

Development of an electrospray-mass spectral database for annotating metabolomics datasets: application to the analysis of the adult human urinary metabolome

Aurelie ROUX¹; Ying XU¹; Jean-François HEILIER²; Marie-Françoise OLIVIER¹; Eric EZAN¹; Jean-Claude TABEL³; Christophe JUNOT¹

¹CEA - Centre d'Etude de Saclay, Gif-Sur-Yvette, France;

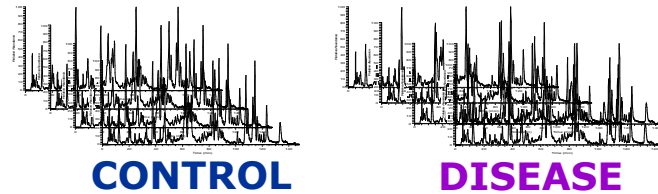
²Institut National de Recherche et de Sécurité, Nancy, France;

³LCSOB, IPCM, UMR-CNRS 7201, UPMC Paris Universitas, Paris, France

Metabolomics, practical approach

énergie atomique • énergies alternatives

1. Analytical chemistry



UPLC / LTQ- Orbitrap



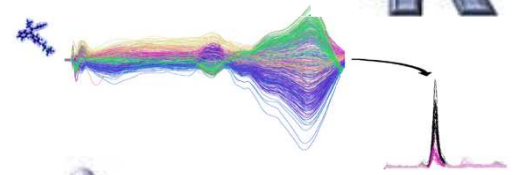
2. Data analysis

Samples →

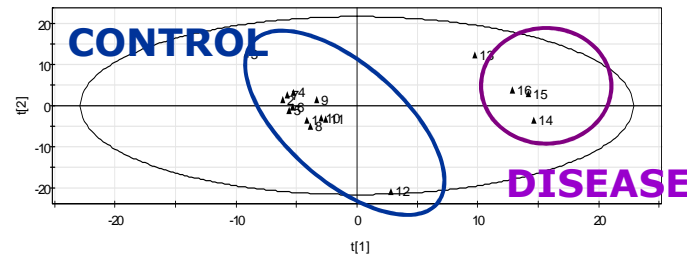
↓ **Variables (RT, m/z)**

Peak No.	Retention	Area	%T	%T	%T	%T	%T	%T	%T
			00000022	00000023	00000024	00000025	00000021	00000022	00000022
1	11.24	800	1.0E+03	1.5E+03	2.4E+03	1.7E+03	2.0E+03	7.9E+04	
2	11.28	1000	1.0E+03	1.5E+03	2.4E+03	1.7E+03	2.0E+03	7.9E+04	
3	16.25	106	2.0E+03	3.7E+04	2.0E+03	1.7E+03	1.8E+03	2.0E+04	
4	16.63	114	2.0E+03	3.7E+04	4.5E+04	1.0E+04	3.5E+05	2.5E+04	
5	16.92	114	2.0E+03	3.7E+04	4.5E+04	1.0E+04	3.5E+05	2.5E+04	
6	17.26	114	2.0E+03	3.7E+04	4.5E+04	1.0E+04	3.5E+05	2.5E+04	
7	17.34	114	2.0E+03	3.7E+04	4.5E+04	1.0E+04	3.5E+05	2.5E+04	
8	18.01	118	2.0E+03	3.7E+04	4.5E+04	1.0E+04	3.5E+05	2.5E+04	
9	18.19	126	0.4E+04	1.6E+05	6.4E+04	1.7E+04	2.3E+05	1.2E+04	
10	18.61	118	2.0E+03	3.7E+04	4.5E+04	1.0E+04	3.5E+05	2.5E+04	
11	18.93	126	2.3E+03	1.0E+05	2.4E+05	1.7E+04	2.3E+05	1.2E+04	
12	18.97	126	1.8E+05	1.6E+05	1.3E+05	1.7E+04	2.3E+05	1.2E+04	
13	18.98	126	1.3E+05	1.6E+05	1.3E+05	1.7E+04	2.3E+05	1.2E+04	
14	18.98	126	1.3E+05	1.6E+05	1.3E+05	1.7E+04	2.3E+05	1.2E+04	
15	19.52	126	1.3E+05	1.6E+05	1.3E+05	1.7E+04	2.3E+05	1.2E+04	
16	19.56	126	1.0E+05	1.6E+05	2.0E+05	9.1E+03	2.3E+05	1.2E+04	
17	19.65	126	1.3E+05	1.6E+05	1.3E+05	1.7E+04	2.3E+05	1.2E+04	
18	19.73	126	1.0E+05	1.6E+05	1.3E+05	9.1E+03	2.3E+05	1.2E+04	
19	19.89	126	1.3E+05	1.6E+05	1.3E+05	1.7E+04	2.3E+05	1.2E+04	
20	19.93	126	1.3E+05	1.6E+05	1.3E+05	1.7E+04	2.3E+05	1.2E+04	
21	19.84	130	1.0E+03	3.0E+04	2.1E+04	1.4E+04	2.3E+05	2.7E+04	
22	19.79	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
23	19.97	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
24	19.97	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
25	19.97	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
26	19.97	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
27	19.97	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
28	19.97	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
29	19.97	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
30	19.97	130	2.3E+03	2.4E+05	2.4E+05	2.4E+05	2.4E+05	2.4E+05	
31	17.34	105	2.3E+03	2.0E+03	2.8E+03	3.0E+03	2.7E+03	3.8E+04	
32	18.47	105	2.3E+03	2.0E+03	2.8E+03	3.0E+03	2.7E+03	3.8E+04	
33	18.42	105	2.3E+03	2.0E+03	2.8E+03	3.0E+03	2.7E+03	3.8E+04	

XCMS
R package



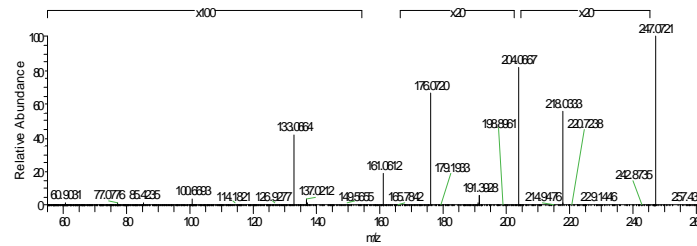
3. Chemometrics



SIMCA P+



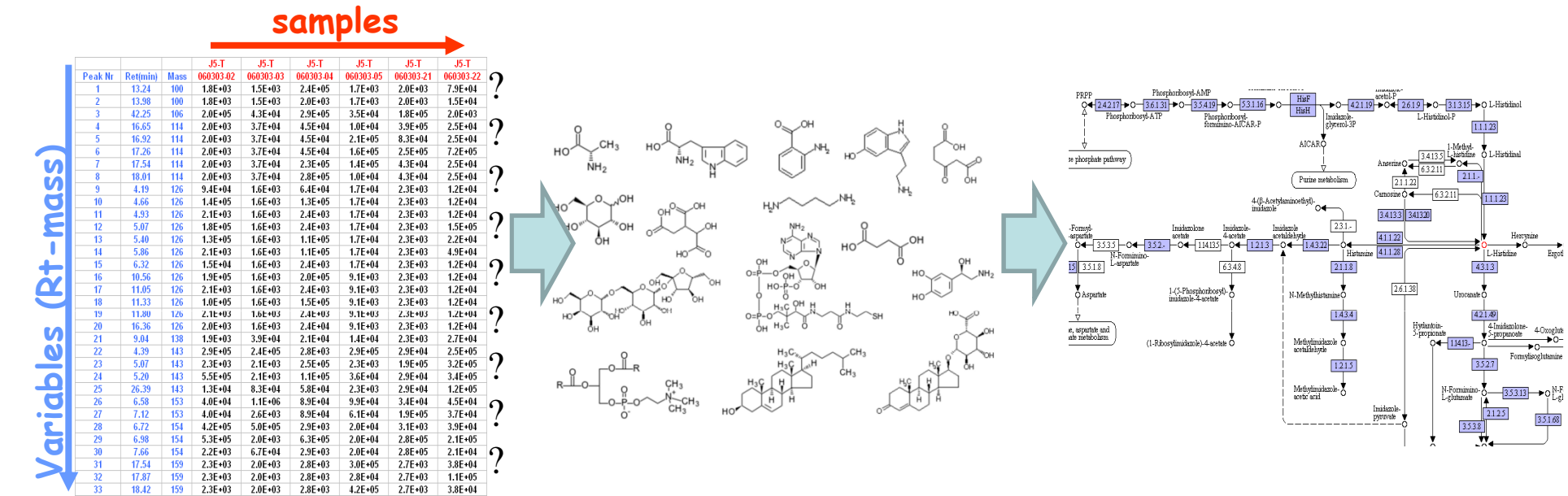
4. Identification



- Peak grouping (CAMERA R package)
- Database annotation
- MSⁿ

ESI-mass spectral database

development for metabolomics: Why ?

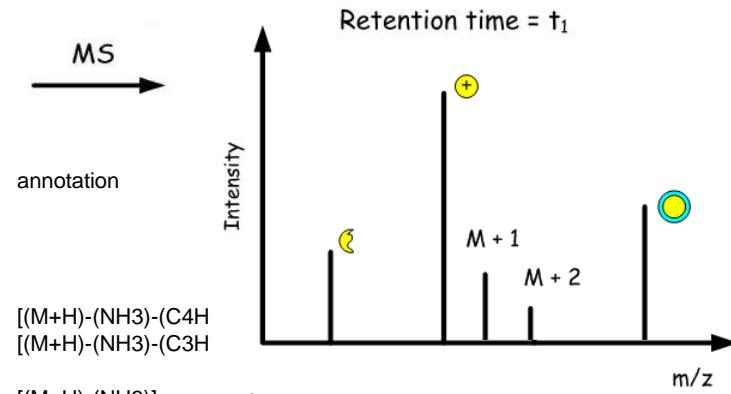
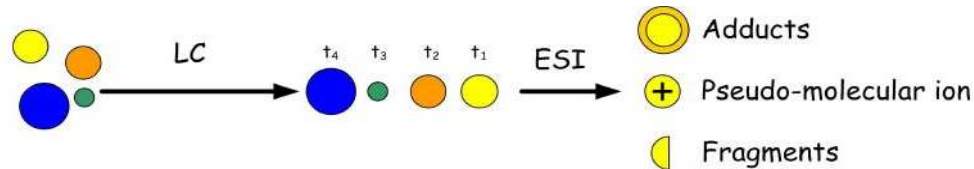


- To annotate easily and quickly peak lists of ions from automatic detection of mass spectra.
- To provide the most comprehensive picture of all the metabolites present in cell extracts or biofluids
- To characterize sources of variability due to physiological factors such as age, gender or body mass index.

ESI-mass spectral database

development for metabolomics : Why ?

énergie atomique = énergies alternatives



var	mzmed	rtmed	mziT	nom
331	94.1035879	41.2395458	93.0963079	1,3-Propanediamine METLIN 3216 109-76-2@1,3-Diaminopropane HMDB HMDB00002 109-76-2@1,3-
164	75.091789	41.4078375	74.084509	Diaminopropane KEGG C00986 109-76-2@
121	72.0808764	41.4121375	71.0735964	3-Buten-1-amine KEGG C12244 IND@
401	102.115188	41.4677549	101.107908	
595	129.139029	41.4682607	128.131749	
404	102.616829	41.5832821	101.609549	
152	74.0879546	41.5832821	73.0806746	
34	58.0649479	41.6414273	57.0576679	Cyclopropylamine KEGG C14150 765-30-0@
327	93.6019358	41.9281549	92.5946558	
775	149.03607	42.3813985	148.02879	
144	73.586346	42.3914349	72.579066	
494	113.115813	42.7321078	112.108533	
834	159.368708	42.7372617	158.361428	
884	164.864699	42.7393498	163.857419	
81	65.072954	42.8461541	64.065674	
848	161.719771	43.2554375	160.712491	
762	147.169448	43.2633375	146.162168	
485	112.112241	43.3123549	111.104961	
870	162.444745	43.3125665	161.437465	
872	162.672941	43.3140799	161.665661	
871	162.582679	43.6324198	161.575399	
235	85.0888656	43.6602673	84.0815856	1-Methylpyrrolinium KEGG C06178 IND@
866	162.299377	43.661028	161.292097	
858	162.081633	43.6615902	161.074353	
371	97.9689677	43.6634508	96.9616877	Spermidine METLIN 254 124-20-9@Spermidine HMDB HMDB01257 124-20-9@Spermidine KEGG C00315 124-20-
756	146.165593	43.6718352	145.158313	9@
857	162.050661	43.7169657	161.043381	
679	133.561403	43.7169657	132.554123	

annotation

[(M+H)-(NH3)-(C4H
 [(M+H)-(NH3)-(C3H

[(M+H)-(NH3)]+ spermidine

[(M+2H)]2+ (13C) spermidine
 [(M+H)-(NH3)-(C4H6)-(NH3)]+ spermidine

[(M+2H)]2+ spermidine
 [(M+H)-2(NH3)]+ (13C) spermidine

[M+H]+ (13C)spermidine
 [(M+H)-2(NH3)]+ spermidine

[M+H]+ spermidine

ESI-mass spectral database development:

Interpretation of mass spectra from reference compounds



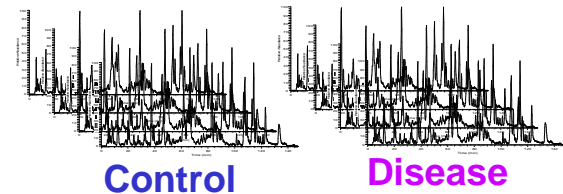
- m/z of molecular ions (adducts, isotopes, in-source fragments)
- Retention time(s)
- MSn experiments

Chemical library containing:

Currently 1200 compounds

500 already analysed

ESI mass spectral
database



Raw Data

Processed data,
Annotated peaklist

Reference compounds analysis

Pantothenic acid [M+H]⁺

Retention time

M/Z	Intensity	Relative	M/Z (theo)	Delta (ppm)	RDBE	Composition	Attribution	Retention time										
72.04431	7.7E+04	0.82	72.04439	-1.11	1.5	C3 H6 O N	[(M+H)-(C6H10O3)-(H2O)]+	1.80	4.77	11.17								
90.05482	3.6E+05	3.78	90.05496	-1.55	0.5	C3 H8 O2 N	[(M+H)-(C6H10O3)]+	1.80	4.77	11.17								
116.034	2.7E+04	0.29	116.03422	-0.22	2.5	C7 H10 O N	[(M+H)-(C2H6O3)-(H2O)]+	1.80	4.77	11.17								
Fragments ions and their isotopes								126	-2.85	2.5	C7 H12 O2 N	[(M+H)-(C2H6O3)]+	1.80	4.77	11.17			
								147	-0.3	1.5	C8 H16 O3 N	[(M+H)-(HCOOH)]+	1.80	4.77	11.17			
								182	0.26	3.5	C9 H14 O3 N	[(M+H)-2(H2O)]+	1.80	4.77	11.17			
								138	-0.46	2.5	C9 H16 O4 N	[(M+H)-(H2O)]+	1.80	4.77	11.17			
								Pseudo-molecular ion and its isotopes								95	0.18	1.5
105	0.2	1.5	C8 13C H18 N O5	[(M+H)]+ (13C)	1.80	4.77	11.17											
194	-2.4	1.5	C9 H18 N O4 18O	[(M+H)]+ (18O)	1.80	4.77	11.17											
149	2.2	1.5	C8 13C H18 N O4 18O	[(M+H)]+ (13C+18O)	1.80	4.77	11.17											
Adducts ions and their isotopes								238.12923	1.76E+03	0.02	238.128515	3	0.5	C9 H20 N O6	[(M+H)+(H2O)]+	1.80	4.77	11.17
								242.09992	6.61E+06	68.15	242.099895	0.1	1.5	C9 H17 N Na O5	[(M+Na)]+	1.80	4.77	11.17
								243.10331	4.44E+05	4.58	243.10325	0.25	1.5	C8 13C H17 N Na O5	[(M+Na)]+ (13C)	1.80	4.77	11.17
								244.10363	1.79E+04	0.18	244.104139	-2.09	1.5	C9 H17 N Na O4 18O	[(M+Na)]+ (18O)	1.80	4.77	11.17
								258.07386	1.29E+05	1.33	258.073833	0.1	1.5	C9 H17 K N O5	[(M+K)]+	1.80	4.77	11.17
								259.07731	1.49E+03	0.02	259.077188	0.47	1.5	C8 13C H17 K N O5	[(M+K)]+ (13C)	1.80	4.77	11.17
								264.082	1.68E+05	1.74	264.08184	0.61	1.5	C9 H16 N Na2 O5	[(M-H+2Na)]+	1.80	4.77	11.17
								265.08498	3.08E+03	0.03	265.085195	-0.81	1.5	C8 13C H16 N Na2 O5	[(M-H+2Na)]+ (13C)	1.80	4.77	11.17
								310.08654	1.10E+03	0.01	310.08732	-2.52	1.5	C10 H18 N Na2 O7	[(M+Na)+(HCOONa)]+	1.80	4.77	11.17
								184.04448	9.45E-05	0.05	184.044569	1.22	2.5	C18 H34 N2 Na O10	[(2M+Na)]+	1.80	4.77	11.17
224	0.99	2.5	C17 13C H34 N2 Na O10	[(2M+Na)]+ (13C)	1.80	4.77	11.17											
279	0.26	2.5	C16 13C2 H34 N2 Na O10	[(2M+Na)]+ (13C2)	1.80	4.77	11.17											
214	0.1	2.5	C18 H33 N2 Na2 O10	[(2M-H+2Na)]+	1.80	4.77	11.17											

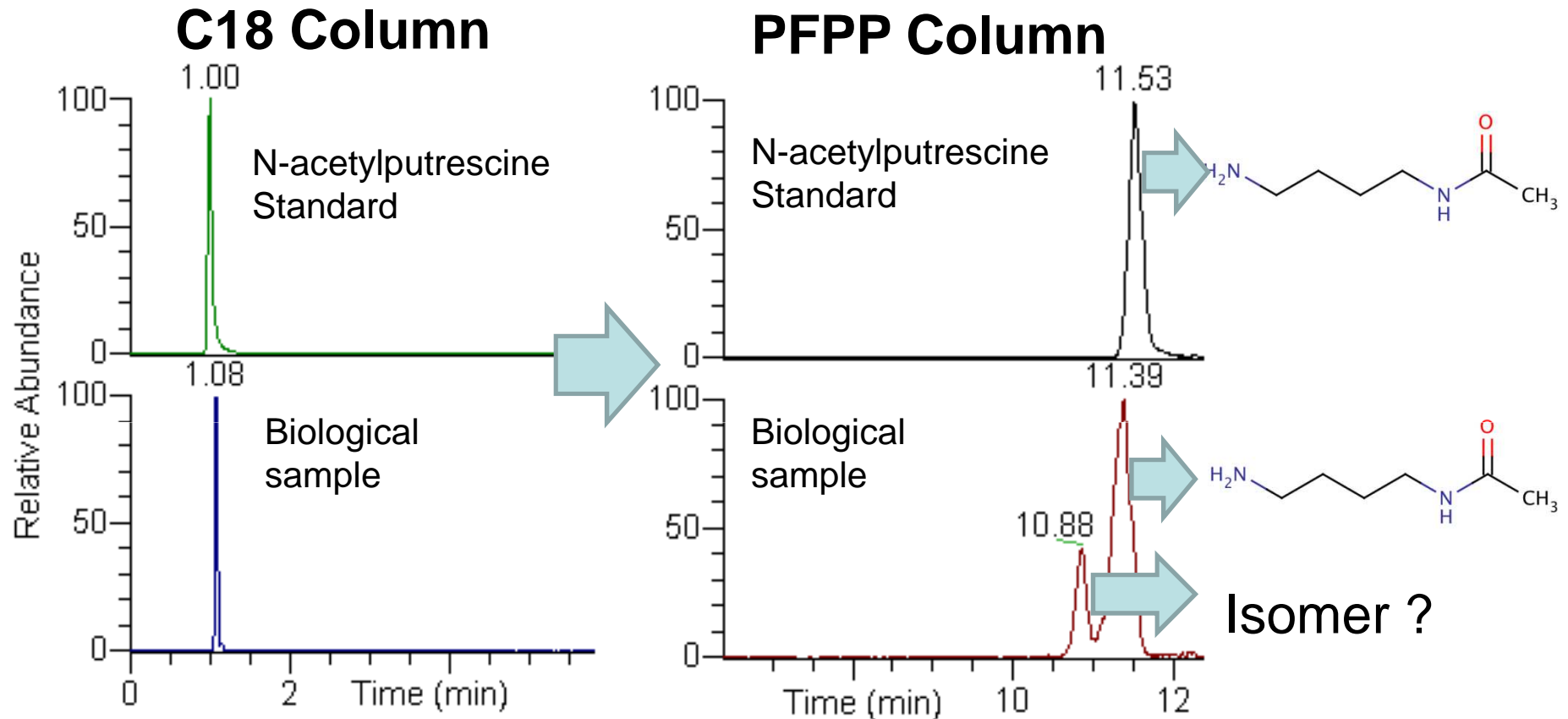


Human urines

RT	MZ	#	Masse	Composé	Composition	Attribution	UPLC (C18)
4.72	202.107624	1	202.10738	pantothenic acid	C9 H16 O4 N	[(M+H)-(H2O)]+	4.77
4.73	90.054915	1	90.05496	pantothenic acid	C3 H8 O2 N	[(M+H)-(C6H10O3)]+	4.77
4.74	220.117542	1	220.11795	pantothenic acid	C9 H18 N O5	[(M+H)]+	4.77
4.74	221.12168	1	221.121305	pantothenic acid	C8 13C H18 N O5	[(M+H)]+ (13C)	4.77
4.74	222.122088	1	222.122194	pantothenic acid	C9 H18 N O4 18O	[(M+H)]+ (18O)	4.77
4.74	242.100637	1	242.099895	pantothenic acid	C9 H17 N Na O5	[(M+Na)]+	4.77
4.74	258.075314	1	258.073833	pantothenic acid	C9 H17 K N O5	[(M+K)]+	4.77

Complexity of identification

énergie atomique • énergies alternatives



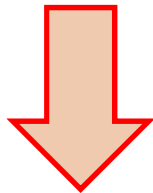
- **38 peaks** (37 in dead volume) with C18 column → **83 peaks** with PFPP column (PentaFluoroPhenylPropyl)
- Many isomers for the same C18 peak = data not interpretable.
- Necessity to use 2 retention systems for identification

ESI-mass spectral database development:

Interpretation of mass spectra from reference compounds



- m/z of molecular ions (adducts, isotopes, in-source fragments)
- Retention time(s)
- MSn



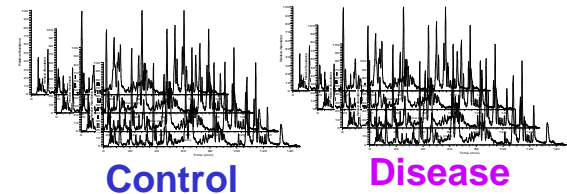
**Not Sufficient = many
detected molecules are
not commercially
available**

ESI mass spectral database

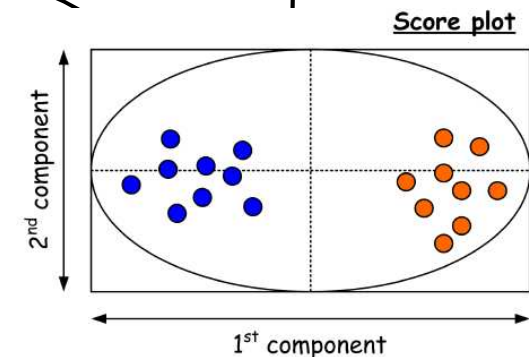
New putatively identified metabolite

- Statistical analysis
- Peak grouping
- Database annotation
- MSn experiments

Metabolite characterization from biological datasets.



Raw Data



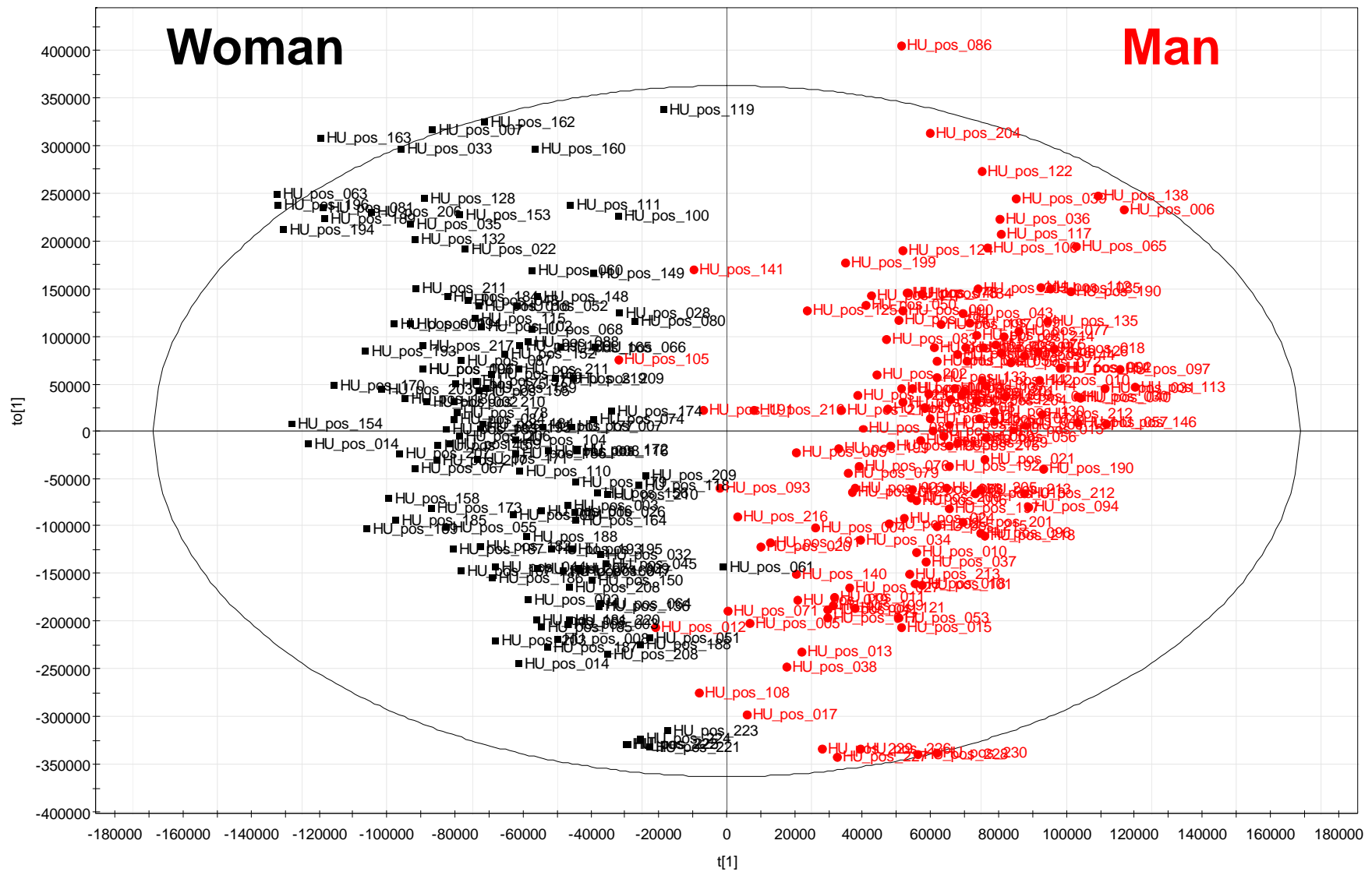
Processed data, Annotated peaklist

Biological dataset analysis: human cohort

- 227 human plasmas and urines (volunteers CEA workers)
- Structuring dataset with biological parameters: age, gender, Body Mass Index

Variables	Classes	Nbr of individual
Sex	Woman	106
	Man	121
Age	18-35 year old	79
	36-45 year old	67
	>46 year old	81
BMI	Normal (18.5-24.9)	150
	Underweight (<18.5)	7
	Overweight (25-29.9)	60
	Obesity (>30)	10

Biological dataset analysis: Statistical Analysis

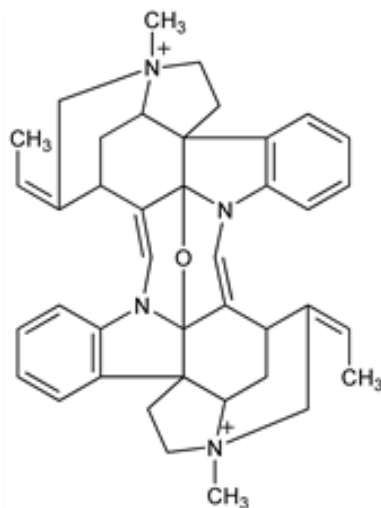


Biological dataset analysis:

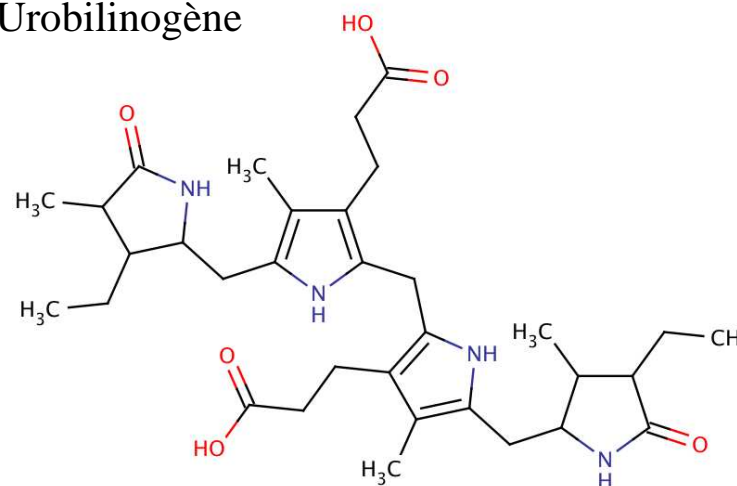
Identification

X	m/z	RT	ANNOTATION CAMERA			Correlation entre échantillons	Annotation par les bases de données publiques
			isotopes	adduct	pcgroup		
1806	303.144298	9.33	**	**	531	NA	**
4663	593.331209	9.34	[681][M]+	**	512	NA	L-Urobilin
4668	594.336833	9.34	[681][M+1]+	**	512	NA	**
4679	595.346349	9.40	[650][M]+	[M-H]-	394	1.00	C-Curarine / L-Urobilinogen
4682	596.351361	9.40	[650][M+1]+	**	394	0.98	**
4878	631.32561	9.40	**	[M+CL]-	394	0.96	**
3797	481.278887	9.46	**	**	552	NA	GPCho(10:0/4:0) / GPCho(12:0/2:0)
2763	381.190985	9.53	**	**	627	NA	**
3834	485.179201	9.61	**	**	544	NA	Rutaevin / Nafenopinglucuronide
1255	253.144008	9.67	**	**	556	NA	**

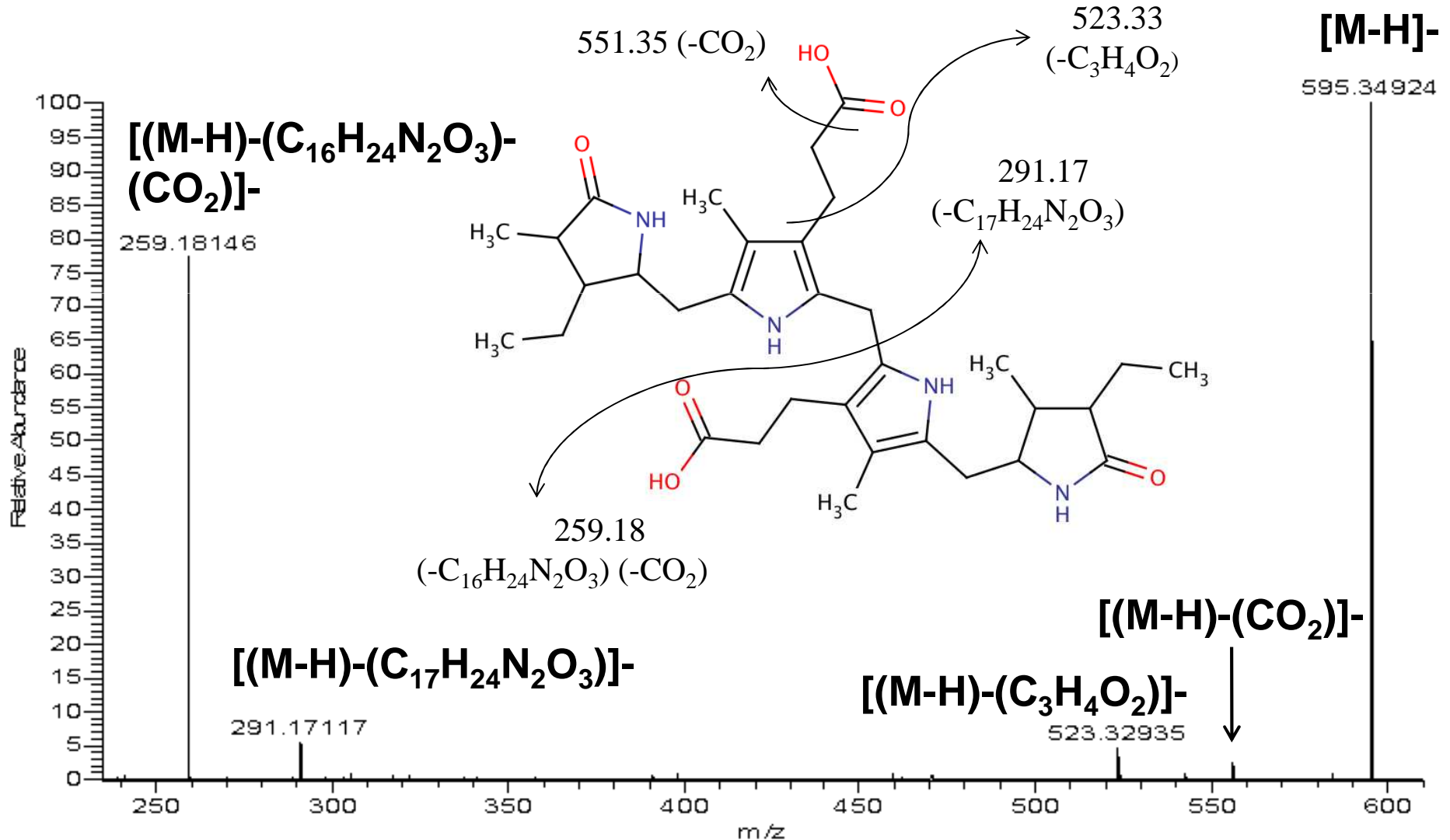
C-Curarine



L-Urobilinogène



Biological dataset analysis: Identification

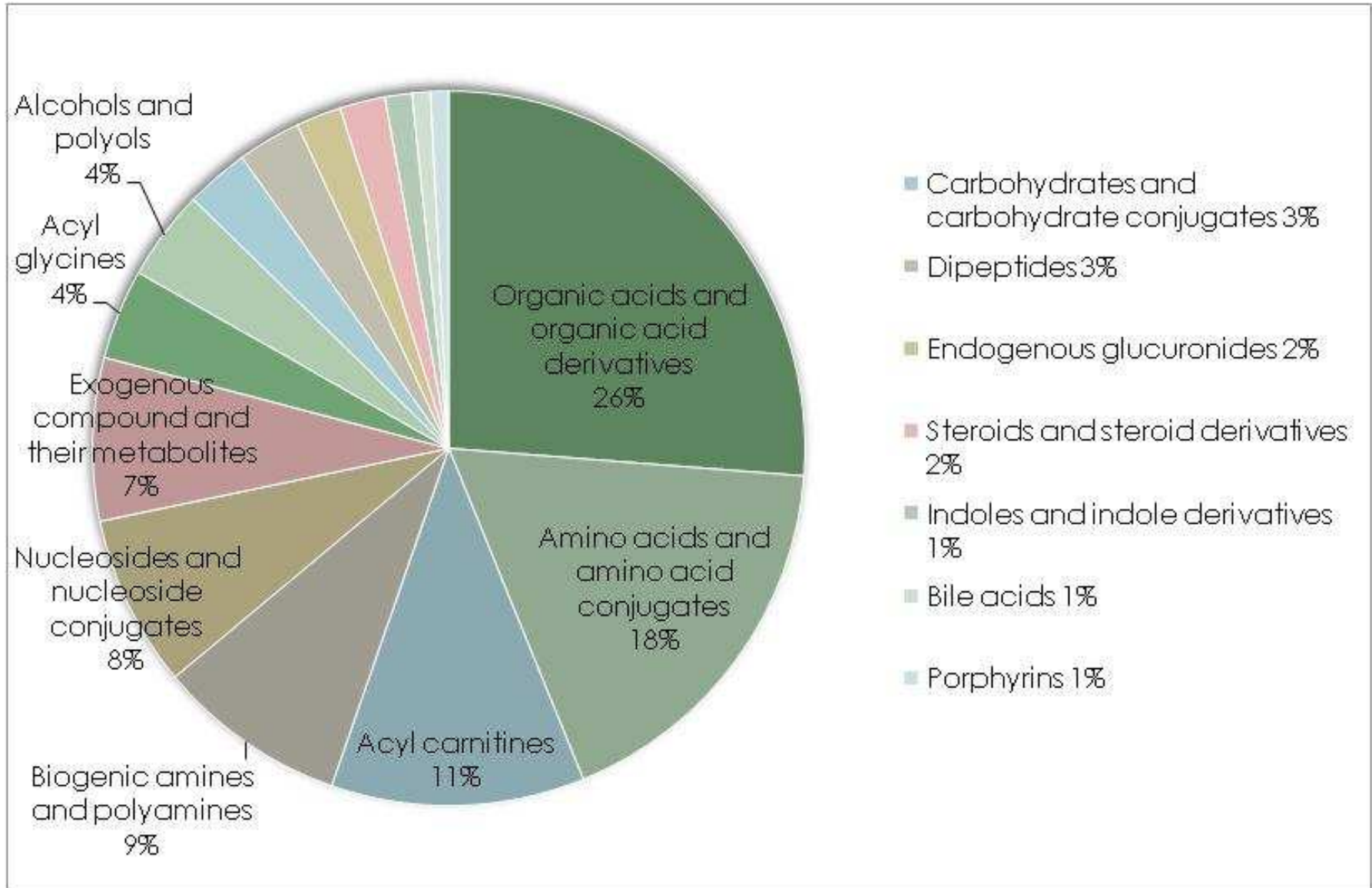


Metabolite identification summary

	Found in urines	Never reported in Human urines
Identified metabolites (from reference compound)	163	19
Putatively identified metabolites (from Biological Dataset)	81	59
Characterized metabolites (based only on accurate mass)	139	/
TOTAL	383	78

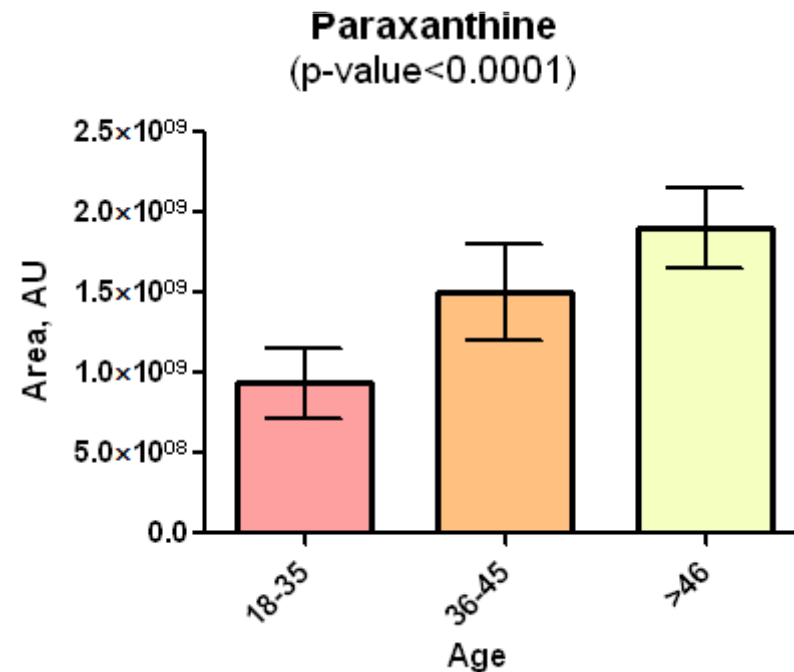
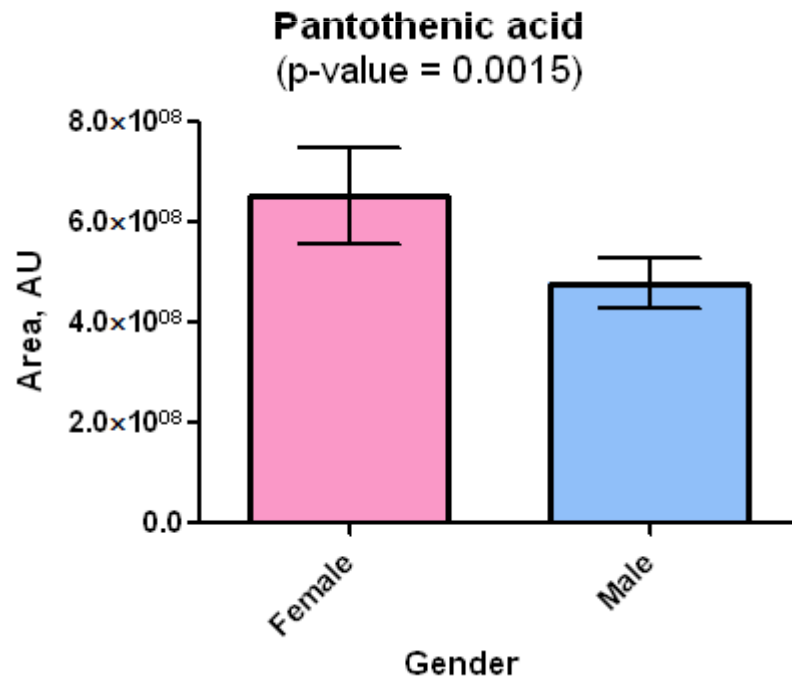
- 244 metabolites were identified or putatively identified, including **78 never reported in Human urines yet**.
- 383 metabolites represent approximately **1000 ions in both ESI (+) and ESI(-)**

Metabolite identification summary



Biological dataset analysis:

	ESI(+)	ESI(-)
Discriminant according to Gender	50	In progress
Discriminant according to age	34	In progress
Discriminant according to BMI	28	In progress
TOTAL OBSERVED	125(/244)	In progress



Perspectives

UPLC / LTQ- Orbitrap



- Chromatographic resolution optimization
- Extension to other mass spectrometers: QTOF, FT-ICR...



ESI mass
spectral
database

Metabolomics relational database

Detection in
biofluids
(urines,
plasmas...)

Physiological
factors (Age,
Gender, BMI...)

Stability
(Bacterial
overgrowth, ...)

Disease,
toxicology



Acknowledgments



CEA/DSV/SPI/LEMM

Jean-François HEILIER

Ying XU

Marie-Françoise OLIVIER

Eric EZAN

Christophe JUNOT

CEA/DRT/LIST/DCSI

Etienne THEVENOT

CEA Occupational

Medicine department

Catherine BOURCIER

Xavier MILLOT



Ministère de la Défense

Délégation Générale pour l'Armement



Laboratoire de Chimie Structurale

Organique et Biologique (Paris 6, CNRS)

Jean-Claude TABET

Thank you for your attention