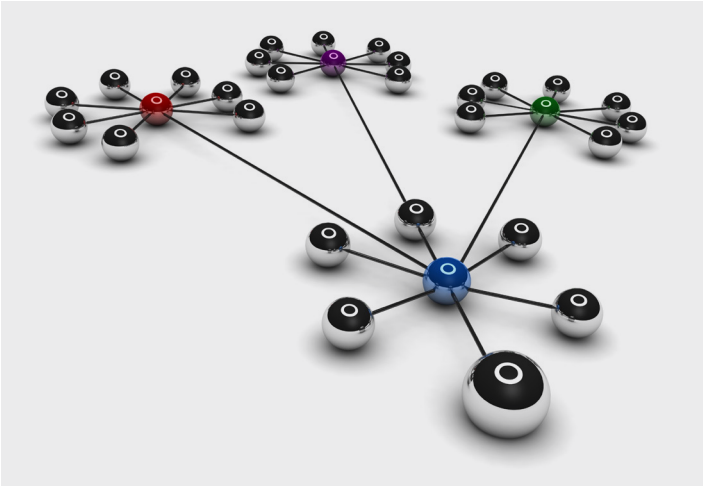
APPLICATION NOTE 73325

Metabolome identification using LC-MS^n Orbitrap-based mass spectrometry

Authors: David Peake, Ioanna Ntai, Amanda Souza and Ralf Tautenhahn, Thermo Fisher Scientific, San Jose, California, USA

Keywords: Metabolomics, untargeted metabolomics, compound annotation, compound identification, Compound Discoverer software, MS^n, mzCloud, mzVault, Orbitrap ID-X Tribrid MS, AcquireX, LC-MS, NIST SRM 1950, human plasma



Goals

Demonstrate the LC-MS data dependent MS^n workflow for the confident untargeted annotation of unknown metabolites and inclusion list-based identification of metabolite standards using a Thermo Scientific™ Orbitrap™ ID-X™ Tribrid™ mass spectrometer and Thermo Scientific™ Compound Discoverer™ software.

Compare the results from standard mixtures, NIST SRM 1950 plasma and spiked plasma extracts to provide unequivocal evidence of highly confident identifications.

Introduction

Annotation of hundreds of unknown metabolites from human plasma is one of the most difficult challenges faced by metabolomics scientists.[1] The degree of difficulty in the identification of unknown metabolites, however, depends on the information that can be extracted from an LC-MS dataset. An “unknown” feature is a single *m/z* and retention time observed during LC-MS analysis. This feature may be related to the chemical background from a blank sample and unrelated to the sample itself. Alternatively, a feature may be a real biological compound and its abundance changes in a control group compared to a treatment group. However, if this compound is not found in any compound database, it is referred to as an “unknown unknown”. When annotating unknowns using high-resolution accurate mass data, one can classify four distinctly different types of unknowns (Table 1):

ThermoFisher SCIENTIFIC

Table 1. Unknown annotation level definitions

Type of Unknown	Level	HR-MS Composition	Database match	Library MS^2 match	Standard Rt match	Standard MS^n match
Known known	0	Full structure confirmed by 2D-NMR, high resolution MS and other spectroscopic methods				
Known known[4]	1	✓	✓	✓	✓	✓
Known unknown[3]	2	✓	✓	✓		
Known unknown[2]	3	✓	✓			
Unknown unknown[1]	4	✓				
Unknowns		Feature: retention time and *m/z*; from chemical background or experimentally unrelated sources				

[1] Unexpected but relevant biological compound with known elemental composition and no match in any compound database
[2] Unknown compound with known elemental composition (database match to multiple isomeric compounds)
[3] Unknown compound with known mass spectral library match (similarity to one or more isomeric compounds)
[4] Known isomer matching retention time and MS^n spectrum of a reference standard (stereochemical configuration not determined)

- **Unknowns**—feature: *m/z* or composition related to background or experimentally unrelated sources
- **Unknown unknowns**—elemental composition of a relevant compound not found in any database
- **Known unknowns**—putative annotation: elemental formula database match; MS^2 library match
- **Known knowns**—identification: formula, retention time and MS^n spectra matches authentic standard

The ease of moving from features to annotation, and ultimately identification, is determined by the quality of analytical measures that are determined from the high-resolution accurate mass data. Confidence in unknown annotation increases as one adds multiple analytical measures to improve the level of annotation. For example, to move from an "unknown unknown" to a "known unknown" the elemental composition of the *m/z* needs to be determined and matched with a metabolite entry in a sample-relevant compound database.

Automated annotation must incorporate accurate mass, isotope pattern and isotopic fine structure to confirm accurate elemental formula for database searching. Unknown MS^2 and MS^3 spectra should be searched for identity or similarity matches against a high-quality MS^n mass spectral library and the identification level reported for each metabolite (Table 1) based on the consensus of the available analytical measures.[2,3] Unequivocal identification of annotated metabolites requires confirmation by spiking authentic standards into the plasma and matching the retention time and fragmentation patterns of the unknown MS/MS spectra.

Workflows using the advantages of the Orbitrap mass analyzer with excellent mass resolution, mass accuracy and accurate mass LC-MS^n capabilities have recently been reported to improve metabolite identification in untargeted metabolomics[4] and lipid identification[5] in untargeted lipidomics. Intelligent acquisition provides comprehensive metabolomics and lipidomics coverage in bacteria, cell media, plants, and mammals.[6,7]

Preliminary analysis of NIST standard reference material SRM 1950 human plasma using the Thermo Scientific™ AcquireX™ acquisition strategy and an Orbitrap ID-X Tribrid mass spectrometer[8,9] showed that more than 5000 compounds were detected (after background removal using Compound Discoverer 3.1 software) in the positive ion reversed-phase LC-MS analysis with 76% of the MS^2 spectra being acquired on the preferred protonated ions. By incorporating the elemental composition of molecular species determined from the high-resolution MS data, more than 4000 ChemSpider database hits were obtained.

The MS^2 product ion spectra were searched against the mzCloud MS^n library (17,392 compounds and 5.97 million spectra), providing 419 identity matches of known metabolites. An additional 65 lipid annotations were provided by searching against a mass list comprised of 350 lipid species identified in the SRM 1950 LC-MS^2 data using Thermo Scientific™ LipidSearch™ 4.2 software. Similarity matches (1174 compounds) were prioritized using the mzLogic algorithm which rank orders annotations by mapping potential structures to known fragment ions in mzCloud.[10] From the preliminary data analysis, a list of over 400 high-quality metabolite annotations was curated, and standards were obtained for 58 of these metabolites.

Once a metabolite's retention time and mass spectral library tree are acquired from a reference standard, a semi-targeted approach[11] is often used to combine untargeted analysis with targeted identification and quantitation of metabolites of interest.

Experimental methods

Sample preparation

SRM 1950 human plasma was purchased from NIST (https://srm1950.nist.gov). Human metabolite reference standards were purchased from MetaSci (https://www.metasci.ca). Lipid reference standards were obtained from Avanti Polar Lipids, Inc. (https://avantilipids.com) and acyl carnitines standards were obtained from Cambridge Isotope Laboratories, Inc. (https://www.isotope.com). Naproxen was obtained from Sigma Aldrich.

Stock solutions of standards (1.00 mg/mL) were prepared in water, methanol, 1:1 methanol-water or 1:1 methanol-chloroform. Some lipid reference standards were supplied as 10 mg/mL solutions in chloroform. Ten different standard mixtures containing 5–10 compounds each were prepared at 10 µg/mL in 1:1 methanol-water or methanol. The ten standard mixtures (Mix01–Mix10) were diluted 1:10 into water to give 1.0 µg/mL per compound for LC-MS^2 analysis.

Aliquots of SRM 1950 human plasma (100 µL) were precipitated using a 3:1 (v/v) methanol to sample ratio (21,000 g at 4 °C for 20 min), and 300 µL of the supernatant was evaporated under vacuum at 7 °C overnight. Each plasma extract was reconstituted in 90 µL of 95% water and 5% methanol, vortexed and the extracts were combined and centrifuged. Ten 90 µL aliquots of plasma extract were spiked with 10 µL of a standard mixture (M01–M10). Four aliquots of unspiked SRM 1950 plasma extracts were prepared as QC samples (P1–P4).

Mass spectrometry

Blanks (2 system blanks), SRM 1950 plasma extracts (4 samples P1–P4), standard mixtures (10 samples, Mix01–Mix10) and spiked plasma samples (10 samples M01–M10) were injected (2 µL) and separated on a Thermo Scientific™ Hypersil GOLD™ C18 column (2.1 × 150 mm, 1.9 µm, 45 °C) using the gradient shown in Table 2. LC-MS analyses were performed using a Thermo Scientific™ Vanquish™ UHPLC system coupled with an Orbitrap ID-X Tribrid mass spectrometer (Figure 1). Mass spectral data were acquired separately in positive and negative ion modes using two different LC-MS methods described in Table 3.

In the first sample injection LC-MS was performed at 120,000 MS resolution (FWHM @*m/z* 200) using internal mass calibration over the mass range *m/z* 67–1000. During the second sample injection LC-MS was performed with data dependent HCD MS^2 (30,000 res.) using stepped collision energy for metabolite characterization. An accurate mass inclusion list was employed to ensure that reference compounds (Table 4) were prioritized for MS^2 analysis; untargeted metabolites were acquired during the remaining cycle time.

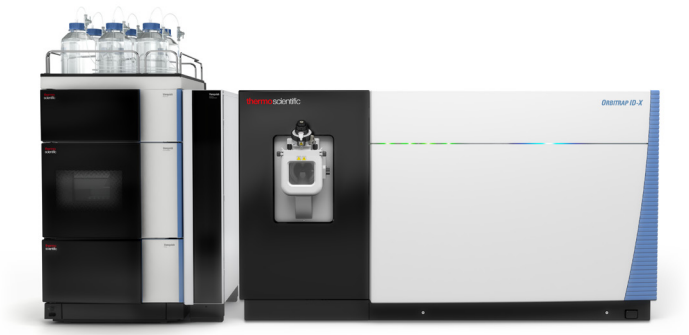


Figure 1. Thermo Scientific Vanquish UHPLC and Orbitrap ID-X Tribrid LC-MS^n System

Table 2. UHPLC gradient method

Time, min	% A	% B
-4.00	100	0
0.00	100	0
8.00	50	50
9.00	2	98
14.00	2	98
14.10	100	0
15.00	100	0

Table 3. Orbitrap ID-X Tribrid MS conditions

Ion Source	Orbitrap MS
ESI pos. = 3500V	**Method 1—LC-MS**
ESI neg. = 3000V	MS R = 120,000
Sheath gas = 40	Max IT = 50 msec
Aux gas = 8	AGC = 1E5
Sweep = 1	**Method 2—LC-$ddMS^2$**
Vaporizer = 320 °C	MS R = 60,000, MS^2 R = 30,000
IT temp = 275 °C	IW = 1.5 Da
RF lens = 35%	AGC = 5E4
Internal Calibration = Easy-IC	Max IT = 54 msec
Method time = 15 min	Collision energy = 20, 35, 50%

Data processing

Analysis of SRM 1950 plasma spiked with metabolite standards was conducted to validate confidence in unknown annotations. Ultra-high resolution MS data was analyzed with Compound Discoverer version 3.1 software by first annotating unknown metabolites based on a formula search of selected ChemSpider databases (BioCyc, DrugBank, HMDB, KEGG, LipidMaps, PubMed) and Metabolika pathways. A mass list generated from the reference standards in Table 4 was used to perform targeted data processing to confirm annotation of the reference compounds in SRM 1950 plasma and for high confidence unknown annotations (Figure 2). All 58 reference compounds were annotated by a formula search, with a grand average mass measurement error of <0.1 ppm, and all but one compounds in the plasma samples matched the retention time of the standards (Table 5).

A fragmentation search was conducted of the unknown MS^2 spectra against the MS^n mzCloud library and a search using the Thermo Scientific™ mzVault™ library constructed from MS^2 spectra obtained from the reference compounds in the standard mixtures (M01–M10). The combination of mzCloud and mzVault searches unequivocally identified 56 out of 58 compounds in the SRM 1950 and spiked plasma extracts. The MS^2 match score reported in Table 5 is the highest match score from mzCloud or mzVault for the plasma/spiked plasma samples. Stearoyl ceramide was not identified due to a very low abundance M+H ion.

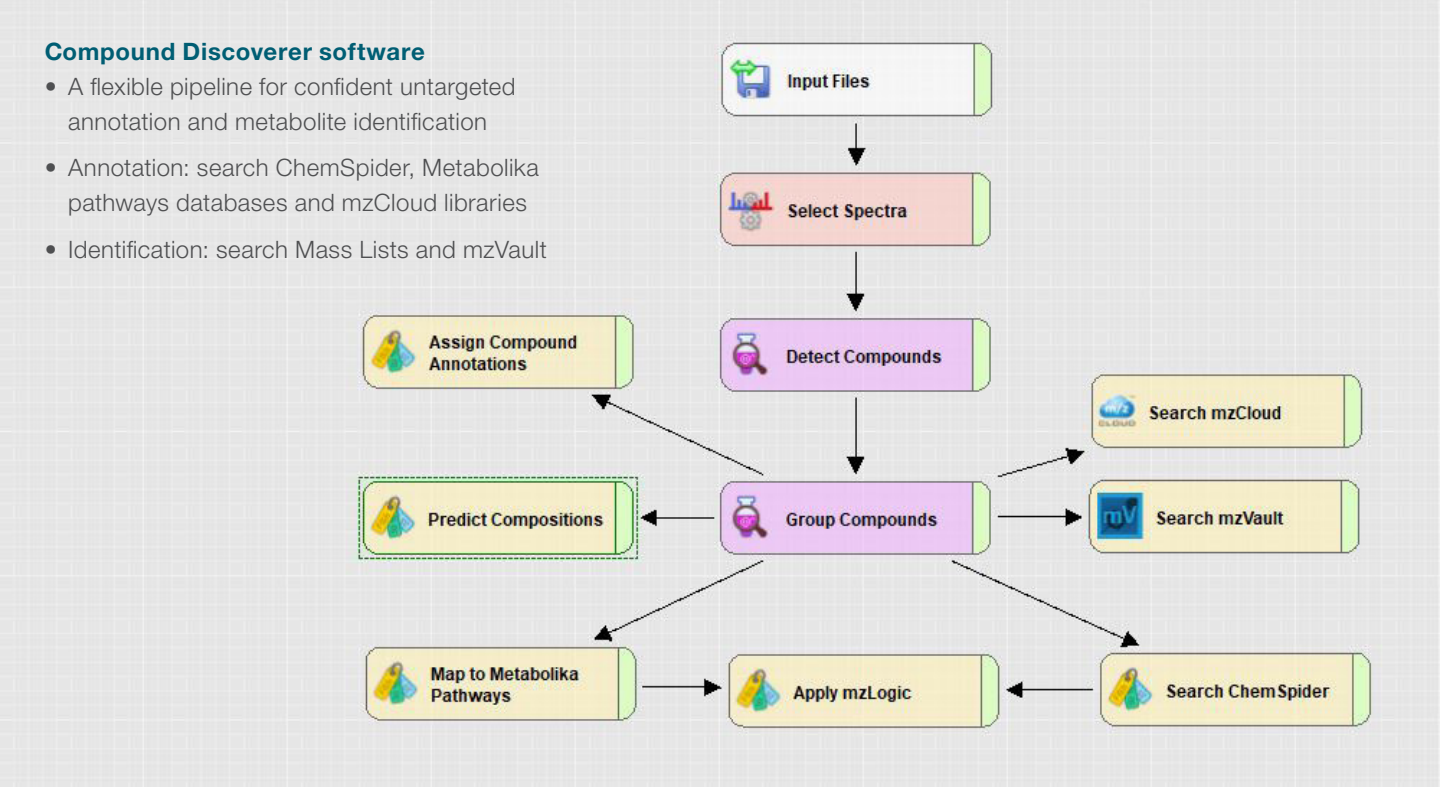


Figure 2. Data analysis workflow for confirming metabolite identification and untargeted annotation

Table 4. Reference standards for confirming metabolite annotation

#	Compound Name	HMDB ID	Location / Cat. #	Formula	µg/mL	Std Mix
1	Choline	HMDB0000097	MetaSci 3J3	$C_5H_{13}NO$	1	M10
2	Serine	HMDB0000187	MetaSci 3E5	$C_3H_7NO_3$	1	M09
3	Creatinine	HMDB0000562	MetaSci 3C3	$C_4H_7N_3O$	1	M08
4	Proline	HMDB0000162	MetaSci 3F6	$C_5H_9NO_2$	1	M09
5	Levulinic acid	HMDB0000720	MetaSci 3H2	$C_5H_8O_3$	1	M06
6	Betaine	HMDB0000043	MetaSci 3A1	$C_5H_{11}NO_2$	1	M07
7	Butyric acid, 2-OH-3-Me	HMDB0000407	MetaSci 4G3	$C_5H_{10}O_3$	1	M02,06
8	Pyroglutamic acid	HMDB0000267	MetaSci 5J10	$C_5H_7NO_3$	1	M03
9	Valeric acid, 3-methyl-2-oxo	HMDB0000491	MetaSci 5E5	$C_6H_{10}O_3$	1	M03
10	Isocaproic acid, α-keto	HMDB0000695	MetaSci 10D7	$C_6H_{10}O_3$	1	M06
11	Proline, 4-hydroxy	HMDB0000725	MetaSci 6G9	$C_5H_9NO_3$	1	M04
12	Salicylic acid	HMDB0015470	MetaSci 5I5	$C_7H_6O_3$	1	M03
13	Glutamine	HMDB0000641	MetaSci 3A2	$C_5H_{10}N_2O_3$	1	M08
14	Lysine	HMDB0000182	MetaSci 4D1	$C_6H_{14}N_2O_2$	1	M02
15	Glutamic acid	HMDB0000148	MetaSci 3B10	$C_5H_9NO_4$	1	M08
16	Methionine	HMDB0000696	MetaSci 6F9	$C_5H_{11}NO_2S$	1	M04
17	Acetminophen	HMDB0001859	MetaSci 5I6	$C_8H_9NO_2$	1	M03
18	Gentisic acid	HMDB0000152	MetaSci 3G7	$C_7H_6O_4$	1	M06
19	Histidine	HMDB0000177	MetaSci 3G3	$C_6H_9N_3O_2$	1	M09
20	Carnitine	HMDB0000062	CIL NSK-B-US	$C_7H_{15}NO_3$	24.5	M01,05
21	Phenylalanine	HMDB0000159	MetaSci 3E6	$C_9H_{11}NO_2$	1	M09
22	Uric acid	HMDB0000289	MetaSci 3D7	$C_5H_4N_4O_3$	0.5	M06
23	Arginine	HMDB0000517	MetaSci 4C10	$C_6H_{14}N_4O_2$	1	M02
24	Indole-3-acetic acid	HMDB0000197	MetaSci 3A6	$C_{10}H_9NO_2$	1	M08
25	Hippuric acid	HMDB0000714	MetaSci 4A9	$C_9H_9NO_3$	1	M02,06
26	Paraxanthine	HMDB0001860	MetaSci 4B2	$C_7H_8N_4O_2$	1	M02
27	Theobromine	HMDB0002825	MetaSci 7C9	$C_7H_8N_4O_2$	0.5	M05
28	Theophylline	HMDB0001889	MetaSci 1B9	$C_7H_8N_4O_2$	1	M07
29	Tyrosine	HMDB0000158	MetaSci 3A7	$C_9H_{11}NO_3$	1	M08
30	Azelaic acid	HMDB0000784	MetaSci 3G10	$C_9H_{16}O_4$	1	M09
31	3-Indolepropionic acid	HMDB0002302	MetaSci 10A7	$C_{11}H_{11}NO_2$	1	M07
32	Caffeine	HMDB0001847	MetaSci 4H9	$C_8H_{10}N_4O_2$	1	M03
33	Hippuric acid, 2-hydroxy	HMDB0000840	MetaSci 9F4	$C_9H_9NO_4$	1	M06
34	Carnitine, 2:0	HMDB0000201	CIL NSK-B-US	$C_9H_{17}NO_4$	7.72	M01,05
35	Tryptophan	HMDB0000929	MetaSci 3H8	$C_{11}H_{12}N_2O_2$	1	M04
36	Carnitine, 3:0	HMDB0062514	CIL NSK-B-US	$C_{10}H_{19}NO_4$	1.65	M01
37	Naproxen	HMDB0001923	Sigma	$C_{14}H_{14}O_3$	1	M05
38	Carnitine, 4:0	HMDB0002013	CIL NSK-B-US	$C_{11}H_{21}NO_4$	1.76	M01
39	Carnitine, 5:0 (Isovaleryl)	HMDB0000688	CIL NSK-B-US	$C_{12}H_{23}NO_4$	1.86	M01
40	Atenolol	HMDB0001924	MetaSci 7C4	$C_{14}H_{22}N_2O_3$	1	M05
41	Piperine	HMDB0029377	MetaSci 9G5	$C_{17}H_{19}NO_3$	1	M07
42	Carnitine, 8:0	HMDB0000791	CIL NSK-B-US	$C_{15}H_{29}NO_4$	2.18	M01
43	Sphingosine, d18:1	HMDB0000252	Avanti 860490	$C_{18}H_{37}NO_2$	1	M10
44	Cortisone	HMDB0002802	MetaSci 9E5	$C_{21}H_{28}O_5$	1	M05
45	Cortisol (Hydrocortisone)	HMDB0014879	MetaSci 6E6	$C_{21}H_{30}O_5$	1	M04
46	Carntine, 14:0	HMDB0005066	CIL NSK-B-US	$C_{21}H_{41}NO_4$	2.82	M01
47	Cholesterol	HMDB0000067	Avanti 700100	$C_{27}H_{46}O$	1	M10
48	Carnitine, 16:0	HMDB0000222	CIL NSK-B-US	$C_{23}H_{45}NO_4$	6.07	M01
49	Carnitine, 18:1(9Z)	HMDB0005065	Avanti 870852	$C_{25}H_{47}NO_4$	10	M10
50	Glycoursodeoxycholic acid	HMDB0000708	MetaSci 9B1	$C_{26}H_{43}NO_5$	1	M07
51	Glycochenodeoxycholic acid	HMDB0000637	MetaSci 6D2	$C_{26}H_{43}NO_5$	1	M04
52	Glycocholic acid	HMDB0000138	MetaSci 3I10	$C_{26}H_{43}NO_6$	1	M06
53	Lyso PC, 1-18:1(9Z)	HMDB0002815	Avanti 845875	$C_{26}H_{52}NO_7P$	10	M06
54	Lyso PC, 1-18:0	HMDB0010384	Avanti 855775	$C_{26}H_{54}NO_7P$	10	M10
55	Ceramide, d18:1/16:0	HMDB0000790	Avanti 860516	$C_{34}H_{67}NO_3$	1	M10
56	Ceramide, d18:1/18:0	HMDB0000829	Avanti 860518	$C_{36}H_{71}NO_3$	1	M10
57	Bilirubin	HMDB0000054	MetaSci 4A8	$C_{33}H_{36}N_4O_6$	1	M10
58	Sphingomyelin, d18:1/16:0	HMDB0061712	Avanti 860584	$C_{39}H_{79}N_2O_6P$	1	M10

Table 5. Compounds identified in SRM 1950 human plasma

ID Level	Compound Name	Rt, min	Rt Δ, min	Formula confirmed	Ion	*m/z*, plasma	Δ, ppm	MS^2 match	ID conf
1	Choline	1.13	0.01	$C_5H_{13}NO$	Pos	104.1069	-0.9	90.9	☑
1	Serine	1.09	0.02	$C_3H_7NO_3$	Pos	106.0499	0.3	84.3	☑
1	Creatinine	1.18	0.02	$C_4H_7N_3O$	Pos	114.0661	-0.8	90.3	☑
1	Proline	1.27	0.00	$C_5H_9NO_2$	Pos	116.0705	-0.9	89.0	☑
3	Levulinic acid	3.76	0.3	$C_5H_8O_3$	Neg	115.0401	0.3	84.6	☒
1	Betaine	1.17	0.00	$C_5H_{11}NO_2$	Pos	118.0862	-0.5	90.4	☑
1	Butyric acid, 2-OH-3-Me	5.13	0.03	$C_5H_{10}O_3$	Neg	117.0557	-0.1	92.9	☑
1	Pyroglutamic acid	2.40	0.02	$C_5H_7NO_3$	Pos	130.0498	-0.5	89.2	☑
1	Valeric acid, 3-methyl-2-oxo	5.93	-0.02	$C_6H_{10}O_3$	Neg	129.0557	-0.1	85.1	☑
1	Isocaproic acid, α-keto	6.35	-0.03	$C_6H_{10}O_3$	Neg	129.0557	-0.1	88.0	☑
1	Proline, 4-hydroxy	1.12	0.01	$C_5H_9NO_3$	Pos	132.0655	-0.1	90.2	☑
1	Salicylic acid	9.15	0.00	$C_7H_6O_3$	Neg	137.0244	-0.1	95.9	☑
1	Glutamine	1.11	0.02	$C_5H_{10}N_2O_3$	Pos	147.0763	-0.8	87.4	☑
1	Lysine	0.97	-0.01	$C_6H_{14}N_2O_2$	Pos	147.1128	0.0	89.7	☑
1	Glutamic acid	1.13	0.01	$C_5H_9NO_4$	Neg	146.0459	0.1	88.5	☑
1	Methionine	1.90	0.01	$C_5H_{11}NO_2S$	Pos	150.0583	-0.2	91.8	☑
1	Acetaminophen	4.26	0.00	$C_8H_9NO_2$	Pos	152.0706	0.0	94.9	☑
1	Gentisic acid	5.33	-0.02	$C_7H_6O_4$	Neg	153.0193	-0.2	90.0	☑
1	Histidine	1.09	0.09	$C_6H_9N_3O_2$	Pos	156.0768	0.3	94.1	☑
1	Carnitine	1.15	0.01	$C_7H_{15}NO_3$	Pos	162.1123	-1.0	98.1	☑
1	Phenylalanine	3.91	-0.01	$C_9H_{11}NO_2$	Pos	166.0862	-0.3	95.4	☑
1	Uric acid	1.80	0.00	$C_5H_4N_4O_3$	Neg	167.0210	-0.4	91.4	☑
1	Arginine	1.09	0.07	$C_6H_{14}N_4O_2$	Pos	175.1190	0.3	88.6	☑
1	Indole-3-acetic acid	8.53	0.00	$C_{10}H_9NO_2$	Pos	176.0706	0.0	92.5	☑
1	Hippuric acid	5.85	-0.03	$C_9H_9NO_3$	Neg	178.0510	-0.1	89.2	☑
1	Paraxanthine	5.03	0.00	$C_7H_8N_4O_2$	Pos	181.0720	0.0	93.4	☑
1	Theobromine	4.35	0.01	$C_7H_8N_4O_2$	Pos	181.0720	0.0	95.7	☑
1	Theophylline	5.21	0.01	$C_7H_8N_4O_2$	Pos	181.0720	0.0	93.1	☑
1	Tyrosine	2.54	-0.03	$C_9H_{11}NO_3$	Pos	182.0811	-0.4	96.7	☑
1	Azelaic acid	9.68	0.00	$C_9H_{16}O_4$	Neg	187.0975	-0.4	96.1	☑
1	3-Indolepropionic acid	9.72	0.00	$C_{11}H_{11}NO_2$	Pos	190.0863	0.2	96.2	☑
1	Caffeine	6.13	0.02	$C_8H_{10}N_4O_2$	Pos	195.0879	1.3	97.3	☑
1	Hippuric acid, 2-hydroxy	6.98	-0.01	$C_9H_9NO_4$	Neg	194.0458	-0.4	93.4	☑
1	Carnitine, 2:0	1.83	0.01	$C_9H_{17}NO_4$	Pos	204.1230	-0.2	93.2	☑
1	Tryptophan	4.91	-0.01	$C_{11}H_{12}N_2O_2$	Pos	205.0971	-0.3	97.8	☑
1	Carnitine, 3:0	2.92	-0.02	$C_{10}H_{19}NO_4$	Pos	218.1388	0.5	92.0	☑
1	Naproxen	10.45	-0.01	$C_{14}H_{14}O_3$	Pos	231.1016	0.1	86.7	☑
1	Carnitine, 4:0	4.15	0.01	$C_{11}H_{21}NO_4$	Pos	232.1543	-0.1	92.0	☑
1	Carnitine, 5:0 (Isovaleryl)	5.50	-0.01	$C_{12}H_{23}NO_4$	Pos	246.1700	0.1	91.8	☑
1	Atenolol	4.17	0.00	$C_{14}H_{22}N_2O_3$	Pos	267.1704	0.3	96.6	☑
1	Piperine	10.54	0.00	$C_{17}H_{19}NO_3$	Pos	286.1438	0.1	96.6	☑
1	Carnitine, 8:0	9.97	0.00	$C_{15}H_{29}NO_4$	Pos	288.2170	0.2	89.9	☑
1	Sphingosine, d18:1	10.45	-0.01	$C_{18}H_{37}NO_2$	Pos	300.2899	0.6	89.1	☑
1	Cortisone	10.13	0.00	$C_{21}H_{28}O_5$	Pos	361.2013	1.0	82.1	☑
1	Cortisol	10.21	-0.01	$C_{21}H_{30}O_5$	Pos	363.2168	0.5	92.3	☑
1	Carnitine, 14:0	10.44	-0.01	$C_{21}H_{41}NO_4$	Pos	372.3111	0.7	87.5	☑
1	Cholesterol	13.25	-0.01	$C_{27}H_{46}O$	Pos	369.3520	1.1	94.7	☑
1	Carnitine, 16:0	10.50	0.00	$C_{23}H_{45}NO_4$	Pos	400.3426	1.1	87.7	☑
1	Carnitine, 18:1	10.51	0.00	$C_{25}H_{47}NO_4$	Pos	426.3580	0.5	88.2	☑
1	Glycoursodeoxycholic acid	10.46	0.01	$C_{26}H_{43}NO_5$	Neg	448.3069	0.1	87.2	☑
1	Glycochenodeoxycholic acid	10.64	-0.01	$C_{26}H_{43}NO_5$	Neg	448.3068	-0.1	81.9	☑
1	Glycocholic acid	10.54	-0.01	$C_{26}H_{43}NO_6$	Neg	464.3019	0.2	84.9	☑
1	Lyso PC, 1-18:1(9Z)	11.30	-0.01	$C_{26}H_{52}NO_7P$	Pos	522.3558	0.7	89.9	☑
1	Lyso PC, 1-18:0	11.55	-0.04	$C_{26}H_{54}NO_7P$	Pos	524.3715	0.8	95.4	☑
1	Ceramide, d18:1/16:0	13.46	-0.02	$C_{34}H_{67}NO_3$	Pos	538.5199	1.0	91.4	☑
3	Ceramide, d18:1/18:0	14.25	0.01	$C_{36}H_{71}NO_3$	Pos	566.5510	0.6	ND	☒
1	Bilirubin	11.81	0.00	$C_{33}H_{36}N_4O_6$	Pos	585.2710	0.4	91.6	☑
1	Sphingomyelin, d18:1/16:0	13.77	-0.01	$C_{39}H7_9N_2O_6P$	Pos	703.5750	0.2	94.0	☑

Discussion

Analytical measures for confident compound annotation

What constitutes confident annotation of unknown compounds? To move from an "unknown unknown" feature to a "known unknown" requires unequivocal determination of elemental composition. This is illustrated in Figure 3a by the accurate mass, isotope pattern and fine structure of a molecular ion at *m/z* 150.05830 from human plasma measured in a single Orbitrap scan at 120,000 resolution. The mass measurement fits the elemental composition $C_5H_{11}NO_2S$ within a mass error of -0.15 ppm (part-per-million). Furthermore, the fine structure at the A+1 peak (*m/z* 151) fits within 1.5 ppm the expected accurate masses and relative abundances of ^{15}N (*m/z* 151.05534), ^{33}S (*m/z* 151.05775), ^{13}C (*m/z* 151.06160) and ^{2}H isotopes (*m/z* 151.06438). The A+2 peak (*m/z* 152) fits within 0.9 ppm of the expected mass for ^{34}S (*m/z* 152.05405) and ^{18}O (*m/z* 152.06244) and the A+3 peak (*m/z* 153) fits within 0.8 ppm for ^{13}C, ^{34}S (*m/z* 153.05736).

Since there were seven isomeric compounds from a search of the ChemSpider database with the assigned elemental composition $C_5H_{11}NO_2S$ of this "known unknown", we absolutely need fragmentation data to annotate the correct isomer. A mzCloud MS^n mass spectral library search of the MS^2 spectrum (Figure 3b) gives a match for reference library entry methionine with a High Res match score of 93.4, which is based on an identity algorithm that measures the weight distance between spectra. Typically, mzCloud library match scores ≥85 are quite reliable and this narrows the search for "known unknown" annotations to one or more potential isomers.

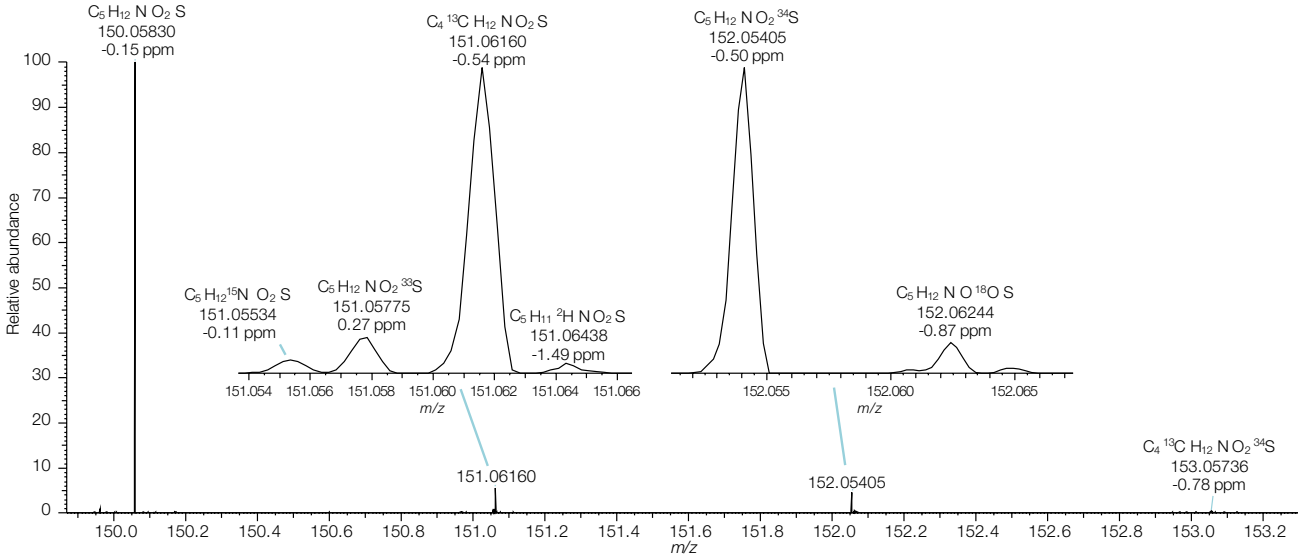


Figure 3a. Mass measurement of *m/z* 150.05830 and isotopic fine structure gives unequivocal assignment of elemental composition, $C_5H_{11}NO_2S$

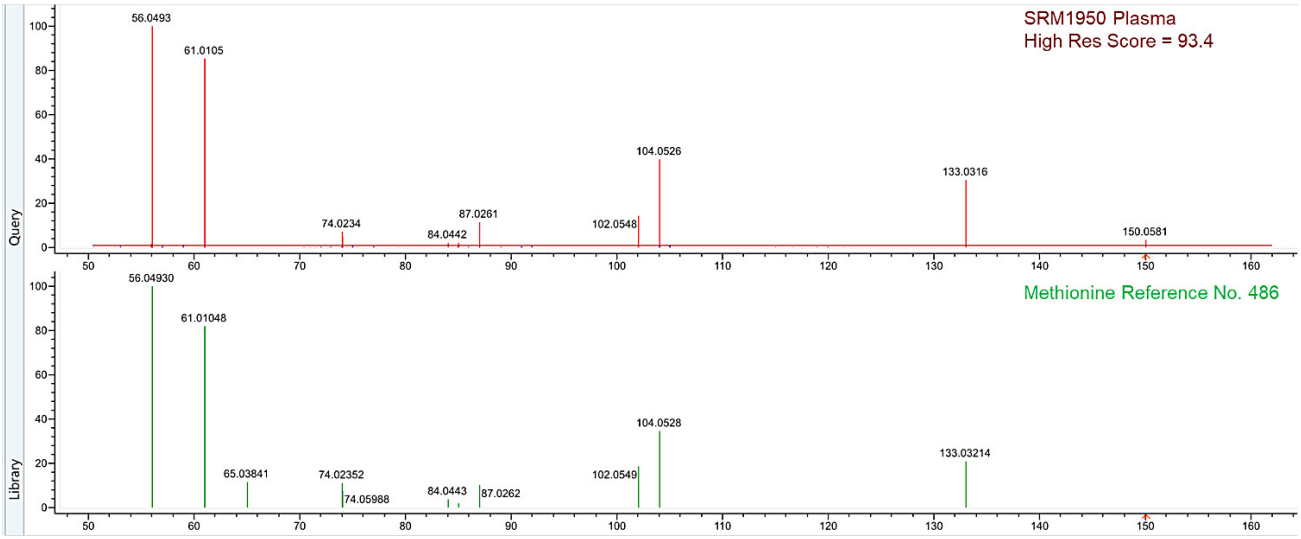


Figure 3b. MS^2 spectral match of $C_5H_{11}NO_2S$ to methionine in mzCloud spectral library

In order to identify this “known unknown” the data for the reference standards and spiked plasma were analyzed. As shown in Figure 3c, the retention time of the methionine standard (1.89 min) matches the *m/z* 150.05833 at 1.90 min found in SRM 1950 plasma. Furthermore, the MS^2 spectrum at 1.91 min in plasma spiked with the Methionine reference standard matches the mzVault reference spectrum with a match score of 88.0. The matching retention time and library fragmentation thus establishes that this compound is identified as a “known known” methionine. Note that it is not possible to determine absolute stereochemistry using achiral chromatography and LC-MS detection even though L-methionine is found in human plasma.

Results

The purpose of this study is to demonstrate that using the analytical measures discussed above and high quality Orbitrap LC-MS^n data it is possible to get high confidence in automated annotation of “known unknowns”. In order to confirm that many of the highly confident annotations are indeed correct metabolite identifications, 58 metabolite standards (Table 4) were spiked into SRM 1950 plasma extracts.[12] During data processing a mass list and mzVault library for the 58 reference standards was added as shown in Figure 2.

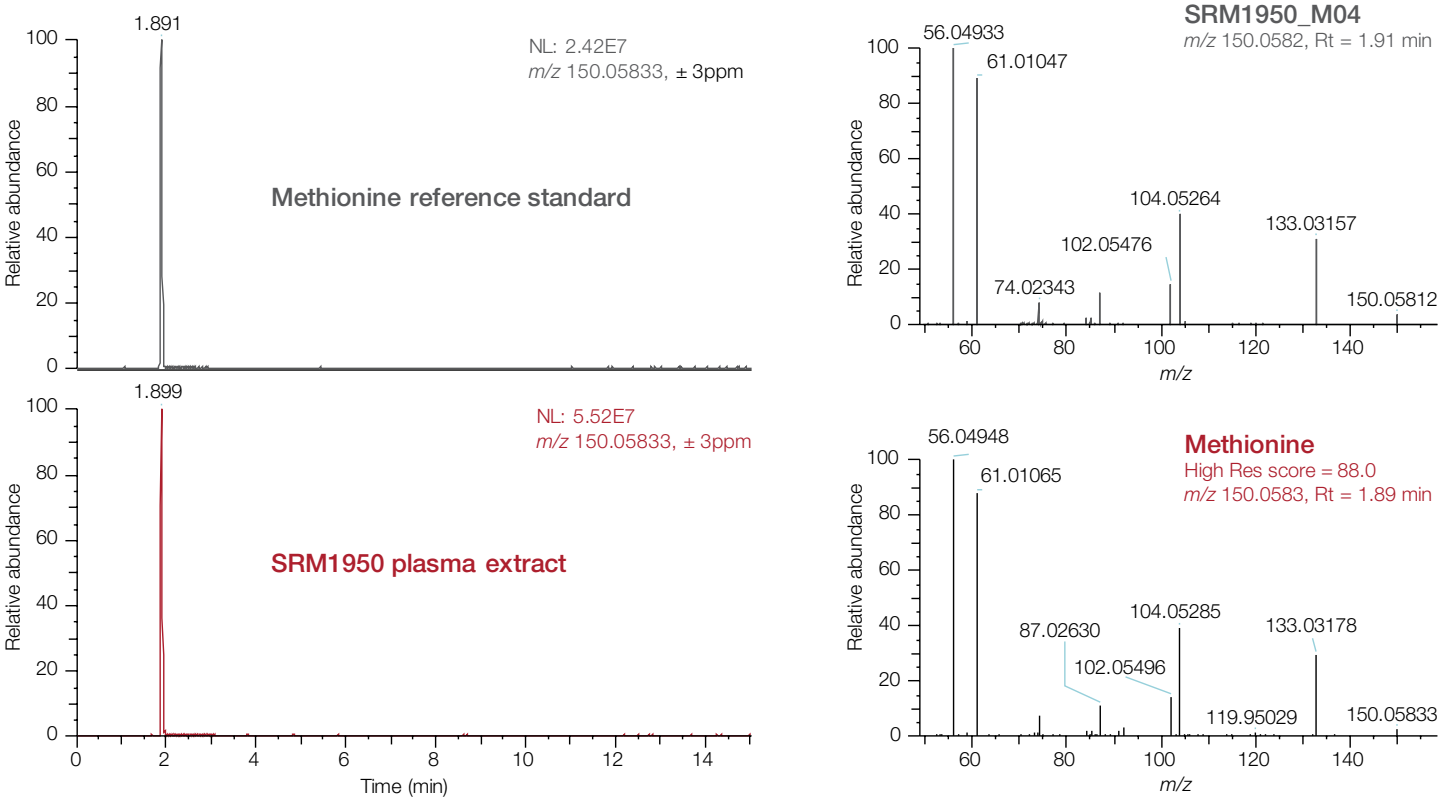


Figure 3c. a) Retention time match (1.90 min) in human plasma to methionine standard (1.89 min) and b) mzVault library match in human plasma spiked with methionine (match score = 88.0)

The “known knowns” that were positively identified (56 out of 58 reference standards) in SRM 1950 are summarized in Table 5. The compound name, retention time (RT) of the standard and retention time difference in the plasma sample are reported as well as the mass measurement error (part-per-million, ppm) from the calculated *m/z*. The best library match score from the plasma/spiked plasma was reported from mzVault or mzCloud. The compound identity in plasma was confirmed if the compound matched retention time (RT difference <0.1 min), *m/z* (mass tolerance ≤1.1 ppm), predicted elemental composition (including fine structure and isotopic pattern) and library match score ≥80 of the standard spiked into the plasma extract. The ID level “1” as described in Table 1 was assigned if the above criteria were met.

Uric acid identity confirmed in SRM 1950 plasma extracts using Compound Discover

Figure 4 confirms the presence of uric acid in SRM 1950 plasma by matching retention time (1.80 min) and *m/z* 167.0210 (-0.4 ppm mass error) by ultra-high resolution (120K @ *m/z* 200) LC-MS, including a good fit to the fine isotopic structure that confirms the elemental composition is correctly assigned by Compound Discoverer software. Figure 5 illustrates the automated identity search of the MS^2 spectrum from the protonated molecular ion of uric acid matching the mzVault and mzCloud mass spectral libraries spectra, with very good match scores obtained from a single high resolution MS^2 spectrum. Only two compounds did not match at ID level 1 using this automated data processing approach.

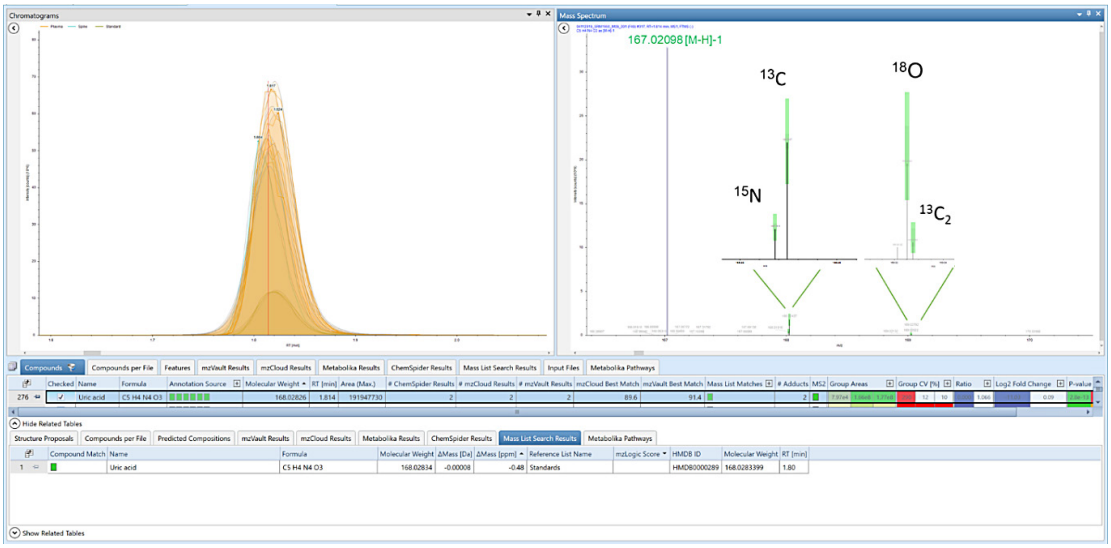


Figure 4. Confirmation of uric acid $[M-H]^-$ ion, *m/z* 167.0210, 1.81 min, $C_5H_4N_4O_3$ matching isotopic fine structure (A+1 peak with ^{15}N and ^{13}C isotopes, A+2 with ^{18}O and $^{13}C_2$ isotopes), in human plasma

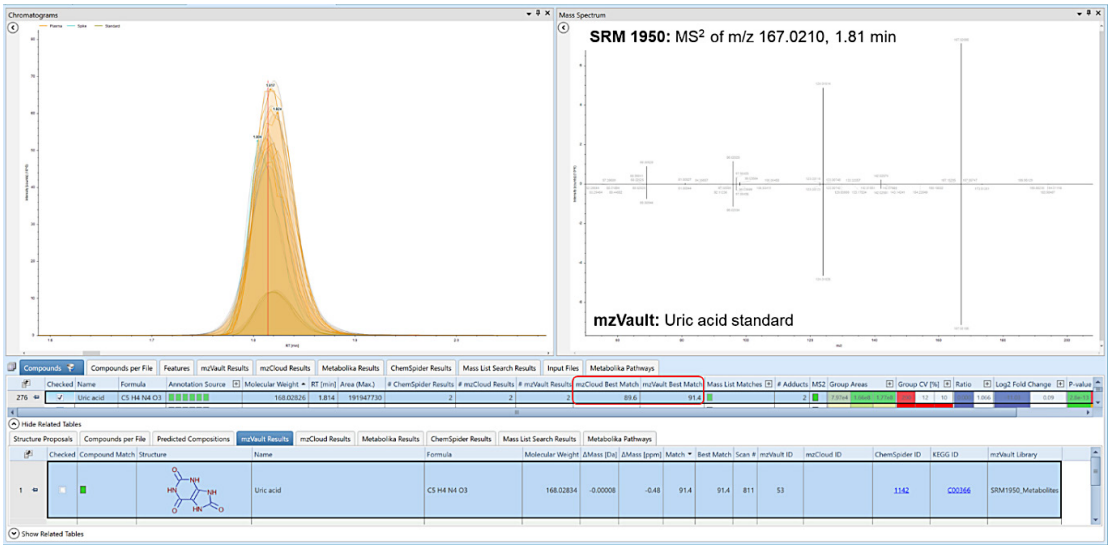


Figure 5. Identification of uric acid in SRM 1950 human plasma extracts with high quality MS^2 spectral matches to the mzVault compound library (91.4) and mzCloud online library (89.6).

An "known unknown" with a composition match and significant retention time difference in the plasma compared to the standard reference compound is shown in Figure 6. An isomer of levulinic acid (ID Level 3) was found in the plasma and spiked plasma at *m/z* 115.0401 ($C_5H_8O_3$) and retention time of 3.76 and 3.78 min, respectively (Figure 6c and 6b) whereas, the retention time of levulinic acid was 3.48 minutes (Figure 6a).

Ceramide (d18:1/18:0, ID Level 3) gave an M+H ion at *m/z* 566.5507 and RT 14.25 min matching the reference compound. However, there was insufficient abundance in the plasma to confirm the MS^2 spectrum.

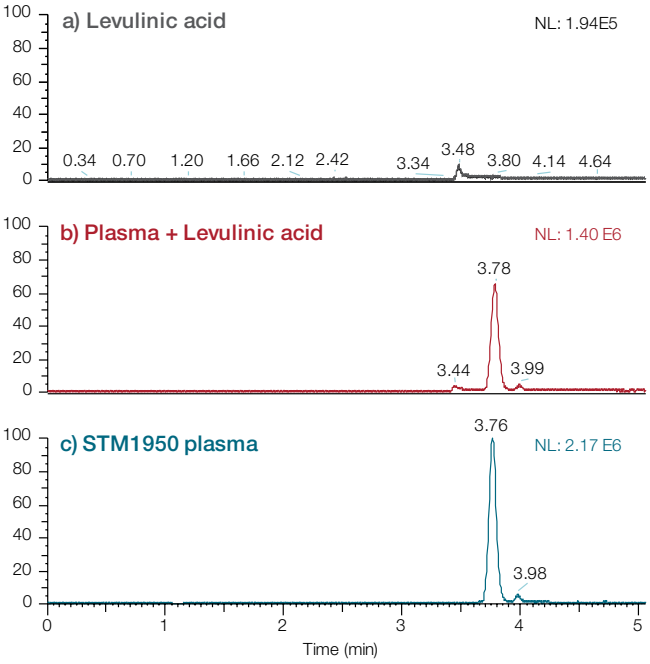


Figure 6. Negative ion XIC of $[M-H]^-$ anion *m/z* 115.0400 with composition $C_5H_8O_3$ in: a) standard mixture containing levulinic acid, b) SRM 1950 plasma extract spiked with levulinic acid and c) SRM 1950 plasma.

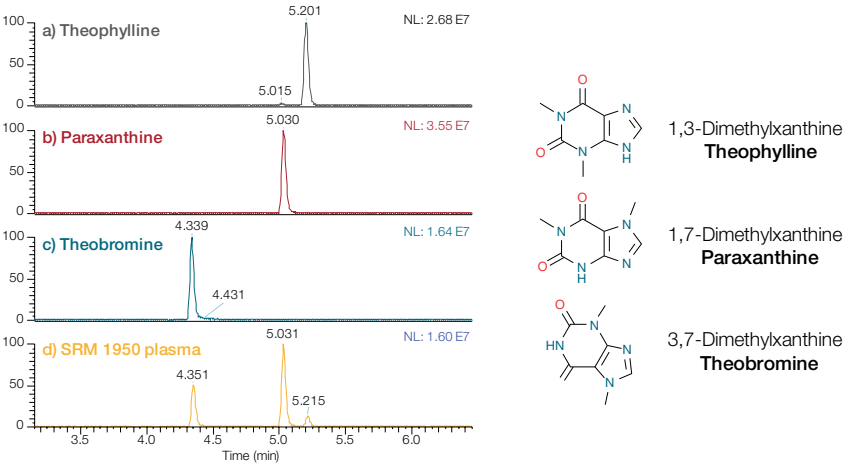


Figure 7. XIC of $[M+H]^+$ ion *m/z* 181.07202 with composition $C_7H_8N_4O_3$ in sample: a) Mix07 – theophylline, b) Mix02 – paraxanthine, c) Mix05 – theobromine and d) P01 – SRM 1950 plasma extract

Dimethylxanthine isomers identified in SRM 1950 plasma extracts

Figure 7 shows mass chromatograms of *m/z* 181.07202 and the retention time of three different standards: theophylline (5.20 min), paraxanthine (5.03 min) and theobromine (4.35 min) matching the retention times of $C_7H_8N_4O_3$ isomers observed in SRM 1950 human plasma.

Figures 8a, 8b and 8c show high quality mzVault library matches for theophylline, paraxanthine and theobromine in SRM 1950 plasma extract, confirming identification. The data were also searched against mzCloud and provided very high match scores confirming differentiation of these isomers by their MS^2 spectra. Note that match score differences between mzVault and mzCloud libraries may reflect the more extensive curation of the mzCloud library.

Phenylalanine identity confirmed in SRM 1950 plasma extracts

Figure 9a shows the confirmation of phenylalanine in SRM 1950 at retention time of 3.89 min and *m/z* 166.0862 (-0.2 ppm) matches that of the standard reference compound, including a good fit to the fine isotopic structure that confirms the correct elemental composition ($C_9H_{11}NO_2$). Figure 9b shows identification by matching the fragmentation of the protonated molecular ion with mzVault and mzCloud mass spectral library entries for phenylalanine with excellent match scores of 100.0 and 89.8, respectively.

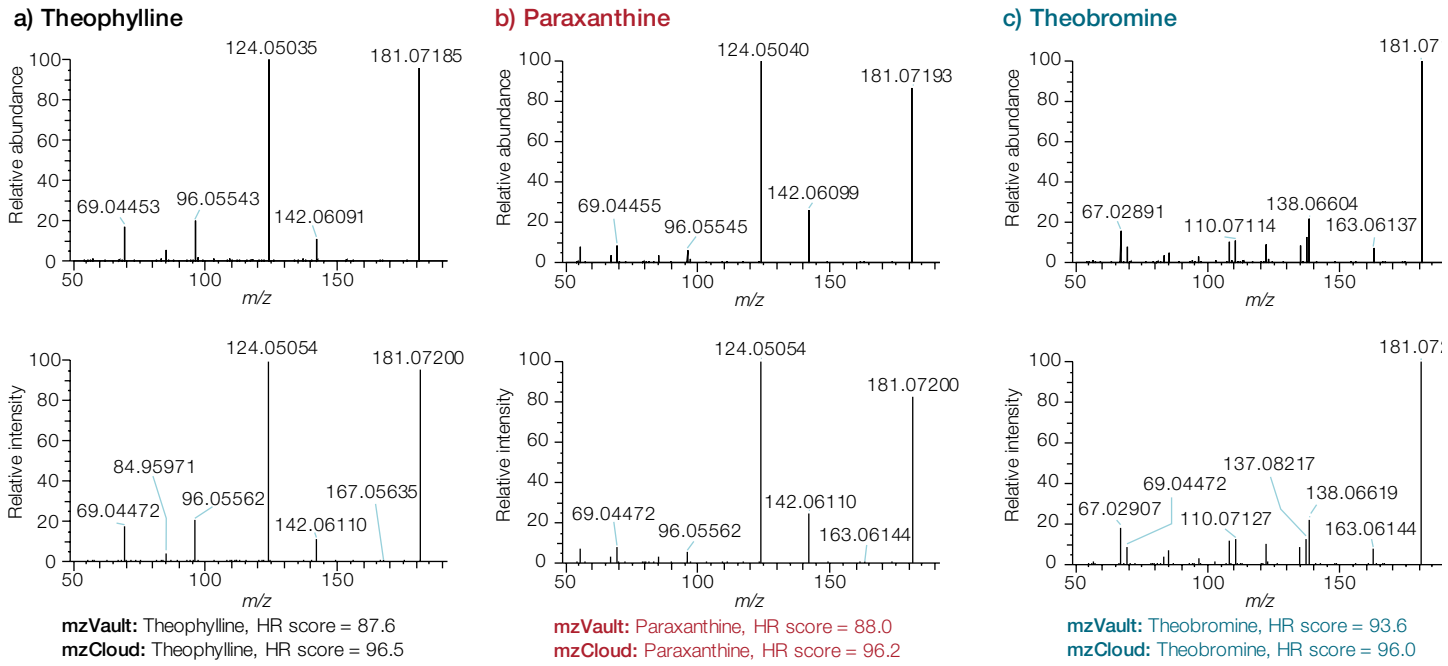


Figure 8. Library search results for MS² of *m/z* 181.07202; a) RT 5.23 min, theophylline spiked into plasma, b) RT 5.03 min, paraxanthine spiked into plasma, and c) RT 4.35 min, theobromine spiked into plasma

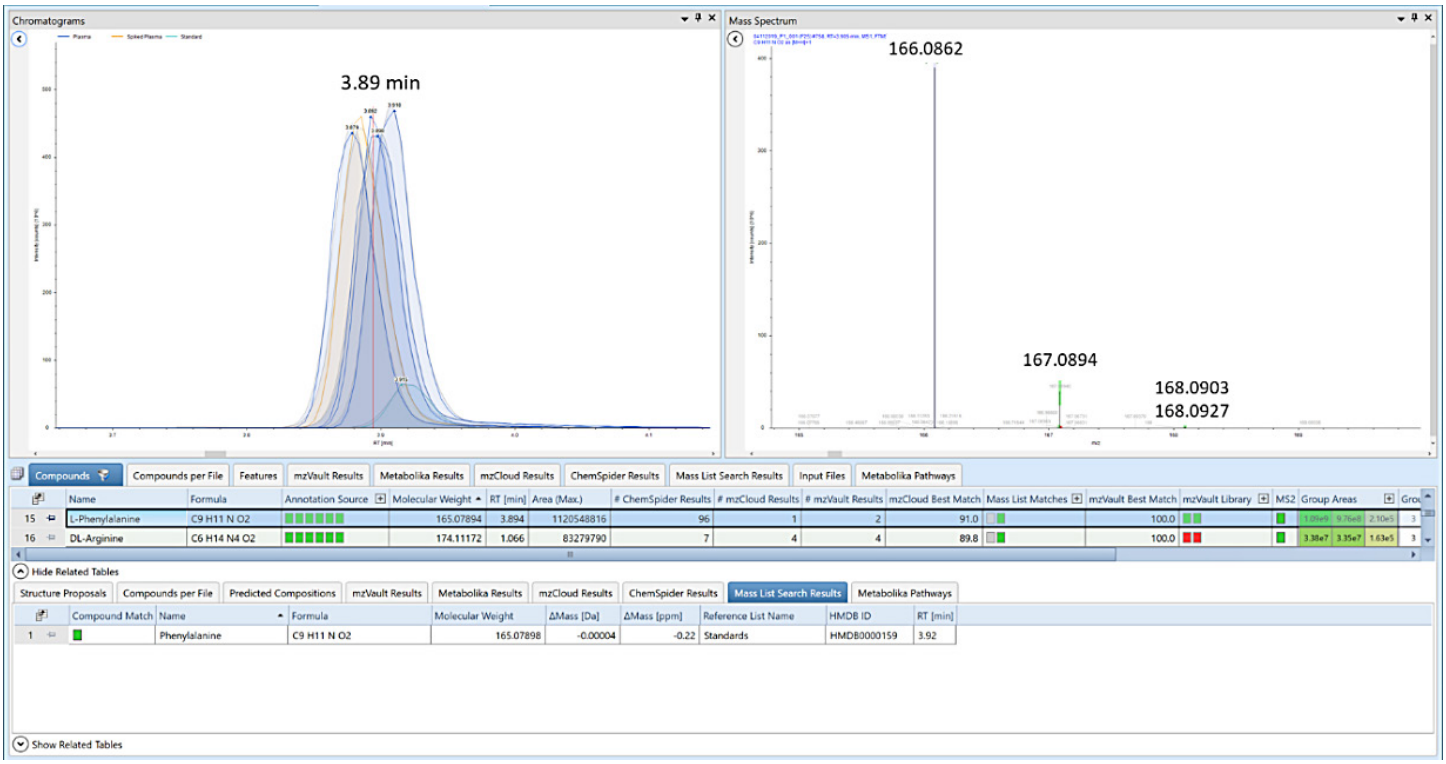


Figure 9a. Confirmation of $[M+H]^+$ ion, *m/z* 166.0862, retention time 3.89 minutes in SRM 1950 spiked plasma extracts with composition $C_9H_{11}NO_2$ matching D/L-phenylalanine in ChemSpider and Mass List databases.

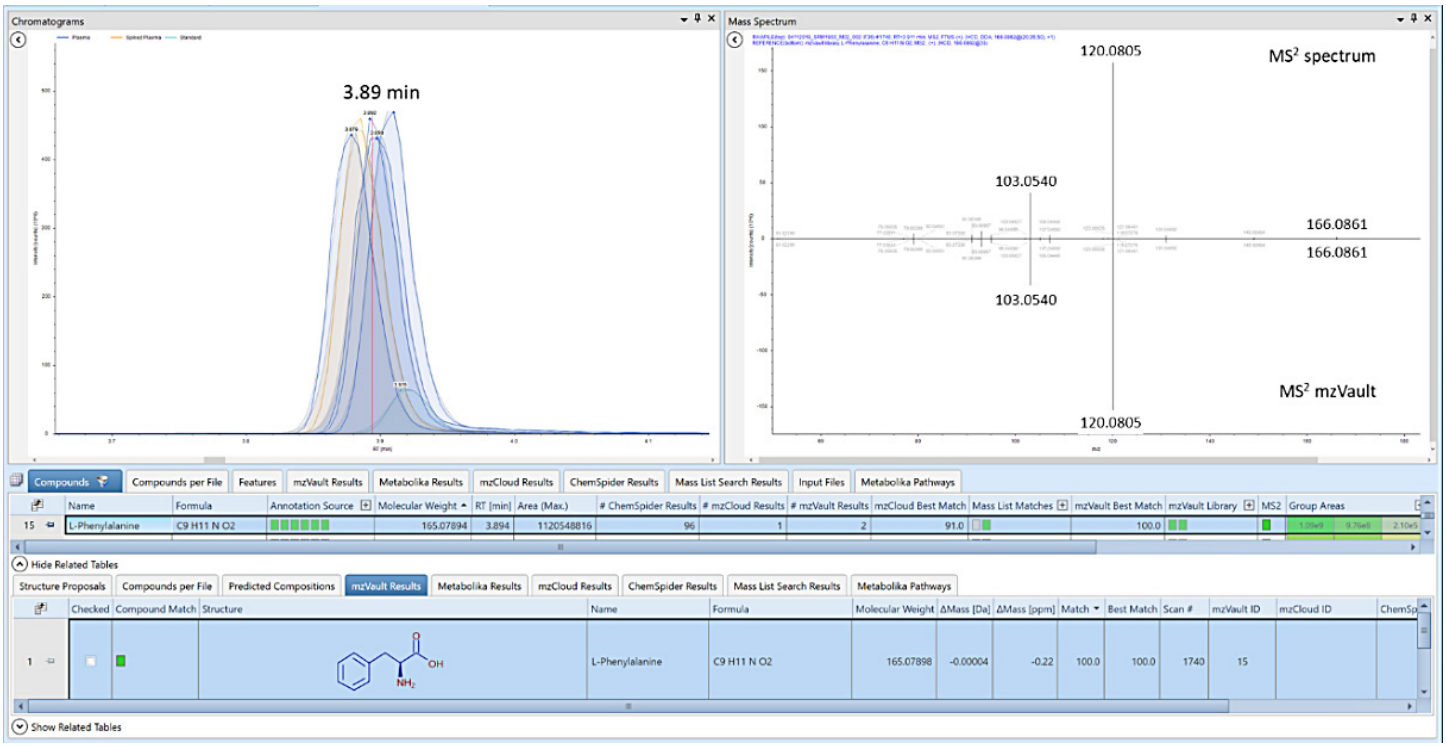


Figure 9b. Identification of phenylalanine *m/z* 166.0862 in SRM 1950 plasma extracts with the MS² spectrum matching D/L-phenylalanine with high-quality match scores (mzVault, 100.0; mzCloud, 91.0).

Sphingomyelin 16:0 identity confirmed in SRM 1950 plasma extracts

Figure 10a shows annotation of d18:1/16:0 sphingomyelin in SRM 1950 at 13.78 min and *m/z* 703.5753 (+0.6 ppm) matching the composition ($C_{39}H_{79}N_2O_6P$) of the reference standard from Mass List and ChemSpider databases. Figure 10b illustrates identification by matching fragmentation of the protonated molecular ion with the mzVault and mzCloud mass spectral libraries (match score of 100.0 and 90.8, respectively).

Compounds confirmed in both positive and negative ion results

Glutamic acid, hippuric acid and glycoursodeoxycholic acid were confirmed in both the positive and negative ion modes, adding to the high confidence in annotation and identification. The negative ion LC-MS result confirms the corresponding anion at 5.85 min and *m/z* 178.0509 ($C_9H_8NO_3$) as shown in Figure 11a. As shown in Figure 11b, positive ion LC-MS confirms the retention time (5.85 min) and *m/z* 180.0659 ($C_9H_{10}NO_3$) of protonated hippuric acid. The library search results for protonated hippuric acid gave a match score of 100.0 from mzVault, while the hippuric acid anion was identified with an mzVault match score of 89.7. Thus, combining results from positive and negative ion provides very high confidence in metabolite identification.

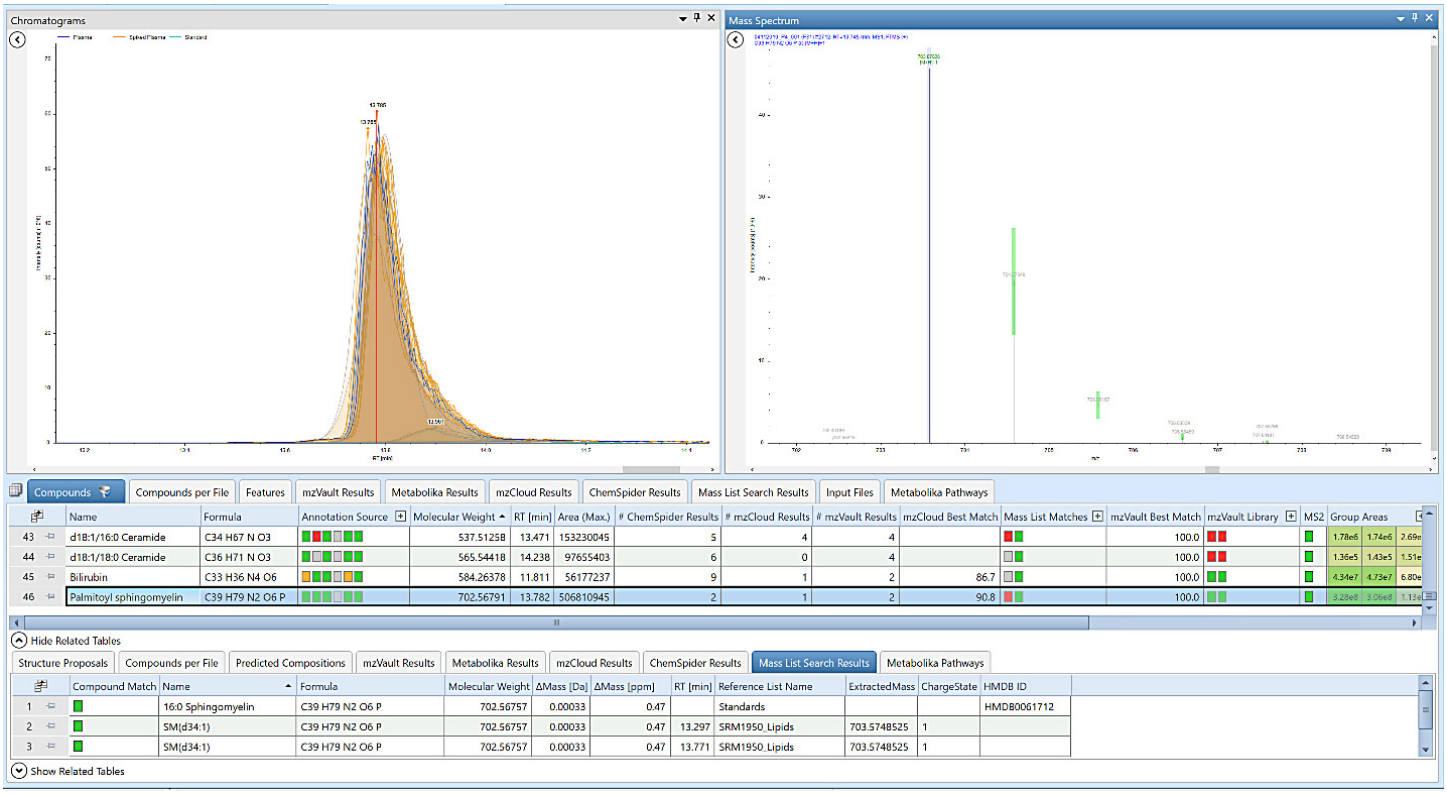


Figure 10a. Confirmation of $[M+H]^+$ ion, *m/z* 703.5753 ($C_{39}H_{79}N_2O_6P$) from SRM 1950 human plasma extracts at retention time 13.78 minutes matching d18:1/16:0 sphingomyelin in ChemSpider and Mass List databases.

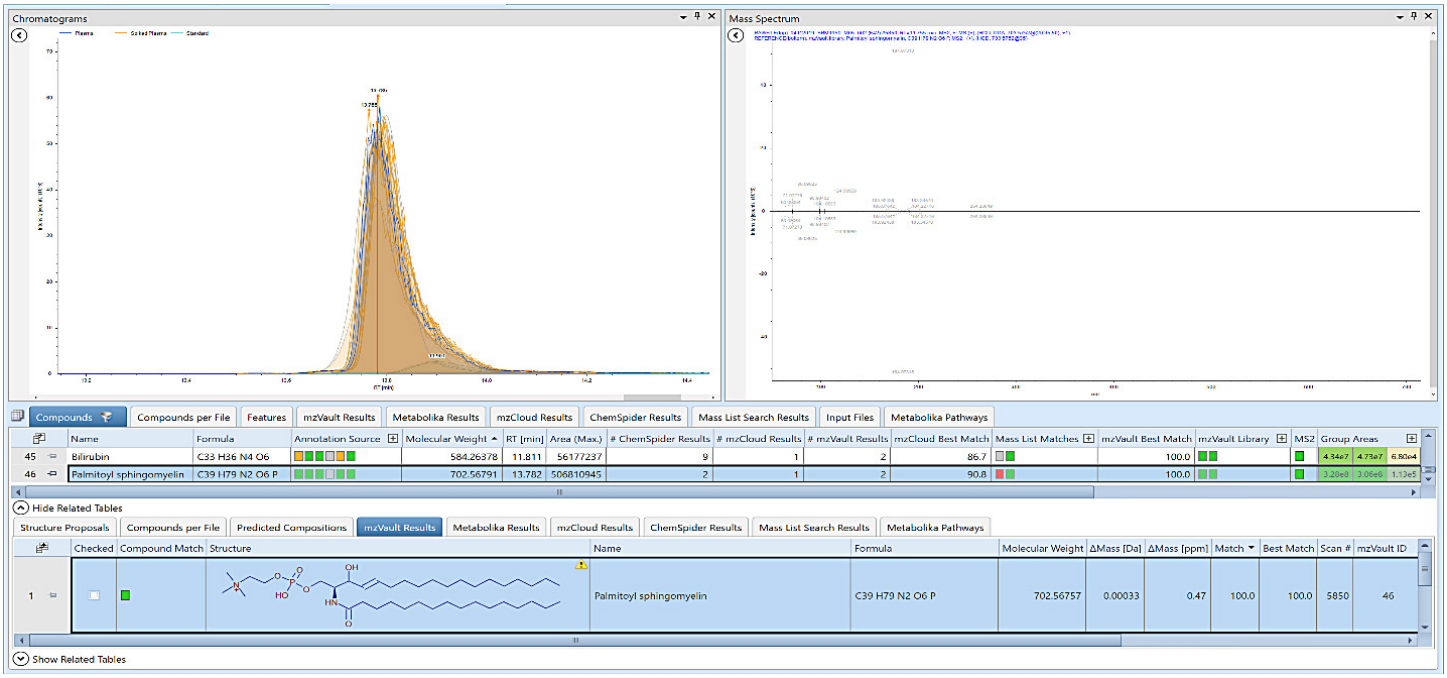


Figure 10b. Identification of *m/z* 703.5753 ($C_{39}H_{79}N_2O_6P$), retention time 13.78 min from SRM 1950 human plasma with a high quality MS^2 spectrum matching sphingomyelin d18:1/16:0 (mzVault, 88.7; mzCloud, 93.7).

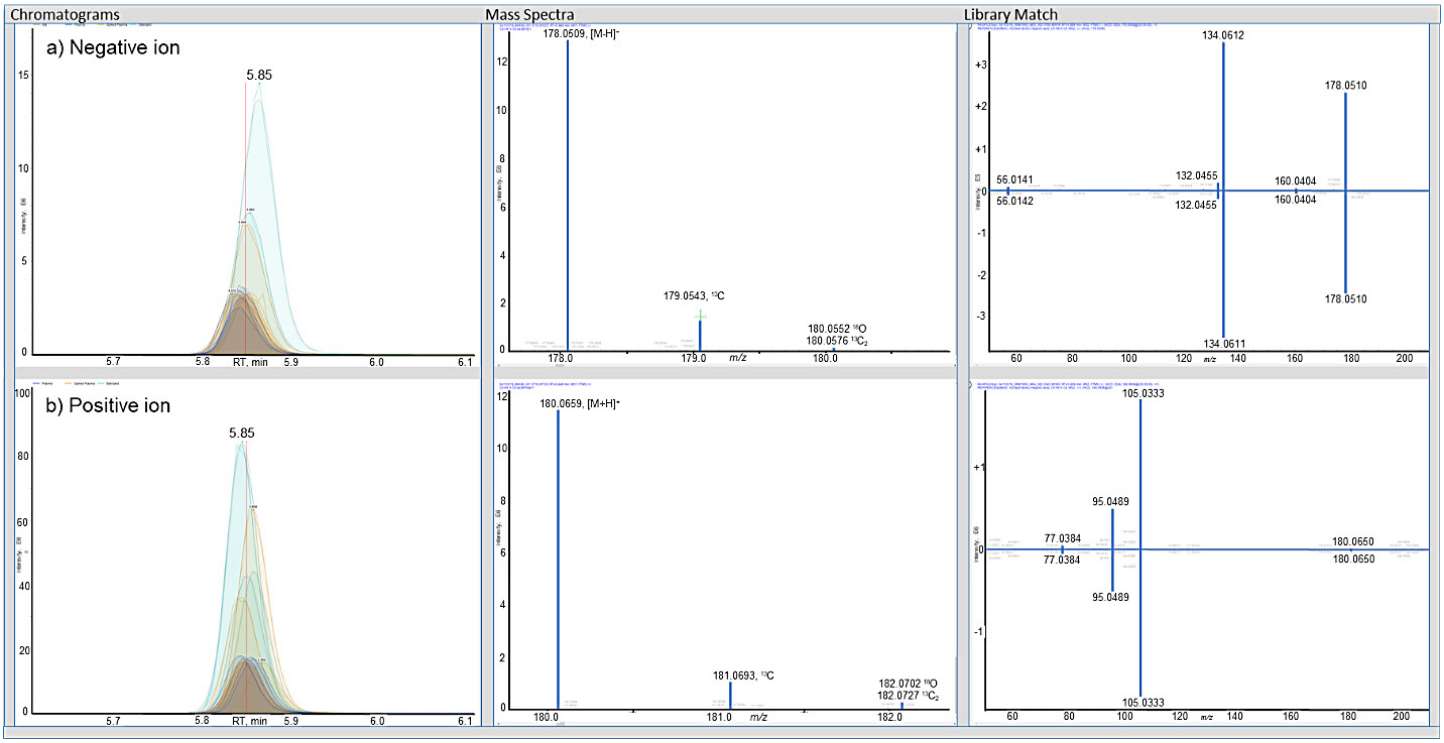


Figure 11. LC-MS positive and negative ion confirmation of hippuric acid in SRM 1950 human plasma extracts. a) Hippuric acid $[M-H]^-$ ion, *m/z* 178.0509, retention time 5.85 min, mzVault match score = 89.7 and b) Hippuric acid $[M+H]^+$ ion, *m/z* 180.0659, retention time 5.85 min, mzVault match score = 100.

Conclusions

- Untargeted annotation and confident identification were demonstrated using the Orbitrap ID-X Tribrid mass spectrometer and automated processing of SRM 1950 human plasma data.
- Compound Discoverer software supports annotation of untargeted metabolites and unequivocal identification of targeted metabolites in a single workflow.
- Ultra-high resolution LC-MS (120 K) and high-resolution data dependent MS^2 (30 K) spectra provide confirmation of targeted metabolite identity with retention time, elemental composition (including fine structure) and high quality matching of local mzVault and on-line mzCloud mass spectral libraries.
- Reference standards spiked into SRM 1950 confirm the identities of 56 out of 58 high-quality annotations, adding confidence to the automated annotations provided using this approach.

thermo scientific

References

1. D Peake, “High-resolution compound identification in metabolomics: a review of current practices” review paper 65356, PlanetOrbitrapA2299.
2. A Souza *et al.*, “Accelerated unknown compound annotation with confidence: from spectra to structure in untargeted metabolomics experiments” application note 65362, PlanetOrbitrapA2319.
3. L W Sumner *et al.*, “Proposed minimum reporting standards for chemical analysis”, *Metabolomics*, 2007, 3, 231–241. DOI: 10.1007/s11306-007-0082-2.
4. I Ntai *et al.*, “Advantage of high resolution accurate mass spectrometry for metabolite identification in untargeted metabolomics studies” scientific poster, PlanetOrbitrapA2210.
5. D Peake *et al.*, “Increased confidence of insect lipidome annotation from high-resolution Orbitrap LC/MS^n analysis and LipidSearch software” application note 72942, PlanetOrbitrapA2365.
6. T Talamantes *et al.*, “Intelligent acquisition for comprehensive metabolome coverage in plants, mammals and bacteria” scientific poster, PlanetOrbitrapA2347.
7. S Hackbusch *et al.*, “Improved lipid annotation depth using automatically generated inclusion and exclusion lists on an Orbitrap-based mass spectrometer” scientific poster PlanetOrbitrapA2343.
8. I Ntai *et al.*, “Improved metabolome coverage and increased confidence in unknown identification through novel automated acquisition strategy combining sequential injections and MS^n” scientific poster, PlanetOrbitrapA2262.
9. “Grant application resource: Thermo Scientific Orbitrap ID-X Tribrid mass spectrometer for metabolomics, lipidomics and structural elucidation” whitepaper 65364, PlanetOrbitrapA2315.
10. T Stratton *et al.*, “Improved ranking of putative candidates through a hybrid in silico/real fragmentation technique” scientific poster, PlanetOrbitrapA2260.
11. I Ntai *et al.*, “High-throughput metabolite profiling of cell media for improved antibody production utilizing a dual separation/mass spectrometry system with intelligent MS^n acquisition” scientific poster, PlanetOrbitrapA2337.
12. D Peake *et al.*, “A complete workflow for improved untargeted metabolome annotation and identification using ultra high-resolution accurate mass and LC-MS^n Orbitrap-based mass spectrometry” scientific poster, PlanetOrbitrapA2355.

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

ThermoFisher SCIENTIFIC