Creating High Quality Metabolite Libraries for Fast Metabolomics Screening and Identification

Gina Tan, Junhua Wang, Yingying Huang, Andreas FR Hühmer[1]
[1]Thermo Fisher Scientific, San Jose, CA, USA

Poster Note 90118

Introduction

The study of endogenous metabolites has brought about new possibilities in the field of metabolomics. 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 of 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 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. 25 standards were combined into a single batch and a total of 15 batches were prepared. Standards were individually weighed at 1mg each and diluted to a final concentration of 0.5mg/ml after constitution with 50/50 methanol/water. All 15 batches were put through sonication followed by a filtration step to remove any undissolved salt particles.

Liquid Chromatography 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. (Refer to Table 1 for LC conditions).

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 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 processed using TraceFinder 3.2. A compound database was created that contains compound information, chemical formulae 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.

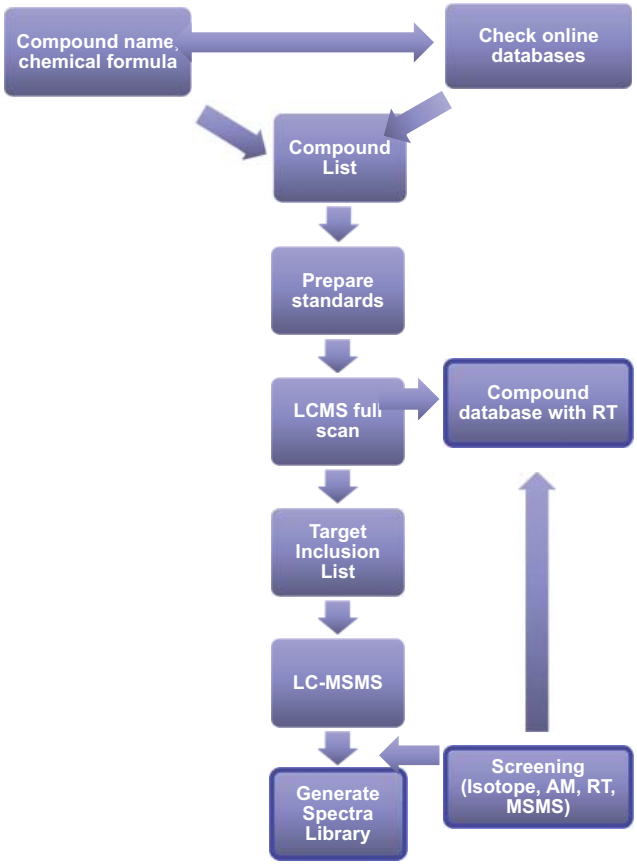
TABLE 1. UHPLC conditions.

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 in water; B: 0.1% Formic acid in methanol

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

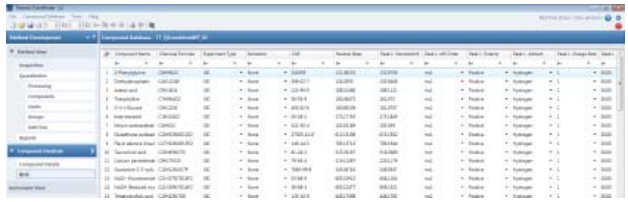
A preliminary list of endogenous, non-lipid metabolites was compiled from important metabolic pathways, which include the catabolic metabolism pathway, glycolysis and citric acid (Krebs') cycle to name a few. The shortlisted compounds were validated against online databases to ensure their endogenous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

Retention time of each compound was recorded based on the standardized reverse-phase chromatographic method used for this experiment. These information was combined with the related chemical information and used to create the endogenous metabolite compound database within TraceFinder (Figure 2). This database serves as a good starting point for metabolite profiling experiments to screen for possible metabolites. Samples will be screened against this compound database first and foremost by the accurate mass. Retention time is a secondary criteria taken into consideration to provide additional confirmation for a metabolite's identity. 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 relevant candidates are narrowed down.

The retention time information 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.

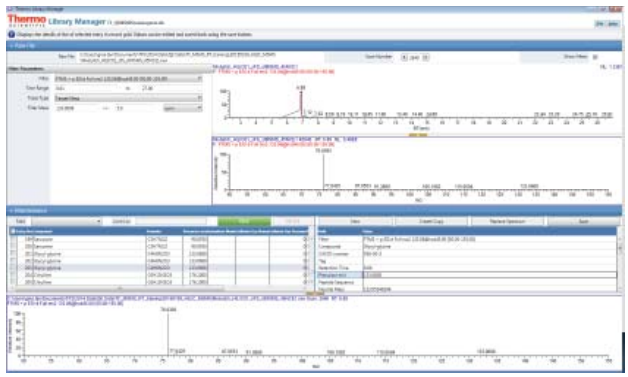


MS2 Spectra Library Creation

MS^2 information of the metabolites was acquired to build the high quality accurate mass MS^2 spectral library. Individual MS^2 spectra were extracted at the three collision energies 10, 30 and 45 in the respective polarities to give a wider coverage. With the import function, the software was able to pick the desired MS^2 spectra with reference to a compound list and add it into the spectral library. Addition of spectra was done by importing and viewing the .raw file in the library creation software. Theoretical accurate mass values of metabolites of interest were specified and the desired MS^2 spectra from the .raw file would be added into the library (Figure 3). Relevant 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 confirmation criteria of the metabolites' identity. MS^2 spectra matching allowed confident identification and complemented 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.

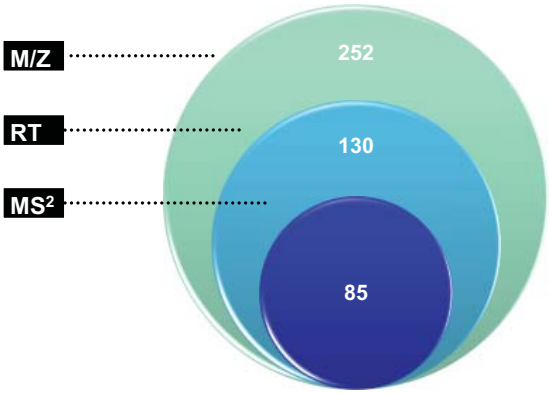


Performance Validation of the Compound Database and MS2 Spectral Library

The performance of the metabolite screening workflow 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%. 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 obtained from the processing software (Figure 4).

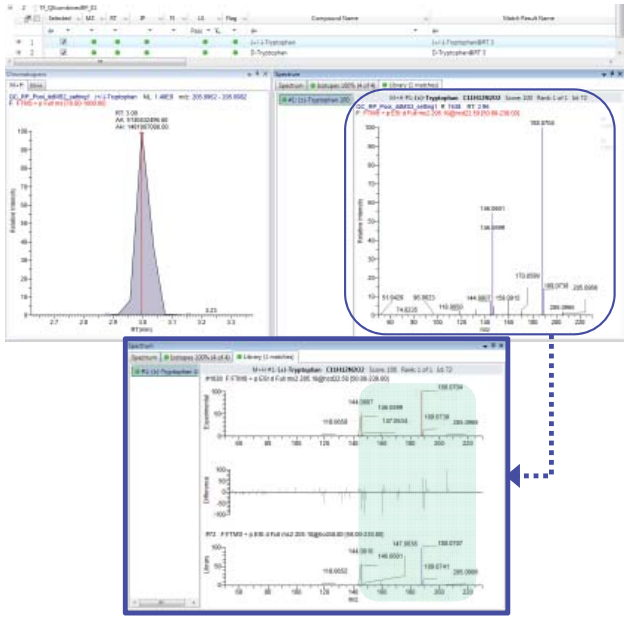
Generated within a short processing time, the results demonstrated that the use of multiple screening features narrowed the identification of the endogenous metabolites in the ZDF rat plasma considerably. Additional search criteria reduces the redundancy and gave improved confidence to the profiling experiment.

FIGURE 4. Coverage of identification results using the 3 search criteria m/z, RT and MS^2 spectra.



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 experimental MS^2 spectra of the sample was matched against the library spectra with emphasis on the mass accuracy of the individual fragment masses and the overall fragmentation pattern unique to the metabolite.

FIGURE 5. Identification of Tryptophan that has matched by screening criteria of m/z, RT and MS^2 spectra library matching.



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