Creating High Quality Metabolite Libraries for Fast Metabolomics Screening and Identification

Gina Tan, Junhua Wang, Yingying Huang, Andreas FR Hühmer¹ ¹Thermo Fisher Scientific, San Jose, CA, USA

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 gaschromatography 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 MS² 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 MS² in both positive and negative polarities. An inclusion list for all the respective batches was included in the full MS data-dependent MS² method. All MS² 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 MS² spectra of the compounds were extracted using Library Manager 2.0 (part of TraceFinder 3.2) to generate the MS² spectra library.

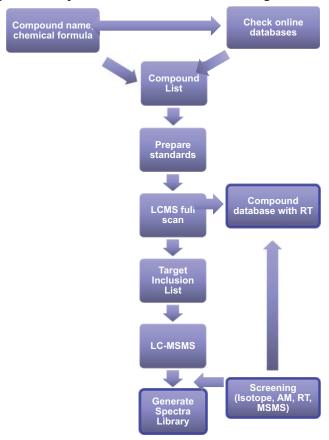
TABLE 1. UHPLC conditions.

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



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 MS² 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.

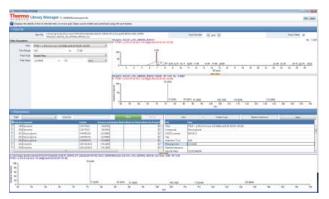
the Contract	**	-	11 (5)																	
terior/tree		(imposit bers	(Flatina Petrole	haarner to	, because		1206		Sector Street	Test t Secure	ind, , i	and in the little of		Park II II o	**	Face 6-4	the state		Face in change for	. 700
manner:				+	+	4	+	-		+ 1				44			-		de a	-
Supplieden		1Pleybore	dental	(K	See.		Julies.		25,810	FAXTRA				Tradet:		- 1000				- 84
Processing	- 12	Smirghaphan	Carcage	-	Sunt		1000		19,189	TA INC.				No. Roy		* Spinge				- 60
	- 1	Automobile .	CHIEF	100	Store		23400		221500	38112		mi .	+	Frederic		+ Total				- 800
Composing.		Teaching	CHOICE.	in:	. See a	0.4	20144		DOMEST	SELECT		ed.		Pestire		· hybrigh				- 310
desir.		Markhoot	090206	TE	test.	- 00	4000		post re-	25,750		4		Product.		· tump		+		- 100
Seeper .		tree-research	Carolet	100	State 1		dries .		DUTH	23140		mit i	40	laster.		+ tones		+		 los
Septime .		Management	CHEST	100	Street 1		12.64		22.22.00	300.000	-	mi.	÷.	Perdont		+ Follows		+		- 80
	- 1	Guattiera school	CHEMICAL	ar.	See		ranne		STILLS.	cure		4		India						- 34
NAME .	_	Own March Steel	consumato.	100	Serie .	100	101.141		961451	78144	- 1	mi .	ď	Partie.		· tolina				- 100
		Secretary and	CRANCO	-	tree		41.41		6.5 part	0.00m		-		State of		+ tubuja				- 40
	- 11	States province	OKCHES	100	Seed.		1648-6		DALDET	DRIFE	- 7	-		France		+ Follows		+		
INDIGETERS.	- 1	Gueste I Fast.	CHENNY	DE.	See .		basine		MARK	1400				India		• Roberton				- 310
14		NO former		16.	best .		2010		461367	44170				Penter		- Political				- 100
and the contract of		hade being on			See .		drift's		miller.	04155				lane.		- Notice		23		- 10
The same of the sa	- 1			100	her		17.01		441709	48170				Feeter		- france				- 10

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. MS² 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 MS² 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 MS² spectra to create the MS² spectra library for metabolites screening using Library Manager 2.0, TraceFinder 3.2.

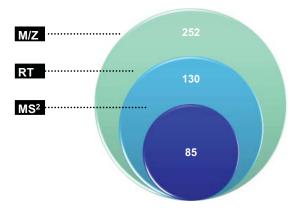


Performance Validation of the Compound Database and MS2 Spectral Library

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 MS² 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 MS² spectra matched exactly obtained from the processing software (Figure 4).

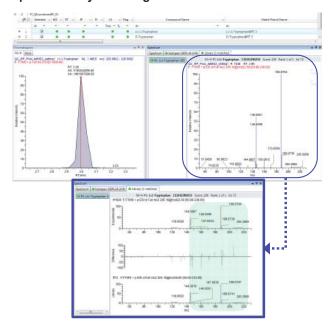
Generated within a short processing time, the results demonstrated that the use of multiple screening features narrowed the identification of the endogenous metabolites in the ZDF rat plasma considerably. Additional search criteria reduces the redundancy and gave improved confidence to the profiling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

FIGURE 4. Coverage of identification results using the 3 search criteria m/z, RT and MS² 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 MS² spectra library matching. The experimental MS² 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 MS² 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 MS² 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

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

www.thermoscientific.com

©2015 Thermo Fisher Scientific Inc. All rights reserved. ISO is a trademark of the International Standards Organization. All other 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



A Thermo Fisher Scientific Brand