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 MS² 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 MS² spectra of each individual compound. All data were processed using the Thermo Scientific TraceFinder 3.2 software. The compound database, MS² spectra library and screening methods were also created with the TraceFinder software. (Figure 1.)

Results: A compound database and MS² 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 MS² 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 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 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-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. 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	%В
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 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 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 MS² spectra of the compounds were extracted using Library Manager 2.0 (part of TraceFinder 3.2) to generate the MS² spectra library.



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 $\rm MS^2$ 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

File Compound Database Tools	Help It In Br																cal tie	se status User: gina.ts	
Method Development	• I Cor	rpound Database	11_QI combinedR	P.01															
 Method View 	8	Compound Name	Chemical Fermula	Eperime	nt Type	Ionization	-	CAS	_	Neutral Mas	Peak 1 Extract	d M. Peak:	MS Order	Peak 1 Polarity	-	Peak 1 Adduct		Peak 1 Charge State	Peak1
Acquisition		6+ ·	a	84		de .		d=		b*	<i>b</i> *	• bi		b .		64		s	51
Quantitation	1	2-Phenylglycine	C8H9N02	XOC	•	None	•	341659		151.06333	152.0706	msi		Positive	٠	Hydrogen	•	1. •	30.00
Bearenies	2	Diethylphosphate	C4H1104P	XOC		None		598-02-7		154.0395	155.0468			Positive	٠	Hydrogen		1 .	30.00
rocenny	- 3	Azelaic acid	C9H16O4	XOC		None		123-99-9		188.10486	189,1121	ms1		Positive	٠	Hydrogen	•	1 .	30.00
Compounds	- 4	Theophylline	C7H8N402	XOC		None		58-55-9		180.06473	181.072	mc1		Positive	٠	Hydrogen		1 .	30.00
Limits	5	D-(+)-Glucese	C6H12O5	XOC		None		492-62-6		180-06339	181.0707	ms1		Positive		Hydrogen		1 -	30.00
Groups	6	beta-estradiol	C18H2402	XOC		None		50-28-2		272.17763	273.1849	ms1		Positive	•	Hydrogen	-	1 .	3000
Intel Seq	7	Rhium acetoacetate	C4H603	200	-	None		541-50-4		102.03169	103.039	ms1		Positive	•	Hydrogen	-	1 -	30.00
Reports	8	Clutathione oxidized	C20H32N601252	XOC	-	None		27025-41-8		612.15296	613.1592	ma1		Positive		Hydrogen		1 -	30.00
Neporo		Ravin adenine dirud	C27H33N9O15P2	XOC		None		146-14-5		785.15713	785.1644	mi1		Positive	٠	Hydrogen	-	1 -	30.00
Compound Database	> 10	Taurocholic acid	C26H45N075	X9C		None		81-34-3		515.29067	515.2989	es1		Positive		Hydrogen		1 -	30.00
Companyed Datally	11	Calcium pantothenat	C9H17NO5	XOC		None		79-83-4		218.11067	220.1179			Positive		Hydrogen		1 .	30.00
compound ortana	12	Guanosine 3:5-cylic	C10H12N507P	XOC		None		7665-99-8		345,04743	345.0547	es1		Positive		Hydrogen		1 .	30.00
Grid	13	NAD+ (Nicotinamide	C21H27N7014P2	XOC		None		\$3-64-9		663.10912	664.1154	msl		Fostive		Hydrogen		1 -	30.00
Instrument View	- 14	NACH (Reduced nicc	C21H29N7014P2	XOC	10	None		58-68-4		665.12477	666.1321	mal		Positive		Hydrogen		1 -	30.00
	15	Tetrahydrofolic acid	C19H23N706	200	-	None		135-16-0		445,17096	445.1783	ma1		Positive		Hidrogen		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 MS² spectral library. Individual MS² 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 MS² 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 MS² 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 MS² 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

O Thems Library M	lanager						_						le	- #
Thermo	Library Manag	ger TT_QEMSMS	rainingtest.db											<u>[4</u> 290
O Displays the deta	ails of the of selected er	ntry in record grid. I	/alues can be edited an	nd seved back usi	ing the save button.									
- Raw File														
	Raw File Cristers Webers	riginatari/Document ah_H0JC01_FF5_PfM	s17FS12004 Datal/QE Da JMS_45NCE1/8w	MULTE INSING REL	training/20150108_HE3C_	MSMS	_		Scan Number	1 2840 K			Show Filters	R
Filter Parameters				_	F: FTM3 + p ESt d Fut m	INSME ASNOFT	00 (50)	00-155.00						NC 1336
Fiter	FTMS = p ESE d Full r	ms2133.06@hod453	30 (50.00-155.00)											
Time Range	031	50	27.00		100.3		1.89							
Trace Type	Taroet Mass			•			1							
forer Mass	133.0000	+1- 30	(66m)		Museuch JHLICOT #78 - F FTHE - p ESI dFull m	HIMEME CONCELL	7 12 7 82543 60 (50	7.62 6.59 9.19 10.11 6 9 10 RT 6.89 NL 6.4688 90-155.00 76.0003	12,25 11,26 11 12 12	13,40 14,40 14,55 0 14 15 15 15 17 RT(min)	16 19 20 21	22.44 23.20	<u>3475 2</u> 24 25	19 28.91 26
					a extran			77.6425 87	0553 91 3855	100 1002 115	MA 13	1 0103		
- Maintenance		• Leve Le			Postsien is	6 6	70	77.5425 87 75 80 85	0553 91,3650 90 95	105.1062 115 106 166 116 1 m/g Create Copy	2004 132 6 120 125 130 Restars Sectors	1 0000	145 1	155
Maintenance Field This is constant		Unit Up	formula	Provenue at from	Band Band	6 6 6	70	77.5425 87 75 80 85	91,3000 90 95 90 95	106.1082 115. 108 106 116 1 108 116 118 1 m/z Create Copy	504 132 6 129 126 130	1.0505 136 140	125 1 Save	
Maintenance Field Toty Instrument Style Saresin	*	Limit Lip	formula C3H01N02	90.0550	Pandala Pa	e e	70 01 -	77.5425 87 75 80 85 New Filter	4053 91,3050 90 95 Value Fitads = p 851 d P	100 1002 115 100 105 110 1 miz Create Capy ur mol 133.06@nexe45.00 (50.00 c	504 133 6 129 125 130 Regiace Spectrum 55.001	10008	145 1 Save	
Maintenance Field 199 Sarresin 200 Sarresin	4 6 1	• Limit Lip	Caninata CBH11002 CBH11002 CBH11002	90.0350	a weather	6 6	8	77.9425 87 75 80 85 New Riter Compand Compand	9000 19 3000 90 95 Votes FTMS = p ESI d F Glyop(-g)cine State 0:1	100.1082 115. 500 105 118 1 mi2 Create Copy will mi2 133.06@mos45.00 (50.00 -)	5004 133 5 129 125 130 Replace Spectrum 55.001	10000	145 1 Save	0 156
Maintenance Net Tete Tota Inscreption 200 Sarcein 200 Sarcein 200 Sarcein 200 Sarcein	e e goine holos	Lawe Lar	Колимала СЭНТНОО2 СЭНТНОО2 САНЯКДОЭ САНЯКДОЭ	90.0550 90.0550 133.0608	RND	e 6		77.6425 87 75 80 85 New Fater Compound CAS ID number Tas	(865) 91,3000 90 95 Velan PTat5 + p.551,4 P Disp(-g)-circle 556-50-3	100 1082 115. 100 100 110 1 miz Create Copy ur mu2 133.06@nce43.00 (50.00 c	7094 132 6 120 125 130 Replace Spectrum 55 001	3.0608 135 140	145 1 Save	
Mainformance Red Total Screen 200 Screen	e e bolne bolne bolne	• Limit Lar	 сылыс сылого сылого сылого сылого сылого сылого сылого 	90.0550 90.0550 133.0608 133.0608	RND	cerni	2	77.5425 EP 75 ED E5 New Riter Compound C451D number Tag Resention Time	6653 91,3850 90 95 1900 1900 1900 1900 1900 1900 1900 19	100 1082 115 100 100 119 1 m/2 Create Copy ut wa 133.06@nexet5.00 (00.00 c	004 122 6 129 125 150 7 Replace Spectrum 95 00]	10000	125 1 Save	
Marconarce Feld Top Sectors T	e Ne Ngine Ngine Ngine Ngine	• Lawe Lar	Famada CBH/1N02 CBH/1N02 CBH/NO3 CBH/NO3 CBH/NO3 CBH/NO3 CBH/NO3 CBH/NO3	90.0550 90.0550 133.0608 133.0608 135.0608 176.1000	RND	COLUMN	2	72,843 87 Ys 80 85 New New Company Com	91,3000 90 95 View FTINS = p.55 d F 90;q2-9)cire 555-50-3 649 132:0608	100.002 115 00 10 119 1 miz Create Copy uit ms2 133.06@next5.00 (50.00 c	0004 132 6 120 125 130 	10000	145 ti	0 156
Mainformance Reis The Serverse The Serverse	e N M Apone Volke Volke	• Lent Ly	Famada C3H1N02 C3H1N02 C4H1N03 C4HN033 C4HN033 C4H3N039 C4H3N039 C4H3N039	90.0550 90.0550 133.0608 133.0608 133.0608 135.0008 176.1000	RND	e 6	2000000	77,843 87 35 80 85 New Case Compound Case Tag Remotion Time Propose m/s Papeter Securico	(455) 91,3000 (6) 95 (7) (7) (7) (7) (7) (7) (7) (7) (7) (7)	105.1082 115. 100 105 119 1 mit Create Copy 41 mit 135.04@next5.00 (50.00 -	004 132 6 128 125 150 Replace Spectrum 05.003	1 0000	145 ti	116
Mainformance Real Top Increase Top Incre	e Pone Apone Pone Pone Pone Pone Pone Pone Pone P	• Limit La 	Fremade CEHTNO2 CEHTNO3 CHTNO3 CH	90.0550 90.0550 133.0608 133.0608 133.0608 135.1000 176.1000 176.1000		6 6 00.00	2	77,5425 87 75 80 85 Nen Padel Rater Compound CAS D number Reproduce m/s Reproduce m/s Reproduce for Reported laters n, 2440 KT 6.60	4853 91,3890 90 95 7005 19 55 47 9005 90 55 47 9005 90 50 555-53 4.89 132,053,48004	100 1002 115 50 mi2 002 Croste Coy 41 mi2 133.06@mests.00 (00.00 2	554 132 6 129 125 136] Register Spectrum 55.00]	10000	145 ti	2 The
Mainformance Real Torophic Company Torophic Comphany Torophic Company Torophic Company	n Ngcine	Limit Lie	/ СРИТИО2 СРИТИО2 СИНТИО3 СИНИОЗ СИНИОЗ СОССОС	90.0550 90.0550 133.000 133.000 135.000 176.1000	Record Jacobi (Fig.	Cectra Cectra (Cectra (Cectra) (Cectra) (Cectra)		77,5425 87 75 80 85 74ter Congound Cot30 number Tagi Ramanjer m/c Populot Experience Tagiotel Hame Ramanjer m/c	0053 913000 00 90 70 70 70 70 70 70 70 70 70 70 70 70 70	199 1992 115 199 1992 115 199 199 199 199 Cean Capy ut mc2 1330/d-well3.01 (50:00 -	9594 153 5 405 455 155] Replace Spectrum 55.500	3.0000 550 - 140 	145 ti	156
Matternance Nee Nee Second	e N Sypone Applice Pytone Spread Sector (1970) 2014 Data Sector (1970) 2014 Da	Canot Lan	Female CHRIN02 CHRIN02 CHRIN03 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN02 CHRIN03	90.0550 90.0550 133.0658 135.0008 10008 100008 10008 10008 10008 10008 10008 10008 1				77,5425 87 35 ab ab Nen Rear Compand CASD number Tag Branston Tana Branston Tana Brans	0053 913000 00 90 70 70 70 70 70 70 70 70 70 70 70 70 70	100 1002 115 100 102 115 101 102 102 102 102 Create Capy of mo2 133.010-010 102 102 101 102 102 102 102 101 102 102 102 102 101 102 101 102 102 101 102 102 101 102 101 102 102 101	994 113 9 109 109 109 10 Regian Spectrum 5.00	3 0000 550 - 140	145 11 Seve	n 195
Multiconstance Real Tors - Constance 202 (Charling - Constance) 202 (Cha	e e bytes bytes e automatific2016 Care automatific2016 Care	•) Law Lar (20 Dearty Mission (20 Dearty Mis	700000 (CH1N22 (CH1N22 (CH1N22 CH1N2	9000000 1330000 1330000 1761000 1761000 1761000	17 (253) \$1,300			77.5405 EP 75 e0 65 Non Pater Compand Cospand Cospand Restor Tag T	6653 91,3000 90 96 PTUS = p 82(47) 2000/01/2000 122.05340004 122.05340004 122.05340004	199 1992 115 199 1992 115 199 1992 199 199 199 Cease Capy 0 m m2 135 040-04550 500 0 0 00 000 000 0 0 000 000 0 0	133 000 133 000 100 100 100 100 100 100	0 0000 15 160	145 1 Save	

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 MS² 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 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%. 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 MS² 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 MS^2 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 MS² spectra library matching. The MS² 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 MS^2 spectra.



Conclusion

A metabolomics library comprising of high quality accurate mass MS² 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

- 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 (ree call domestic) 400 650 5118 PN64467-EN 0616S 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 33 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

