Poster Note 64467

Creating High Quality Metabolite Libraries for Fast Metabolomics Screening and Identification

Gina Tan, Junhua Wang, Yingying Huang, Andreas FR Hühmer1
Thermo Fisher Scientific, San Jose, CA, USA

Overview

Purpose: To create a comprehensive and reliable metabolites screening library of 300 commercially available metabolite standards, containing retention time and high quality MS^2 spectra information.

Methods: 300 metabolite standards were prepared in mixtures and analyzed on a Thermo Scientific Q Exactive™ mass spectrometer coupled to an Ultimate 3000 UHPLC system. Full MS and full MS - data dependent MS/MS analysis in both positive and negative polarities was performed to obtain retention time information and MS^2 spectra of each individual compound. All data were processed using the Thermo Scientific TraceFinder 3.2 software. The compound database, MS^2 spectra library and screening methods were also created with the TraceFinder software. (Figure 1.)

Results: A compound database and MS^2 spectral library of 300 endogenous metabolites was created and incorporated into a metabolite screening workflow. A ZDF rat plasma dataset was screened against the library and a total of 85 metabolites were confidently identified based on the accurate mass, retention time and MS^2 spectra.

Introduction

The study of endogenous metabolites has brought about new possibilities in the field of metabolomics, especially in the area of life sciences research. Compared to genomics and proteomics, identification of endogenous metabolites continues to pose great limitations and challenges to many researchers. This is due to the absence of a reliable screening library to provide accurate information and high quality data needed for metabolic profiling studies. Multiple online databases are available for metabolite identification, but often provide numerous candidates that require further analysis to remove redundancy. Most of the early metabolite libraries were derived from gas-chromatography mass spectrometry (GC-MS) based methods. Application of liquid-chromatography mass spectrometry (LC-MS) in the field of metabolomics research has grown exponentially in recent years and has become the method of choice. The number metabolite entries in LC-MS based metabolite libraries however remains limited. In order to overcome the obstacles in metabolite identification, we present a compound spectral library of endogenous metabolites. This library contains a repository of accurate masses, retention times and LC-MS^2 spectra information, which are acceptance criteria that can be used to improve the confidence in metabolite identification.

Methods

Sample Preparation

300 commercially available metabolite standards were used for this experiment. 20 to 25 standards were combined into a single batch and a total of 15 batches were prepared. Standards were individually weighed at 1mg each and gave a final concentration of 0.5mg/ml after constitution with 50/50 methanol/water. The standard mixtures for all 15 batches were put through sonication followed by a filtration step to remove any undissolved salt particles.

Liquid Chromatography

The liquid chromatographic separation was done on the Thermo Scientific™ Dionex™ UltiMate™ 3000 RS system . A Hypersil™ GOLD C18, 150 X 2.1, 1.9µm reverse-phase column was used for the separation of the metabolite analytes at a flow rate of 450 µl/min.

TABLE 1. UHPLC Conditions.

Time	%A	%B
0.0	99.5	0.5
5.5	50.0	50.0
6.0	2.0	98.0
12.0	2.0	98.0
13.0	99.5	0.5
15.0	99.5	0.5

A:0.1% Formic acid

Mass Spectrometry

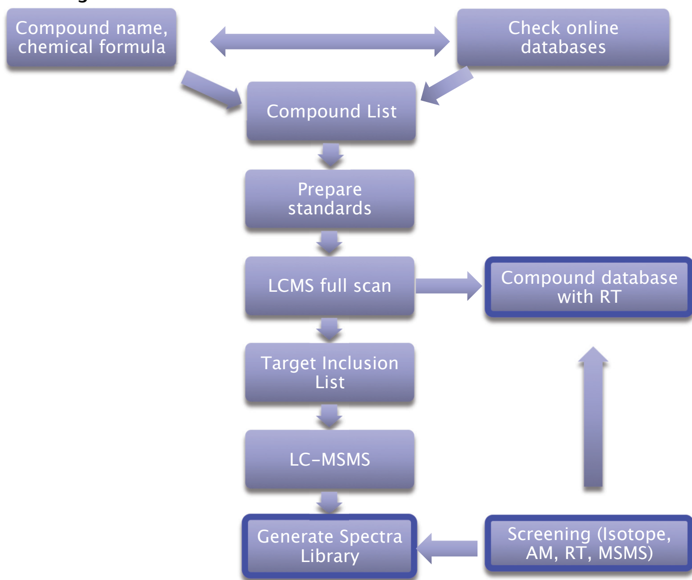
All samples were analyzed on a Thermo Scientific Q Exactive™. Two acquisition modes were used. A full MS scan at 35,000 resolution with positive and negative switching. The other acquisition mode was a full MS at 70,000 resolution followed by data-dependent MS^2 in both positive and negative polarities. An inclusion list for all the respective batches was included in the full MS data-dependent MS^2 method. All MS^2 spectra of the compounds were acquired at three fixed collision energies, 10, 30 and 45.

Data Analysis

All data were was processed using TraceFinder 3.2. A compound database was created that records compound information, chemical formulas and retention times. A screening method for processing the data was also created in TraceFinder 3.2. Raw files were imported and individual MS^2 spectra of the compounds were extracted using Library Manager 2.0 (part of TraceFinder 3.2) to generate the MS^2 spectra library.

Thermo
SCIENTIFIC

FIGURE 1. Overview of experimental workflow from sample preparation, data acquisition, compound and spectra library creation for metabolites screening



Results

Generation of The Metabolites List

A preliminary list of endogenous, non-lipid metabolites was compiled from important metabolic pathways, which include the catabolic metabolism pathway, glycolysis and citric acid (Krebs') cycle to name a few. The shortlisted compounds were validated against online databases to ensure their endogenous nature. Final choice on the metabolites list was also dependent on the commercial availability of the metabolite standards.

Compound Database Creation With Inclusion of Retention Time Information

From the first set of full MS scans acquired, the retention time of each compound was recorded based on the standardized reverse-phase chromatographic method used for this experiment. All the retention time information was combined with the related chemical information and used to create the endogenous metabolite compound database within TraceFinder (Figure 2.). The compound repository includes the 300 metabolite compound names, chemical formulas, positive and negative m/z values, as well as CAS IDs. This database serves as a good starting point for metabolite profiling experiments to screen for possible metabolites that could be present. Samples will be screened against this compound database and the likely metabolite candidates will be identified, first and foremost by the monoisotopic mass.

The retention time is a secondary criteria taken into consideration in order to provide additional confirmation for a metabolite's identity in the sample. The use of retention time selection criteria removes the redundancy often associated with results obtained from larger databases. Greater confidence is achieved as the number of likely and relevant candidates are narrowed down.

The retention time information of each metabolite was also subsequently included in the target inclusion list to be used for the full MS data-dependent MS^2 acquisition method.

FIGURE 2. Compound database containing information of compound names, chemical formulas, CAS IDs, m/z values and retention times in the table grid view and the detailed view

Compound Database - TF_QEcombinedRP_01

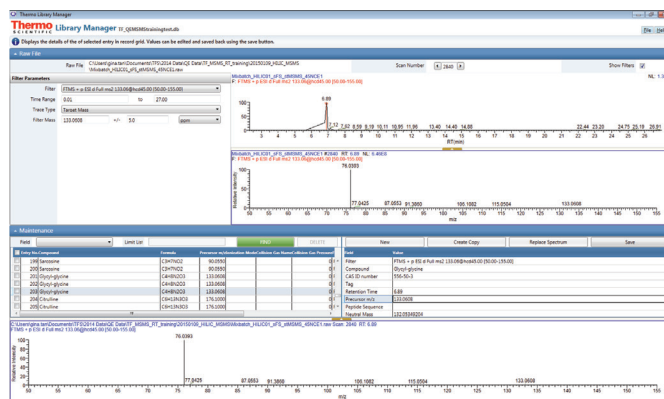
	Compound Name	Chemical Formula	Experiment Type	Ionization	CAS	Neutral Mass	Peak 1 Extracted M.	Peak 1 MS Order	Peak 1 Polarity	Peak 1 Adduct	Peak 1 Charge State	Peak 1
1	2-Phenylglycine	C8H9NO2	XIC	None	341659	151.06333	152.0706	ms1	Positive	Hydrogen	1	30.00
2	Diethylphosphate	C4H11O4P	XIC	None	598-02-7	154.0395	155.0468	ms1	Positive	Hydrogen	1	30.00
3	Azelaic acid	C9H16O4	XIC	None	123-99-9	188.10486	189.1121	ms1	Positive	Hydrogen	1	30.00
4	Theophylline	C7H8N4O2	XIC	None	58-55-9	180.06473	181.072	ms1	Positive	Hydrogen	1	30.00
5	D-(+)-Glucose	C6H12O6	XIC	None	492-62-6	180.06339	181.0707	ms1	Positive	Hydrogen	1	30.00
6	beta-estradiol	C18H24O2	XIC	None	50-28-2	272.17763	273.1849	ms1	Positive	Hydrogen	1	30.00
7	lithium acetoacetate	C4H6O3	XIC	None	541-50-4	102.03169	103.039	ms1	Positive	Hydrogen	1	30.00
8	Glutathione oxidized	C20H32N6O12S2	XIC	None	27025-41-8	612.15196	613.1592	ms1	Positive	Hydrogen	1	30.00
9	Flavin adenine dinucl	C27H33N9O15P2	XIC	None	146-14-5	785.15713	786.1644	ms1	Positive	Hydrogen	1	30.00
10	Taurocholic acid	C26H45NO7S	XIC	None	81-24-3	515.29167	516.2989	ms1	Positive	Hydrogen	1	30.00
11	Calcium pantothenat	C9H17NO5	XIC	None	79-83-4	219.11067	220.1179	ms1	Positive	Hydrogen	1	30.00
12	Guanosine 3',5'-cylic	C10H12N5O7P	XIC	None	7665-99-8	345.04743	346.0547	ms1	Positive	Hydrogen	1	30.00
13	NAD+ (Nicotinamide	C21H27N7O14P2	XIC	None	53-84-9	663.10912	664.1164	ms1	Positive	Hydrogen	1	30.00
14	NADH (Reduced nicc	C21H29N7O14P2	XIC	None	58-68-4	665.12477	666.1321	ms1	Positive	Hydrogen	1	30.00
15	Tetrahydrofolic acid	C19H23N7O6	XIC	None	135-16-0	445.17098	446.1783	ms1	Positive	Hydrogen	1	30.00

MS2 Spectra Library Creation

The second part of the experiment was to acquire MS^2 information of the metabolites to build the high quality accurate mass MS^2 spectral library. Individual MS^2 spectra of the metabolites at the three collision energies 10, 30 and 45 in the respective polarities were extracted to give wider coverage for profiling experiments. The data analysis software has the added functionality to aid in the creation of a MS^2 spectral library. A combination of approaches were used to build the library. The first approach was to use the compound list that consisted information of the specified .raw files where the spectra scans of the compounds were present. The chemical formulae and the adduct type was supplemented as well. With the import function, the software was able to pick the desired MS^2 spectra with reference to the compound list and add it into the spectral library. The second approach used was for the addition of spectra by importing and viewing the .raw file in the library creation software. Theoretical accurate mass values of metabolites of interest were specified and the corresponding scan within the .raw file would be displayed (Figure 3.). The required MS^2 spectra would next be added into the library. Relevant information such as the extracted precursor m/z value and scan filter information were automatically populated from the acquired .raw file into the library.

The complete spectral library was specified in the processing method and used as an additional dimension for confirmation of the metabolites' identity. MS^2 spectra matching allows confident identification and complements the result matches by filtering the redundancy otherwise generated only by precursor m/z matching. A total of 1500 high quality accurate mass MS^2 spectra for the 300 metabolites were collated into this metabolite library as part of the screening solution for metabolic profiling workflow.

FIGURE 3. Addition of individual MS^2 spectra to create the MS^2 spectra library for metabolites screening using Library Manager 2.0, TraceFinder 3.2



Performance Validation of the Compound Database and MS2 Spectral Library

The performance of the metabolite screening workflow, which comprised of the compound database and the high quality accurate mass MS^2 spectral library, was validated against a ZDF rat plasma dataset. The ZDF rat plasma sample was ran under the same LC gradient conditions as stated in the "Methods" section. The sample dataset was processed to identify the possible endogenous metabolites present based on the criteria of accurate mass, retention time and MS^2 spectra matches. Of the 300 metabolites, 252 metabolites were identified with matching accurate mass values. 152 metabolites were identified with isotopic pattern matched scores of 100%. With retention time information, 130 metabolites showed positive identification with both exact mass and retention time matches. Lastly, of these metabolites, 85 metabolites were further identified to have the MS^2 spectra matched exactly. An overview of the number of metabolites identified is shown in Figure 6. Generated within a short processing time, the results demonstrated that the use of multiple screening features narrowed the identification of the endogenous metabolites in the ZDF rat plasma considerably. Additional search criteria reduces the redundancy and gave improved confidence to the profiling experiment. Figure 4. shows the metabolite profiling screening result pbtained from the processing software.

FIGURE 4. Screening results of ZDF rat plasma dataset processed using TraceFinder 3.2

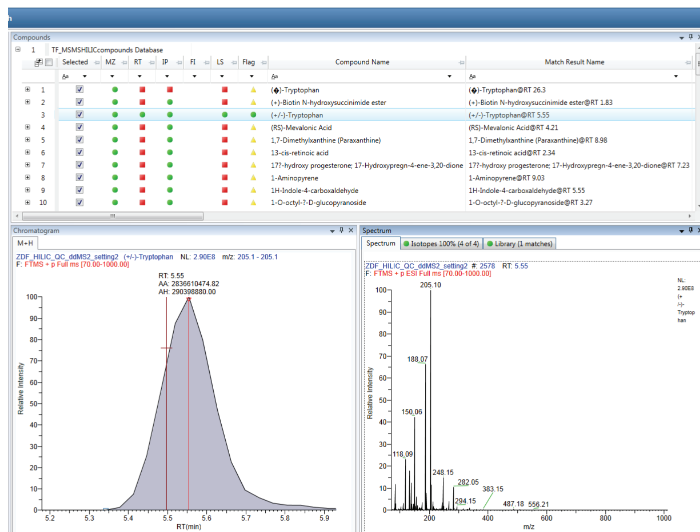
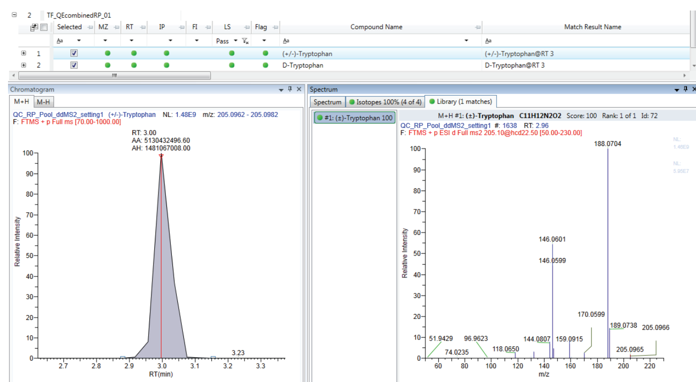
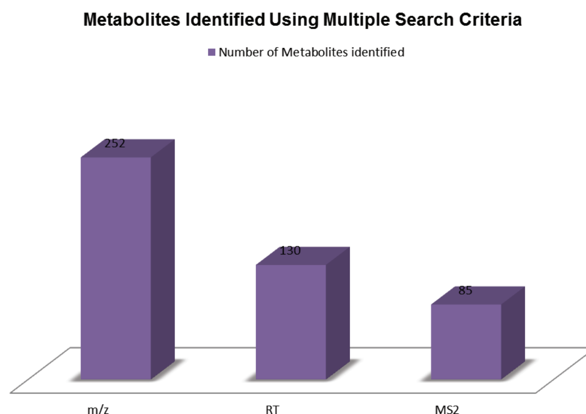


FIGURE 5. Identification of Tryptophan that has matched by screening criteria of m/z, RT and MS² spectra library matching.



The screening results demonstrate confident metabolite identification. Here, an example of Tryptophan is shown (Figure 5.). Tryptophan was matched based on all the search criteria with a mass accuracy of 1.7ppm and a maximum score of 100 for the MS^2 spectra library matching. The MS^2 spectra of the sample was matched against the standard library spectra with emphasis on the mass accuracy of the individual fragment masses and the overall fragmentation pattern unique to the metabolite.

FIGURE 6. Coverage of identification results using the 3 search criteria m/z, RT and MS² spectra.



Conclusion

A metabolomics library comprising of high quality accurate mass MS^2 spectra combined with retention time information enables fast and confident metabolite screening with the use of multiple search criteria.

- Improved confidence in metabolite identification and reduced occurrence of redundancy often associated with current metabolomics databases search methodology.
- HRAM quality MS^2 spectral entries obtained using a high resolution mass spectrometer.
- Future addition of more metabolites to the existing library to increase the identification coverage in screening workflows for metabolic profiling experiments.

References

1. Metabolomics Fiehn Lab, http://fiehnlab.ucdavis.edu/Metabolite-Library-2007/.
2. Xu, G., Yin P. Journal of Chromatography A. 2014, 1374, 1-13
3. Human Metabolome Library, http://www.hmdb.ca/hml/metabolites

www.thermofisher.com

©2016 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries. This information is presented as an example of the capabilities of Thermo Fisher Scientific products. It is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others. Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
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 45 453 9100
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
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

PN64467-EN 0616S

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand