GC-HRMS processing tools for untargeted metabolomics

Axel Raux, Lauriane Rambaud, Anne-Lise Royer, Justine Massias, Yann Guitton, Bruno Le Bizec

GC-HRMS PROCESSING TOOL FOR UNTARGETED METABOLOMICS

Axel RAUX, Lauriane RAMBAUD, Anne-Lise ROYER, Justine MASSIAS, Yann GUITTON, Bruno LE BIZEC

LABERCA, Oniris, INRAE, Nantes, France
Fax : +33 (0)2 40 68 78 78 - Tél : +33 (0)2 40 68 78 80
laberca@oniris-nantes.fr

Introduction

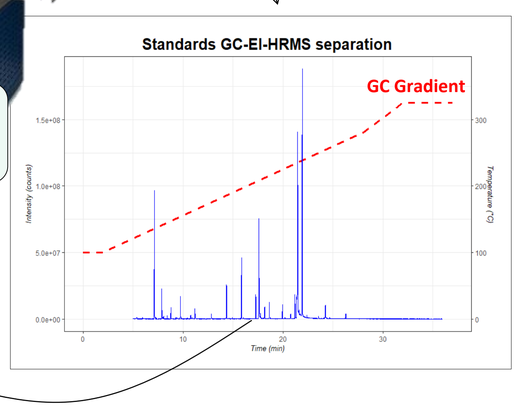
Metabolomics and related sciences use a combination of analytical and statistical approaches to qualitatively and quantitatively analyze small molecules present in a biological system to answer biological questions. Metabolomics and untargeted studies require the maximization of compounds detected in the analysis, while the results must be reproducible and robust. Gas chromatography-mass spectrometry (GC-MS) is a commonly used technology in metabolomics research thanks to the high sensitivity and high throughput, as well as the wealth of databases. The use of an internal database in essential in metabolomics to improve and fasten biomarker annotation, as well as giving biological meaning to biomarkers. Database creation can be done with mainly two ways : either manual, or automatic, with the use of software, like TraceFinder that we have evaluated.

Experimental

1 GC-HRMS acquisition

Mix Stds → MOX/MSTFA Derivatization

MS parameters :
- GC-Q-Exactive
- Full scan EI-MS
- 140k resolution

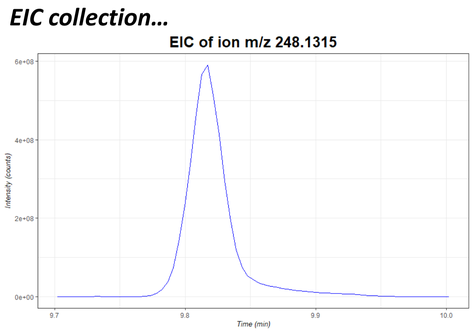


GC-EI-HRMS spectra collection of injected standards

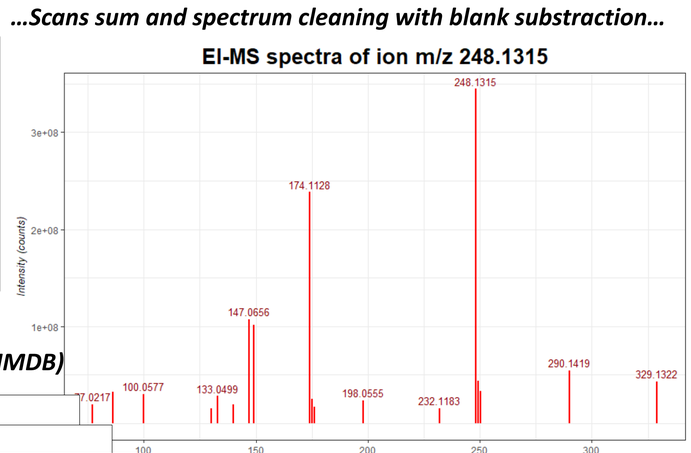
WHAT ARE THE NEXT STEPS ?

2 Manual spectral matching with external GC-EI-MS databases

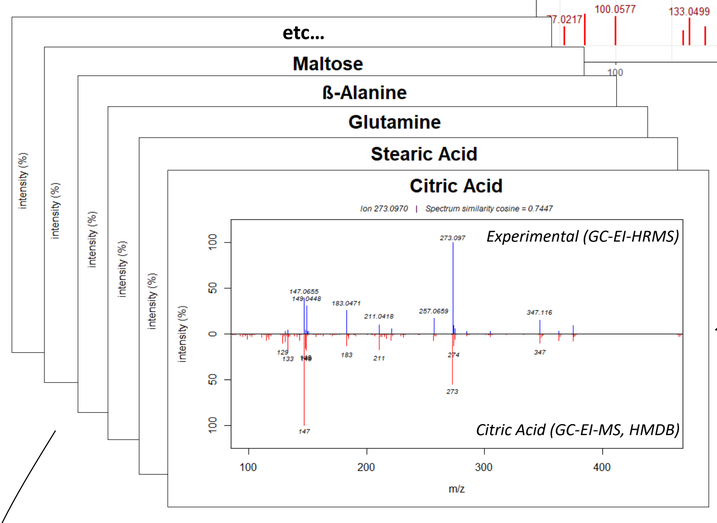
EIC collection…



…Scans sum and spectrum cleaning with blank substraction…



Spectral matching with cosine scores (via HMDB)



hmdb The Human Metabolome Database

NIST National Institute of Standards and Technology U.S. Department of Commerce

Internal database

PROS	CONS
• No software knowledge • Human eye assessment	• Fastidious • Time consuming • Reproducibility

NEED OF ANOTHER SOLUTION

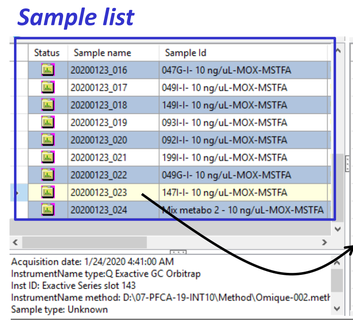
3 Automatic spectral matching with TraceFinder™ software

Thermo Scientific software for LC-MS or GC-MS data processing

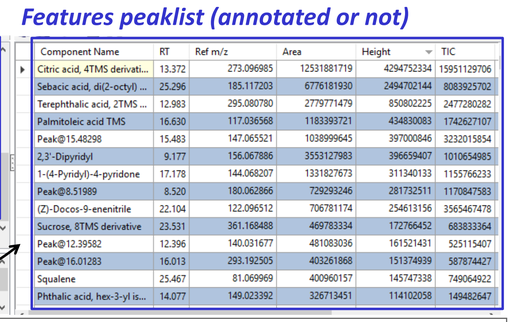
TraceFinder™ The Power to Know

TraceFinder interface

Sample list



Features peaklist (annotated or not)



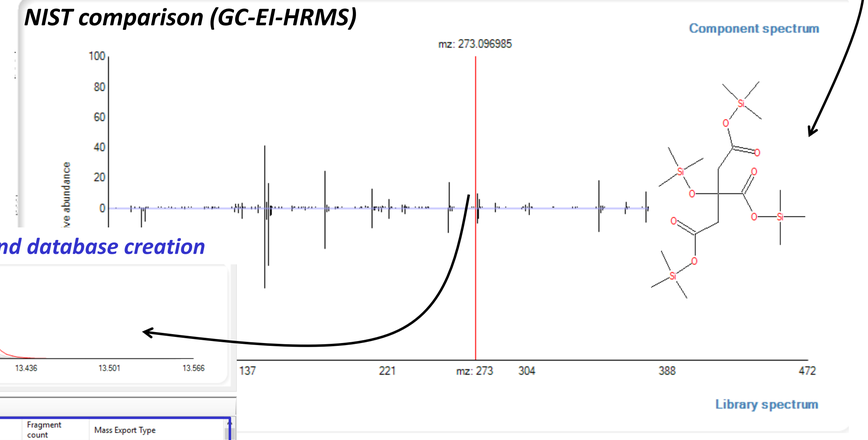
Component Name	RT	Ref m/z	Area	Height	TIC
Citric acid, 4TMS derivati...	13.372	273.096985	12531881719	4294752334	15951129706
Sebacic acid, di(2-octyl) ...	25.296	185.117203	6776181930	2494702144	8083925702
Terephthalic acid, 2TMS ...	12.983	295.080780	2779771479	850802225	2477280282
Palmitoleic acid TMS	16.630	117.036568	1183393721	434830083	1742627107
Peak@15.48298	15.483	147.065521	1038999645	397000846	3232015854
2,3'-Dipyridyl	9.177	156.067886	3553127983	396659407	1010654985
1-(4-Pyridyl)-4-pyridone	17.178	144.068207	1331827673	311340133	1155766233
Peak@8.51989	8.520	180.062866	729293246	281732511	1170847583
(Z)-Docos-9-enenitrile	22.104	122.096512	706781174	254613156	3565467478
Sucrose, 8TMS derivative	23.531	361.168488	469783334	172766452	683833364
Peak@12.39582	12.396	140.031677	481083036	161521431	525115407
Peak@16.01283	16.013	293.192505	403261868	151374939	587874427
Squalene	25.467	81.069969	400960157	145747338	749064922
Phthalic acid, hex-3-yl is...	14.077	149.023392	326713451	114102058	149482647

Peak identification and NIST library comparison

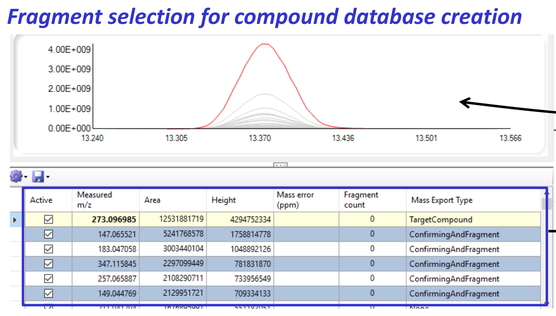
Score	Matched Compound	mula	CAS	SI	RSI	Library	Library ID Number
97.6	**Citric acid, 4TMS deriva...**	**H40O7Si4**	**14330-97-3**	**882**	**889**	**orbitrap_meta...**	**7**
94.9	Citric acid, 4TMS derivati...	H40O7Si4	14330-97-3	745	745	mainlib	239

Spectra

NIST comparison (GC-EI-HRMS)



Fragment selection for compound database creation

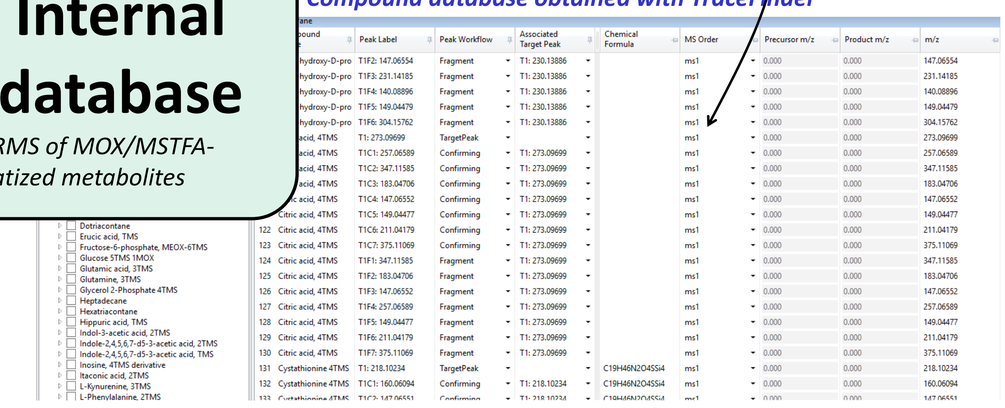


Active	Measured m/z	Area	Height	Mass error (ppm)	Fragment count	Mass Export Type
☑	**273.096985**	12531881719	4294752334		0	TargetCompound
☑	147.065521	5241768578	1758814778		0	ConfirmingAndFragment
☑	183.047058	3003440104	1048892126		0	ConfirmingAndFragment
☑	347.115845	2297099449	781831870		0	ConfirmingAndFragment
☑	257.065887	2108290711	733956549		0	ConfirmingAndFragment
☑	149.044769	2129951721	709334133		0	ConfirmingAndFragment

LABERCA **Internal database**

GC-EI-HRMS of MOX/MSTFA-derivatized metabolites

Compound database obtained with TraceFinder



	Compound	Peak Label	Peak Workflow	Associated Target Peak	Chemical Formula	MS Order	Precursor m/z	Product m/z	m/z
	Dotriacontane								
	Erucic acid, TMS								
	Fructose-6-phosphate, MEOX-6TMS								
	Glucose 5TMS 1MOX								
	Glutamic acid, 3TMS								
	Glutamine, 3TMS								
	Glycerol 2-Phosphate 4TMS								
	Heptadecane								
	Hexatriacontane								
	Hippuric acid, TMS								
	Indol-3-acetic acid, 2TMS								
	Indole-2,4,5,6,7-d5-3-acetic acid, 2TMS								
	Indole-2,4,5,6,7-d5-3-acetic acid, TMS								
	Inosine, 4TMS derivative								
	Itaconic acid, 2TMS								
	L-Kynurenine, 3TMS								
	L-Phenylalanine, 2TMS								
122	Citric acid, 4TMS	T1C6: 211.04179	Confirming	T1: 273.09699		ms1	0.000	0.000	211.04179
123	Citric acid, 4TMS	T1C7: 375.11069	Confirming	T1: 273.09699		ms1	0.000	0.000	375.11069
124	Citric acid, 4TMS	T1F1: 347.11585	Fragment	T1: 273.09699		ms1	0.000	0.000	347.11585
125	Citric acid, 4TMS	T1F2: 183.04706	Fragment	T1: 273.09699		ms1	0.000	0.000	183.04706
126	Citric acid, 4TMS	T1F3: 147.06552	Fragment	T1: 273.09699		ms1	0.000	0.000	147.06552
127	Citric acid, 4TMS	T1F4: 257.06589	Fragment	T1: 273.09699		ms1	0.000	0.000	257.06589
128	Citric acid, 4TMS	T1F5: 149.04477	Fragment	T1: 273.09699		ms1	0.000	0.000	149.04477
129	Citric acid, 4TMS	T1F6: 211.04179	Fragment	T1: 273.09699		ms1	0.000	0.000	211.04179
130	Citric acid, 4TMS	T1F7: 375.11069	Fragment	T1: 273.09699		ms1	0.000	0.000	375.11069
131	Cystathionine 4TMS	T1: 218.10234	TargetPeak		C19H46N2O4SSi4	ms1	0.000	0.000	218.10234
132	Cystathionine 4TMS	T1C1: 160.06094	Confirming	T1: 218.10234	C19H46N2O4SSi4	ms1	0.000	0.000	160.06094
133	Cystathionine 4TMS	T1C2: 147.06551	Confirming	T1: 218.10234	C19H46N2O4SSi4	ms1	0.000	0.000	147.06551

PROS	CONS
• Fast • Reproducible • Automatic • All-in-one solution	• Nothing better than human experience 😉 • Consequent data volume needed • Not free of charge

Conclusion

Internal databases are of great interest for annotation pipeline in metabolomics and lipidomics. The addition of GC-EI-HRMS spectra appended a further annotation level for metabolites identification. In order to create the database, the « manual » solution seems to be the most available and easy to use, but also the most fastidious one. On the other hand, many softwares now offer help in database creation from raw data. TraceFinder software, provided by Thermo Scientific, was evaluated here and helped us to fastly and easily build our internal metabolites database. The GC-EI-HRMS database, now composed of 51 standards, which keeps expanding, will be used later for features annotation in different metabolomics projects involving several matrices (urine, blood, milk, tissue, etc…). Other softwares, preferably free of charge, will be assessed for database creation.