Poster Note 64746

Unknown Profiling of Drinking Water Using High Resolution LC-MS/MS and New Software

Caroline Ding[1], Anastasia Kalli[1], Charles Yang[1], Tim Stratton[1], Hans Grensemann[2]
[1]Thermo Fisher Scientific, San Jose, CA, USA; [2]Thermo Fisher Scientific, Bremen, Germany

ABSTRACT

Purpose: Unknown compound profiling using high resolution LC-MS/MS and new software to confidently and quickly identify unknown compounds.

Methods: A treated tap water sample was collected from a city in China and stored in a plastic bottle before analysis. Mobile phase (5mM ammonium formate with 0.1% formic acid in water) was used as a blank to generate a data-dependent exclusion list for acquisition and background removal during data processing. LC-MS analysis was performed on the blank and water samples in positive modes with two replicate injections. The MS analysis employed a 70k HRMS full scan followed by top 10 data dependent ms2 collected on a Q Exactive mass spectrometer. Data analysis was performed with Thermo Scientific™ Compound Discoverer™ software using a single unknown data processing workflow.

Results: Data was processed using Compound Discoverer with one single workflow. The processing workflow included automatic unknown component detection, unknown elemental composition, library searching against mzCloud™ HRAM fragmentation library, ChemSpider database search, mass list search against built-in HRAM EFS library, automatic blank removal and structure interpretation using Custom Explanations and FISh Scoring on the fly. Batch searching against mzCloud online fragmentation library proved to be the most productive and confident way for unknown compound identifications. ChemSpider database search provided more hits which complements mzCloud search, however there were too many false positives from ChemSpider search. ChemSpider search using predicted formula helped reduce the number of false positives. Built-in FISh in Custom Explanations was used to verify hits from ChemSpider against MS2 data. All in all, Compound Discoverer software provides an effective and complete workflow for unknown identifications.

INTRODUCTION

Unknown compound profiling of water sample is very challenging due to complexity of contaminants in the water sample. Multiple software and lots of manual interpretations are usually required to identify the unknown compounds. New emerging software and tools shed light on unknown compound identifications. mzCloud is a new online HRAM fragmentation library which contains highly curated MS/MS and MSn spectra from different collision types and collision energies. It provides the fastest and most confident small molecule unknown compound ID. mzCloud search is integrated into Compound Discoverer 2.0 along with other tools like predicted compositions based on high resolution full ms, ChemSpider search that help partially identify the unknowns. This study demonstrates a simple yet powerful workflow for unknown compound profiling using high resolution Thermo Scientific™ Orbitrap™ mass spectrometer and Compound Discoverer software.

MATERIALS AND METHODS

Sample Preparation

Untreated tap water and mobile phase blank (injection volume 50ul) were directly injected on to the column for chromatographic separation and MS analysis.

Analytical Method for LC-MS analysis

Chromatographic separation was performed with a Thermo Scientific™ Dionex™ Ultimate™ 3000 RS LC system using a Thermo Scientific™ Accucore AQ column (100 × 2.1 mm, 2.6 µ particle size). Mobile phase A was 5 mM ammonium formate and 0.1 % F.A in water. Mobile phase B was 5 mM ammonium formate and 0.1 % F.A in MeOH. Mass spectrometric analysis was performed on a Thermo Scientific™ Q Exactive™ Plus Orbitrap mass spectrometer operated in full MS-ddMS2 mode. Analysis was performed in positive ion mode followed by top10 data-dependent MS/MS scans. Resolution for the full MS scan was set at a 70, 000 and at 17500 for the ddMS2 scans. Stepped collision energy was used at 20, 40 and 60. Analysis time, including column equilibration, was 25 min.

Retention time (min)	Flow (ml/min)	%B
0.00	0.300	2.0
0.50	0.300	2.0
2.00	0.300	40.0
20.00	0.300	95.0
22.00	0.300	95.0
22.10	0.300	2.0
25.00	0.300	2.0

Data Analysis

Samples and blanks were grouped based on user defined study factor and processed together. In this study the grouping was based on defined water type "TapWater" (see Figure 1). Sample grouping was persisted into data processing and results display.

The HRAM data was processed by Compound Discoverer software using a single processing workflow (Figure 2). The workflow employed unknown compound detection followed by online ChemSpider, mzCloud™ database search and local EFS HRAM compound database search. The Mark Background Compounds node hides the background compounds in the blank files from the result table.

Figure 1. Sample grouping

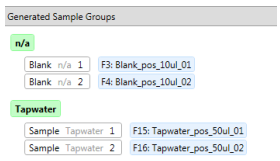
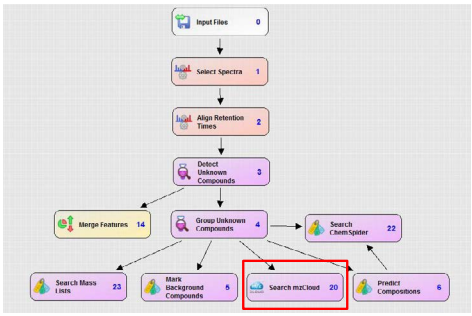


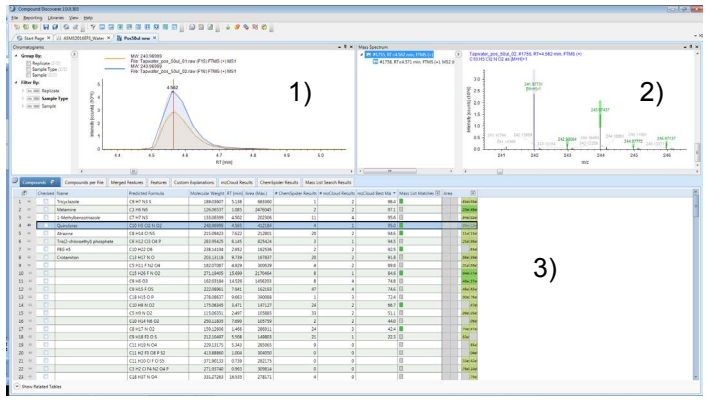
Figure 2. Workflow tree in Compound Discoverer software



RESULTS

The results review in Compound Discoverer is broken into three parts: 1) Chromatogram view which interacts with the result table; 2) Mass Spectrum view which also interacts with the result table and displays the spectral tree for selected compound; 3) Result tables: the most important table is the Compounds table on the far left (See Figure 3). All the views can be docked, repositioned or dragged onto a second monitor.

Figure 3. Result View in Compound Discoverer



Thermo SCIENTIFIC

Result Filtering

185 compounds with unique molecular weight and retention times above 1e6 peak intensity were detected from the positive mode data by Compound Discoverer excluding compounds found in the blanks. Without blank removal, the number of detected MW and RT was 711. Result filters were used to filter out compounds from the table based on user defined conditions, i.e. area threshold (see Figure 4).

Unknown ID with mzCloud

Compound Discoverer 2.0 includes batch compound ID against mzCloud online HRAM fragmentation library which contains high quality curated MS/MS and MSn spectra. The search algorithm allows match with ion activation energy with user adjustable ion activation energy tolerance window (Figure 5). The sophistication in the search algorithm increases the confidence in the identifications for small molecules where fragmentation pattern changes with ion activation energies.

mzCloud hits are indicated in the Compounds table with number of hits and best match scores (Figure 6). For each hit, the spectra comparison between the query spectrum and library match spectrum is visualized in a mirrored plot (Figure 7).

Figure 4. Result Filters

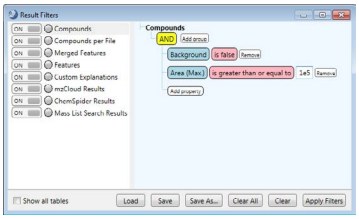


Figure 5. mzCloud node settings

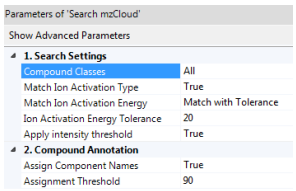


Figure 6. mzCloud hits in Compound Discoverer with match scores

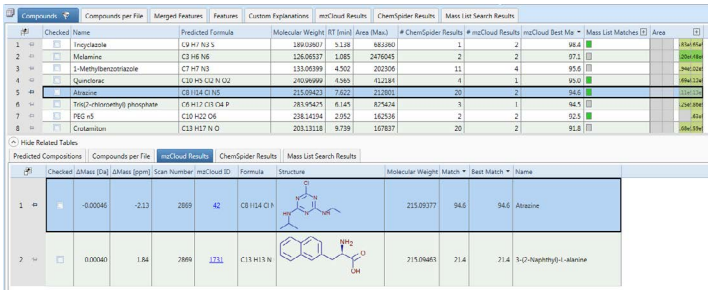
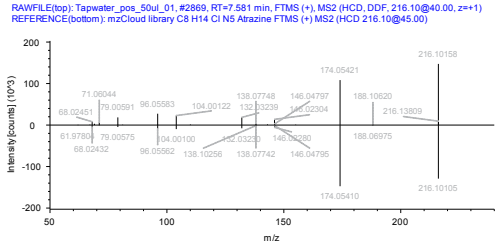


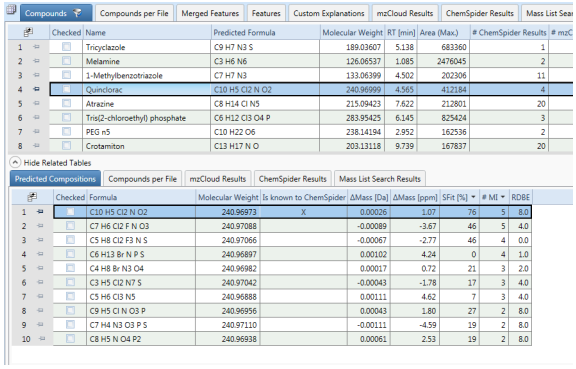
Figure 7. mzCloud hit spectral comparison between query and library spectra



Unknown Compound Formula Prediction

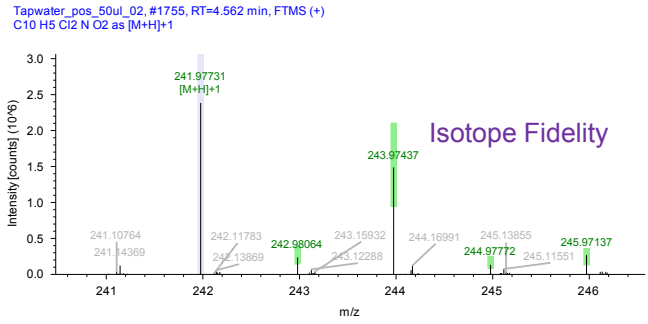
Formula predictions by the Predict Composition node are listed in the sub table for each compound. The one with the best SFit% and most number of matching isotopes (#MI) is listed on the top with information like delta mass ppm. (See Figure 8)

Figure 8. Predicted Compositions for each unknown compounds listed in the sub table.



Spectral fit is visualized for each composition prediction in the spectrum window. (See Figure 9).

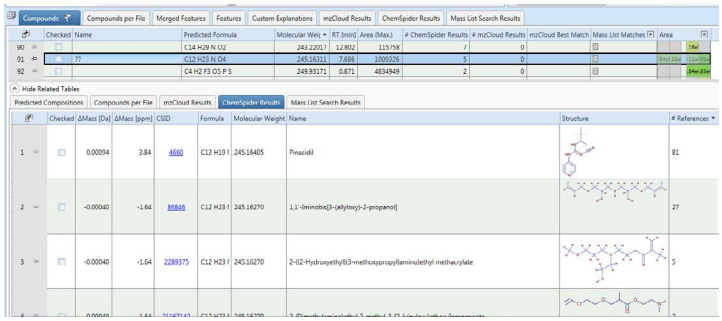
Figure 9. Spectral fit for predicted composition C10H5Cl2NO2 based on resolution



ChemSpider Search and Custom Explanations

For the compounds that did not have match from mzCloud, ChemSpider hits were reviewed. ChemSpider search was performed using predicted formulas. If formula was not available, then accurate mass was used. The databases used were ACToR: Aggregated Computational Toxicology Resource; DrugBank; EAWAG Biocatalysis/Biodegradation Database; EPA DSSTox; FDA UNII – NLM. ChemSpider hits for each compounds are listed in the sub table in the order of # of references (See Figure 10).

Figure 10. ChemSpider hits for each compound listed in the sub table

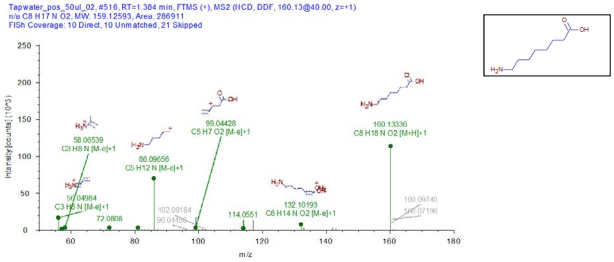


The problem is how do we know if any of the ChemSpider hits is the right answer? FISh Scoring in Custom Explanations in Compound Discoverer was used to verify compound ID against MS/MS data. User proposes a structure for the compound of interest, runs FISh Scoring on the fly based on the proposed structure. FISh coverage score is calculated and fragment structures are automatically annotated. Figure 11 shows an example of FISh annotations on an unknown compound based on ChemSpider proposal.

Scan this QR code on your **mzCloud app** to find out what this compound is!

Download the mzCloud app from your App Store on iphone or androids.

Figure 11. FISh Scoring based on proposed structure for unknown ID



Identified Compounds

From the 185 compounds detected in the water sample, 16 of them were identified by mzCloud automatically (see Table 1). The most dominant identified compound is Melamine. Its peak intensity is about 2e7. Others include drugs, pesticides, herbicides and etc. The ones with match score > 90 are very confident identifications based on MS/MS spectrum match and collision energy match. The other unknown compounds without mzCloud hits were much more difficult to identify. A strategy of combing ChemSpider hits, delta ppm, predicted compositions, custom explanations and FISh Scoring were used to try to identify these unknowns. However, the false positive ID rate was very high.

Table 1. Unknown compounds identified by mzCloud

#	Name	Formula	Molecular Weight	RT [min]	Delta ppm	Area (Max.)	# ChemSpider Results	# mzCloud Results	mzCloud Match Score
1	Melamine	C3 H6 N6	126.06537	1.09	0.17	2476045	2	2	97
2	Dextromethorphan	C18H25NO	271.19405	15.70	-1.62	2170464	8	1	85
3	hydroxycoumarin	C9 H6 O3	162.03184	14.53	-0.89	1456203	8	4	75
4	Tris(2-chloroethyl) phosphate	C6 H12Cl3O4P	283.95425	6.15	-1.33	825424	3	1	95
5	Tricyclazole	C9 H7 N3 S	189.03607	5.14	-0.03	683360	1	2	98
6	Quinclorac	C10 H5 Cl2 N O2	240.96999	4.57	-1.07	412184	4	1	95
7	tri-phenylphophine oxide	C18 H15 O P	278.08637	9.66	-1.14	390088	2	3	72
8	Triethyl Phosphate	C6 H15 O4 P	182.07087	4.93	-0.41	300639	4	2	90
9	8-Aminooctanoic acid	C8 H17 N O2	159.12606	1.47	0.83	286911	24	3	42
10	Atrazine	C8 H14 Cl N5	215.09423	7.62	-2.13	212801	1	2	95
11	1-Methylbenzotriazole	C7 H7 N3	133.06399	4.50	0.06	202306	11	4	96
12	Crotamiton	C13 H17 N O	203.13118	9.74	-0.82	167837	20	2	92
13	PEG n5	C10 H22 O6	238.14194	2.95	-1.26	162536	1	2	93
14	phthalate	C12 H14 O4	222.08961	7.94	-1.79	162193	45	4	75
15	Indole-3-acetic acid	C10 H9 N O2	175.06345	3.47	-0.68	147127	24	2	67
16	Proline	C5 H9 N O2	115.06351	2.50	-1.53	105885	33	2	51

CONCLUSIONS

- Compound Discoverer 2.0 provides a single software solution for HRAM data processing and confident unknown compound identifications
- Unknown compound ID via batch search against mzCloud online HRAM fragmentation library proved to be the most productive and confident way for unknown compound identifications
- ChemSpider search combined with calculated formula from high resolution Orbitrap data complements mzCloud search but has too many false positives
- Structure elucidation using Custom Explanations and FISh Scoring in Compound Discoverer was handy and a nice way to verify ChemSpider hits against MS/MS data
- Quantitation of unknown contaminants is not the focus of this study. However results can be exported from Compound Discoverer to software like TraceFinder for absolute quantitation.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Brazil +55 11 2730 3006
Canada +1 800 530 8447
China 800 810 5118 (free call domestic) 400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 6 6885 1213
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
Taiwan +886 2 8751 6655
UK/Ireland +44 1442 233555
USA +1 800 532 4752

PN64746-EN 0716S

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand