COST STSM: Preparation of results from a ring analysis of analytical standards
Scientific report

1 Locations of work

- iBET/ITQB NOVA, Oeiras, Portugal
- Faculty of Pharmacy, University of Lisbon, Portugal

2 Objective

The aim of this work was to prepare the data from the analyses of phytochemical standards obtained from 13 different laboratories participating in the project. The data would then be used to assist in the identification of these compounds in future research as well as to validate the capability of each platform to detect the compounds from biological samples.

3 Materials and methods

54 chemical standards obtained from the participating platforms were diluted into an appropriate solvent (water:ACN 1:1, methanol, DMSO, chloroform) based on the information about their solubility from vendors and literature. In case the standard was not dissolved into the target concentration of 10 mM, the suspension was diluted into 5 mM with another solvent or into 1 mM with the same solvent, again based on previous knowledge. Ellagic acid was dissolved in methanol at a concentration of 50 ppm, as this was the maximum concentration possible to achieve based on previous knowledge. The standards list and their chemical information are listed in Supplementary Table 1.

From the 10 mM stock solutions, three standard mixtures were prepared into the concentration of 200 µM (except for the less soluble compounds) as follows:

- Mixture A: hydrophilic compounds dissolved in water:ACN 1:1 or methanol, diluted to 200 µM with water:ACN 1:1
- Mixture B: lipophilic compounds dissolved in methanol:DMSO 1:1 or 100 % DMSO, diluted to 200 µM with methanol
- Mixture C: lipophilic compounds dissolved in chloroform, diluted to 200 µM with chloroform

As a proof of concept, the mixtures were analysed with four different mass spectrometry platforms at iBET/ITQB and Faculty of Pharmacy, University of Lisbon. In each analysis, no prior optimization was performed. The platforms were as follows:

- NanoLC and TripleTOF, Sciex 6000 (iBET)
- MALDI-TOF/TOF, Sciex 4800 (ITQB)
- Finnigan LTQ Linear Ion Trap (ITQB)
- Tandem Quadrupole, Waters Quattro micro API (U. Lisboa), will be performed after the end of this STSM

Data from three different MS platforms were analysed to identify the standards based on the *m/z* values and MS/MS spectra. The observed mass was compared with the calculated mass with regard to the mass accuracy of each mass spectrometer. For a positive tentative identification, the difference between the observed and calculated precursor mass had to be within the range of the expected mass error and the MS/MS spectra had to be similar with the published data, such as METLIN database or literature.

3.1 TripleTOF

For the TripleTOF analysis, mixture A was prepared in Milli-Q water to a concentration of 25 nM, according to the instrument's sensitivity. The column used was a C18-CL reversed phase column with dimensions of 75 µm × 15 cm. The nanoLC conditions were the following: the injection volume 5 µl, flow rate 300 nl/min, pressure 800 mbar. For the gradient, water containing 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B) were used as follows: 0–1 min 95% A, 1–18 min 95%→5% A. The MS experiment was carried out using a collision energy of 10 V, positive ionization mode and an injection volume of 5 µl.

3.2 MALDI-TOF/TOF

The MALDI-TOF/TOF experiment was carried out using three different matrices (sinapinic acid, alpha-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid) and two different methods for applying the aliquots on the sample plate: by using a C-18 capillary and a direct application of the sample. The MS experiment was carried out in both positive and negative ion mode, but the negative MS/MS data had not been acquired previously with the device. The standard mixtures A and B were used in this experiment at a high concentration (200 µM). The intensity of the laser used for the ionization was 5000 units.

3.3 LTQ Linear Ion Trap

The analysis with the LTQ Linear Ion Trap was performed using two approaches: a direct infusion method, where the sample was slowly injected directly into the mass spectrometer at 5 µl/min, and a conventional automated LC method. For the LC method, a Waters XBridge C-18 column with a particle size of 5 µm and dimensions of 2.1 × 150 mm was used. The injection volume was 20 µl, the flow rate 100 µl/min, and the solvents water (A) and acetonitrile (B). The gradient was as follows: 0–2 min 95% A, 2–20 min 95%→5% A, 20–22 min 5% A, 22–24 min 5%→95% A, 24–30 min 95% A. Only mixture A was analysed in this platform, dissolved into Milli-Q water at a concentration of 25 µM. The collision energy for the MS/MS fragmentation was varied between 30 and 50 V for the direct infusion method and 35 V for the automated LC method.

4 Results and discussion

4.1 TripleTOF

The TripleTOF data showed 12 distinguishable peaks at the corresponding *m/z* values of the 26 standards included in the mix. MS/MS data was collected in the automated process for 7 peaks, out of which 5 could be positively identified based on the accurate mass of the parent ion and the MS/MS fragmentation pattern (Supplementary Table 2). The error of the observed *m/z* was 0 to 5 ppm, well within the expected resolution of 10 ppm. However, the concentration of many of the standards was too low for MS/MS data to be produced, and there were impurities present in the mix, which dominated in the total ion chromatogram. To optimise the method, the analysis of the standard mix may be performed at a higher concentration, such as 10 µm (as per SOP), and to detect more compounds, it should be also carried out in the negative ion mode. The TripleTOF shows potential as a suitable platform for metabolomics analyses, owing to its high sensitivity and speed.

4.2 MALDI-TOF/TOF

Only two standards (hesperetin and kaempferol 3-glucuronide) were possible to be tentatively identified from the MALDI-TOF/TOF data (Supplementary Table 3). There were limitations and challenges to analyse small compounds with MALDI. The matrix used for the co-crystallisation of the sample aliquot was chemically similar to some of the standards themselves, or even the same compound in the case of sinapinic acid. Mixture B did not crystallise properly with the matrices used. The method was optimised for the analysis of proteins and peptides in the positive ion mode, while most of the standards are more easily observable in the negative mode. In addition, many of the precursor peaks observed were out of the mass accuracy range to be linked with the standards, and the MS/MS data produced from these precursors did not match with

the reported spectra for most compounds. No MS/MS data was produced for precursors with a mass below 200 Da.

4.3 LTQ Linear Ion Trap

From the LTQ Linear Ion Trap data, 16 compounds out of 34 in the mixture A could be tentatively identified based on the MS/MS data in the positive and negative mode (Supplementary Table 4). For another 5 peaks, the identification could not be done at this point due to lack of reference MS/MS data in the negative ion mode. Despite of the low mass accuracy (0.5 Da), a high lower limit to the observed *m/z* (50) and the weak signals obtained for some of the peaks, the MS/MS data produced was of good quality and compared well with the METLIN database, which contains mostly UPLC-QTOF data. For all the tentatively identified compounds, the mass error was below 0.1 Da. To distinguish between compounds with similar masses, the use of the LC method is required instead of the direct infusion method in order to have a second dimension (retention time) for the data. Furthermore, the use of ESI instead of APCI may increase the ionization of the compounds, thus resulting in stronger signals.

4.4 Deviations from the work plan

Due to delays in receiving all the chemical standards, the data from the different platforms was not ready during the STSM. Therefore, the work focused on progressing the experiment by completing the preliminary data on the standards, preparing all the standard mixtures, and performing initial analysis of the mixtures with a variety of platforms available in the location. After the STSM, a sample of each mixture is ready to be sent to each participating lab, where a 10 µm final solution will be prepared with a solvent optimal for the specific platform. The final version of the SOP was also prepared and will be sent to all the platforms involved. After this STSM, the collaboration will continue at the home institute with the analysis of the data received from all the participating platforms.

5 Supplementary material

Supplementary Table 1. The standards used in this experiment, sorted based on their neutral monoisotopic mass, with the chosen solvent and mixture.

ID	Compound	Formula	MM	Solvent	Notes	Mix
INRA_7	trigonelline	C7H7NO2	137.0477	water/ACN		A
INRA_6	stachydrine (proline betaine)	C7H13NO2	143.0946	methanol		A
FFUL_8	vanillin	C8H8O3	152.0473	methanol		A
CIAL_7	4-hydroxyphenylacetic acid	C8H8O3	152.0473	methanol		A
CIAL_3	protocatechuic acid	C7H6O4	154.0266	methanol		A
FFUL_3	hydroxytyrosol	C8H10O3	154.0630	DMSO		B
UEF_5	p-coumaric acid	C9H8O3	164.0473	methanol		A
UEF_4	3-(4-hydroxyphenyl)propionic acid	C9H10O3	166.0630	water/ACN		A
UEF_9	vanillic acid	C8H8O4	168.0423	methanol		A
NIHS_5	gallic acid	C7H6O5	170.0215	methanol		A
NIHS_4	hippuric acid	C9H9NO3	179.0582	methanol		A
NIHS_3	caffeic acid	C9H8O4	180.0423	methanol		A
UEF_3	myo-inositol	C6H12O6	180.0634	water/ACN		A
INRA_2	theobromine	C7H8N4O2	180.0650	water/ACN		A
CIAL_4	dihydrocaffeic acid	C9H10O4	182.0579	methanol		A
FFUL_7	homovanillic acid	C9H10O4	182.0579	DMSO		B
CIAL_5	veratric acid	C9H10O4	182.0579	methanol		A
ITQB_2	catechol-O-sulfate	C6H6O5S	189.9936	DMSO	1 mM; 10 % MeOH	B
NIHS_1	ferulic acid	C10H10O4	194.0579	methanol		A
INRA_11	bergaptol	C11H6O4	202.0270	methanol:DMSO 1:1	5 mM	B
ITQB_1	Pyrogallol-2-O-sulfate	C6H6O6S	205.9885	methanol		A
CEBAS_2	urolithin B	C13H8O3	212.0473	DMSO		B
UEF_6	sinapic acid	C11H12O5	224.0685	methanol	1 mM	A
CEBAS_1	urolithin A	C13H8O4	228.0423	DMSO		B
FFUL_10	resveratrol	C14H12O3	228.0786	methanol		A
ITQB_3	4-O-methylgallic acid 3-O-sulfate	C8H8O8S	263.9940	methanol		A
FFUL_2	genistein	C15H10O5	270.0528	DMSO		B
FFUL_9	phloretin	C15H14O5	274.0841	DMSO		B
FFUL_5	kaempferol	C15H10O6	286.0477	DMSO		B
FFUL_6	luteolin	C15H10O6	286.0477	methanol:DMSO 1:1	5 mM	B
FFUL_1	(-)-epicatechin	C15H14O6	290.0790	DMSO		B
UEF_7	ellagic acid	C14H6O8	302.0063	methanol	50 ppm	A
INRA_4	hesperetin	C16H14O6	302.0790	methanol		A

FFUL_4	isorhamnetin	C16H12O7	316.0583	methanol:DMSO 1:1	0.5 mM	B
INRA_8	cafestol	C20H28O3	316.2038	methanol		A
SLU_1	alkylresorcinol C17:0	C23H40O2	348.3028	DMSO		B
NIHS_2	chlorogenic acid	C16H18O9	354.0950	methanol		A
INRA_1	rosmarinic acid	C18H16O8	360.0845	methanol		A
UEF_1	(+)-lariciresinol	C20H24O6	360.1573	methanol		A
CIAL_2	curcumin	C21H20O6	368.3800	methanol:DMSO 1:1	5 mM	B
INRA_10	tangeretin	C20H20O7	372.1210	DMSO		B
CAS_1	quercetin 3'-O-sulfate	C15H10O10S	381.9995	methanol	1 mM	A
CAS_2	quercetin 4'-O-sulfate	C15H10O10S	381.9995	methanol	1 mM	A
FFUL_11	beta-sitosterol (containing other plant sterols)	C29H50O	414.3862	chloroform		C
FFUL_13	(+-)-alpha-tocopherol	C29H50O2	430.3811	methanol		A
UEF_2	apigenin 7-O-glucoside	C21H20O10	432.1056	methanol:DMSO 1:1	5 mM	B
INRA_3	ursolic acid	C30H48O3	456.3603	methanol		A
SLU_2	alkylresorcinol C25:0	C31H56O2	460.4280	DMSO	5 mM	B
INRA_9	kaempferol 3-glucuronide	C21H18O12	462.0798	methanol		A
CAS_3	quercetin disulfate	C15H10O12S3	477.9330	methanol:DMSO 1:1	0.5 mM	B
FFUL_12	beta-carotene	C40H56	536.4382	chloroform		C
INRA_5	procyanidin A2	C30H24O12	576.1268	water/ACN		A
UEF_8	naringin	C27H32O14	580.1792	methanol		A
CIAL_1	verbascoside	C29H36O15	624.2054	water/ACN		A

Supplementary Table 2. The results from the TripleTOF experiment. The masses of the parent ions outside the reported mass error range are highlighted in red.

Observed mass	Calculated mass	Δppm	Mode	Observed fragments	Collision energy	Matching fragments	Tentative identification
165.0756	165.0551	124	pos	n/a			
225.0764	225.0763	0	pos	207.067, 175.045	10 V	207.065, 175.038	sinapic acid
303.0878	303.0868	3	pos	153.023, 177.061	10 V	153.018, 177.054	hesperetin
317.1968	317.2116	-47	pos	n/a			
361.0934	361.0923	3	pos	163.039	10 V	163.039	rosmarinic acid
463.0888	463.0873	3	pos	287.061	10 V	287.055	kaempferol 3-glucuronide
581.19	581.187	5	pos	273.083, 153.020	10 V	*273, 153*	naringin

Supplementary Table 3. The results from the MALDI-TOF/TOF experiment. The masses of the parent ions outside the reported mass error range are highlighted in red.

Observed mass	Calculated mass	Δppm	Mode	Observed fragments	CID energy	Matching fragments	Tentative identification
303.09021	303.0868	11.25	pos	177.0281 (67), 152.9937 (44), 287.0434 (23), 179.0074 (15), 274.0156 (14), 300.4034 (13), 275.1251 (12)	35 V	177.05, 153.02, 179.03	hesperetin
463.063324	463.0876	-52.42	pos	287.0128 (86), 414.9301 (25), 416.9123 (20)	35 V	287.05, 417.08	kaempferol 3-glucuronide*
317.170746	317.2116	-128.79	pos	287.0687 (93), 297.1503 (62), 285.1458 (41), 288.0654 (38), 298.1150 (32), 286.1140 (22)	35 V	n/a	
478.96405	478.9408	48.54	pos	417.0448 (90), 265.9296 (58), 432.956 (51), 255.9819 (21), 249.9758 (17), 38.9541 (17)	35 V	n/a	
273.016174	273.0763	-220.18	neg	262.5565 (91), …, 144.4066 (11), 79.20494 (10), 188.4648 (10)	35 V	n/a	
353.035156	353.0872	-147.40	neg	343.6856 (101), 351.4691 (90), 339.894 (63), …, 341.8889 (22), 144.3434 (18)	35 V	n/a	
359.074402	359.0767	-6.40	neg	351.6351 (99), 311.5919 (39), 343.6948 (35), 350.7112 (30), 353.5537 (26)	35 V	n/a	
380.972015	380.9917	-51.67	neg	381.08 (94), 287.5533 (87), 381.3497 (84), 380.8501 (82), 375.3777 (34), 380.541 (34)	35 V	n/a	
575.243286	575.119	216.10	neg	407.5274 (64), 554.4589 (52), 571.7621 (48), 604.3528 (45)	35 V	n/a	
623.965698	623.1976	1232.51	neg	576.6002 (90), 431.3836 (42), 213.3539 (7), 142.4147 (6)	35 V	n/a	
262.900085	262.9862	-327.45	neg	261.0654 (62)	35 V	n/a	
289.028595	289.0712	-147.39	neg	287.6317 (62)	35 V	n/a	
315.075195	315.0505	78.38	neg	311.6323 (82), 144.3526 (70), 311.3907 (38), 310.532 (18)	35 V	n/a	
371.044373	371.1132	-185.46	neg	354.7432 (95), 355.761 (78), 311.7228 (45), 371.007 (41)	35 V	reference data to be acquired	
431.007446	431.0978	-209.59	neg	371.7032 (96), 377.6593 (40), 421.6316 (34), 415.7568 (31), 420.6152 (11)	35 V	n/a	

* Mass error slightly above the reported accuracy but considered acceptable

Supplementary table 4. The LTQ Linear Ion Trap results.

Observed mass	Calculated mass	ΔDa	Mode	Observed fragments	CID energy	Matching fragments	Tentative identification	Notes
138.04	138.0555	0.02	pos	110.02 (100), 138.04 (74), 93.97 (51), 121.06 (19)	40 V	94.07	trigonelline	
144.03	144.1024	0.07	pos	144.03 (100), 84.07 (51), 58.04 (20), 101.97 (14), 98.08 (10)	30 V	98.10, 84.08	proline betaine	
152.99	153.0551	0.07	pos	124.99 (100), 152.99 (41), 93.01 (4)	20 V	93.01, 125.06	vanillin	
181.00	181.0501	0.05	pos	137.02 (100), 163.00 (98), 181.00 (76), 137.99 (52), 110.07 (10), 135.05 (4), 96.00 (3)	50 V	163.04	caffeic acid	fragment 110 may be from theobromine!
303.18	303.0868	-0.09	pos	177.00 (100), 178.98 (39), 153.03 (17), 285.03 (15), 261.02 (4), 151.08 (3), 303.18 (3), 219.04 (3)	20 V	177.05, 153.02, 179.03	hesperetin	
383.33	383.0073	-0.32	pos	221.05 (100), 184.94 (58), 203.04 (9), 365.03 (6), 382.66 (3), 339.19 (3)	20 V	n/a	n/a	
581.05	581.187	0.14	pos	563.21 (100), 416.76 (54), 456.74 (37)	20 V	n/a	n/a	
151.04	151.0395	0.00	neg	136.01 (100), 151.09 (3)	30 V	136.02	*vanillin*	151.04 from 4-hydroxyphenylacetic acid?
204.98	204.9807	0.00	neg	124.99 (100), 123.05 (83), 204.98 (23), 80.03 (2)	30 V	reference data?		
262.98	262.9862	0.01	neg	183.02 (100), 168.02 (6), 262.98 (5)	20 V	reference data?		
301.05	301.0712	0.02	neg	286.04 (100), 242.11 (52), 301.05 (38), 283.06 (37), 257.09 (28), 125.01 (17), 258.09 (16), 199.11 (13)	20 V	286.05, 242.06, 199.04, 283.06, 257.08, 125.02	*hesperetin*	
353.06	353.0872	0.03	neg	191.03 (100), 179.03 (7)	20 V	191.06	chlorogenic acid	
359.09	359.0767	-0.01	neg	161.00 (100), 197.10 (26), 179.02 (25), 223 (15)	20 V	161.02, 197.04, 179.03	rosmarinic acid	
380.99	380.9917	0.00	neg	301.04 (100)	20 V	301.03 is the negative ion of quercetin	quercetin sulfate	cannot differentiate between the isomers
461.07	461.072	0.00	neg	285.04 (100)	20 V	285.04 is the negative ion of kaempferol	kaempferol 3-glucuronide	
623.17	623.1976	0.03	neg	461.16 (100)	20 V	reference data?		
153.03	153.0188	-0.01	neg	108.99 (100)	35 V	109.03	protocatechuic acid	
163.06	163.0395	-0.02	neg	119.03 (100)	35 V	*119*	p-coumaric acid	
167.00	167.0345	0.03	neg	123.02 (100), 152.08 (71), 107.98 (13), 167.10 (2)	35 V	*123*	vanillic acid	

169.06	169.0137	-0.05	neg	125.02 (100), 169.05 (3)	35 V	125.02	gallic acid	
178.04	178.0504	0.01	neg	134.06 (100), 178.13 (2)	35 V	134.06	hippuric acid	
179.06	179.0345	-0.03	neg	135.01 (100), 179.07 (3)	35 V	135.04	caffeic acid	
181.07	181.0501	-0.02	neg	137.05 (100), 123.05 (15), 166.10 (7), 119.10 (4), 109.04 (2), 181.12 (2)	35 V	137.06		not possible to distinguish between veratric / homovanillic / dihydrocaffeic acids
193.07	193.0501	-0.02	neg	134.02 (100), 149.03 (45), 178.10 (28)	35 V	134.04, 178.03	ferulic acid	
226.98	227.0345	0.05	neg	112.96 (100)	35 V	reference data?		
455.37	455.3525	-0.02	neg	407.42 (100), 455.36 (7), 439.47 (2)	35 V	reference data?		