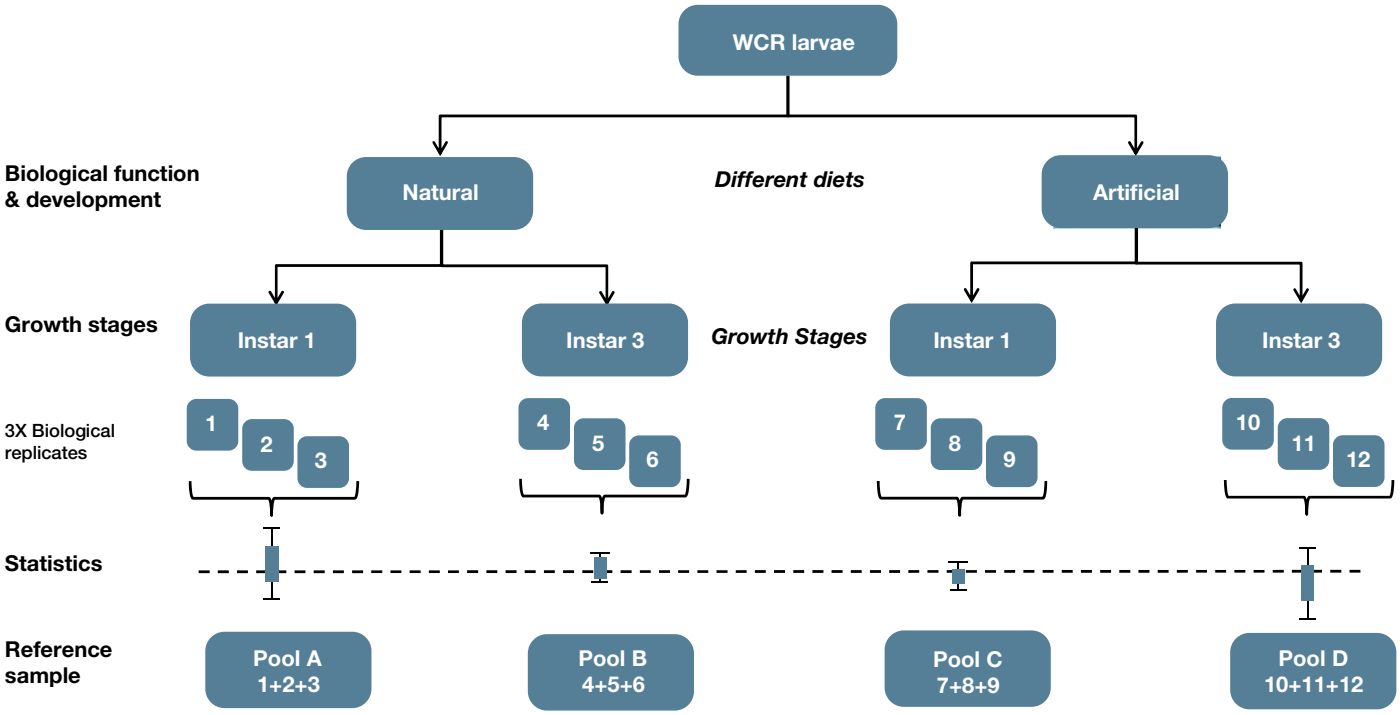


Figure 2. Western corn rootworm lipidomics study: experimental design

Mass spectrometry

Total lipid extracts from western corn rootworm (WCR) larvae were separated using a Thermo Scientific™ Accucore™ 2.1 × 150 mm C30, 2.7 µm column and a Thermo Scientific™ Vanquish™ chromatograph. The chromatograph was operated at a flow rate of 260 µL/min, the 2.1 × 150 mm, 2.7 µm C30 Accucore column maintained at 45 °C, and the injection volume was 2 µL. Chromatographic gradient conditions are shown in Table 1. Mobile phase A was 60:40 acetonitrile/water and mobile phase B was 90:10 isopropanol/acetonitrile. Both mobile phase A and B contained 10 mM ammonium formate and 0.1% formic acid.

LC/MS^n analyses were performed using a Thermo Scientific™ Orbitrap ID-X™ Tribrid™ mass spectrometer (Figure 3) using two different methods: 1) a standard HCD data dependent MS^2 method and 2) an AcquireX experimental workflow[7] for lipid characterization (Figure 4). LC/MS at 120,000 resolution (FWHM @ *m/z* 200) and data-dependent HCD MS^2 experiments (15,000 resolution) were performed in positive and negative ion modes with the instrument conditions summarized in Table 2. During each 1.5 s cycle of the AcquireX dd-MS^n profiling method, additional targeted product ion (*m/z* 184.0733) or neutral loss (fatty acid

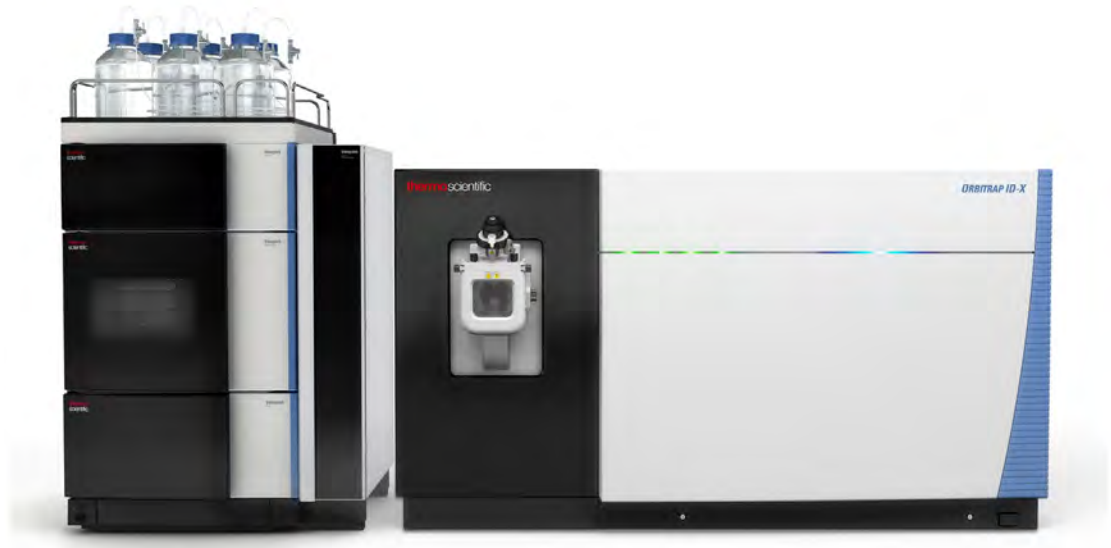


Figure 3. Thermo Scientific Orbitrap ID-X mass spectrometer

Table 1. Chromatographic gradient conditions

Time (min)	% A	% B
0.00	70	30
2.00	57	43
2.10	45	55
12.00	35	65
18.00	15	85
20.00	0	100
25.00	0	100
25.10	70	30
30.00	70	30
31.00	70	30

Table 2. Orbitrap ID-X conditions

Parameter	Value
Spray voltage	3500 V pos, 2400 V neg
Vaporizer temperature	300 °C
Ion transfer temperature	300 °C
RF Lens	45
MS OT resolution	120,000 (FWHM @ *m/z* 200)
MS^2/MS^3 OT resolution	15,000 (FWHM @ *m/z* 200)
HCD CE	25-30-35 pos, 20-40-60 neg
CID MS^2 CE	32%
CID MS^3 CE	35%
Isolation	1.6 Da HCD, 2.0 Da CID
Dynamic exclusion	5 s
Cycle time	1.5 s

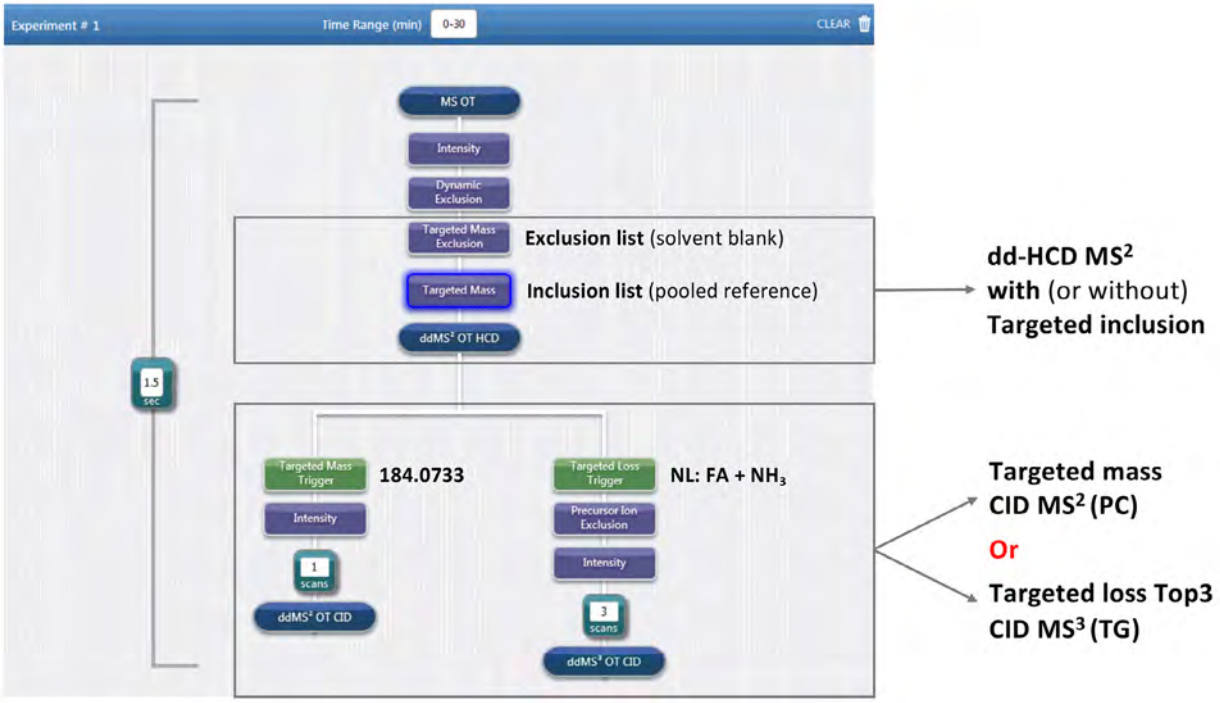


Figure 4. AcquireX instrument acquisition method

plus ammonia, Table 3) CID MS^2 and MS^3 experiments were selectively performed to provide higher quality characterization of phosphatidylcholine (PC) and triglyceride (TG) lipids. Prior to analysis of the biological samples, LC/MS (120,000 resolution) analyses of blank and pooled samples were used to automatically create an exclusion list for background ions and inclusion list for sample-relevant ions. Both lists were dynamically updated following subsequent LC/MS^n analyses of the individual biological replicates.

Table 3 (part 1). Neutral loss inclusion list

FA	Formula	NL (FA + NH_3)
12:0	$C_{12}H_{24}O_2$	217.2042
12:1	$C_{12}H_{22}O_2$	215.1885
14:0	$C_{14}H_{28}O_2$	245.2355
14:1	$C_{14}H_{26}O_2$	243.2198
16:0	$C_{16}H_{32}O_2$	273.2668
16:1	$C_{16}H_{30}O_2$	271.2511
16:2	$C_{16}H_{28}O_2$	269.2355
17:0	$C_{17}H_{34}O_2$	287.2824
17:1	$C_{17}H_{32}O_2$	285.2668
18:0	$C_{18}H_{36}O_2$	301.2981
18:1	$C_{18}H_{34}O_2$	299.2824
18:2	$C_{18}H_{32}O_2$	297.2668
18:3	$C_{18}H_{30}O_2$	295.2511
19:0	$C_{19}H_{38}O_2$	315.3137
20:0	$C_{20}H_{40}O_2$	329.3294
20:1	$C_{20}H_{38}O_2$	327.3137

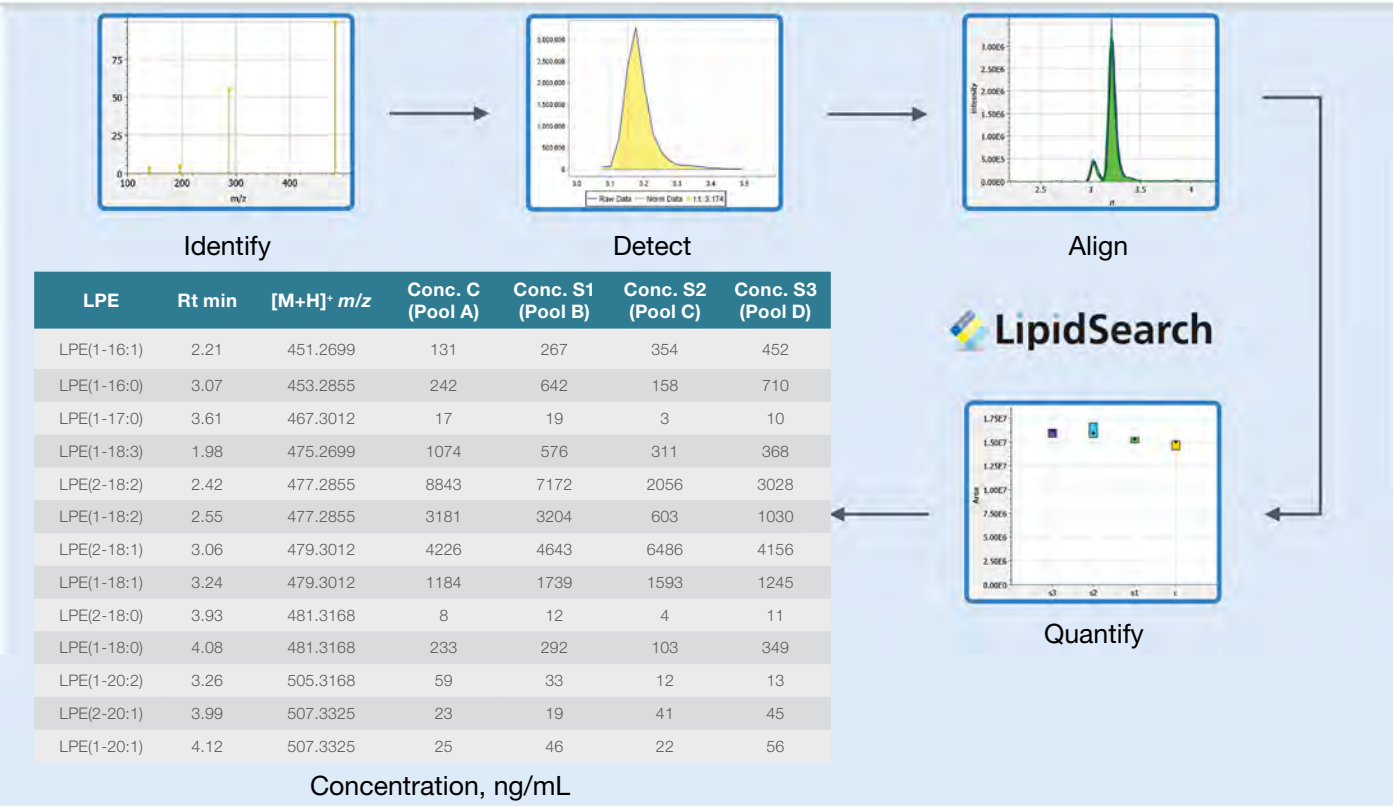
Table 3 (part 2). Neutral loss inclusion list

FA	Formula	NL (FA + NH_3)
20:2	$C_{20}H_{36}O_2$	325.2981
20:3	$C_{20}H_{34}O_2$	323.2824
20:4	$C_{20}H_{32}O_2$	321.2668
20:5	$C_{20}H_{30}O_2$	319.2511
21:0	$C_{21}H_{42}O_2$	343.3450
22:0	$C_{22}H_{44}O_2$	357.3607
22:1	$C_{22}H_{42}O_2$	355.3450
22:2	$C_{22}H_{40}O_2$	353.3294
22:3	$C_{22}H_{38}O_2$	351.3137
22:4	$C_{22}H_{36}O_2$	349.2981
22:5	$C_{22}H_{34}O_2$	347.2824
22:6	$C_{22}H_{32}O_2$	345.2668
23:0	$C_{23}H_{46}O_2$	371.3763
24:0	$C_{24}H_{48}O_2$	385.3920
24:1	$C_{24}H_{46}O_2$	383.3763
26:0	$C_{24}H_{52}O_2$	413.4233

Data processing

LC/MS^n datasets were processed using Thermo Scientific™ LipidSearch™ 4.2 software (Figure 5) with an expanded lipid database, improved peak detection during alignment of annotated peaks, and better rejection of false positives. The database for TG species was modified to include 2:0, 3:0, 5:0 and 7:0 fatty acids in order to annotate some unusual TG species with one very short acyl chain[9]. The data processing parameters used for the WCR lipid samples are summarized in Table 4.

Multiple HCD MS^2, CID MS^2, and MS^3 product ion mass spectra for the same precursor ions were automatically combined to provide more comprehensive annotation for hundreds of lipid molecular species. The search results from positive/negative ion, HCD MS^2 and CID MS^2 and MS^3 spectra were aligned by retention time and the combined annotation was obtained from all the lipid adduct ions. In the search results, isomeric lipid species are scored by how well each predicted fatty acyl combination matches the actual MS/MS spectrum. These match scores are used to rank order these results and the isomer with the best ranking is used to represent the lipid species in the alignment results. Thus, LipidSearch reports the main lipid species based on the best match score, which includes the results obtained for CID MS^2 and MS^3 spectral matches at a specified retention time.



LPE	Rt min	[M+H]$^+$ *m/z*	Conc. C (Pool A)	Conc. S1 (Pool B)	Conc. S2 (Pool C)	Conc. S3 (Pool D)
LPE(1-16:1)	2.21	451.2699	131	267	354	452
LPE(1-16:0)	3.07	453.2855	242	642	158	710
LPE(1-17:0)	3.61	467.3012	17	19	3	10
LPE(1-18:3)	1.98	475.2699	1074	576	311	368
LPE(2-18:2)	2.42	477.2855	8843	7172	2056	3028
LPE(1-18:2)	2.55	477.2855	3181	3204	603	1030
LPE(2-18:1)	3.06	479.3012	4226	4643	6486	4156
LPE(1-18:1)	3.24	479.3012	1184	1739	1593	1245
LPE(2-18:0)	3.93	481.3168	8	12	4	11
LPE(1-18:0)	4.08	481.3168	233	292	103	349
LPE(1-20:2)	3.26	505.3168	59	33	12	13
LPE(2-20:1)	3.99	507.3325	23	19	41	45
LPE(1-20:1)	4.12	507.3325	25	46	22	56

Figure 5. LipidSearch Software Workflow

Table 4. LipidSearch processing conditions

Search Parameter	Setting	Units
Precursor mass tolerance	5	ppm
Product mass tolerance	10	ppm
Product intensity threshold	1.0	%
m-Score threshold/display	2.0/5.0	
Quan *m/z* tolerance	±5.0	ppm
Quan range	±0.5	min
Main isomer peak filter	ON	
ID Quality filter	A, B, C, D	
Adducts (pos. ion)	+H, NH_4, Na, +H-H_2O	
Adducts (neg. ion)	-H, -CH_3, +CH_3CO_2	
Lipid Classes	*Lipids	
Alignment Parameter	**Setting**	**Units**
R.T. tolerance	0.10	min
All isomer peak filter	ON	
m-Score threshold	5.0	
ID Quality filter	A, B, C, D	
Configuration	**Setting**	
Number data points threshold	5	
Intensity baseline	0.05	%
Intensity ratio threshold	3	
S/N ratio threshold	5	

Results

In parallel to the AcquireX method for efficient acquisition of the data-dependent LC/MS^2, neutral loss and product ion directed CID MS^2 and MS^3 experiments were performed in the same instrument cycle to improve the structural characterization of two of the main lipid sub-classes (phosphatidylcholines and triglycerides) expected in the insect larvae. Ultra-high-resolution LC/MS analysis provides lipid annotation at *sum composition level* (Figure 1). The high-resolution positive ion mass spectrum of the control sample 1 from pool A (Figure 6a) shows the molecular ion region for a lipid species with the elemental composition $C_{42}H_{82}NO_8P$. The protonated and sodiated species fit PC 34:1 or PE 37:3 lipids.

For annotation at the molecular lipid level (Figure 1), MS^2 and MS^3 information allows determination of lipid headgroup and fatty acyl information. As illustrated in Figure 6b, the HCD MS^2 spectrum of *m/z* 760.5851 gives a predominant *m/z* 184 product ion, which allows assignment of PC 34:1 as the lipid species. Observation of the *m/z* 184 product ion using the scan filter triggered the acquisition of the CID MS^2 mass spectrum shown in Figure 6c. The neutral loss of the phosphocholine lipid headgroup (183) and the neutral losses of fatty acid 16:0 and 18:1, as well as the loss of the corresponding ketenes, provide unequivocal assignment of the lipid annotation: PC 16:0_18:1.

Note that LipidSearch software version 4.2 has been modified to use the shorthand notation suggested by Liebisch, *et al.*[10] and for lipid species with positional isomers, the use of an underscore indicates that the *sn* configuration is unknown. For example, PC 16:0/18:1 indicates 16:0 in the *sn1* position of the glycerol and 18:1 in the *sn2* position as shown in Figure 7. PC 16:0_18:1 indicates the sn position is 16:0/18:1 or 18:1/16:0.

LipidSearch software was used to search the MS^2 and MS^3 spectra against the predicted product ions and neutral losses for all potential lipid species within 5 ppm precursor and 10 ppm product ion mass tolerances. For each LC/dd-MS^n analysis, potential lipid species were identified separately from the positive or negative ion MS^2 or MS^3 spectra. All the data for each biological replicate were aligned within a chromatographic time window by combining the positive and negative ion annotations and merging these into a single lipid annotation in the results. This approach provides lipid annotations that reflect the appropriate level of MS^n information from the entire dataset giving higher confidence in lipid identifications. The alignment results were filtered by minimum number of data points, signal-to-noise ratio, main adduct ion, and ID quality.

The numbers of the highest quality lipid annotations per sub-class after filtering rejected peaks are summarized in Table 5. Analysis of the insect lipid extracts using data dependent LC/MS^2 gave a total of 866 lipid annotations after filtering, whereas the LC/MS^n approach with AcquireX gave a total of 1045 lipid annotations, representing an overall increase of 179 annotations (+17%). Moreover, the average match scores increased from 31.4 for the HCD results to 38.5 (+7.1) for the AcquireX results. The increase in match scores is more apparent when looking at the targeted PC and neutral lipid species for CID MS^n (ChE, CoQ, DG, LPC, PC, SiE, TG and WE). The average match score for these sub-classes increased from 33.9 for HCD to 46.5 (+12.6) for the AcquireX results. Thus, both the number of lipid annotations and the quality of the spectral matches improves substantially using the AcquireX approach, leading to higher confidence.

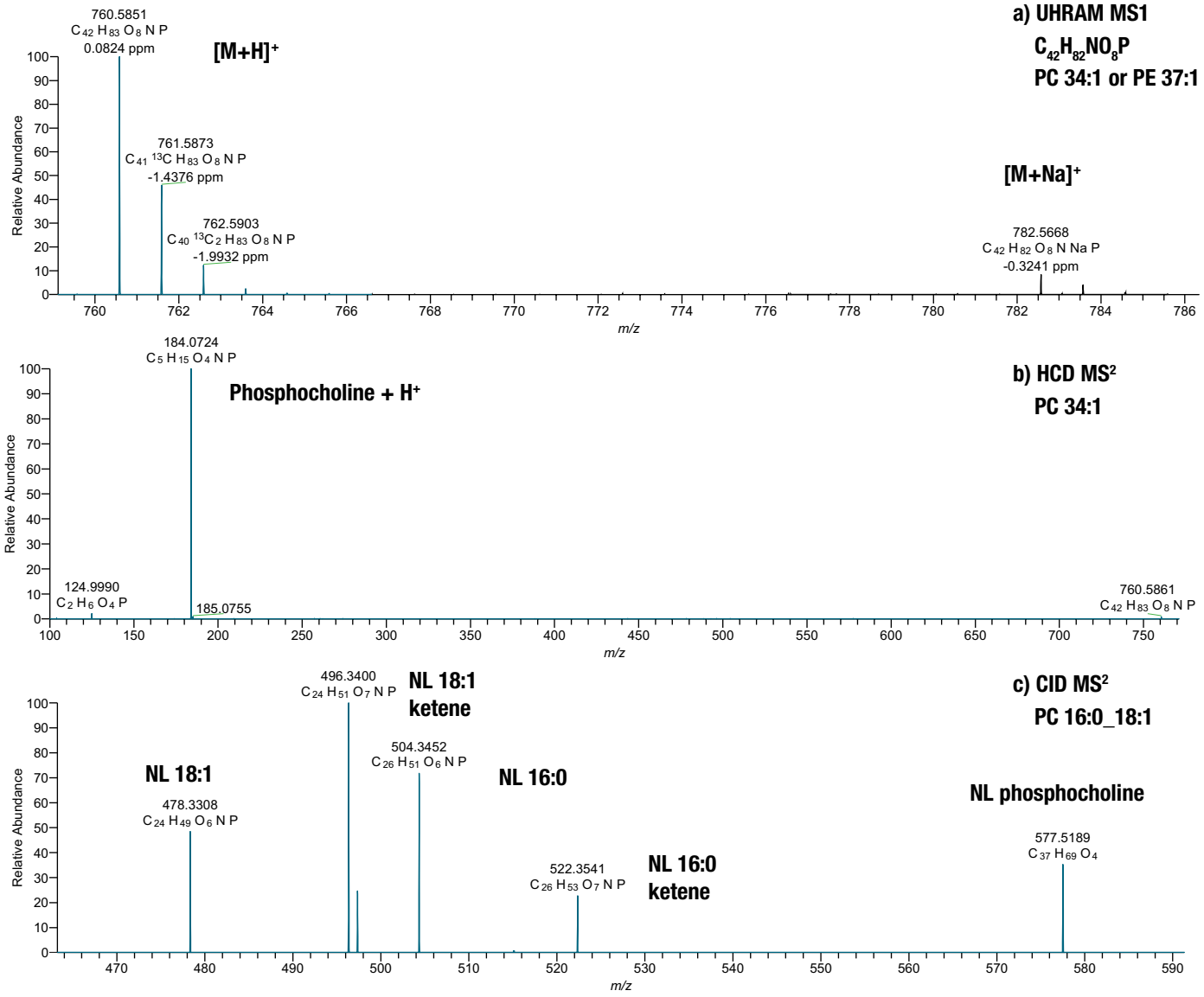


Figure 6. Confident annotation of phosphatidylcholine 34:1

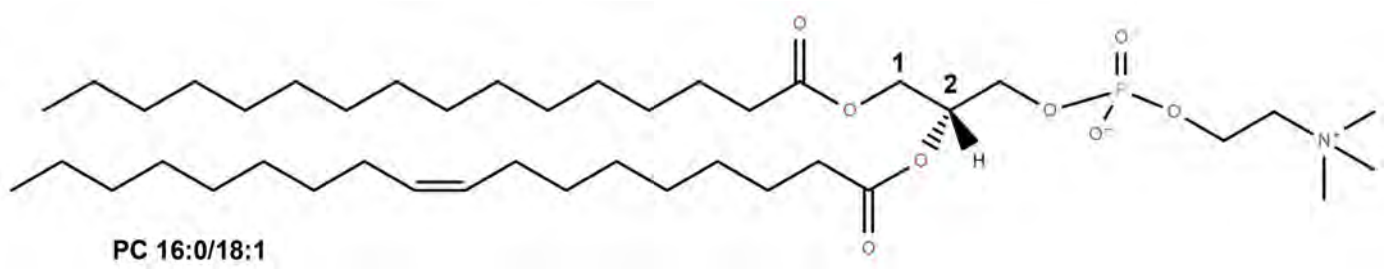


Figure 7. Structure: PC 16:0/18:1

Table 5. Summary of lipid species from WCR extracts

Lipid Species by Sub-class	dd-MS² (filtered)	AcquireX (filtered)	% Change
Acylcarnitine	5	4	-25
Ceramides (Cer)	66	55	-20
Ceramide PE (CerPE)	1	3	67
Cholesterol ester (ChE)	3	4	25
Co-enzyme Q	0	4	
Diacylglycerol (DG)	38	51	25
Hexosyl ceramide (Hex1Cer)	25	23	-9
$Hexosyl_2$ ceramide (Hex2Cer)	2	2	0
$Hexosyl_3$ ceramide (Hex3Cer)	5	3	-67
Hexosyl sphingosine (HexSPH)	1	1	0
Lyso PC (LPC)	29	50	42
Lyso PE (LPE)	27	27	0
Lyso PG (LPG)	6	1	-500
Lyso PI (LPI)	15	6	-150
Lyso PS (LPS)	10	11	9
Phosphatidic acid (PA)	16	15	-7
Phosphatidylcholine (PC)	40	68	41
Phosphatidylethanolamine (PE)	83	70	-19
Phosphatidylglycerol (PG)	10	6	-67
Phosphatidylinositol (PI)	59	34	-74
Phosphatidylserine (PS)	36	29	-24
Sphingomyelin (SM)	52	117	56
Sitosterol ester (SiE)	1	2	50
Triacylglycerol (TG)	329	449	27
Wax ester (WE)	7	10	30
TOTAL	866	1045	17

Discussion

LC/MSⁿ

For lipid identification, a standard HRAM workflow is the LC/data-dependent MS^2 approach along with LipidSearch software for structure annotation (Figure 5). This provides simultaneous untargeted profiling and identification for lipids from cells, plasma, and tissues.

A common issue in lipid software today is the over-reporting of lipid annotations. For example, exact structures are used to represent lipids from LC/MS profiling experiments even though the MS/MS information is often incomplete. Without additional experiments with reference standards it is not possible to unequivocally confirm assignment of double bond location, stereo-chemistry and sn position on the glycerol backbone. Thus, LipidSearch software reports a shorthand notation as suggested by Liebisch[10] to represent a partial structure. Instead of the exact structure for phosphatidylcholine 1-16:0/2-18:1 (9Z), PC 16:0_18:1 is reported indicating that sn position, double bond location, and stereochemistry are not assigned. LipidSearch software reports the sum composition level (PC 34:1, Figure 6a, 6b) to indicate that the fatty acyl information is not known. If CID MS^2 data are also present for the protonated molecular ion, the fatty acyl groups are confirmed and then PC 16:0_18:1 is reported (Figure 6c).

The analysis of complex lipid extracts from insect larvae requires a more sophisticated approach to distinguish coeluting isomers. Structure-specific CID LC/MS^2 or LC/MS^3 experiments (Figure 3) were used to selectively characterize specific lipids during a data-dependent LC/MS^2 run using the Thermo Scientific ID-X Tribrid mass spectrometer and advanced scan filters. WCR larvae contain a high content of triglycerides in addition to phospholipids, sphingolipids, and sterols. In this application, we focused on PC and neutral TG lipids for further LC/MS^n characterization. With HCD MS^2 data, neutral loss of fatty acid and ammonia is the signature fragmentation observed for several classes of neutral lipids including CoQ, DG, TG, and sterol esters. The Orbitrap ID-X Tribrid mass spectrometer provides an intelligent workflow for monitoring of class-specific product ion or neutral losses and automatically conducting a predefined experiment on the same precursor ion during the instrument cycle. This approach efficiently provides additional structure-based information without wasting time scheduling targeted experiments within predefined retention times. The use of AcquireX helps to increase the efficient use of the full capabilities of the Orbitrap ID-X Tribrid mass spectrometer by getting more information on relevant ions resulting in an increase in the number of lipid annotations. In addition, targeting the class specific information increased the lipid match scores for the neutral TG and PC lipid sub-classes of interest in this experiment.

Figure 8a shows the mass chromatogram obtained for protonated PC 34:0 at *m/z* 762.6010 and a retention time of 12.22 min. The combined HCD and CID MS^2 spectrum shown in Figure 8b identifies the lipid as a phosphatidylcholine (HCD, *m/z* 184.0733) with 16:0 and 18:0 fatty acyl (FA) chains, information derived from the CID neutral loss of 16:0 ketene (*m/z* 524.3711) and 18:0 FA (*m/z* 478.3292), respectively. Since PC 34:0 was lacking fatty acyl information in negative ion mode, it would not possible to assign the fatty acyl information using HCD MS^2 alone. By obtaining the positive ion CID MS^2 spectrum, it is feasible to obtain the information needed for more complete annotation. Figure 8c shows the peak areas obtained for the triplicate biological replicates for the four different experimental conditions. The pattern in the peak areas indicates an “Age” phenotype related to the larval developmental stage.

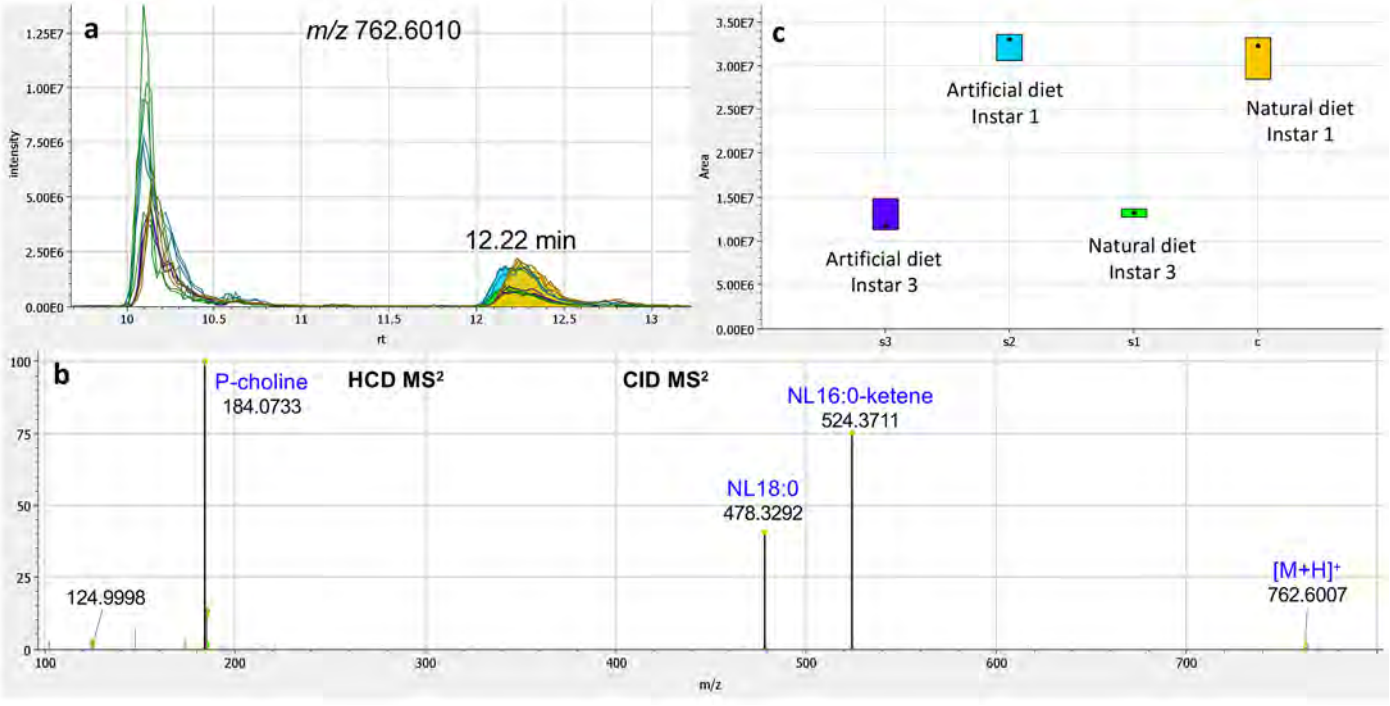


Figure 8. PC 34:0 $[M+H]^+$ a) Mass chromatogram *m/z* 762.6007, b) HCD and CID MS^2 spectra and c) peak area statistics

Figure 9 illustrates the annotation of triglyceride 48:1 ammonium adduct (*m/z* 822.7548 at 20.66 min) found in corn rootworm larvae lipid extracts. Three product ions are observed in the HCD MS^2 spectrum corresponding to neutral loss of 18:1, 16:0, and 14:0 fatty acids. During a single scan cycle, the neutral losses of fatty acid were automatically detected and three additional CID MS^3 scans were performed. The MS^3 spectrum corresponding to loss of 18:1 fatty acid (Figure 9a) produces 14:0 and 16:0 acyl ions giving the assignment TG 18:1-14:0-16:0 (*isomer 1*). Similarly, the MS^3 spectrum corresponding to 14:0 loss (Figure 9b) produces 16:0 and 18:1 acyl ions giving the same assignment (*isomer 1*). However, the MS^3 spectrum from 16:0 loss (Figure 9c) is a mixture consisting mainly of *isomer 1* (14:0 and 18:1 acyl ions) and a lesser amount of *isomer 2,* TG 16:0-16:1-16:0, giving rise to 16:0 and 16:1 acyl ions. This example illustrates the power of LC/MS^n for elucidating the structure of isomeric mixtures.

Figure 10a shows the HCD MS^2 spectrum of diglyceride (DG 34:2) sodium adduct (*m/z* 615.4957, 12.31 min), which gives low abundant product ions formed by the loss of the fatty acid sodium salt. This result gives an annotation of grade "B" since the product ions of the fatty acyl ions are not directly observed from the sodium adduct ion. The diglyceride 34:2 ammonium ion (*m/z* 610.5406) dissociates by neutral loss of water and ammonia to give *m/z* 575.5034 in the MS^2 spectrum, which in turn gives four different CID MS^3 product ions (Figure 10b). The neutral losses of 16:0 FA and 18:0 ketene combined with the 16:0 and 18:2 fatty acyl product ions are entirely consistent with annotation of DG 16:0_18:2.

Figure 11a shows the chromatogram of 18:2 sitosteryl ester $[M+NH_4]^+$ at *m/z* 694.6497 and 21.70 minutes. Figure 11b shows the main HCD product ion is the loss of 18:2 FA and ammonia to give the sitosterol carbonium ion (*m/z* 397.3829). The CID MS^3 spectrum of the *m/z* 397 ion (Figure 11c) gives the same product ions as those observed in the HCD spectrum, confirming these ions come from the sitosterol hydrocarbon backbone.

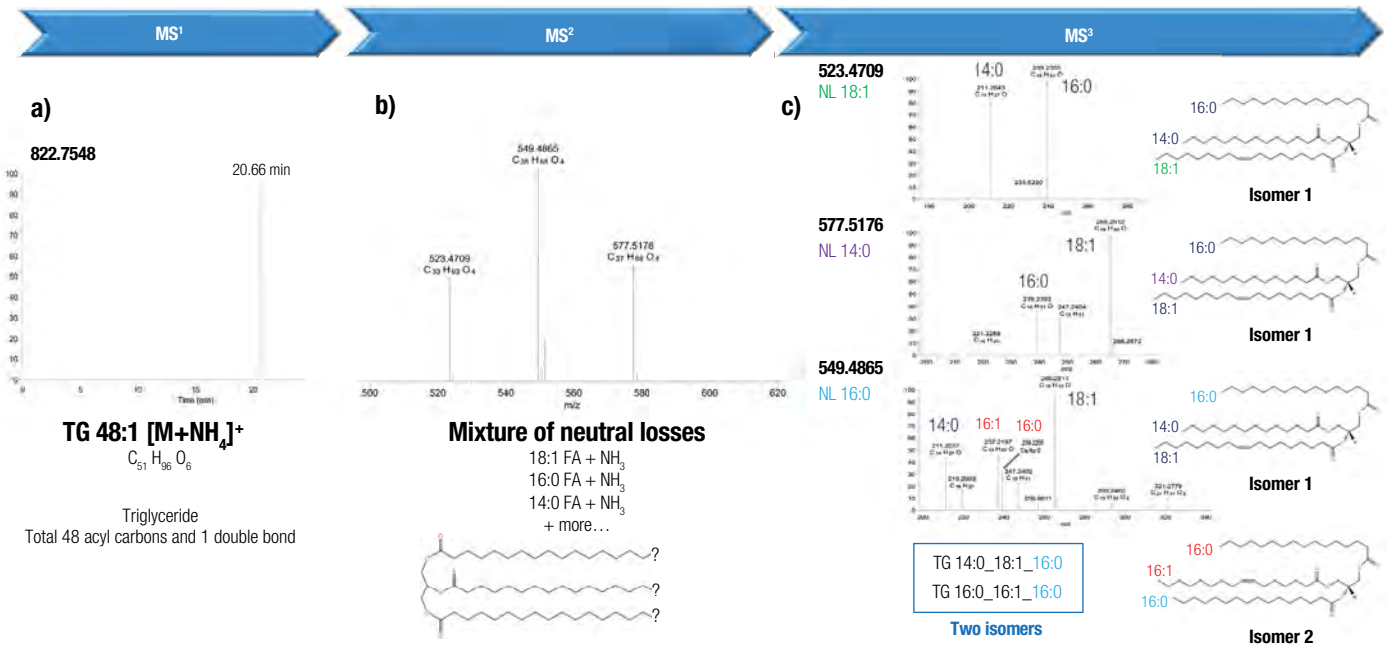
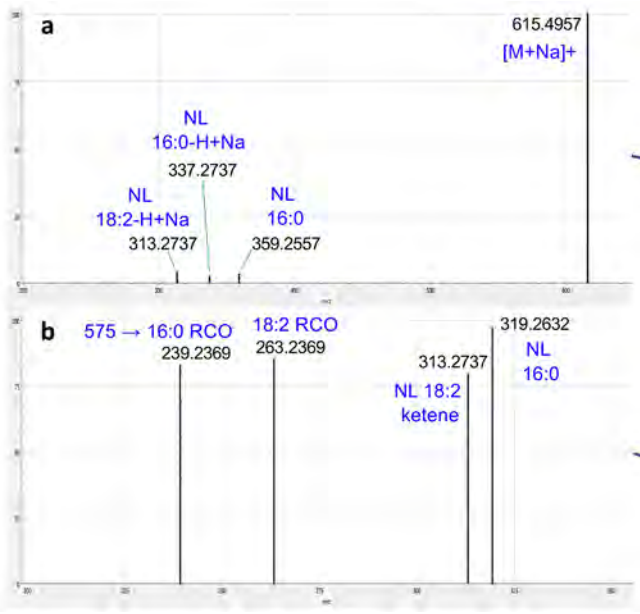


Figure 9. Complete characterization of triglyceride 48:1: a) chromatogram, b) HCD MS^2 spectrum and c) CID MS^3 spectra

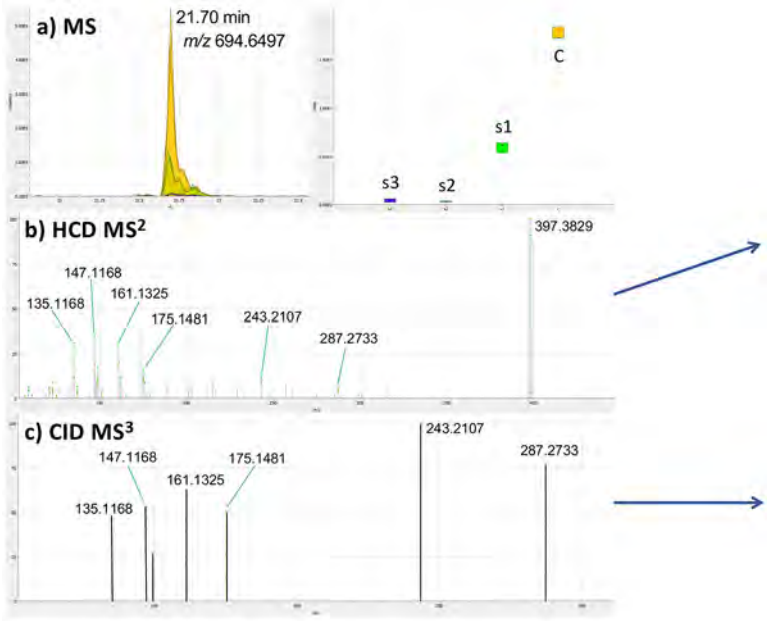


DG 34:2 [M+Na]+ Grade = B			
MS	MS²	Products	Assignment
615.4957	359.2557	NL 16:0	DG 16:0_18:2
	337.2737	NL 16:0-H+Na	
	313.2737	NL 18:2-H+Na	

Combining HCD MS² and CID MS³ provides higher quality lipid annotation

m/z 610. DG 34:2 [M+NH4]+ Grade = A			
MS²	MS³	Products	Assignment
575.5034	319.2632	NL H_2O+NH_3, 16:0	DG 16:0_18:2
	313.2737	NL H_2O+NH_3, 18:2 ketene	
	263.2369	NL H_2O+NH_3, 18:2 RCO	
	239.2369	NL 18:2+NH_3, 16:0 RCO	

Figure 10. a) HCD MS² of DG 34:2 [M+Na]⁺ at 12.31 min and b) CID MS³ spectra of DG 34:2 [M+NH_4]⁺



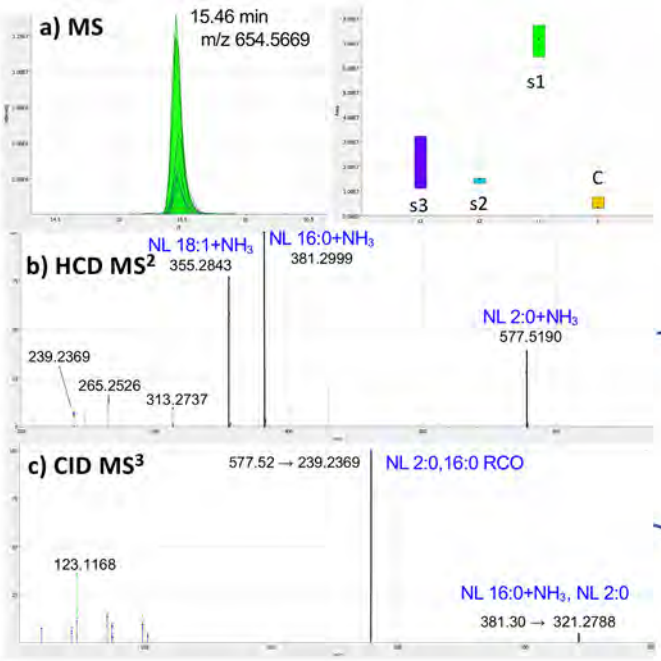
Sitosterol 18:2 – Grade B

MS² *m/z* 694.6497, SiE 18:2 [M+NH_4]⁺		
MS²	Products	Assignment
397.3829	$C_{29}H_{49}$ – NL 18:2+NH_3	Sitosterol 18:2
287.2733	C_21H_{35}	
243.2107	$C_{18}H_{27}$	
175.1481	$C_{13}H_{19}$	
161.1325	$C_{12}H_{17}$	
147.1168	$C_{11}H_{15}$	

MS³ *m/z* 694.6497, SiE 18:2 [M+NH_4]⁺			
MS²	MS³	Products	Assignment
397.3829	287.2733	$C_{21}H_{35}$	Sitosterol+H-H_2O
	243.2107	$C_{18}H_{27}$	
	175.1481	$C_{13}H_{19}$	
	161.1325	$C_{12}H_{17}$	
	149.1325	$C_{11}H_{17}$	
	147.1168	$C_{11}H_{15}$	
	135.1168	$C_{10}H^{15}$	

Figure 11. a) Chromatogram of sitosterol 18:2, [M+NH_4]⁺, b) HCD MS² spectrum and c) CID MS³ spectrum

Figure 12 shows the annotation of an unusual 2:0 fatty acid containing triglyceride. The *m/z* 654.5669 chromatogram (Figure 12a) shows a very abundant ammonium adduct for TG 36:1. The HCD MS² spectrum (Figure 12b) could not be explained using the normal range of fatty acids contained in the LipidSearch database for triglycerides. Manual inspection of the first neutral loss (77.0479 Da) suggested that a 2:0 fatty acid was present. After modifying the database to include 2:0 FA, the assignment of TG 2:0_16:0_18:1 was supported by all the HCD MS² product ions. The MS³ data (Figure 12c) also fully supports this assignment by confirming the neutral loss of 2:0 FA after the initial loss of 16:0 FA and ammonia, adding confidence to the assignment.



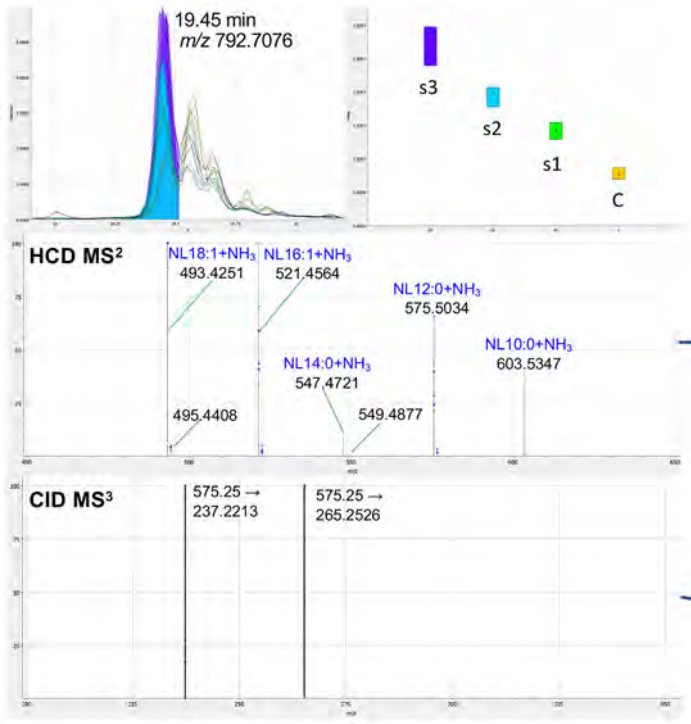
TG 36:1 – 2:0-16:0-18:1
All major product ions are accounted for with a single isomer containing 2:0 fatty acid.

MS^2 Results, *m/z* 654.5669, TG 36:1 $M+NH_4$		
MS^2	Products	Assignment
577.5190	NL 2:0+NH_3	2:0-16:0-18:1
381.2999	NL 16:0+NH_3	
355.2843	NL 18:1+NH_3	
313.2737	NL 2:0+NH_3 + NL 18:1 ketene	
265.2526	18:1 RCOO	
239.2369	16:0 RCOO	

MS^3 Results, *m/z* 654.5669, TG 36:1 $M+NH_4$			
MS^2	MS^3	Products	Assignment
355.2843	239.2369	NL 18:1+NH^3, 16:0 RCO	2:0-16:0-18:1
381.2999	321.2788	NL 16:0+NH_3, NL 2:0	2:0-16:0-18:1
577.5190	239.2369	NL 2:0+NH_3, 16:0 RCO	2:0-16:0-18:1

Figure 12. HCD MS^2 and CID MS^3 of 36:1 triglyceride $[M+NH_4]^+$ from insect larvae

Figure 13 illustrates the difficulty of annotating a complex mixture of co-eluting triglyceride species. The information provided by the HCD MS^2 analysis for TG 46:2 (*m/z* 792.7076) gives 12:0_16:1_18:1 as the predominant species and this is confirmed by the MS^3 results in LipidSearch software as shown by the assignments in blue. However, there are at least 4 other possible combinations based on fatty acid neutral losses.



Mixture of TG 46:2 isomers – 12:0-16:1-18:1 and 10:0-18:1-18:1, 14:0-16:1-16:1, 14:0-14:1-18:1, 14:0-14:0-18:2? MS^2 Results, *m/z* 792.7076, TG 46:2 $[M+NH_4]^+$

MS^2 Results, *m/z* 792.7076, TG 46:2 $[M+NH_4]^+$		
MS^2	Product Ions	Possible Assignments
603.5347	NL 10:0+NH_3	10:0-18:1-18:1
575.5034	NL 12:0+NH_3	12:0-16:1-18:1
549.4877	NL 14:1+NH_3	14:0-14:1-18:1
547.4721	NL 14:0+NH_3	14:0-16:1-16:1
521.4564	NL 16:1+NH_3	12:0-16:1-18:1 or 14:0-16:1-16:1
495.4408	NL 18:2+NH_3	14:0-14:0-18:2
493.4251	NL 18:1+NH_3	12:0-16:1-18:1 or 10:0-18:1-18:1

MS^3 Results *m/z* 792.7076, TG 46:2 $[M+NH_4]^+$			
MS^2	MS^3	Product Ions	Assignment
575.5034	265.2526	NL 12:0+NH_3, 18:1 RCO	12:0-16:1-18:1
	237.2213	NL 12:0+NH_3, 16:1 RCO	12:0-16:1-18:1
521.4564	265.2526	NL 16:1+NH_3, 18:1 RCO	12:0-16:1-18:1
493.4251	237.2213	NL 18:1+NH_3, 16:1 RCO	12:0-16:1-18:1

Figure 13. HCD MS^2 and CID MS^3 of 46:2 triglyceride $[M+NH_4]^+$ from insect larva

Figure 14 shows the three MS^3 spectra obtained for the most abundant species in the MS^2 spectrum. The MS^3 spectra for *m/z* 493 and 521 also confirm there are at least two additional isomers (TG 10:0-18:1-18:1 and 14:0-16:1-16:1, shown in the black assignments) and rule out a substantial amount of the remaining isomers (TG 14:0-14:1-18:1 and 14:0-14:0-18:2). Thus, LC/MS^3 analysis is absolutely required to confirm correct annotations when mixtures of isomeric triglycerides are present.

Conclusions

- This LC/MS^n lipidomics workflow can be applied to any complex biological sample including plasma, plants, tissues, cells, and whole organisms such as insects.
- The Orbitrap ID-X Tribrid mass spectrometer provides a highly sophisticated and customizable workflow for lipidomics and improves lipid structure characterization.
- The AcquireX intelligent LC/MS^n workflow provides 17% more lipid annotations compared to using data-dependent MS/MS analysis alone. DG, PC, and TG annotations, targeted for CID MS^2/MS^3 using product ion or neutral loss scan filters, increased by 25%, 41% and 27%, respectively.
- Furthermore, 275 CID MS^2 / MS^3 spectra were merged with the HCD data to give higher confidence in annotations by an increase in the LipidSearch match score.

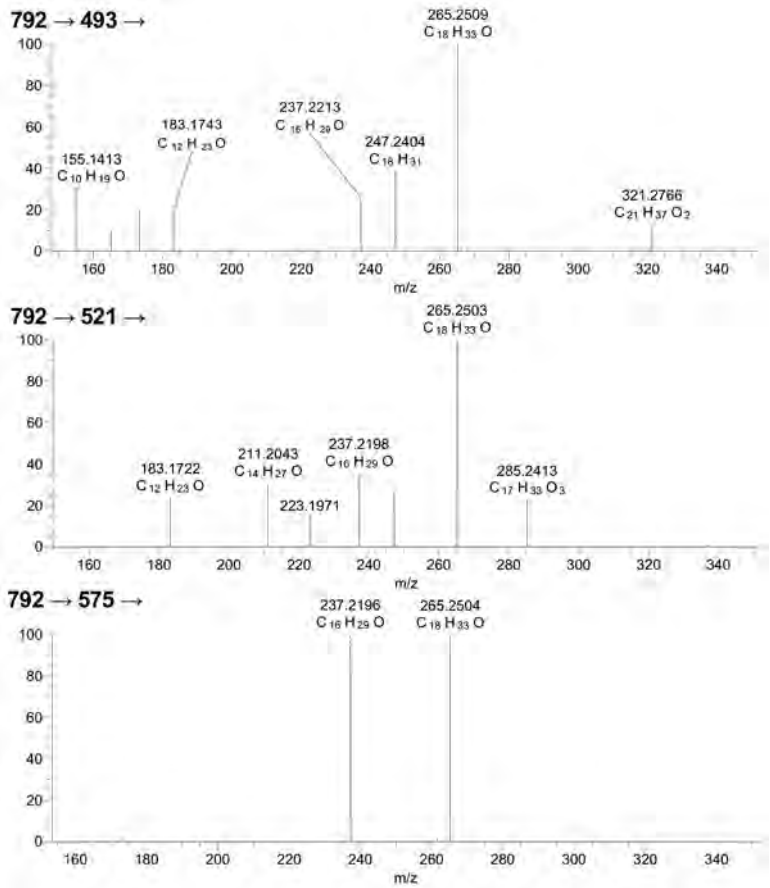


Figure 14. CID MS^3 spectra of 46:2 triglyceride $[M+NH_4]^+$ from insect larva

Mixture of TG 46:2 isomers – 12:0-16:1-18:1 and 10:0-18:1-18:1, 14:0-16:1-16:1, ~~14:0-14:1-18:1, 14:0-14:0-18:2~~

MS^3 Results *m/z* 792.7066, TG 46:2 $[M+NH_4]^+$			
MS^2	MS^3	Product Ions	Assignment
493.4251	321.2788	NL 18:1+NH_3, NL 10:0	10:0-18:1-18:1
	265.2526	NL 18:1+NH_3, 18:1 RCO	10:0-18:1-18:1
	237.2213	NL 18:1+NH_3, 16:1 RCO	12:0-16:1-18:1
	183.1743	NL 18:1+NH_3, 12:0 RCO	12:0-16:1-18:1
	155.1430	NL 18:1+NH_3, 10:0 RCO	10:0-18:1-18:1
521.4564	285.2424	NL 16:1+NH_3, NL 16:1 ketene	14:0-16:1-16:1
	265.2526	NL 16:1+NH_3, 18:1 RCO	12:0-16:1-18:1
	237.2213	NL 16:1+NH_3, 16:1 RCO	14:0-16:1-16:1
	211.2056	NL 16:1+NH_3, 14:0 RCO	14:0-16:1-16:1
	183.1743	NL 16:1+NH_3, 12:0 RCO	12:0-16:1-18:1
575.5034	265.2526	NL 12:0+NH_3, 18:1 RCO	12:0-16:1-18:1
	237.2213	NL 12:0+NH_3, 16:1 RCO	12:0-16:1-18:1

thermo scientific

References

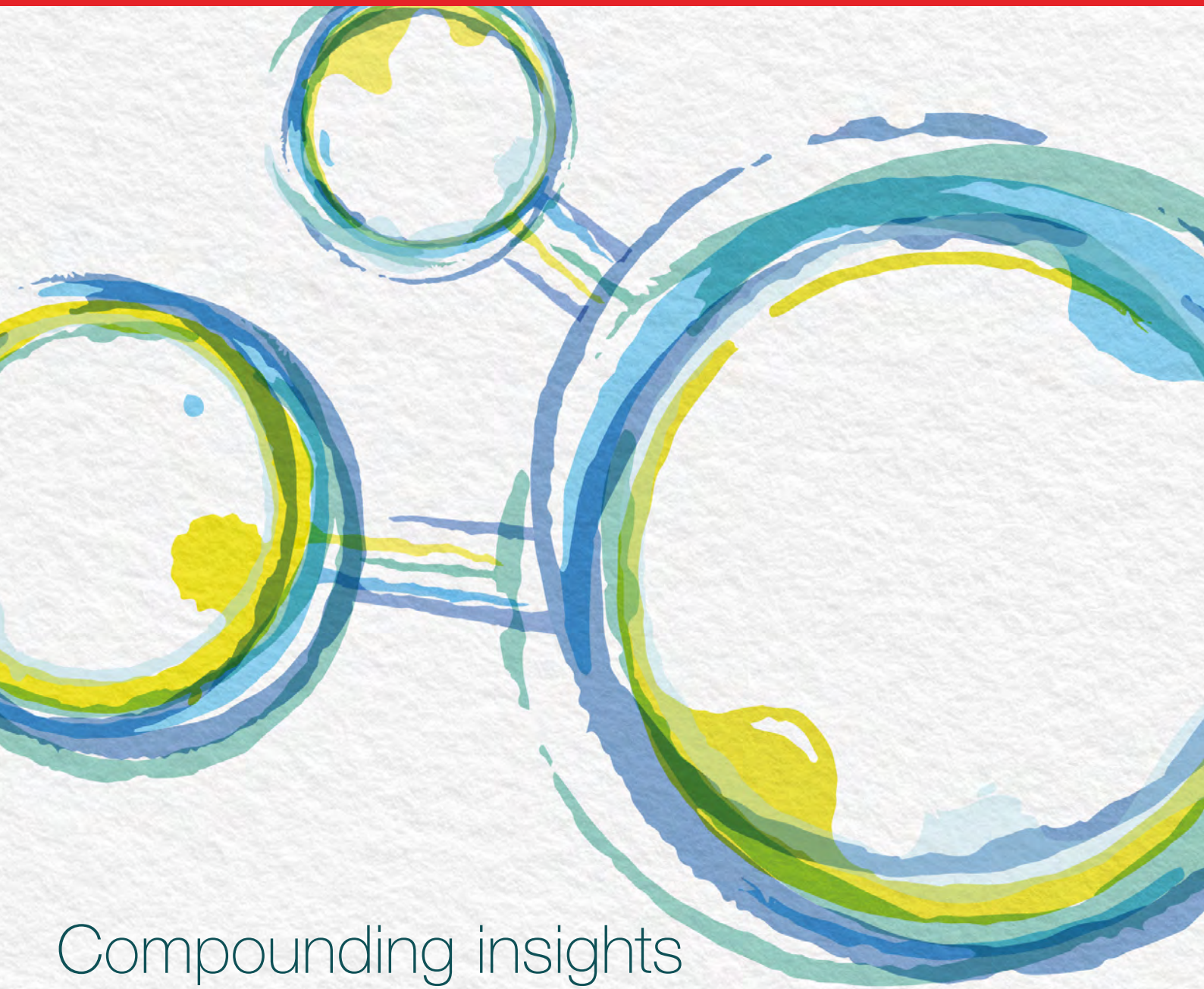
1. The, H.F. et al. Review: Omics and Strategic Yield Improvement in Oil Crops. *J Am Oil Chemists Soc.* **2017,** *94,* 1225–1244.
2. Tian, H. et al. Metabolomics, A Powerful Tool for Agricultural Research. *Int J Mol Sci.* **2016,** *17,* 1871.
3. Dixon, R.A. et al. Perspective – Applications of metabolomics in agriculture. *J Agr Food Chem.* **2006,** *54,* 8984–8994.
4. Gray, M.E. et al. Adaptation and Invasiveness of Western Corn Rootworm: Intensifying Research on a Worsening Pest. *Ann Rev Entomology,* **2009,** *54,* 303 –321.
5. Wesseler, J. and Fall, E.H. Potential damage costs of Diabrotica virgifera virgifera infestation in Europe – the 'no control' scenario. J Applied Entomology 2010, 134, 385–394.
6. Ryan, E. and Reid, G.E. Chemical Derivatization and Ultrahigh Resolution and Accurate Mass Spectrometry Strategies for "Shotgun" Lipidome Analysis. *Acc Chem Res.* **2016,** *49,* 1596–1604.
7. Grant application resource: Thermo Scientific Orbitrap ID-X Tribrid mass spectrometer for metabolomics, lipidomics and structural elucidation, White paper #65364.
8. Welti, R. et al. Profiling membrane lipids in plant stress responses. Role of phospholipase D alpha in freezing-induced lipid changes in Arabidopsis. *J Biol Chem,* **2002,** *277*:31994–32002.
9. Brody, T. (1999) Nutritional Biochemistry (2nd ed.). Academic Press. p. 320. ISBN 0121348369.
10. Liebisch, G. et al. Shorthand notation for lipid structures derived from mass spectrometry, *J Lipid Res,* **2013,** *54,* 1523–1530.

Find out more at **www.thermofisher.com**

AN72942-EN 0719S

ThermoFisher SCIENTIFIC

thermoscientific



Compounding insights

Integrated, complete, toolset solves small-molecule analysis challenges

Thermo Scientific™ Orbitrap™ mass spectrometers produce information-rich data. To efficiently extract high-confidence understanding from this comprehensive data, usually requires an expert user and many unsupported third party software packages. Thermo Scientific™ Compound Discoverer™ software solves this challenge for small-molecule analysis. With a comprehensive, integrated set of libraries, databases, and statistical analysis tools linked in customizable workflows, the software streamlines unknown identification, determination of real differences between samples, and elucidation of biological pathways.

Find out more at **thermofisher.com/compounddiscoverer**

ThermoFisher SCIENTIFIC

thermo scientific

APPLICATION NOTE 65363

Novel structure-based profiling and annotation workflow—high-throughput analysis of flavonoids using the Thermo Scientific Orbitrap ID-X Tribrid MS

Authors

Reiko Kiyonami[1], Iwao Sakane[2], Seema Sharma[1], Graeme McAlister[1], Caroline Ding[1] and Andreas Huhmer[1]

[1]Thermo Fisher Scientific,
[2]ITO EN, LTD, Tokyo, Japan

Keywords

Flavonoid, flavonoid class, MS^n spectral tree, sugar neutral loss, Mass Frontier 8.0 software, sub-tree search, mzCloud spectra library, structure annotation, Compound Discoverer 3.0 software, FISh score

Introduction

Untargeted metabolomics aims to detect and compare as many metabolites as possible from a sample set. One bottleneck for the untargeted metabolomics approach is how to annotate the identification of unknown compounds from the sample set. Often there are classes of compounds which are difficult to identify and annotate due to the limited availability of authentic standards and structural diversity of these compounds, such as steroids, phospholipids, endocannabinoids and flavonoids. These compounds share basic core structure, but have various modifications in different structure positions of the basic structure, yielding a class of compounds with different molecular weights and diversified structures.

Flavonoids are a good example of this. Flavonoids are widely found in fruits, vegetables, grains, bark, roots, stems, flowers, tea, cocoa and wine, and are powerful antioxidants with anti-inflammatory and immune system benefits. Flavonoids have been widely studied for over a decade because of their diverse and important biological roles. One of the main approaches for flavonoid studies is untargeted flavonoid profiling because it provides insights into their biological functions and potential health benefits for humans. However, comprehensive identification of flavonoids remains challenging because of the structural diversity of this class. Chemically, flavonoids have the general structure of a 15-carbon skeleton, which consists of two phenyl

ThermoFisher SCIENTIFIC

rings (A and B) and heterocyclic ring (C). They can be subdivided into several classes based on the structural features of the C ring (Figure 1)[1].

Flavonoids are one class of secondary plant metabolites and are often modified in positions 3, 5, 6, 7, 8, 3', 4", and 5' of the basic flavonoid structure with hydroxylation, methylation, acylation, prenylation, and O- and C-glycosylation via secondary metabolic pathways. This diversity of modifications yields a wide range of isomeric and isobaric species (Figure 2). Currently, more than 10,000 flavonoids have been reported. However, the number of authentic flavonoid standards that can be used to generate reference spectra for unknown flavonoid identification is very limited. Authentic references for flavonoids with multiple glycoside modifications are especially hard to produce. As such, many unknown flavonoid compounds do not have an exact spectral library match. As a result, the majority of published flavonoid structure characterization studies have been carried out through manual assignment of fragment ions generated from MS^2 and higher order of MS^n data[2,3]. This type of manual analysis requires extensive knowledge of the flavonoid fragmentation rules and a tremendous amount of time to interpret the data. Plus, for most flavonoid glycoconjugates, MS^2 does not provide sufficient structurally relevant fragment ion information for characterizing aglycone structures[4].

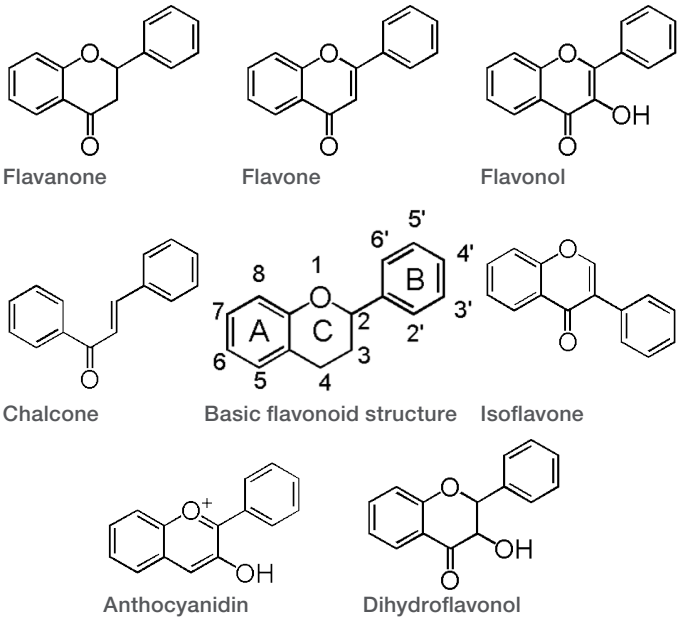


Figure 1. Basic structure of the flavonoid and major structure classes.

MS^n (multiple stage mass spectrometry) can provide more structurally relevant fragment ion information by the systematic breakdown of a compound and may be employed to generate a spectral tree to facilitate the unknown compound annotation. Here we present a new class or structure-based flavonoid profiling workflow analyzing fruit and vegetable juices that uses comprehensive fragment ion information from both HCD (higher-energy collisional dissociation) and CID (collisional induced dissociation) Fourier Transform (FT) MS^2, and higher order CID FT MS^n, for rapid flavonoid annotation on a Thermo Scientific™ Orbitrap ID-X™ Tribrid™ mass spectrometer. This workflow is demonstrated on glycoconjugates, but works on other transformation products of secondary metabolism.

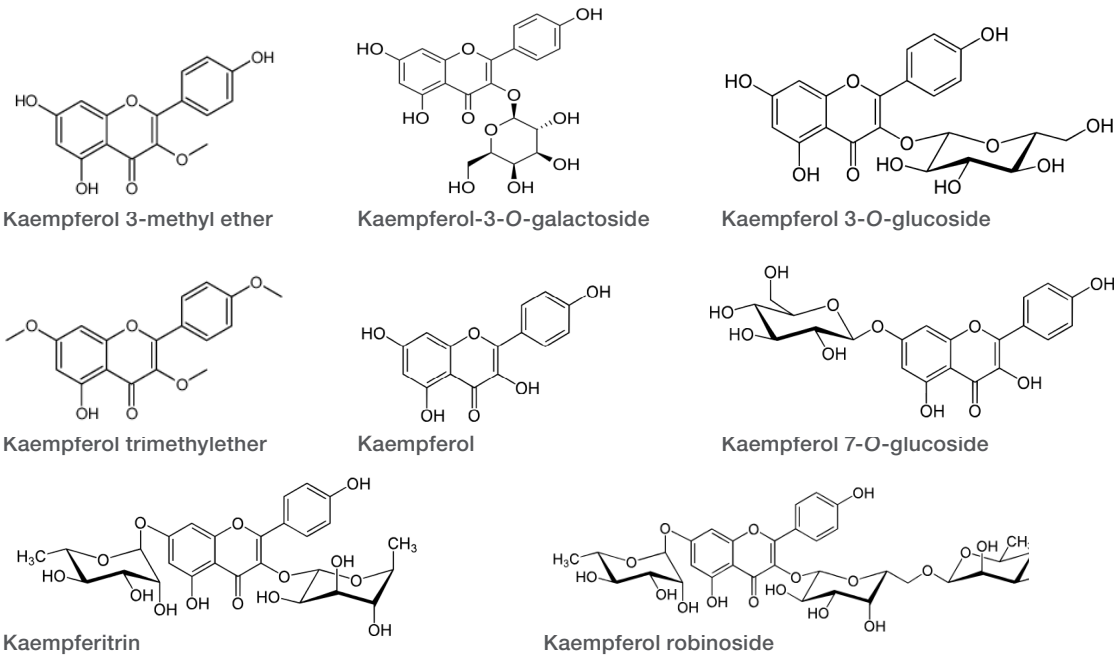


Figure 2. Various modifications on kaempferol aglycone generate complex secondary metabolites.

Materials And Methods

Sample preparation

Three commercially available fruit and vegetable juice samples (Naked® Kale Blazer, Odwalla® Berries Gomega®, and Odwalla® Red Rhapsody®) were analyzed in this study. Each juice sample was filtered and diluted two times with methanol.

HPLC conditions

A Thermo Scientific™ Vanquish™ UHPLC system performed separations using the gradient conditions shown in Table 1. Mobile phase A was water with 0.1% formic acid and mobile phase B was methanol with 0.1% formic acid. The column was a Thermo Scientific™ Hypersil Gold™ (2.1 × 150mm, 1.9µm), which we operated at 45 °C and a flow rate of 200 µL/min. The injection volume was 2 µL. Each sample was analyzed in triplicate.

Table 1. HPLC gradient.

Vanquish UHPLC		
Time (min)	**%A**	**%B**
0	99.5	0.5
1	90	10
10	70	30
18	50	50
22	1	99
25	1	99
25.1	99.5	0.5
30	99.5	0.5

MS conditions

All the data was collected on an Orbitrap ID-X Tribrid mass spectrometer. The mass spectrometer source conditions and method are shown in Table 2. The instrument template—structure specific MS^4 (monosaccharide loss) was used. This instrument template aims to collect an MS^n spectral tree (intuitively organized multi-stage tandem mass spectra) by systematically breaking down fragment ions from multiple stages in order to get structure relevant fragmentation pathways for unknown flavonoid structure annotation. For the purpose of establishing a fragmentation pathway, it is preferred to have fragment ions generating from a precursor ion remain stable without further fragmentation. One of the unique benefits that the Orbitrap ID-X Tribrid MS offers is multiple fragmentation techniques, including HCD providing relatively higher collision energy and CID providing relatively lower collision energy. In the case of flavonoid glycoconjugates, the softer CID is able to preserve the fragment ions with sugar neutral loss, providing a systematic fragmentation pathway to facilitate the structure annotation.

Table 2. Orbitrap ID-X Tribrid MS instrument set up.

ESI source	Orbitrap ID-X
Sheath gas 35	Pos ion (150-1200 amu)
Aux gas 5	MS: R=60K (FWHM at *m/z* 200)
Spray volt. 3.4 kV	MS^n: R=15K (FWHM at *m/z* 200)
RF-Lens 40	Cycle time: 1.2 second
Cap. temp. 300 °C	MS^2 Isolation width: 1.6 Da
Heater temp. 300 °C	MS^n Isolation width: 1.6 Da (MS^2) → 2.0 Da (MS^n)

As our goal is to carry out flavonoid identification using FT MS^n spectral tree data and simultaneous flavonoid quantitation using FT MS data, the instrument template is designed to collect the maximum amount of meaningful MS^n spectral tree data for the unknown flavonoid structure annotation within a short cycle time (1.2 seconds) to maintain enough MS scan points across a peak for precise quantitation. Because HCD MS^2 already provides sufficient fragment ions for structure annotation when the flavonoid compounds do not have glycol modifications, for the precursor ion mass range between 150–420 *m/z*, only HCD MS^2 data are collected. For the precursor ion mass range between 420–1200 *m/z*, glycol modifications are expected and an intelligent product ion-dependent MS^n approach was used, in which a high-resolution accurate mass (HRAM) full MS scan was followed by CID MS^2 scans. The product ions generated from each MS^2 scan were monitored by the mass spectrometer and an MS^3 scan was triggered if one or multiple pre-defined neutral sugar molecules were detected. An additional MS^4 scan was triggered if pre-defined neutral sugar molecules were detected from the MS^3 scan. Figure 3 shows the flowchart of the developed product ion-dependent MS^n data acquisition instrument method.

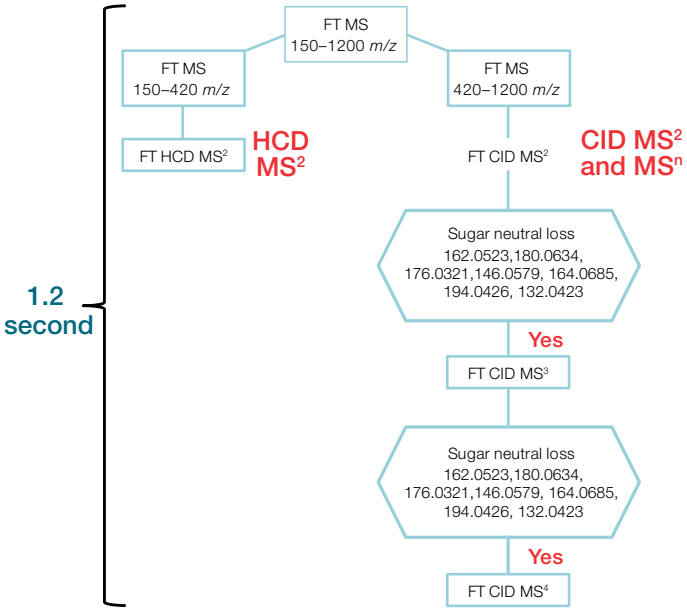


Figure 4 shows an example MS^3 spectral tree generated from the reference standard for the flavonoid rutin. Two major fragment ions in the CID MS^2 spectrum were detected. The *m/z* 465.1022 fragment ion represents the loss of one rhamnose sugar moiety from the intact structure while the *m/z* 303.0495 fragment ion represents the loss of two sugar moieties, rhamnose and hexose. At the MS^3 stage, the major fragment ion from the *m/z* 465.1022 precursor ion was the *m/z* 303.0495, confirming that the 303.0495 did come from the loss of two sugar moieties (rhamnose and hexose) from the starting structure, while the fragment ions of the precursor *m/z* 303.0495 provided sufficient structure relevant fragment ions which can be used for the aglycon structure annotation.

Saccharide	Neutral Loss	Composition
Pentose (xylose, arabinose)	132.04226	$C_5H_8O_4$
Deoxyhexose (rhamnose)	146.05791	$C_6H_{10}O_4$
Hexose (glucose, galactose)	162.05282	$C_6H_{10}O_5$
Glucuronide	176.03209	$C_6H_8O_6$
Glucuronic acid	194.04265	$C_6H_{10}O_7$

Figure 3. Flowchart of sugar neutral loss triggered high order MS^n data acquisition on the Orbitrap ID-X Tribrid MS instrument.
The associated name and composition for each targeted sugar neutral loss are described in the table.

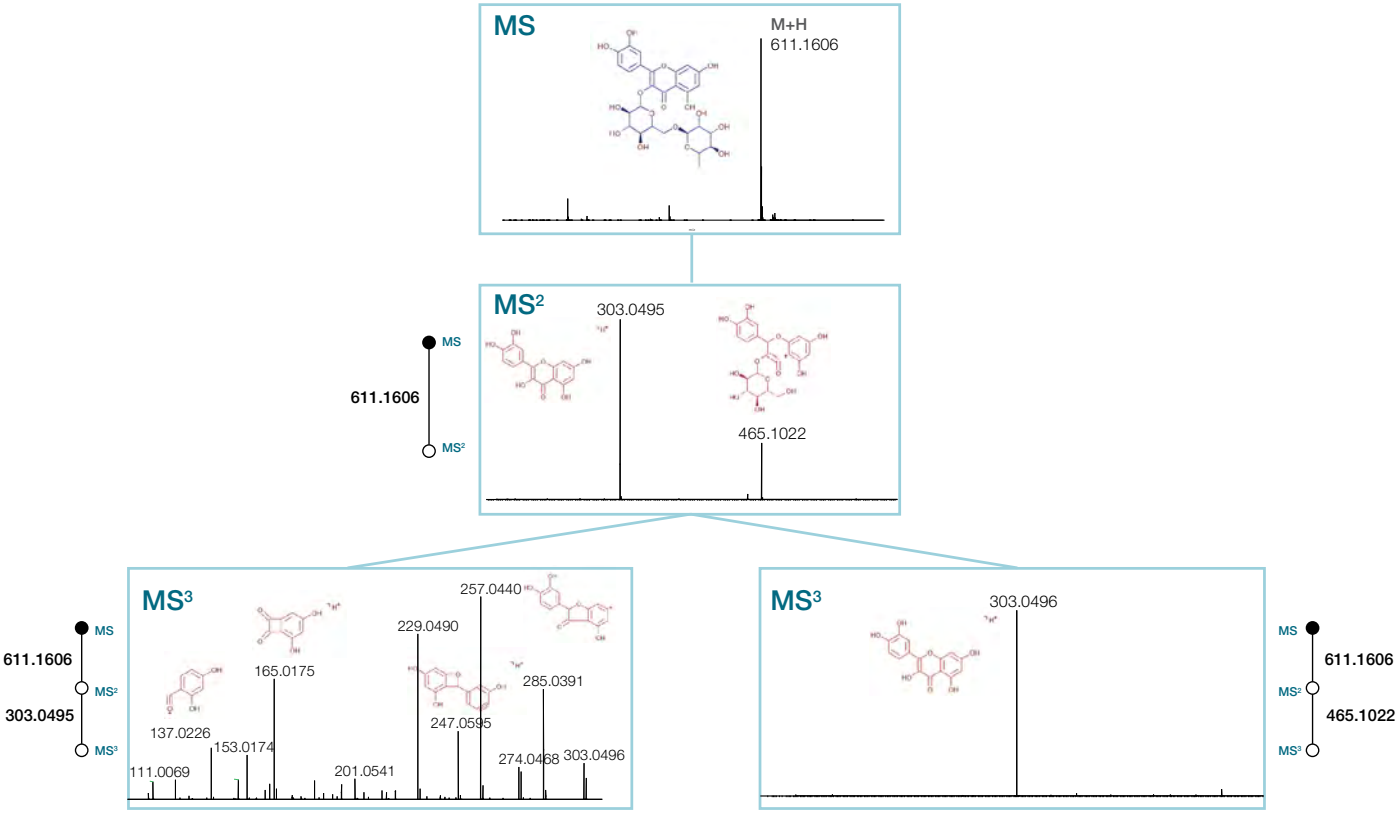


Figure 4. CID MS^3 spectral tree generated from the rutin reference standard.

Results and discussion

Addressing the flavonoid annotation challenges using a structure-based MS^n approach

The MS^n fragmentation approach provides a systematic breakdown of a compound and generates a spectral tree, yielding more structurally relevant fragment ion information compared to MS/MS fragmentation (Figure 3). Figure 5 shows an MS^3 spectral tree collected from an unknown compound detected from the Kale Blazer juice sample. The MS^2 spectrum alone for the *m/z* 641.1720 did not provide a library match to the mzCloud spectral library. More structurally relevant fragment ions were detected by further fragmenting the MS^2 product ion at *m/z* 317.0657, which matched the reference compound Isohamnetin in the mzCloud spectral library, providing confident sub-structure annotation (aglycon) of this unknown compound. As a result, this unknown compound can be annotated as a flavonoid class-related compound whereby the aglycon sub-structure was detected by using MS^3 spectral data.

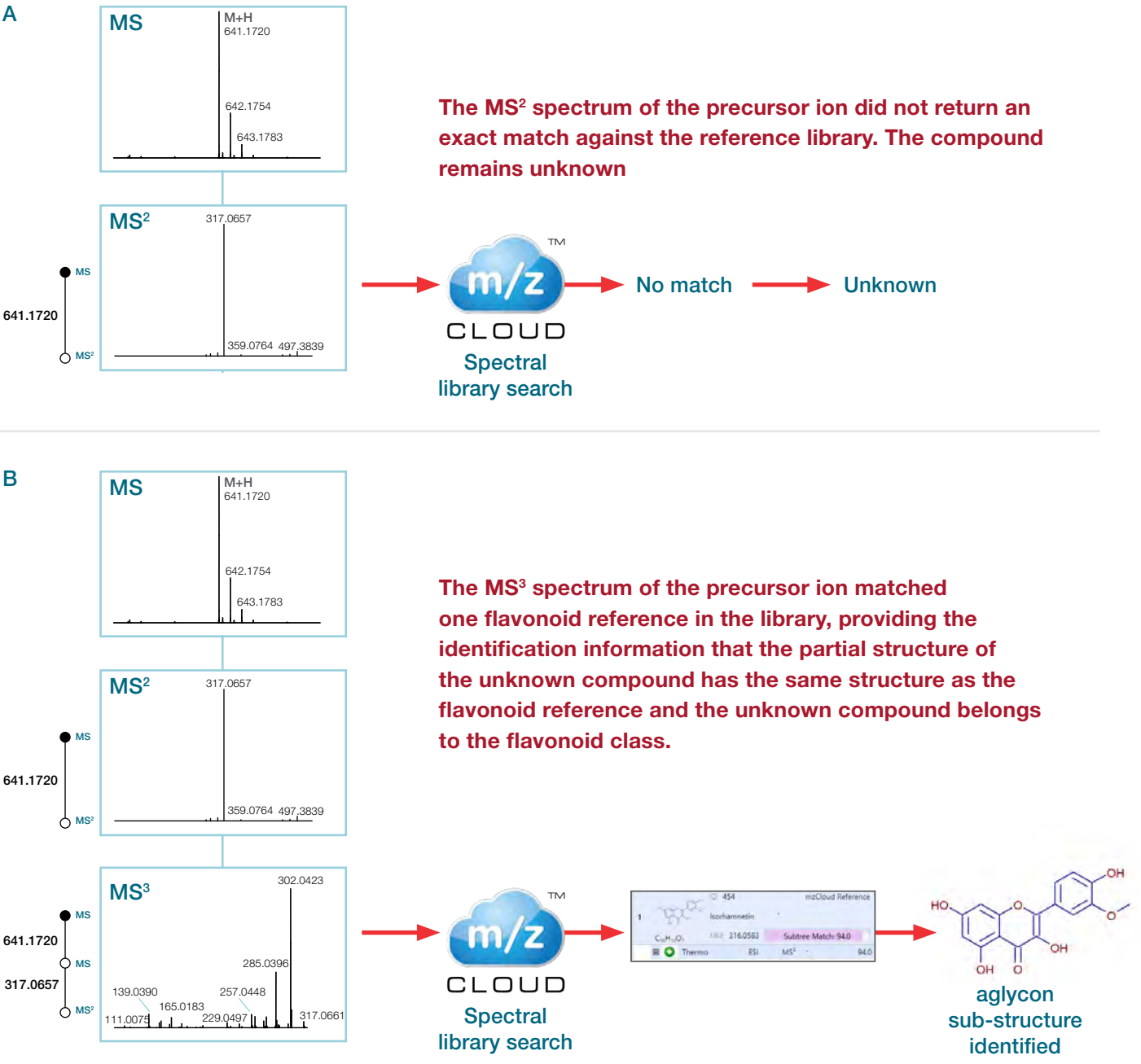


Figure 5. MS^2 (A) and MS^3 (B) spectral trees generated on an unknown compound detected (M+H: 641.1720) from kale blazer juice sample.

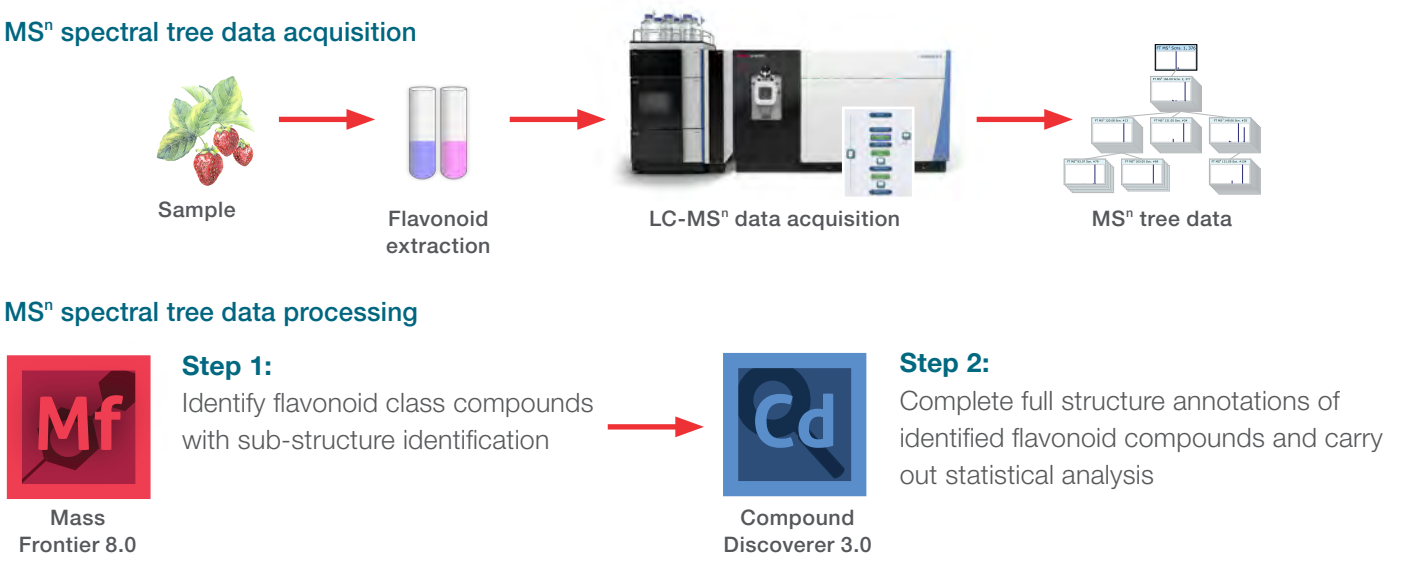


Figure 6. Workflow of structure-based MSⁿ approach to address the flavonoid annotation challenges.

Taking advantage of the fact that MSⁿ fragmentation enables sub-structure identification, as shown in above example, we developed a structure-based MS workflow to address the flavonoid annotation challenges. Figure 6 shows the workflow of this approach, which acquires the MSⁿ spectral tree data using the instrument template discussed in the MS conditions section. The collected MSⁿ spectral tree data are first processed by Thermo Scientific™ Mass Frontier™ 8.0 software to determine which compounds include the basic flavonoid structure thus belonging to the flavonoid class. These detected flavonoid-related compounds are then further annotated using a flavonoid structure database and structure ranking tools within Thermo Scientific™ Compound Discoverer™ 3.0 software.

Flavonoid class compound identification using Mass Frontier 8.0 spectral interpretation software

Mass Frontier 8.0 software was used to process the MSⁿ tree data collected from juice samples. The software detects unknown compounds from each juice raw file using the Joint Components Detection (JCD) algorithm. All detected compounds and associated spectral trees were then queried against the mzCloud MSⁿ spectral library containing mass spectra generated from authentic reference material (Box 1) using the "Subtree Search" approach. A Subtree search compares the experimental MSⁿ tree against MSⁿ trees in the mzCloud library (Figure 7). By selecting "Search All", all detected compounds are automatically searched against the references in the mzCloud library in batch mode (Figure 8).

Box 1.

mzCloud MSⁿ spectral library

The mzCloud library is an extensively curated database of high-resolution tandem mass spectra that are arranged into spectral trees. MS^2 and multi-stage MSⁿ spectra were acquired at various collision energies, precursor *m/z*, and isolation widths using CID and HCD. Each raw mass spectrum was filtered and recalibrated. It is a fully searchable library. mzCloud's website https://www.mzcloud.org/ is freely accessible to query MSⁿ spectra data and retrieve the library reference match manually. The users of Compound Discover (2.1, 3.0) software and Mass Frontier 8.0 software can search all MSⁿ spectra data automatically to get library reference match in a batch mode.

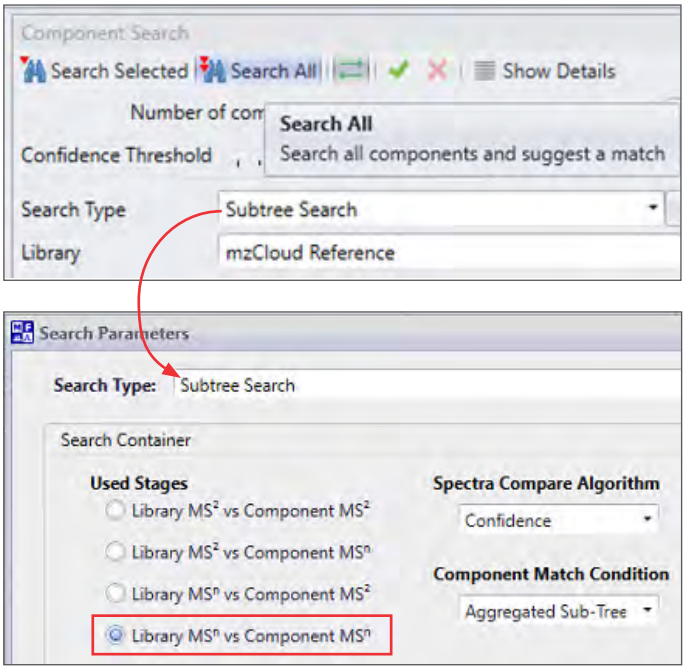
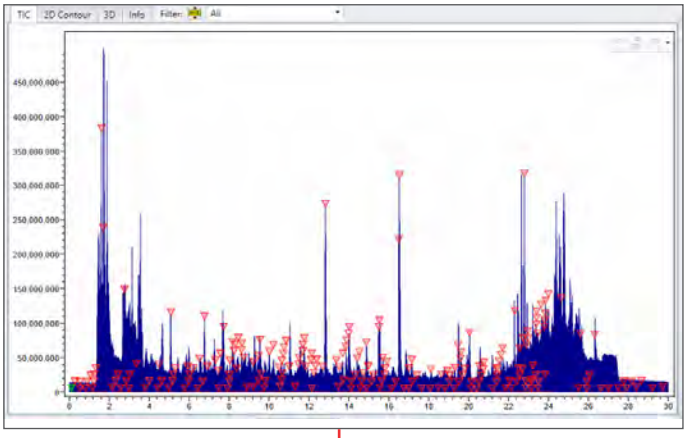


Figure 7. Partial MSⁿ spectral tree search results from the Kale juice sample.

Unknown compound detection



Match result

Name	Scan No.	Precursor m/z	Match	MSⁿ	t_R (min)	Abundance
Components						
Component 2766	11336	595.2028	94 Poncirin	4	20.040	30,625,312
Component 2423	9553	641.1722	94 Isorhamnetin	4	17.005	4,016,400
Component 2358	9287	611.1976	93 Neohesperidin	4	16.516	136,944,496
Component 2268	8931	611.1980	73 Neohesperidin	4	15.897	375,730
Component 2213	8718	581.1872	87 Naringin	4	15.520	41,592,528
Component 1985	7884	611.1611	98 Robinin	4	14.033	16,899,672
Component 1934	7698	597.1822	79 Hesperidin	4	13.695	14,091,042
Component 1913	7603	611.1965	86 Hesperidin	4	13.522	5,888,135
Component 1740	6954	627.1563	98 Rutin	4	12.389	2,455,858
Component 1705	6785	741.2244	91 Rhoifolin	4	12.114	707,318
Component 1660	6584	581.1873	84 Naringin	4	11.763	14,305,817
Component 1585	6325	787.2299	94 Isorhamnetin	4	11.304	3,313,007
Component 1548	6164	803.2244	94 Isorhamnetin	4	11.010	1,965,657
Component 1523	6085	757.2191	96 Kaempferitrin	4	10.856	814,939
Component 1505	6004	935.2670	76 Kaempferitrin	4	10.709	1,112,348
Component 1496	5971	773.2142	92 Robinin	4	10.644	2,863,759
Component 1285	5148	773.2141	93 Rutin	4	9.217	2,942,259
Component 1281	5147	188.07077	97 Rutin	4	9.215	2,378,850
Component 1165	4721	935.2456	84 Kynurenic acid	4	8.461	11,755,922

Figure 8. Partial MSⁿ spectral tree search results from the Kale juice sample.

Breakdown of MSⁿ subtree search results:

Subtree search calculates the largest overlap between the potentially large component spectral tree against the library. It not only provides exact compound matches based on an MSⁿ tree match, but also provides substructure/subtree matches when the unknown compound does not exist in the reference library. Subtree search results can include the following two types:

Match result type 1 (exact MSⁿ tree match): The MS² precursors of the unknown compound and library reference matches and the spectral tree match between the unknown compound (MS², MS³ and MS⁴) and reference (confidence score >60). The unknown compound is confidently identified with full annotation of tree match (Figure 9).

Match result type 2 (partial MSⁿ tree match): In most cases, the MS² precursor and MS² spectra of the unknown compound do not match any library references because of limited reference flavonoid standards. The Subtree search is not limited by this factor. Partial MSⁿ spectral tree match provides valuable substructure information for true unknown compounds. When subtree matches between unknown and reference, the sub-structure of the unknown compound is identified to match the reference structure or its substructure, in this case, the flavonoid reference structure (Figure 10). With subtree search, Mass Frontier 8.0 software was able to detect true unknown compounds that belong to the flavonoid compound class with molecular weight, retention time and sub-structure information.

All detected flavonoid class compounds can be exported to a cvs file using the “Export to Grid” tool in the Mass Frontier 8.0.

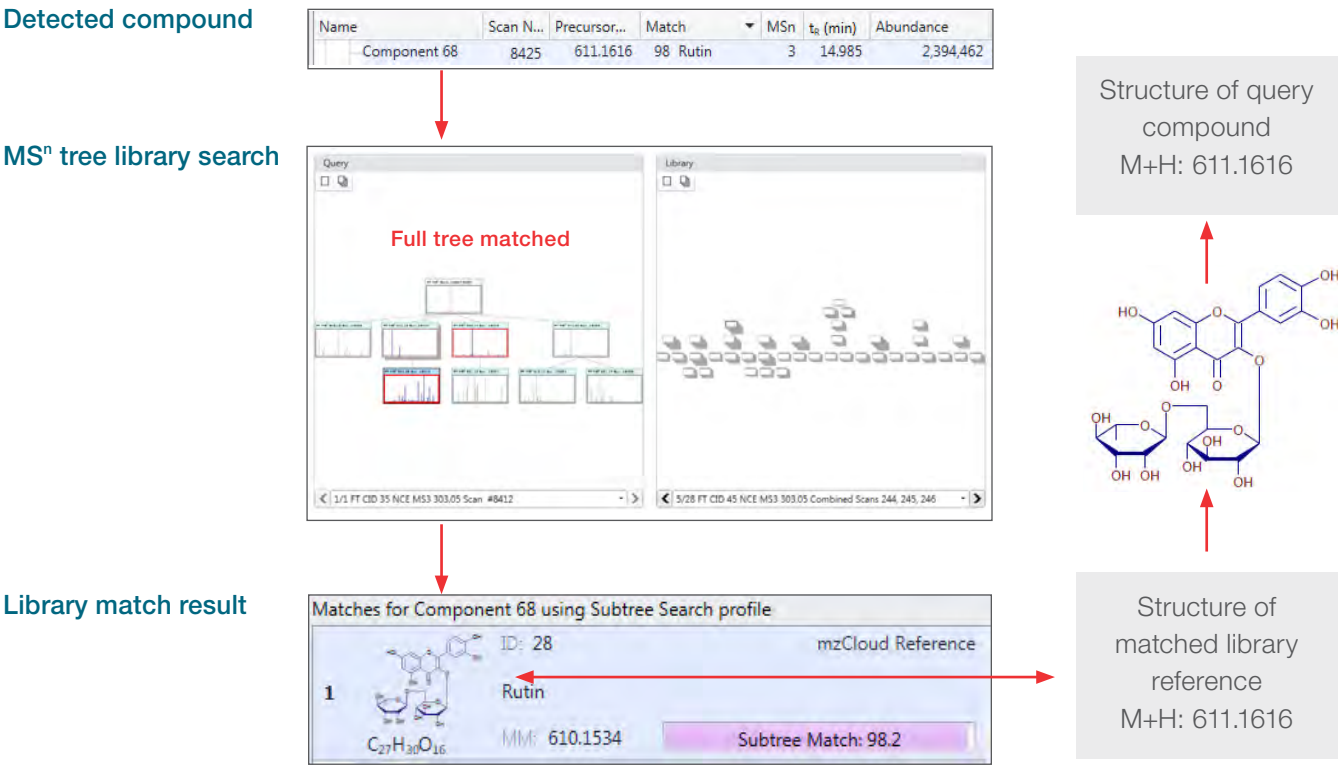


Figure 9. Full tree match using the Mass Frontier software. The unknown compound ([M+H]$^+$: *m/z* 611.1616) from the juice sample was identified as rutin, with complete structure annotation by an exact MS^n spectral tree match with the flavonoid rutin reference in the library.

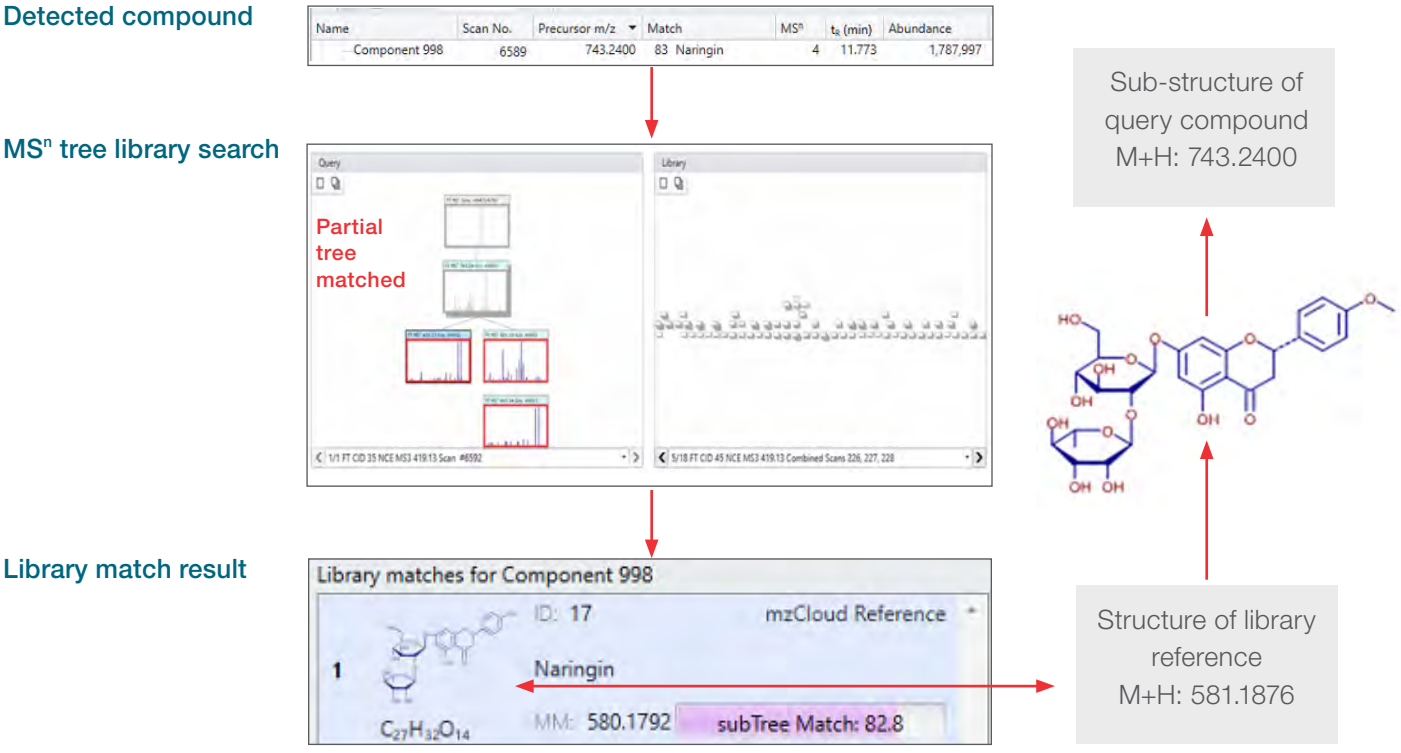


Figure 10. Partial tree match using Mass Frontier software. The unknown compound ([M+H]$^+$: *m/z* 743.2394) from the juice sample was identified as a member of the flavonoid class by sub-structure annotation from a partial MS^n spectral tree match with the flavonoid naringin reference in the library.

Box 2.

Arita Lab 6549 flavonoid structure database

A dedicated flavonoid structure database built from 6549 Mol files contributed by Professor Masanori Arita from National Institute of Genetics, Japan.

Professor Arita's laboratory has worked extensively to build a web-searchable, flavonoid database. This database collects original references that report the identification of flavonoid in various plant species. The database consists of three major resources: (flavonoid) compounds, plant species, and references. Currently, 6961 flavonoid structures, 3961 plant species, and 5215 references describing a total of 19,861 metabolite-species relationships are registered. More details can be found on the website http://metabolomics.jp/wiki/Main_Page.

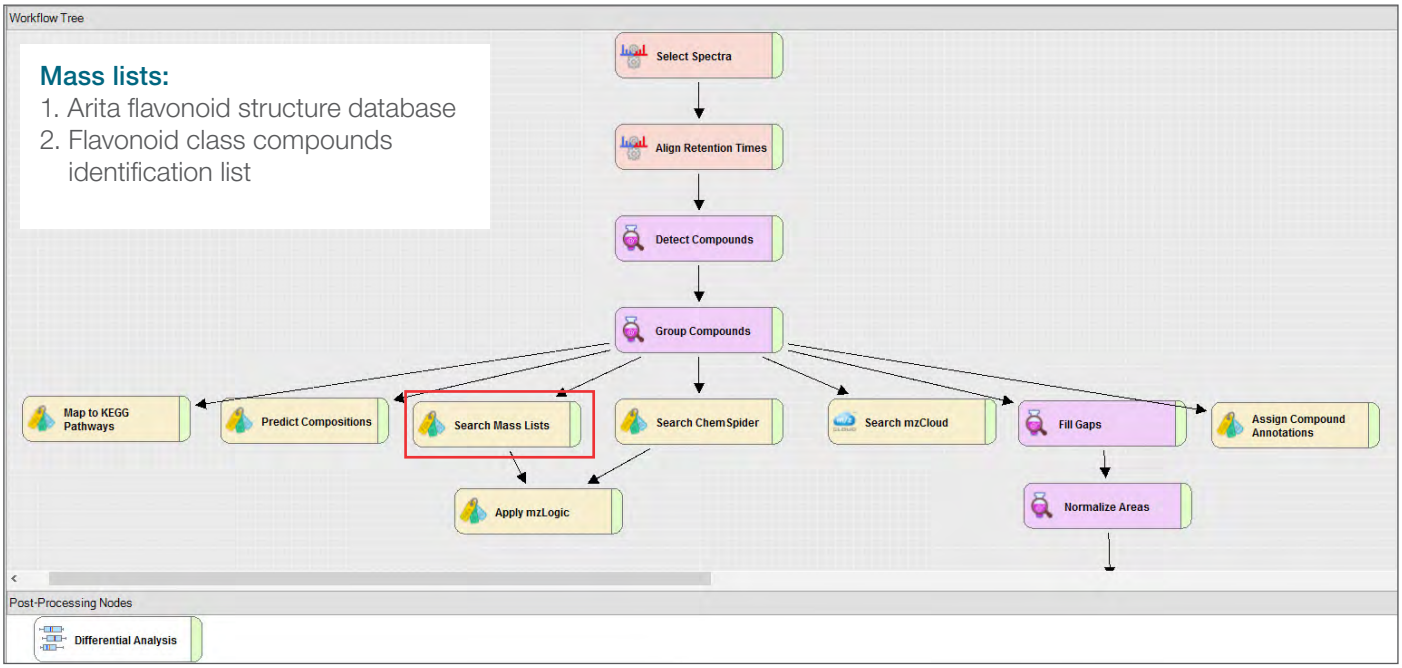


Figure 11. The Compound Discoverer 3.0 software processing workflow used for flavonoid annotation and statistical analysis.

The detected compounds that matched both mass lists (molecular weight from both lists and retention time from juice flavonoid class compound identification result list) are selected for further flavonoid structure annotation. Figure 12 shows that a detected compound with the molecular weight of 742.2320 matched both mass lists. The two isomeric flavonoid structures from the Arita lab 6549 Flavonoid structure database were selected as structure candidates of the compound. In addition, three isomeric flavonoid structures from ChemSpider database were also selected as structure candidates (Figure 13). The ranking of total five structure candidates was carried out using the FISh (Fragment Ion Search) scoring algorithm (Box 3). To rank the isomeric structures, the software predicts fragmentation of five structure candidates based on known fragmentation rules first, then calculates the FISh scores by matching the predicted fragment ions with the observed fragment ions from MS^n data. The proposed structure with the highest FISh score represents the best match with the observed fragment ions from the MS^n data and is the best structure candidate for the unknown flavonoid class compound. Figure 14 shows that this flavonoid is annotated as Narirutin 4'-glucoside based on the FISh score ranking.

Box 3.

FISh scoring algorithm included in Compound Discoverer 3.0 software

The FISh scoring algorithm attempts to match the fragment structures in a list of predicted fragments to the centroids in the fragmentation scans of the precursor ions.

When a precursor ion scan is followed by only one fragmentation scan, it calculates the FISh coverage score as follows:

$$\text{FISh coverage score} = \frac{\text{\# matched centroids}}{\text{\# used centroids}} \times 100$$

where:

used (matched + unmatched) centroids represents the number of centroids in the fragmentation scan that are above the user-specified signal-to-noise threshold. The algorithm skips centroids below the user-specified signal to noise threshold.

When a precursor scan is followed by more than one fragmentation scan, it calculates a composite score as follows:

$$\text{FISh coverage score} = \frac{(\Sigma_{\text{per all scans}} \text{\# matched centroids})}{(\Sigma_{\text{per all scans}} \text{\# used centroids})} \times 100$$

The FISh scoring algorithm annotates the centroids in the fragmentation scans with the matching fragment structures. It also provides a FISh Coverage score for data-dependent scans in the Mass Spectrum view legend and a FISh Coverage score in the results table.

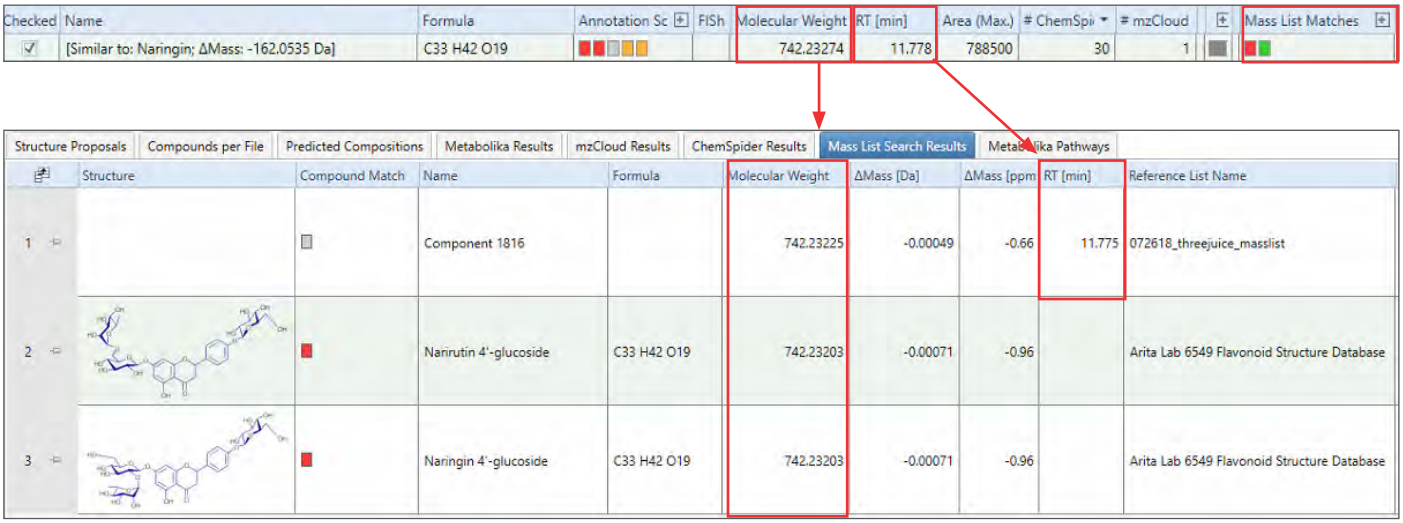


Figure 12. An example showing unknown flavonoid compound which matched both mass lists.

Information from MS^n tree data search results

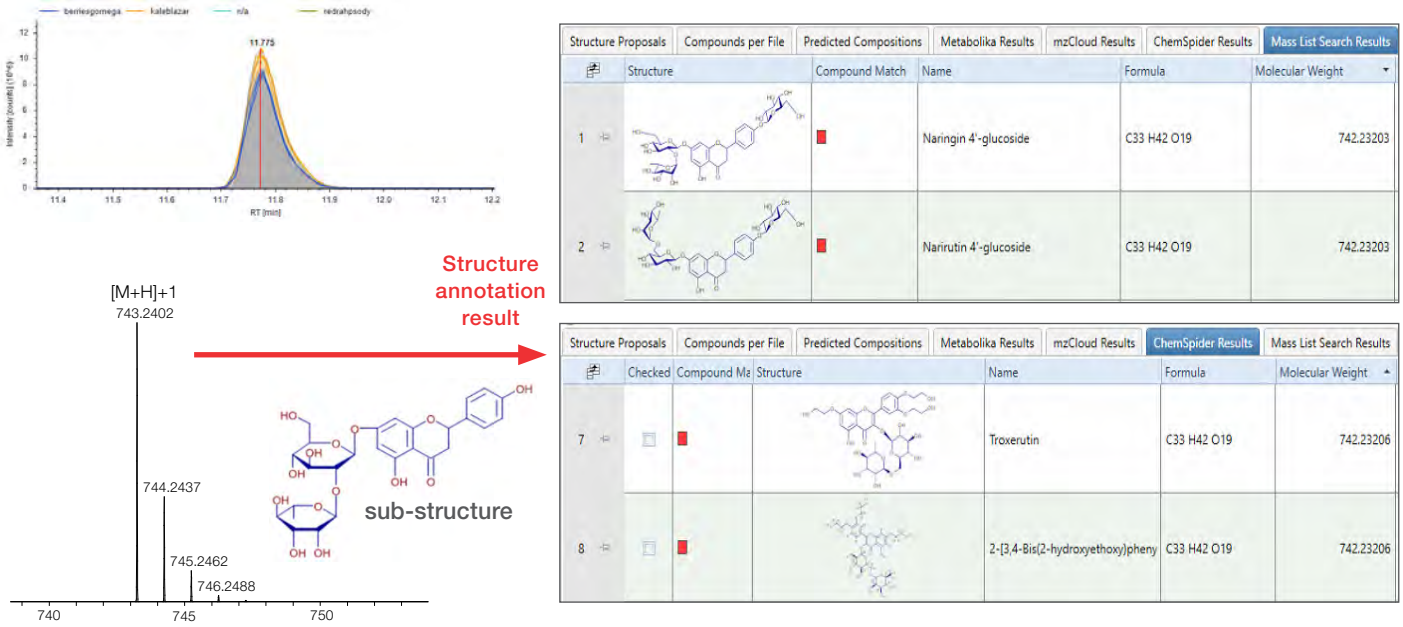


Figure 13. Structure candidates proposed using the Arita flavonoid structure database and the ChemSpider database for the identified flavonoid class compound (MW: 742.2320).

Information from MS^n tree data search results

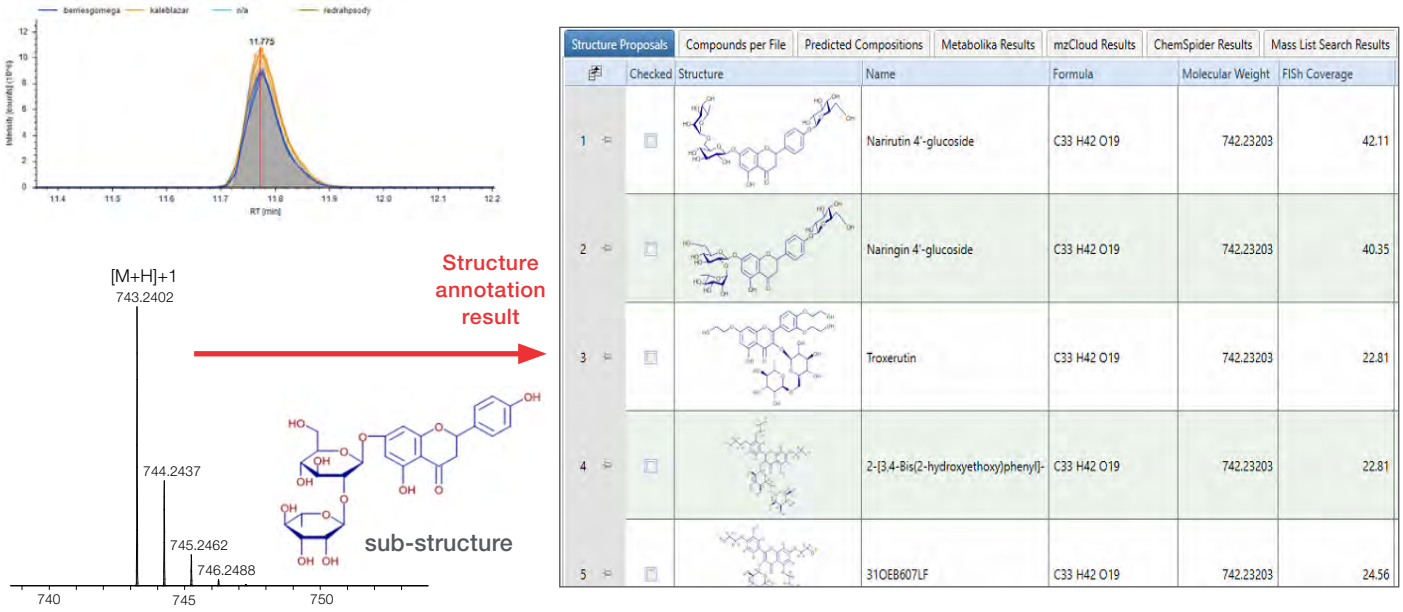


Figure 14. The detected flavonoid class compound (MW: 742.2320) is annotated as Narirutin 4'-glucoside based on the FISh score ranking.

Discussion

The new structure-based MS^n approach enables increased flavonoid identification coverage

Although it has been known that MS^n spectral tree data provide deeper and more detailed fragmentation pathways, thus enabling more structural information for flavonoid annotation[4], the use of an MS^n workflow for flavonoid annotation was historically hindered by two factors: (i) the MS^n instrument method setup was difficult for non-mass-spectrometer-expert users, and (ii) the MS^n spectral tree data processing was a bottleneck because it required manual fragment ion assignments that were based upon extensive expert knowledge about flavonoid chemical structure and fragmentation rules. Our newly developed structure-based MS^n approach addresses both of these challenges and allows unknown flavonoid annotation with increased throughput and coverage.

Firstly, by offering a ready-to-use structure-specific MS^n instrument template, everyone can easily acquire high-quality of MS^n data using this template.

Secondly, by offering new software tools including the sub-tree search process in Mass Frontier 8.0 software, new Mass List format allowing structure attachment and FISh score calculated using an MS^n spectra, the rich structure relevant fragment ion information from MS^n spectra tree can be processed in an automatic fashion without need to know any specific fragmentation rules.

Combining the new instrument template and software tools, the new approach takes full advantage of deeper and more structurally relevant fragment ion information offered by MS^n, enabling more flavonoid compounds to be annotated compared to an MS^2 only approach. Partial MS^n spectral tree match results provided valuable sub-structure information for true unknown compounds. With subtree search, Mass Frontier 8.0 software identified true unknown compounds that belong to the flavonoid compound class which do not have exact references in the mzCloud library. Table 3 shows that six unknown flavonoids which are rutin and its secondary metabolites with different modifications were identified from the juice samples using MS^n tree data, while MS^2 data only identified rutin. Table 4 shows that five unknown flavonoids which are isorhamnetin and its secondary metabolites with different modifications were identified using MS^n tree data while MS^2 data only identified isorhamnetin and one of its secondary metabolites. Figure 15 shows a comparison of the number of annotated flavonoids from the three fruit and vegetable juice samples using MS^2 only data and MS^n spectral tree data with 2D column and Venn diagram formats. The MS^2 only approach annotated sixty- two flavonoids, while the MS^n approach annotated the same sixty-two flavonoids and plus sixty-seven more flavonoids. As a result, the MS^n approach showed a two-fold increase in annotations to the MS^2 only approach.

Table 3. Identified rutin and its secondary metabolites using MS^2 vs MS^n.

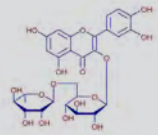
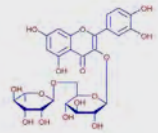
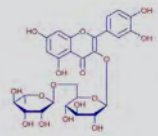
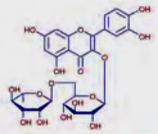
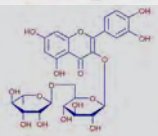
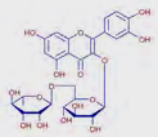
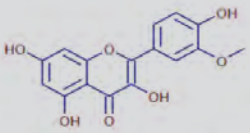
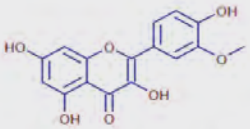
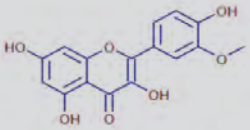
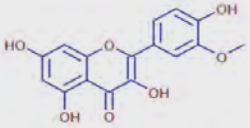
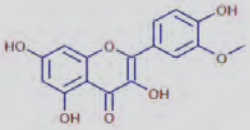
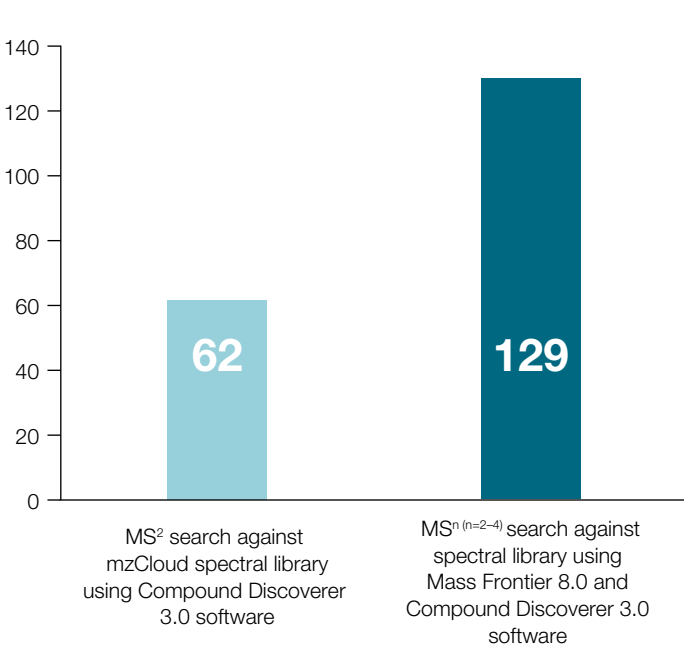
Molecular Weight	ID Structure/Substructure in MF 8.0	Identified with MS^2 in CD 3.0	Identified with MS^n and FiSh score in CD 3.0
610.1539	C27H30O16 MW: 610.1534 	●	●
626.1490	C27H30O16 MW: 610.1534 	✕	●
756.2120	C27H30O16 MW: 610.1534 	✕	●
772.2071	C27H30O16 MW: 610.1534 	✕	●
788.2023	C27H30O16 MW: 610.1534 	✕	●
950.2328	C27H30O16 MW: 610.1534 	✕	●

Table 4. Identified isorhamnetin and its second metabolites using MS^2 vs MS^n.

Molecular Weight	ID Structure/Substructure in MF 8.0	Identified with MS^2 in CD 3.0	Identified with MS^n and FISh score in CD 3.0
316.0590	C16H12O7 MW: 316.0583 	●	●
478.1122	C16H12O7 MW: 316.0583 	✕	●
624.1698	C16H12O7 MW: 316.0583 	●	●
640.1652	C16H12O7 MW: 316.0583 	✕	●
786.2226	C16H12O7 MW: 316.0583 	✕	●



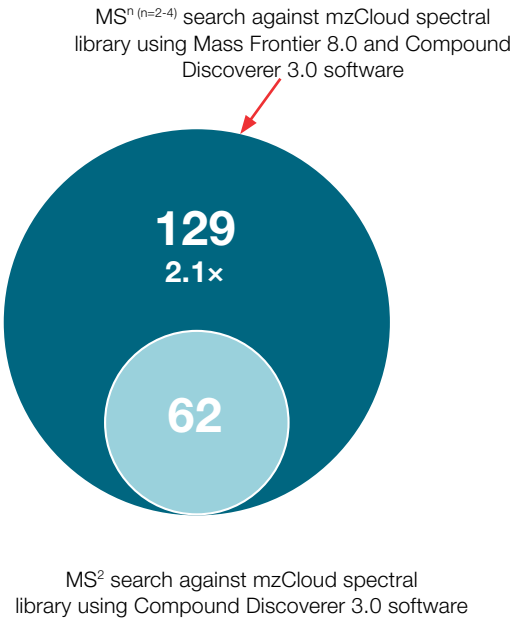


Figure 15. Comparison of the number of detected flavonoids from the juice samples with full structure annotation obtained using MS^2 only and MS^n tree spectral data.

Advantage to enable simultaneous quantitation and statistical analysis

Our newly developed structure-based MS^n approach also enables simultaneous quantitation of identified flavonoid compounds and statistical analysis. We designed the instrument template using a short cycle time (1.2 second) in order to get enough scan points across the chromatographic peak for precise quantitation while collecting MS^n spectral tree data in the same LC-MS run. Higher annotation coverage of flavonoid compounds using the new developed structure-based MS^n approach enabled more data points for precise statistical analysis. As shown from the heat map (Hierarchical Cluster Analysis) of detected flavonoids (Figure 16), more high abundance flavonoids were detected from Kale Blazer and Berries Gomega juice samples.

On another hand, most flavonoids detected from the Red Rhapsody juice sample were less abundant. Figure 17 depicts an extracted ion chromatogram for the quantitation of detected neoeriocitrin. Figure 18 shows that the three juice samples are clearly differentiated by principal component analysis.

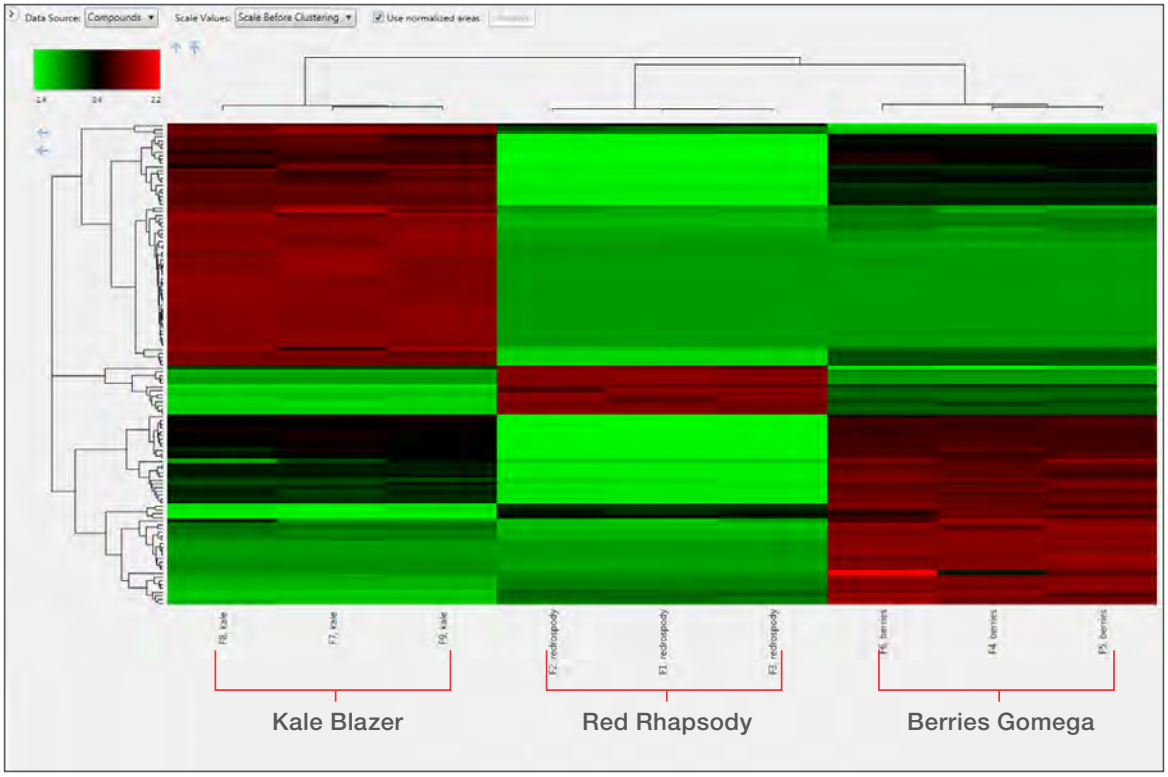
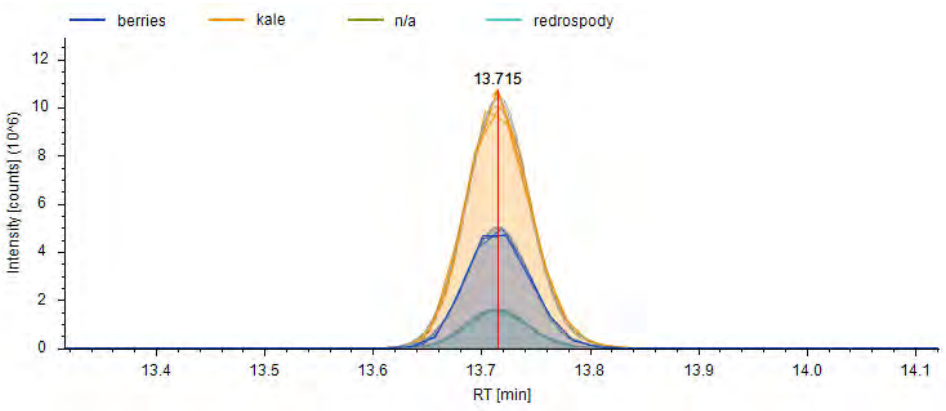


Figure 16. Hierarchical cluster analysis.

Extracted ion chromatogram



Trend Chart

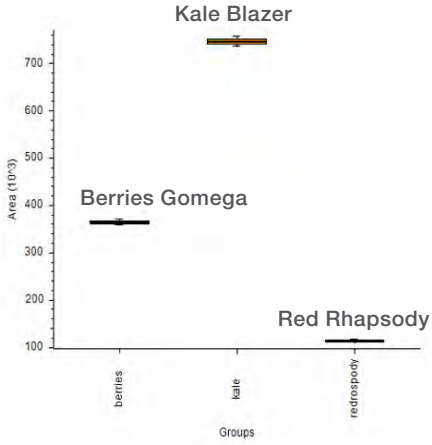


Figure 17. Extracted ion chromatograms and trend chart of neoeriocitrin.

thermo scientific

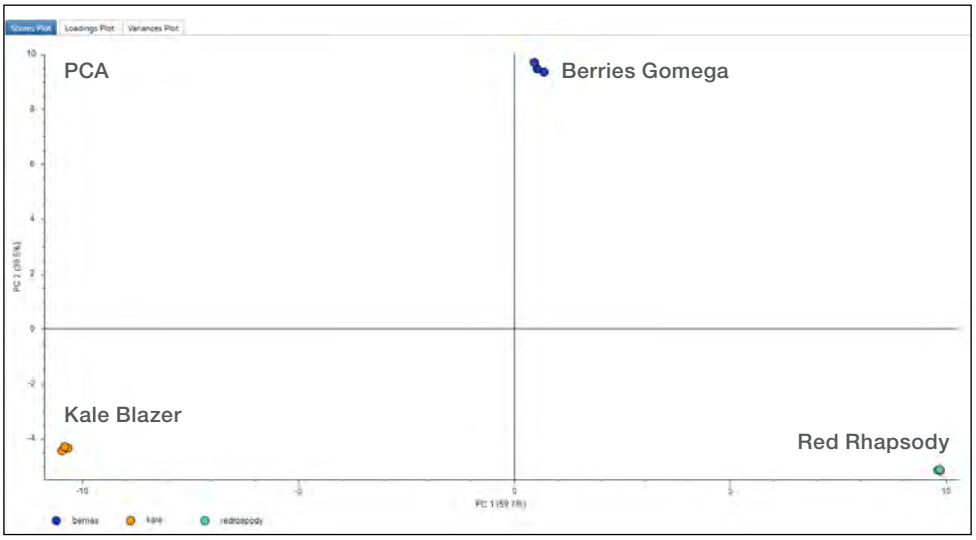


Figure 18. PCA of flavonoid compounds identified from the three juice samples.

Conclusions

We have developed a new structure-based MS^n approach to address the challenges of unknown flavonoid annotation. This new approach enables rapid flavonoid class compound detection and further structure annotation without the need to have expertise in flavonoid chemical structure and fragmentation rules. Two times more flavonoids were annotated from the three fruits and vegetable juice samples using this structure-based MS^n approach compared to MS^2 only approach. This new structure-based MS^n approach also allowed quantitation of annotated flavonoids simultaneously in a single LC-MS run. In addition to the class of flavonoid compounds, the concept of this structure-based MS^n approach can be applied to other classes of compounds, such as steroids and endocannabinoids.

References

1. http://metabolomics.jp/wiki/Category:FL

2. P. Kachlicki, A. Piasecka, M. Stobiecki, L. Marczak. Structural Characterization of Flavonoid Glycoconjugates and Their Derivatives with Mass Spectrometric Techniques, *Molecules* **2016**, 21, 1494.

3. D. Tsimogiannis, M. Samiotaki, G. Panayotou, V. Oreopoulou. Characterization of Flavonoid Subgroups and Hydroxy Substitution by HPLC-MS/MS, *Molecules* **2007**, *12*, 593–606.

4. J.J.J. van der Hooft, J. Vervoort, R. J. Bino, J. Beekwilder, R.C.H. de Vos. Polyphenol Identification Based on Systematic and Robust High-Resolution Accurate Mass Spectrometry Fragmentation, *Anal. Chem.* **2011**, *83*, 409–416.

5. M. Arita, K. Suwa. Search extension transforms Wiki into a relational system: A case for flavonoid metabolite database. *BioData Mining* **2008 1**:7.

thermoscientific

APPLICATION NOTE 65722

A multiresidue method for pesticide profiling using an Orbitrap Tribrid mass spectrometer

Authors: Ed George, Seema Sharma, Scott Peterman, and Richard Fussell

Thermo Fisher Scientific, San Jose, CA

Keywords: Orbitrap ID-X mass spectrometer, Tribrid, pesticide residues screening, non-targeted, high-resolution accurate-mass, Vanquish UHPLC, identification points, part per billion (ppb), Compound Discoverer, TraceFinder, quantitation of pesticide residue, intelligent MS, data dependent acquisition, dynamic exclusion, AcquireX

Goal

Develop a method based on full scan acquisition that enables the detection and quantitation of multiple targeted pesticide residues and simultaneous post-acquisition screening analysis of unknown/unexpected pesticide residues. The resulting method must be able to acquire compound-specific precursor and product ions with sufficient sensitivity, mass accuracy, and resolution to selectively extract ion intensity attributed to compounds of interest. Meeting these requirements will enable accurate quantitation and help to maintain a high degree of reproducibility and robustness for the duration of the study. In addition, the overall LC-MS^n workflow must be easy to use, applicable to most sample types and matrices, and extremely efficient.

Introduction

Pesticides are routinely applied to crops for preventing, destroying, or controlling pest activity. In order to protect consumers and ensure they are not being exposed to pesticide levels harmful to their health, pesticides are regulated in many food sources and several countries have established maximum residue levels (MRLs) or tolerances. Given the large number of pesticides used and the globalization of the food supply, multiresidue methods offer a great advantage, allowing analysis of hundreds of pesticides in a single experiment.

A maximum residue level (MRL) is the highest level of a pesticide residue that is legally tolerated in or on food or feed when pesticides are applied correctly in accordance with Good Agricultural Practice. When a pesticide has been registered on a particular crop, the MRL for the pesticide in/on the crop is usually set at a valued determined from “supervised field residue trials”. However, if a pesticide has not been approved for use on a crop, the MRL can be set at the limit of detection (LOD). In the EU, the default LOD MRL is 0.01 mg/kg.

ThermoFisher SCIENTIFIC

Liquid chromatography coupled to mass spectrometry (LC-MS) is the preferred method for performing multiresidue analysis of LC-amenable pesticides.[1-4] Traditionally, panels of SRM transitions associated for each pesticide are monitored using triple quadrupole mass spectrometers for rapid and accurate quantitation of targeted pesticides.[2-5] Each method contains a specific set of precursor/product ion mass pairs per pesticide for data acquisition and uses known product ion ratios, internal standards, and/or known retention times for identification and quantitation. While these methods meet the required analytical performance metrics, expanding target panels with additional SRM transitions must be individually evaluated to ensure target selectivity in new matrices. Also, targeted methods cannot detect or identify the presence of unknown/unexpected pesticides, their metabolites, or degradation products.

Recently, high-resolution accurate-mass (HRAM) MS data was used to demonstrate effectiveness for non-targeted pesticide detection and identification in the presence of complex matrices.[1,4,6-8] Instrument methods using full scan HRAM MS acquisition facilitated post-acquisition extracted ion chromatographic (XICs) analysis to detect pesticides with high mass accuracy (±5 ppm). Furthermore, automated tandem mass spectral acquisition routines complemented HRAM MS data by acquiring full scan HRAM product ion spectra, resulting in confident pesticide identification based on measuring at least one compound-specific product ion.[4] (See the EU SANTE Guidelines 12682/2019 or the FDA Guidelines describing the acceptance criteria for residue detection and identification.) The benefit of acquiring full scan HRAM MS^n data enabled retrospective data mining for not only known target compounds, but also suspected compounds.

With the recent introduction of the Thermo Scientific™ Orbitrap ID-X™ Tribrid™ mass spectrometer, a new paradigm of multiresidue characterization and screening has become available.[7] The Orbitrap ID-X mass spectrometer has redefined intelligent MS workflows by implementing a new approach (Thermo Scientific™ AcquireX™ method)[4] to automate the generation of a comprehensive background exclusion list to enhance the detection and identification of pesticide residues at low levels. The AcquireX workflow effectively manages HRAM MS and data-dependent MS^n data acquisition by performing real-time identification of precursor features associated with the matrix and those associated with compounds of interest. This directs the mass spectrometer to acquire full scan HRAM MS/MS (or higher order MS^n spectra) on only the latter. By automating the generation of the exclusion list, the AcquireX workflow maximizes the instrument cycle time to be spent acquiring higher quality precursor and product ion spectra for enhanced confidence in detection, identification, and quantitation.

An AcquireX workflow implemented on the Orbitrap ID-X Tribrid mass spectrometer was evaluated for the analysis of 250 pesticides spiked into strawberry matrix to demonstrate the efficiency of the workflow for identifying the pesticides spiked at different levels. Detection and identification were determined by matching experimentally acquired MS and MS/MS data to validated spectral libraries. The efficiency of the AcquireX workflow was evaluated against a standard data-dependent acquisition and dynamic exclusion method that did not utilize an exclusion list.

Experimental

Sample preparation

Strawberry samples were obtained from a local retail store. Following homogenization using an IKA ULTRA-TURRAX homogenizer (Sigma-Aldrich, St. Louis, MO), strawberry sub-samples were extracted using a QuEChERS approach.[2,7] Briefly, 10 g of sample was weighed into the ready-to-use QuEChERS extraction tubes with 4 g of $MgSO_4$, 1 g of trisodium citrate dehydrate and 0.5 g sodium citrate for buffered extraction. A total of 10 mL of acetonitrile was added and the sample was then mixed using a vortex mixer. Samples were shaken and centrifuged and an aliquot of the supernatant retained for analysis. Matrix-matched standards were prepared by spiking the 250 pesticide standards into the extracted matrices at concentration levels ranging from 0.05 to 200 ng/mL (equivalent to ng/g in the samples). The spiked levels are lower than the reported MRLs for many pesticides on fruit as listed within the CODEX database.

Liquid chromatography

Chromatographic separation was performed on a Thermo Scientific™ Vanquish™ Flex Binary UHPLC system using a Thermo Scientific™ Accucore™ aQ C18 column with dimensions of 100 × 2.1 mm and 2.6 µm particles. Mobile phase A consisted of 98:2 water/MeOH containing 5 mM ammonium formate and 0.1% formic acid. Mobile phase B consisted of 98:2 MeOH/water containing 5 mM ammonium formate and 0.1% formic acid. The flow rate was 300 µL/min and the column temperature was set to 25 °C. Analysis time was 15 min including 3 min equilibration time, and each experiment was performed using a 1 µL injection volume.

Mass spectrometry

All experiments were performed on an Orbitrap ID-X Tribrid mass spectrometer using the AcquireX workflow for automated generation of the background exclusion list and management of data-dependent acquisition (DDA) and dynamic exclusion (DE). The exclusion list was automatically generated from the matrix blank and exported to the data acquisition methods for subsequent analyses. The automatic gain control (AGC) target value was set at 2e5 for the full MS and 5e4 for the MS/MS spectral acquisition. The mass resolution was set to 60,000 (@*m/z* 200) for full scan MS and 15,000 for MS/MS events. All full scan MS/MS spectra were acquired using a Top 7 DDA method with dynamic exclusion implemented. High-energy collision dissociation (HCD) was performed with a stepped collision energy of 20, 40, and 70%. All samples were analyzed using a second mass spectral method using the same DDA/DE acquisition parameters but without the AcquireX generated exclusion list.

Data processing

Full scan data processing was performed with Thermo Scientific™ TraceFinder™ software. Quantitation and reproducibility analysis were performed on the precursor ion with mass extraction tolerance settings of ±5 ppm. The limits of detection were determined based on reproducibly measuring precursor response and the S/N ≥ 3. The limit of quantitation was based on S/N ≥ 10 and the coefficient of variance (%CV) and relative standard deviation (%RSD) were less than 20%. Qualitative data analysis and spectral library matching were performed using Thermo Scientific™ Compound Discoverer™ 3.1 software in which the experimentally acquired product ion spectra were matched against Thermo Scientific™ mzCloud library spectra, and successful matching was based on dot-product correlation coefficients.

Results and discussion

The role of HRAM MS^n analysis in efficient, confident, comprehensive non-targeted pesticide screening

Comprehensive, non-targeted pesticide screening requires data acquisition strategies to generate compound-specific mass spectral data, enabling post-acquisition detection and identification. For non-targeted full scan acquistion LC-MS methods, identification consists of detecting two ions, preferably the precursor and at least one product ion, with high mass accuracy (≤5 ppm). Full scan MS data acquired at resolving powers of 50,000 or greater can eliminate interference from background ions in the matrix, enabling selective extraction of precursor ions for qualitative and quantitative analysis. High resolving power also enables confident isotope detection and more accurate intensity measurements to help discern chemical formula of unknowns. High mass measurement accuracy maximizes confidence in precursor/product ion assignments based on spectral matching routines for known pesticides as well as significantly reducing false positive detections.

The Orbitrap ID-X mass spectrometer is capable of acquiring high resolution MS, MS/MS, and higher order MS^n data while maintaining high mass measurement accuracies (±1 ppm when using internal standard) across a wide dynamic range and for all MS^n spectra. This performance characteristic is critical for implementation of narrow mass extraction tolerance settings to enhance automated processing routines at MRLs or lower levels for unauthorized or unknown toxic compounds. For example, the EU apply a default MRL of 10 ppb (0.01 mg/kg) to pesticide-sample type combinations for which substantive MRLs (those based on field trial data) do not exist.

The AcquireX workflow uniquely facilitates non-targeted pesticide screening studies by automating intelligent MS^n data acquisition. The AcquireX workflow uses two different data acquisition methods, one for LC-MS precursor mapping and the second for LC-MS^n acquisition based on DDA/DE. The workflow automatically creates the acquisition sequence for blanks, matrix background, and samples as well as manages data storage. The workflow first acquires the precursor map of a representative blank matrix, processes the data, and creates the exhaustive exclusion list consisting of precursors and corresponding retention times. The resulting exclusion list is automatically imported into the LC-MS^n method enabling more selective acquisition of tandem MS spectra by real-time data

analysis to bypass matrix peaks. The workflow is easy to set up, amenable to any matrix (provided there are matrix blanks), and can effectively profile large numbers of pesticides across a wide dynamic range.

A stock solution mixture for the 250 pesticides was used to perform spiking levels to mimic different residue concentrations on strawberries. The lowest spiking level used in this is 100-fold lower than the EU default MRL. Figure 1 shows the post-acquisition data analysis for the pesticide ametryn. The resulting quantitation curve is used to evaluate the Orbitrap ID-X mass spectrometer performance and stability over the course of the study using a narrow mass tolerance for precursor extraction and integration. A 1/× weighting scheme was applied to the curve, resulting in a linear regression of 0.99. The inset shows the expanded curve covering the spiking levels of 0.5 to 5 ppb. The coefficient of variance calculated for the three replicate injections at 0.5 ppb was 4.5% and the relative standard deviation was 4.4%.

The parameters used for post-acquisition data processing were ideal for both known and unknown organic compounds due to the unbiased HRAM MS^n data acquired. Evaluating the data showed average chromatographic peak widths around 6 s wide. Despite the narrow peak widths, an average of seven DDA cycles were acquired enabling reproducible area under the curve measurements and acquisition of full scan MS/MS. The resulting data generated a high number of identification points (IPs) measured per compound per spiking level.[9-10] Identification points are achieved based on variance between empirical and reference measurements for chromatographic retention times, precursors, and product ion detection and relative abundance values. For example, mass errors higher than 10 mDa for a precursor scored 1.0 IPs and 1.5 for product ions as compared to mass errors below 2 mDa scoring 2.0 and 2.5 for precursors and product ions, respectively. This criteria has been updated by the EU SANTE guidelines (SANTE/12682/2019) requiring a minimum of a precursor and product ion measured with a mass error ≤5 ppm. Increased confidence is achieved with each additional precursor and product ion. Confident acquisition of product-ion rich tandem mass spectra facilitates spectral matching.

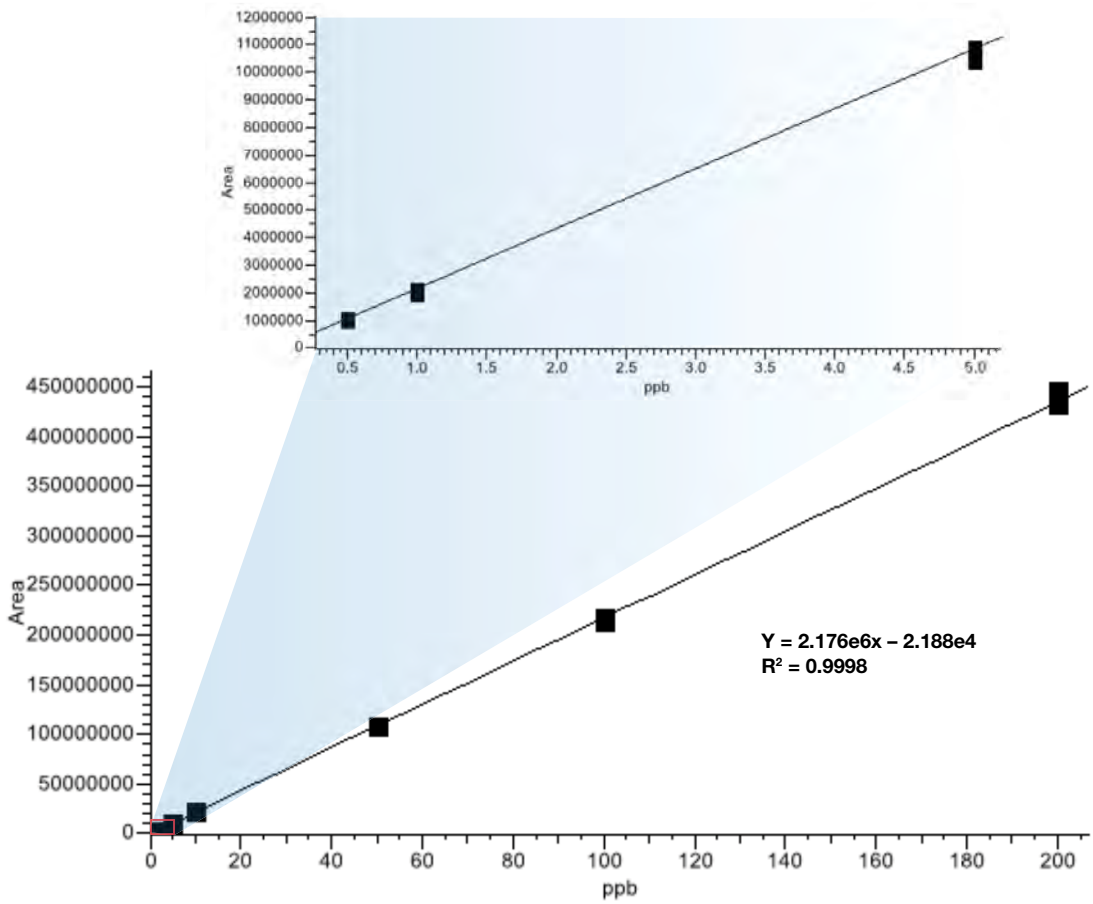


Figure 1. Quantitation curve for the pesticide ametryn in strawberry matrix across the spiking range of 0.5 to 200 ppb. The inset shows the expanded curve at the low end of the spiking range. A mass tolerance of ±5 ppm was used for post-acquisition data extraction, integration, and analysis using an automated processing software routine.

For the current work, a precursor *m/z* extraction tolerance of ±5 ppm was used for all post-acquisition data processing and analyses. For the ametryn example, the mass measurement error was less than 2 mDa consistently measured for all spiked levels. Since a single data processing routine was used for the analysis of all spiking levels, a stable chromatographic and mass measurement accuracy is achieved across the entire study. In addition, full scan product ion spectra were acquired and matched against spectral library entries (examples shown below). Product ion spectral matching utilized ±5 ppm tolerance achieving the highest IP scores per pesticide studied.

Non-targeted pesticide analysis

The detection efficency and quantitative ranges for all spiked pesticides were evaluated using the same automated, post-acquisition data processing workflows defined above. To maintain the quantitative accuracy at the low spiking levels for all 250 pesticides, the full scan detection capabilities of the Orbitrap ID-X mass spectrometer must have a high dynamic detection range to measure the ion signal attributed to the spiked pesticides in the presence of the background matrix. In addition, the full scan mass spectra must be acquried with high resolution to ensure selectivity for target extraction, and the mass measurement accuracy must remain stable across the entire study. Table 1 lists the reproducibility measurements for the three technical replicates at the LOQ for the selected pesticides. The representative 18 pesticides had LOQ levels between 0.5 and 1 ppb and measured %RSDs and %CVs were less than 8%. To provide context, the LOQ for both carbaryl and dinotefuran was 1.0 ppb compared to published Codex MRLs of 800 and 500 ppb, respectively.

Evaluation of the measured LODs and LOQs for all pesticides was performed using the same post-acquisition data processing described above. Automated data processing was performed on all pesticides across all levels and replicates to first determine the linear regression and %CV/RSD per level. The LOD/LOQ levels per pesticide were determined and manually evaluated for S/N determination. Figure 2 shows the pesticide distribution for the respective LOD and LOQ levels as a function of the spiking level. The presented workflow successfully detected and quantified all spiked pesticides at 10 ppb (ng/g) and 96% were detected 100-fold lower. Almost 94% of the pesticides analyzed had LOQ values at the 0.5 ppb (ng/g) level, demonstrating excellent detection and quantitation using HRAM MS data.

Table 1. List of representative pesticides, respective LOQ levels, and reproducible measurements across the three replicates

Pesticide residue	LOQ (ppb)	% Difference injection 1	% Difference injection 2	% Difference injection 3	%RSD	%CV
Ametryn	0.5	-0.07	-6.27	-8.01	4.38	4.48
Carbaryl	1.0	4.14	-6.79	3.43	6.10	5.10
Chloridazon	1.0	-0.13	-3.33	8.17	5.84	5.86
Clomazone	0.5	0.54	4.46	-9.50	7.31	7.24
Cyanazine	0.5	-4.15	-1.42	-1.71	1.54	1.53
Cyazofamid	0.5	-4.48	-4.40	-7.92	2.13	2.01
Dicrotophos	0.5	3.66	2.87	9.13	3.24	3.62
Dinotefuran	1.0	2.14	8.90	1.67	3.88	4.88
Fensulfothion	0.5	1.08	-4.20	5.61	4.87	4.74
Fuberidazole	0.5	-1.24	7.81	1.97	4.46	4.47
Hexazinone	0.5	-5.17	-0.35	-2.01	2.51	2.55
Heptonophos	1.0	-0.37	3.72	-0.37	2.34	2.40
Methabezthiazuron	0.5	3.38	6.83	2.54	2.18	2.13
Metosulam	1.0	-1.21	-1.81	-2.98	0.92	0.72
Ofurace	0.5	6.41	-1.51	-0.11	4.16	4.14
Tebufebpyrad	1.0	1.39	-5.94	6.06	6.02	5.60
Thiabendasole	0.5	1.49	-2.71	-5.99	3.84	3.81
Tricyclazole	0.5	-1.04	-4.31	-2.49	1.68	1.73

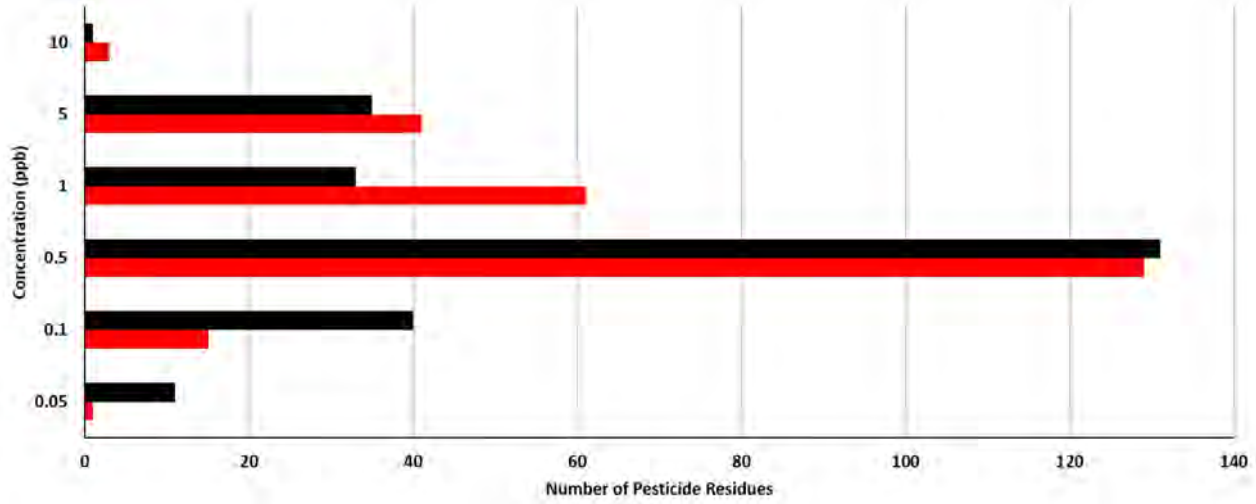


Figure 2. Comparative histogram for the measured LOD and LOQ levels for the pesticides tested. The distribution of pesticides for LOD are shown in black and LOQ in red.

The variance threshold was used as the first criteria for establishing LOD/LOQ for the measured pesticide response with the cutoff set to 20%. Figure 3 shows that all pesticides were measured with acceptable %CVs down to 10 ppb, which meets the EU requirements for pesticides. As further demonstration of the precursor detection capabilities in full scan MS data performance of the Orbitrap ID-X mass spectrometer, reproducible measurements were maintained for 205 pesticides down to 1 ppb.

Increasing unknown pesticide characterization using the AcquireX workflow and automated spectral library matching

The second challenge to effectively screen known and unknown pesticides is to acquire product ion spectra for structural confirmation or characterization. Data-dependent acquisition with dynamic exclusion (DDA/DE) has been routinely implemented to handle unknown targets in the presence of background matrix features. Thus a level of intelligent data acquisition must be performed by the

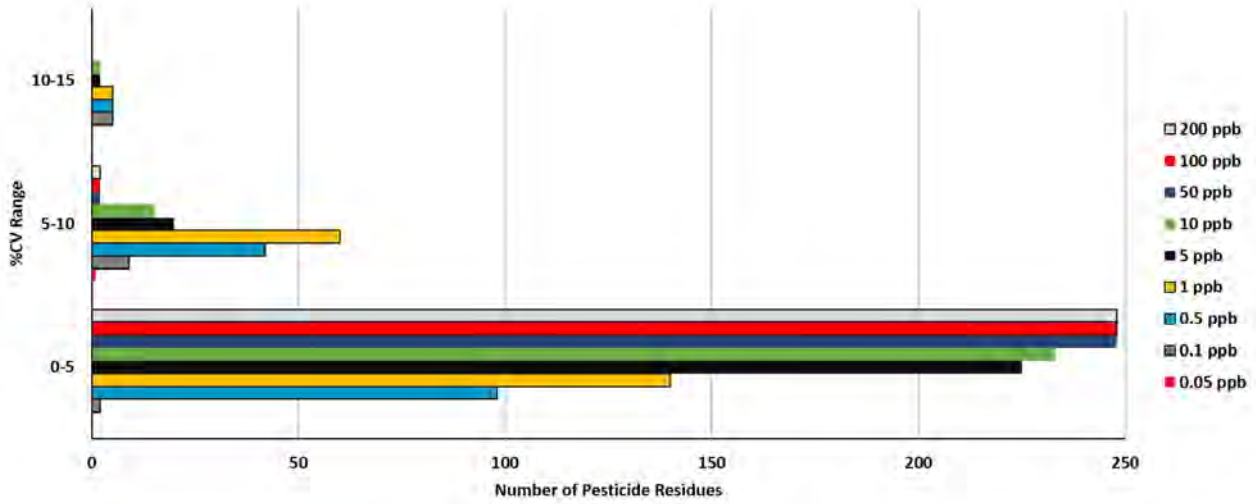


Figure 3. Distribution profile for all pesticides as a function of spiking level based on measured %CV

mass spectrometer to determine which features should be targeted for tandem mass spectral analysis and which ones should be avoided due to the precursor *m/z* value being previously interrogated within the user-defined time. Generally, the DDA/DE method prioritizes the more intense precursor *m/z* values, requiring a greater number of MS^2 spectral acquisitions to overcome matrix precursors to target potential pesticides. This could result in using longer cycle times, reducing the number of HRAM MS data points across the chromatographic peak and potentially compromising relative quantitative accuracy. Another option used to maintain a sufficient number of DDA/DA acquisition cycles utilizes lower precursor resolution settings (30,000) to devote more of the cycle time for DDA MS/MS spectra acquisition. Both options may still be insufficient for comprehensive non-targeted pesticide sampling.

Automating the exclusion list using the AcquireX workflow enhances the intelligent MS^n acquisition, resulting in a greater potential for interrogating primarily features associated with the sample, even at very low spiking levels. Coupling the AcquireX methods with traditional DDA/DE substantially increases DDA/DE efficiency by significantly reducing the number of precursors considered for tandem mass spectral analysis due to the extensive exclusion list. Therefore, implementing a static cycle time for the DDA/DE method can ensure enough full scan MS data points for post-acquisition quantitation as well as manage tandem mass spectral acquisiton through the setting of the AGC target value and maximum ion fill times. The resulting HRAM MS and MS^2 spectra are used to confirm the pesticide structure through spectral library matching or identify putative matches through chemical database matching.

The non-targeted pesticide selection efficiency was evaluated across all spiking levels. The same exclusion list generated at the onset of the study by the AquireX routine was used for all levels. Precursor *m/z* values not on the exclusion list but measured with sufficient intensities were then selected based on DDA/DE routines. Using this approach provides three advantages:

- Bypassing the background matrix features enables lower precursor ion intensity thresholds used to target compounds of interest.
- Increasing the maximum ion fill times enhances product ion spectral quality.
- Maintaining adequate cycle times allows for precursor quantitation.

All of these extend the dynamic range for non-targeted screening and relative quantitative analysis.

The resulting product ion spectra were searched against the mzCloud spectral library using Compound Discoverer 3.1. software. Figure 4 shows four pesticides selected for tandem mass spectral acquistion and the resulting mzCloud spectral library match. The product ion spectra were acquired at the LOQ values. Of particular interest is the quality of the spectra acquired at 0.5 and 1 ppb. The measured product ions contained in the spectral libraries are accounted for as well as the product ion distribution, resulting in a Pearson dot product correlation coefficient of almost 1. In addition, the product ion spectra were measured in the Orbitrap mass analyzer with a resolving power of 15,000 and mass accuracies less than 3 ppm for maximum confidence.

The mzCloud library has the world's largest HRAM LC-MS spectral library and continues to grow weekly. Over 16,000 compounds covering wide chemical diversity support over 17 different small molecule markets. A total of 745 pesticides and herbicides compounds are validated with 1289 ion trees and almost 144,000 spectra. Each compound has product ion spectra acquired at multiple collision energies and at MS^3 and MS^4 stages for increased matching and sub-structural matching routines. In addition, the mzCloud library is used in a unique unknown search strategy to help identify putative unknown structures not in the spectral library.

The spectral library matching shown in Figure 4 underscores the fragmentation and product ion detection efficiency of the Orbitrap ID-X Tribrid mass spectrometer. The empirical data was compared against validated spectral libraries for scoring. As stated above, the greater the number of measured product ions, the greater the IP score. Despite the low spiking level, each product ion spectrum showed at least four product ions with mass errors of ca. 0.4 mDa, each of which contributes 2 IPs per pesticide. In addition, the relative product ion distribution ratios fit additional requirements for increased IPs.

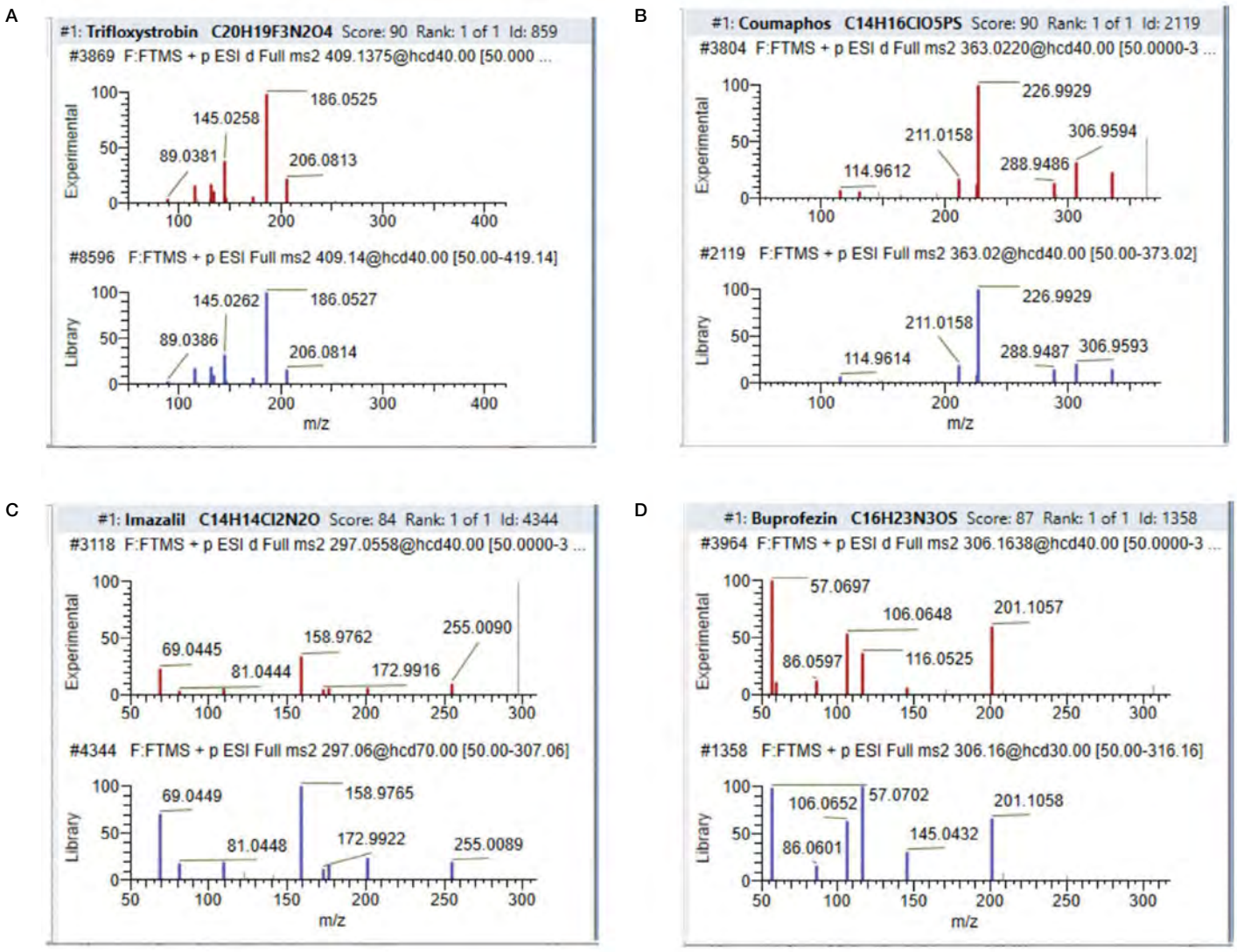


Figure 4. Evaluation of the MS² product ion spectra for four different pesticides and comparative spectral library match. The product ion spectra were acquired at 0.5 ppb for (A) trifloxystrobin and (b) coumaphos and 1 ppb for (c) imazalil and (d) buprofeszin.

Figure 5 shows the success of the intelligent tandem mass spectral acquisition at each spiking level as well as the reproducibility between two replicate studies. As demonstrated above, the resulting product ion spectra acquired per spiked pesticide resulted in a high-quality tandem mass spectrum that could be successfully matched against the reference spectrum for pesticide confirmation. About 92% of the spiked pesticides were interrogated and matched at 10 ppb or less, which is approximately an order of magnitude below the reported MRLs for the pesticides involved in the study. The success of the precursor selection method remains almost 90% down to the 5 ppb level and almost 60% at the 0.5 ppb level. In addition, the second study shows excellent reproducibility at all levels differing only by 1 pesticide at any level without manually creating a targeted precursor inclusion list. Note the success of the automated data acquisition scheme to reproducibly quantify and interrogate 12 and 13 pesticides at the 0.1 ppb level, which is 100-fold lower than the default thresholds established in the EU for pesticides.

The same samples were re-evaluated using the standard DDA/DE methods to assess automated pesticide selection and spectral matching. The loop count and cycle times were identical to that used for the AcquireX workflow, ensuring similar performance on post-acquisition HRAM MS quantitation. The difference is the lack of a comprehensive, automated exclusion list to enhance data-dependent selection of the compounds of interest. Figure 6 shows the comparative results for the two methods. The AcquireX method demonstrated better pesticide interrogation and spectral matching success at every level. The difference in the number of matched pesticides was much greater at the 5 ppb level and lower using the AcquireX workflow, and the standard DDA/DE method did not interrogate any of the spiked pesticides at the lowest spiking level. This was expected as the number of possible precursors surpassing the user-defined triggering thresholds was too great to enable automated selection of the low-intensity precursors associated with the spiked pesticides.

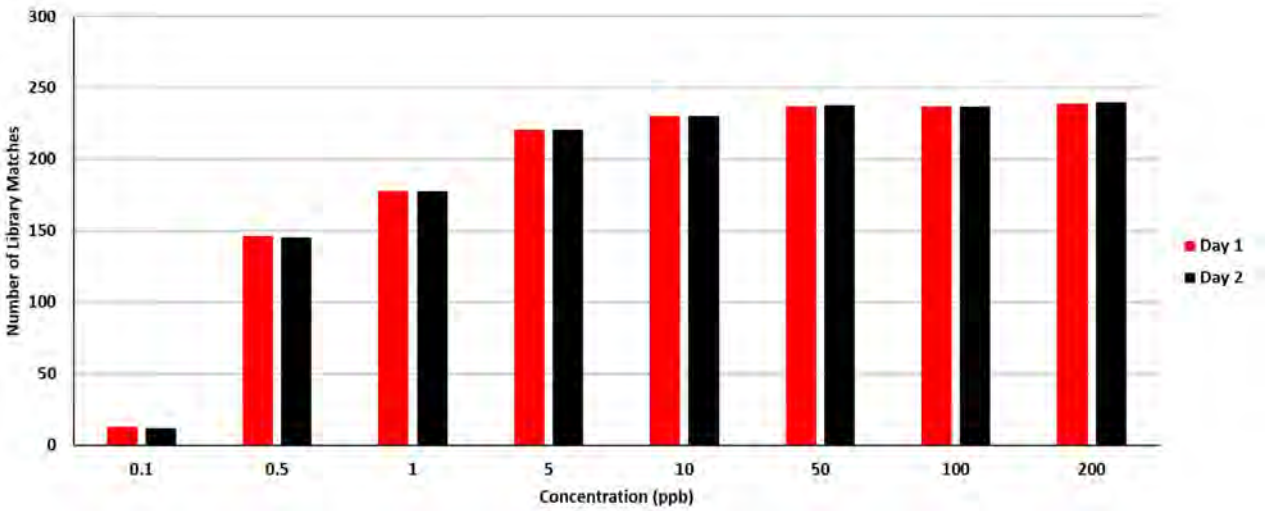


Figure 5. Comparative analysis of the number of pesticides that were confidently matched through mzCloud library searching for the same study on successive days. The entire AcquireX routine was evaluated on each day for the same samples.

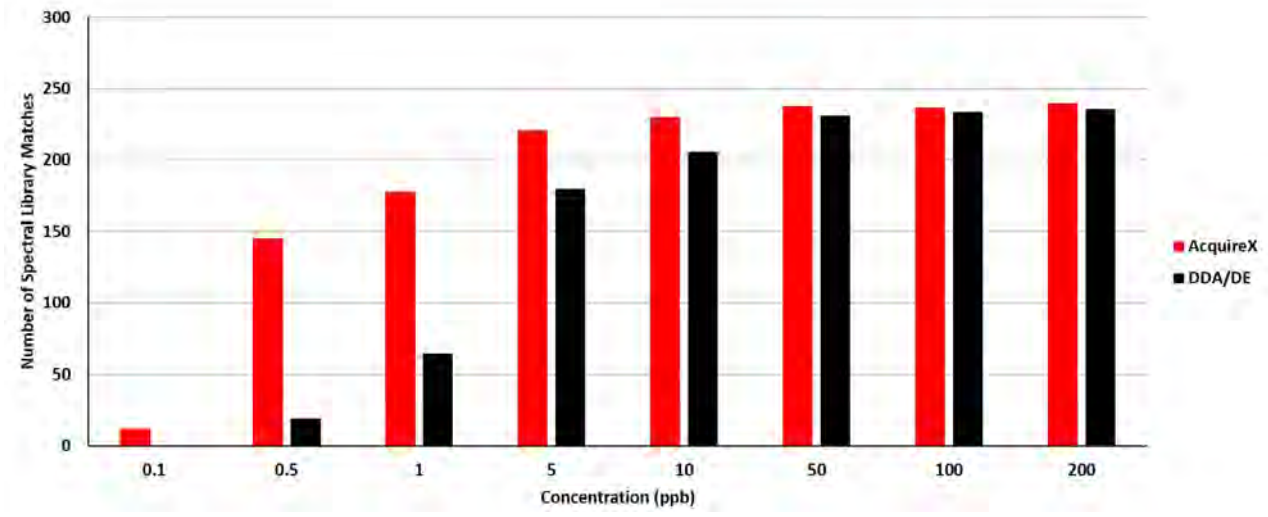


Figure 6. Comparative results for the two methods

thermo scientific

Conclusions

The AcquireX workflow operated on the Orbitrap ID-X Tribrid mass spectrometer presents a new paradigm for non-targeted, multiresidue LC-MS pesticide profiling. The incorporation of intelligent MS routines managed by the AcquireX workflow removes the burden of manually creating inclusion/exclusion lists previously needed to ensure pesticide detection and confirmation at LOD/LOQ levels. The results presented demonstrate the enhanced profiling capability for large numbers of pesticides across a wide range of residue levels substantially lower than existing MRLs. In addition, the HRAM MS^n data analysis in the Orbitrap mass analyzer results in high resolution and mass measurement accuracy for both MS and MS^2, enabling selective and sensitive automated post-acquisition data identification and extraction and processing routines to further enhance the workflow efficiency without sacrificing confidence. By automating the exclusion list generation and implementation, the AcquireX workflow is easy-to-use, amenable to any matrix, and ideal for post-acquisition analysis of known and unknown pesticides or sample-specific compounds.

References

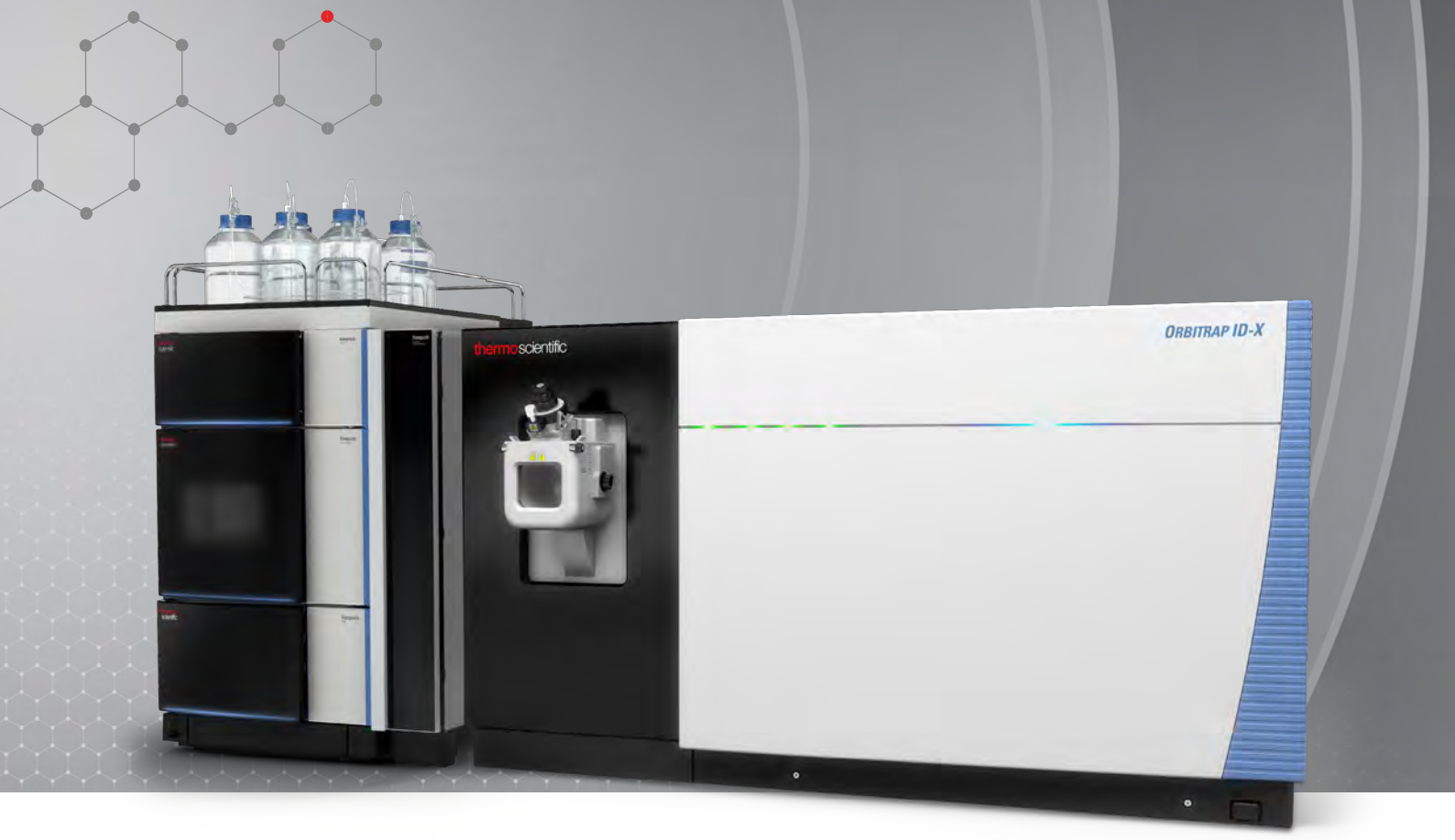
1. del Mar Gomez-Ramos, M. R.-A. Liquid chromatography Orbitrap mass spectrometry with simultaneous full scan and tadem MS/MS for highly selective presticide residue analysis. *Anal. Bioanal. Chem.* **2015,** *407*(21), 6317–6326.
2. Wong, J. Development and interlaboratory validation of a QuEChERS-based liquid chromatography-tandem mass spectrometry method for multiresideue pesticide analysis. *J. Agric. Food Chem.* **2010,** *58,* 5897–5903.
3. Niessen, W. M. Matrix effects in quantitative pesticide analysis using liquid chromatography-mass spectrometry. *Mass Spectrom. Reviews*, **2006,** *25,* 881–899.
4. Ferrer, I. G.-R.-A. Multi-residue pesticide analysis in fruits and vegtables by liquid chromatography-time-of-flight mass spectrometry. *J. Chrom. A* **2005,** *1082,* 81–90.
5. Barbieri, M. V.-A.-A. Analysis of 52 pesticides in fresh fish muscle by QuEChERS extraction followed by LC-MS/MS determination. Sci. of the Total Environment, 653, 958-967.Hernandez, F. I. (2004). Comparison of different mass spectrometric techniques combined with liquid chromatography for confirmation of pesticides in environmental water based on the use of identification points. Anal. *Chem.* **2019,** *76,* 4349–4357.
6. Martinez-Dominguez, G. R.-G. Multi-class methodology to determine pesticides and mycotoxins in green tea and royal jelly supplements by liquid chromatography coupled to Orbitrap high resolution mass spectrometry. *Food Chemistry* **2016,** *197,* 907–915.
7. Kalli, A. S. A multiresidue method for pesticide analysis using an Orbitrap Tribrid mass spectrometer and automatic background exclusion. American Society of Mass Spectrometry. Atlanta, GA: *J. Am. Soc. Mass Spectrom.* **2019.**
8. Shi, F. G. Application of a high resolution benchtop quadrupole-Orbitrap mass spectrometer for the rapid screening, confirmation, and quantification of illegal adulterated phosphoidesterase-5 inhibitors in herbal medicines and dietary supplements. *J. Chromatography A* **2014,** *1344,* 91–98.
9. Koplin, D. W. Response to comment on “pharamceuticals, hormones, and other organic wastewater contaminants in U. S. streams, 1999-2000: a national reconnaissance”. *Environ. Sci. Technol.* **2002,** *36*(18), 4007–4008.
10. Hernandez, F. I. Comparison of different mass spectrometric techniques combined with liquid chromatography for confirmation of pesticides in environmental water based on the use of identification points. *Anal. Chem.* **2004,** *76,* 4349–4357.

Find out more at **thermofisher.com**

ThermoFisher SCIENTIFIC

thermo scientific

Go beyond
with technology and intelligence



Transform small molecule identification and characterization

The Thermo Scientific™ Orbitrap ID-X™ Tribrid™ mass spectrometer combines industry-leading mass analyzer technology dedicated for small molecule analysis with the Thermo Scientific™ AcquireX intelligent MS^n data acquisition strategy, to collect more meaningful data. By capturing more low-abundance analytes and leveraging streamlined data analysis and mass spectral prediction tools, the system's intuitive, automated workflow accelerates results to let you go beyond faster.

Advance quicker at **thermofisher.com/orbitrapID-X**

ThermoFisher SCIENTIFIC

Top tips

Set your acquisition cycle time to fit the chromatography type using method templates such as ID MS³. Establish the overall cycle time range between 0.6 and 1.0 seconds to acquire high-resolution and high-mass accuracy MS, MS^2 and MS^3 spectra.

Perform collision-induced dissociation (CID) in place of higher-energy collisional dissociation (HCD). Use CID at the MS^2 or MS^3 stage to increase the intensity of subsequent tandem MS spectra, specifically for small molecules with labile functional groups, such as glucuronide and sulfate conjugation, in drug metabolism studies and sugar side chains in natural products.

Ensure appropriate maximum ion accumulation times are set for the intended experimental goals. The Orbitrap ID-X mass spectrometer incorporates the ion routing multiple to accumulate precursors and product ions prior to detection and can be determined for each MS^n spectrum acquired, enhancing intra- and inter-scan dynamic range. This becomes critical for low-level detection or increased compound interrogation down to MS^3, MS^4, or higher without sacrificing spectral quality.

Automatically manage replicate injections. Implement more in-depth MS^n methods, including assisted CID to enhance the product ion spectra for unpredicted compounds, through automation.

Leverage intelligent MS capabilities and perform real-time decision making to enhance MS^n acquisition. The Thermo Scientific method editor enables user-defined input parameters to enable on-the-fly spectral processing for the determination of optimum, higher-order MS data.

Select the iterative precursor exclusion method to automatically update the exclusion list following each replicate injection. Spend more cycle time on each compound of interest and remove the need to sample all plausible compounds in only one or two injections.

Maintain ultra-high mass accuracy for the precursor and product ions using EASY IC. The greater the mass accuracy, the fewer the potential number of chemical formulae that are considered for a possible match – EASY IC promises to maintain mass measurement accuracy ≤1 ppm across the detected *m/z* range.

Featured Products

Thermo Scientific™ Orbitrap ID-X™ Tribrid™ Mass Spectrometer

SCIENTISTS' CHOICE AWARDS 19 NOMINEE

The Orbitrap ID-X Tribrid Mass Spectrometer uses industry-leading mass analyzer technology, dedicated for small molecule analysis, to collect more meaningful data.

"Best mass spectrometer ever!"

Ease of use: ★★★★★ **After sales service:** ★★★★★ **Value for money:** ★★★★★

Rating: 5 ★★★★★

Application Area: HD Exchange

"I have been using Thermo Fisher Orbitrap Mass Spec to find out the dynamics of the proteins using hydrogen-deuterium exchange. It is user friendly and I am able to produce high-quality HD exchange data for publications. I strongly recommend this for anyone who is working on proteomics."

Venkata Machha, Mayo Clinic

AcquireX Intelligent Data Acquisition Workflow

The AcquireX intelligent MS^n data acquisition workflow enables you to acquire high-quality, in-depth LC-MS^n data for every analyte of interest.

Mass Frontier 8.0 Spectral Interpretation Software

The Mass Frontier software uses high-quality, fully curated experimental spectral and fragmentation libraries in order to provide greater insights, understanding, and confidence in your data.

Compound Discoverer Software

The Compound Discoverer software offers a full suite of tools aimed to address small molecule data analysis challenges. The software is designed to streamline identification of unknowns, determine real differences between samples, and elucidate biological pathways.

Additional Resources

Product information

AcquireX intelligent data acquisition technology

mzLogic data analysis algorithm

Orbitrap ID-X Tribrid Mass Spectrometer

On-demand webinars

Advanced metabolome annotation with the Orbitrap ID-X Tribrid mass spectrometer

Addressing the "identity crisis" in small molecule compound identi ication

More LC-MS webinars

EB65635-EN0520S