Metabolomics

Orbitrap Tribrid Real-Time similarity searching of mzCloud library for guided unknown characterization of tea metabolites

Bills, Brandon[1]; Yedla, Sunandini[1]; Deshpande, Rahul[1]; Amer, Bashar[1]; Melani, Rafael[1]; Bird, Susan[1], [1]Thermo Fisher Scientific, 355 River Oaks Parkway, San Jose CA, USA, 95134

Abstract

Purpose: To automatically trigger ddMS^n data collection on unknown tea metabolites in real time based on similarities to the closest similarity library entry

Methods: Real-Time Library Search as part of a data dependent MS^3 experiment

Results: In real time the instrument found the closest library match for eluting features and triggered MS^3 on matching and non-matching fragments. This MS^3 data was then used to narrow down potential structure annotations for the unknowns through fragment ion search or FISh scoring.

Introduction

In metabolomics LC-MS data is processed to generate a list of detected compounds with a level of confidence in the identity of each compound based on reference data. This reference data may be spectral data and/or retention time information stored in a library or in silico information generated based on fragmentation rules. These annotated compounds are then used to evaluate biological changes between normal and disease states. One of the major hurdles between collected data and coming to reliable conclusions about the biology is the limited number of standards available to build these libraries. It is common to have hundreds to thousands of compounds detected in a data set that have no library match because there is no commercial standard for that unknown. Instead, there is just an elemental formula and matches from a structure database. These structures can be compared using in silico fragmentation techniques to determine which isomers plausibly match the data. MS^n data can help reduce the list of possible matches, but without guidance, ddMS^3 may not consistently collect data on the most relevant MS^2 fragments to accomplish this.

In this work we use Real-Time Library Search (RTLS) in a ddMS^3 method to help guide the instrument in real time to collect the most relevant data for confirming structural matches.

Materials and methods

Sample preparation

Tea samples were extracted both in boiling water for 3 minutes and with a mixture of 1:1 methanol water with 0.1% formic acid. Extracts were centrifuged prior to analysis.

Test method(s)

1 µL aliquots were analyzed using a Thermo Scientific™ Hypersil GOLD™ column on a Thermo Scientific™ Vanquish™ Horizon UHPLC coupled with a Thermo Scientific™ Orbitrap™ IQ-X™ Tribrid™ mass spectrometer and Thermo Scientific™ Orbitrap™ Tribrid™ Apex Small Molecule mass spectrometer. RTLS was run in similarity mode using the offline version of Thermo Scientific™ mzCloud™ library.

Data analysis

Data was processed using Thermo Scientific™ Freestyle™ and Compound Discoverer™ software

Figure 1. Thermo Scientific Orbitrap Tribrid Apex Small Molecule Mass Spectrometer

Discussion

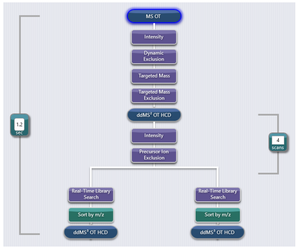
Real-Time Library Search

RTLS is a filter within the method editor that allows the instrument to compare experimental spectra to library spectra and generate similarity scores in real time. To learn more about how the filter works, please scan the QR code shown in figure 2.

Figure 2. Link to video on RTLS

In this work the instrument was using the offline version of the mzCloud library running a similarity type experiment. During this similarity experiment the instrument found the entry with the closest MS^2 spectrum to the experimental spectrum (regardless of whether the precursor m/z matches) and triggered MS^3 collection on both the highest m/z matched fragment and the highest m/z unmatched fragment as shown in figure 3. A second version of this method was also tested where MS^3 were collected in the ion trap.

Figure 3. RTLS Method



By targeting both matched and unmatched fragments from the spectrum the instrument was guided towards the highest m/z fragment that was shared with the closest match in the library to confirm substructure similarity and the highest unique fragment to investigate the unknown portion of the structure.

The mzCloud version used in this experiment had over 20,000 entries with over 2.5 million spectra. To allow a similarity experiment on such a large library, a setting called “Fragment Ion Index” was used. This pre-indexes the library so that scoring only takes place if out of the top 100 fragments there are at least 4 matching fragments with at least 5% relative abundance in the library. This reduces time spent generating scores between dissimilar spectra.

Triggering on a compound

During the analysis of tea extract there was an unknown at m/z 217.1183 with no matching result in the library. However, there was a compound with a similarity score of 59 that the system determined as the closest matching library entry. These entries were displayed in the spreadsheet that was generated by the instrument because RTLS was part of the method. A partial version of the spreadsheet is shown in figure 4.

Figure 4. Entry in the RTLS spreadsheet for the closest matching compound for the unknown at m/z 217.1183

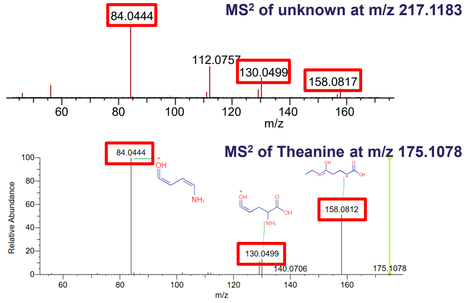
Isolation m/z	Formula	Compound Name	Compound Class	Search Time (ms)	Cosine Score
217.1183	$C_7H_{14}N_2O_3$	L-Theanine	Endogenous Metabolites	25	59

Results

Comparing spectra

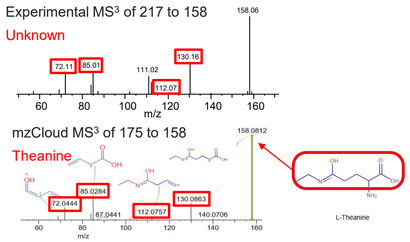
The spectra between theanine (m/z 175.1078) and the unknown with m/z at 217.1183 show numerous shared fragments as shown in figure 5:

Figure 5. Experimental HCD spectra of unknown 217.1183 (top) and annotated mzCloud spectra for theanine (bottom). Fragments outlined in red are in both spectra



This suggests that the structure of the unknown is potentially similar to that of theanine. Because the method was set to acquire the MS^3 on the highest m/z MS^2 fragment that matched between the two spectra, that similarity can be investigated by comparing the MS^3 spectra of the fragment at m/z 158 between the experiment and the mzCloud online library (figure 6).

Figure 6. Experimental HCD MS^3 spectra of unknown 217.1183 to 158.0817 transition (top) and annotated mzCloud MS^3 spectra for theanine (bottom). Fragments outlined in red are in both spectra. The substructure circled in red is the annotation for m/z 158.0812 from theanine.

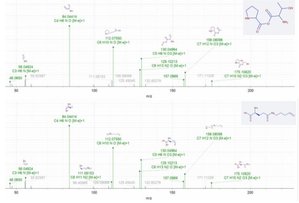


The matching MS^3 spectra further corroborates the structural similarity between Theanine and the unknown.

Processing results

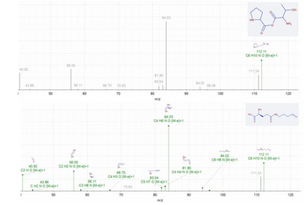
Data was processed in Compound Discoverer software and the unknowns with no library match were compared to the structure database ChemSpider. FISh scoring, a comparison of an in-silico fragmentation spectra, was used to evaluate matches based on MS^2 data (figure 7) and MS^3 data (figure 8)

Figure 7. FISh scoring at the MS^2 level for two potential structures for the unknown at m/z 217.1183



Using only the MS^2 spectra leads to an ambiguous annotation between the two structures shown. However, FISh scoring uses MS^3 as well as MS^2 spectra and RTLS was set to trigger on the highest non-matching m/z fragment between the experimental spectrum and the library spectrum. This gives additional information on the unique part of the structure. In this case it was m/z 112.0757. Figure 8 shows the FISh annotation of this spectra for the two putative structures.

Figure 8. FISh scoring at the MS^3 level for two potential structures for the unknown at m/z 217.1183



Conclusions

Tea extracts were analyzed using RTLS to find the most similar compound in the offline version of the mzCloud library and trigger MS^3 on the highest m/z matching and non-matching fragments. This provided additional information to confirm similarity and provide insight on shared sub-structures to the library compound as well as additional information on the portions of the structure that were unique to the experimental compound.

Trademarks/licensing

Conflict of Interest: The authors BB, SY, RD, BA, SB, and RM are employees of Thermo Fisher Scientific whose instrumentation and software were used to acquire and process the data. PO004849-2026-EN