Supporting Information

Systematic isolation and structure elucidation of urinary metabolites optimized for the analytical-scale molecular profiling laboratory

Luke Whiley [1,2 #], Elena Chekmeneva [1 #], David J Berry [1], Beatriz Jiménez [1], Ada H. Y. Yuen [1], Ash Salam [1], Humma Hussain [1], Matthias Witt [3], Zoltan Takats [1], Jeremy Nicholson [1, †], Matthew R. Lewis [1*]

[1] The MRC-NIHR National Phenome Centre and Imperial BRC Clinical Phenotyping Centre, Imperial College London, London, United Kingdom

[2] UK Dementia Research Institute, Burlington Danes Building, Imperial College London, Hammersmith Hospital, London, W12 0NN, United Kingdom

[3] Bruker Daltonik GmbH, MRMS Solutions, Bremen, Germany

[†] Present address: Australian National Phenome Centre, Murdoch University, Harry Perkins Building, Perth, Western Australia, Australia

# These authors contributed equally

[*] Corresponding author (s)

Dr. Matthew R Lewis, email: matthew.lewis@imperial.ac.uk, telephone: +44 (0) 20 7594 3108

(ORCID ID 0000-0001-5760-5359)

This supporting information document contains an extended **Materials and Methods** section consisting of: Secondary and tertiary isocratic fractionation; MRMS analysis; NMR analysis

The supporting information also contains four tables: **Table S1** contains the chromatographic conditions used for the secondary and tertiary isocratic separations of each features of interest; **Table S2** shows the NMR signal assignment of Feature A, tetrahydropentoxyline; **Table S3** shows the NMR signal assignment of Feature B, indole-3-acetic-acid-O-alpha-glucuronide; **Table S4** shows the NMR signal assignment of Feature D, pregnanediol-3-glucuronide.

The supporting information also contains eighteen figures: **Figure S1** shows photographs and metabolite feature heatmaps of desalted urine and non-desalted control urine; **Figure S2** shows a principal component analysis scores plot comparing repeat LC-MS profiles of repeat injections of desalted and untreated urine; **Figure S3** presents the fraction bank protocol cycle; **Figure S4** shows an MS/MS spectra of Feature A; **Figure S5** shows an ES+ MRMS with an isotopic fine structure confirmation of elemental composition ($C_{17}H_{23}N_2O_7$) of purified Feature A; **Figure S6** shows the ^{1}H-

^{1}H COSY NMR spectrum of Feature A; **Figure S7** shows the ^{1}H-^{13}C HMBC NMR spectrum of Feature A; **Figure S8** Extracted ion chromatograms (EIC) of the m/z 350.088 and the co-eluting feature m/z 187.007; **Figure S9** Extracted ion chromatograms (EIC) and MS/MS spectra of features with m/z 350.088 (ES-); **Figure S10** shows an ES- MRMS with an isotopic fine structure confirmation of elemental composition ($C_{16}H_{16}NO_8$) of purified Feature B; **Figure S11** shows evidence of the degradation of purified Feature B when stored in phosphate buffer; **Figure S12** shows the ^{1}H-^{1}H COSY NMR spectrum of Feature B; **Figure S13** shows the ^{1}H-^{13}C HMBC NMR spectrum of Feature B; **Figure S14** Negative MS/MS spectrum and statistical heterospectroscopy analysis of feature C; **Figure S15** shows EIC chromatogram (ES-) comparing urine pool with analytical standard of pregnanediol-3-glucuronide; **Figure S16** shows an ES- MRMS with an isotopic fine structure confirmation of elemental composition ($C_{27}H_{43}O_8$) of purified Feature C; **Figure S17** shows the ^{1}H-^{1}H COSY NMR spectrum of Feature D; **Figure S13** shows the ^{1}H-^{13}C HMBC NMR spectrum of Feature D;

MATERIALS AND METHODS

Secondary and tertiary isocratic fractionation

To generate the library of isocratic chromatography conditions, four reversed-phase C18 chemistries were assessed (Waters X-Bridge BEH C18; Waters X-select CSH C18, Waters, Sunfire C18, Waters Atlantis T3 C18)). Columns of dimensions 4.6mm x 150mm,3µm were used in each case. Two solvent combinations were employed: the first consisted of water (0.1% formic acid) and acetonitrile (0.1% formic acid), whilst the second replaced acetonitrile with methanol as the organic modifier. Overall, eleven sets of reversed phase isocratic conditions were assessed using two solvent gradients (99% A, 95% A and then at 5% interval decreases to 50% A).

It should also be noted that an additional Waters X-Bridge BEH C8 column with selectivity similarities with the X-Bridge C18 was used in cases where the X-Bridge C18 required further bespoke modification.

Two HILIC phase chemistries were additionally assessed for the database (Waters X-Bridge HILIC; Waters Atlantis HILIC). The column dimensions were 4.6mm x 150mm, 3µm in each case. For the HILIC database only one mobile phase combination was employed consisting of 20mM ammonium formate in water (A) and 0.1% formic acid in acetonitrile (B). Ten sets of HILIC isocratic conditions were assessed (5% A and then at 5% interval increases up to 50% A).

MRMS analysis

Magnetic resonance mass spectra (MRMS) were acquired with a Bruker solariX 2xR (Bruker Daltonics, Billerica, MA, US) using electrospray ionization (ES) and direct infusion with syringe pump. The fraction samples were diluted 1:9 in 50% acetonitrile + 0.1% formic acid for ES+. Fraction samples were diluted 1:9 in 50% acetonitrile ES-. Samples were measured with direct infusion using a flow of 2 µl/min. Mass spectra were acquired with a mass resolution of 1.350.000 at m/z 200 using quadrupolar detection. 64 single scans were added for the final mass spectrum. Spectra were externally mass calibrated with NaTFA cluster.

NMR analysis

A range of 2D NMR spectroscopy experiments including J-resolved, ^{1}H,^{1}H-COSY, ^{1}H,^{1}H-TOCSY, 2D ^{1}H,^{1}H-NOESY, ^{1}H,^{13}C-HSQC, and ^{1}H,^{13}C-HMBC were performed for structural elucidation of the purified metabolites. The relaxation delays were set to 2 s to reduce the experimental time, and all pulse sequences included a pre-saturation period to suppress water signal. Non-uniform sampling acquiring 25% or 50% of the points in the indirect dimension was applied in some of the 2D experiments in order to further reduce experimental time. The ^{1}H spectral window was set to 12.0 ppm and either 190 or 230 ppm for the ^{13}C windows of HSQC or HMBC respectively. The centre of the ^{1}H window was centered on the water signal and either at 85 or 105 ppm for the indirect dimension of HSQC and HMBC, respectively.

Table S1. Specific chromatographic conditions for the secondary and tertiary isocratic separations of the features of interest. FA – formic acid

	Feature A	**Feature B**	**Feature C**	**Feature D**
	Secondary fractionation – Isocratic conditions			
Column	Atlantis T3 C_{18}	Atlantis T3 C_{18}	Atlantis T3 C_{18}	Atlantis T3 C_{18}
Mobile phase A	Water (0.1% FA)	Water (0.1% FA)	Water (0.1% FA)	Water (0.1% FA)
Mobile phase B	Methanol (0.1% FA)	Acetonitrile (0.1% FA)	Acetonitrile (0.1% FA)	Acetonitrile (0.1% FA)
Isocratic solvent composition	90% A	80% A	80% A	65% A
RT (min)	24.55	7.25	7.15	22.58
	Tertiary fractionation – Isocratic conditions			
Column	X-Select C_{18}	X-Bridge C_8	X-Bridge C_8	Atlantis T3 C_{18}
Mobile phase A	Water (0.1% FA)	Water (0.1% FA)	Water (0.1% FA)	Water (0.1% FA)
Mobile phase B	Acetonitrile (0.1% FA)	Acetonitrile (0.1% FA)	Acetonitrile (0.1% FA)	Acetonitrile (0.1% FA)
Isocratic solvent composition	99% A	90% A	90% A	35% A
RT (min)	9.60	20.70	17.87	14.64

Table S2. NMR signal assignment of Feature A, tetrahydropentoxyline.

ID	δppm		# of Hs	Type	Multiplicity		Connectivity Correlations		
	1H	13C			Type	J (Hz)	HMBC (ppm)	COSY	TOCSY
1	7.54	114.7	1	CH	d	8.53	122.7, 129.8	7.35	7.75, 7.32, 7.22
2	7.32	125.7	1	CH	t	7.6	121.6, 138.9	7.75, 7.22	7.75, 7.54, 7.22
3	7.23	122.7	1	CH	t	7.6	114.7, 129.8	7.54, 7.32	7.75, 7.54
4	7.75	121.7	1	CH	d	7.87	125.7, 138.9	7.22	7.54, 7.32, 7.21
5		138.9		C					
6		129.8		C					
7		136.1		C					
8		111.0		C					
9	3.56	28.7	1	CH	dd	15.6, 5.2	136.1, 129.8, 111.0	4.04	4.04, 3.36
9	3.36	28.7	1	CH	dd	15.2, 9.1	136.1, 129.8, 111.0, 58.0	4.04	4.04, 3.57
10	4.04	58.1	1	CH	dd	9.6, 5.2	111.0, 178.0	3.56, 3.35	3.57, 3.36
11		178.0		COOH					
12	5.18	68.9	1	CH	d	8.53	70.44, 111.0, 136.1	4.44	4.44, 4.14
1'	4.44	70.5	1	CH	dd	8.53, 3.4		5.18, 4.14	5.18, 4.14, 3.96
2'	4.14	73.2	1	CH	t	4		4.44, 3.96	5.18, 4.44, 3.96
3'	3.96	71.8	1	CH	t	4.3		4.13, 3.9	4.44, 4.27, 4.14, 3.75
4'	3.91	82.0	1	CH	dt	8.9, 3.5		3.97	4.27, 4.14, 3.75
5'	4.27	61.9	1	CH	dd	12.8, 9.13	82	3.90, 3.75	3.90, 3.75
5'	3.75	61.9	1	CH	dd	12.7, 3.1		4.27	4.27, 3.91

doublet (d); doublet of doublets (dd); triplet (t); doublet of triplets (dt)

Table S3. NMR signal assignment of Feature B, indole-3-acetic-acid-O-alpha-glucuronide.

ID	δppm		# of	Type	Multiplicity		Connectivity Correlations		
	13C	1H	Hs		Type*	Coupling (Hz)	HMBC (ppm)	COSY	TOCSY
1		10.18	1	NH	br.s			2	2
2	127.9	7.36	1	CH	d	2.46	109.2; 129.3; 139.3	1	1; 10
3	109.2		0	C					
4	129.3		0	C					
5	139.2		0	C					
6	121.4	7.66	1	CH	ddd	8.23, 2.46, 0.9	125.1; 139.2	7	7
7	122.1	7.2	1	CH	t	7.57 (1.86, 1.5)	114.6; 129.3	6	6
8	125.1	7.27	1	CH	t	7.60 (1.86, 1.5)	121.4; 139.2	9	9
9	114.6	7.54	1	CH	dt	7.87, 1.86	122.1; 129.3	8	8
10	33	4.03	2	CH2	d	1.86	109.2; 127.9; 129.3; 176.5		2
11	176.5		0	CO					
1'	96.6	5.62	1	CH	d	7.87	176.5	2'	2'; 3'
2'	74.4	3.55	1	CH	t	9.49	96.6	1'	
3'	74.7	3.6	1	CH	t	9.19	74.4		1'; 5'
4'	74.6	3.56	1	CH	t			5'	5'
5'	79.3	3.84	1	CH	d	9.79		4'	3'; 4'
6'			0	COOH					

broad singlet (br.s); doublet (d); doublet of doublet of doublets (ddd); triplet (t); doublet of triplets (dt)

Table S4. NMR signal assignment of Feature D, pregnanediol-3-glucuronide.

ID	δppm		# of Hs	Type	Multiplicity		Connectivity Correlations		
	1H	13C			Type	Coupling (Hz)	HMBC (ppm)	COSY (ppm)	NOE (ppm)
1	0.96/1.84	37.5	2	CH2	td / dt	14.3, 3.3	42.9	1.77, 1.35	
2	1.35/1.77	29.2	2	CH2	m			0.96, 1.84	
3	3.83	83	1	CH	m		102.7	1.34, 1.88	
4	1.58/1.88	36.4	2	CH2	m		83	1.425	1.58 ppm with 1.27 ppm
5	1.425	44.7	1	CH	m				
6	1.27/1.87	29.6	2	CH2	m			1.14	
7	1.14/1.43	29	2	CH2	m		38.02		
8	1.44	38.02	1	CH	m				
9	1.44	42.9	1	CH	m				
10		40.2		C					
11	1.25/1.44	23.1	2	CH2	m				
12	1.20/188	41.9	2	CH2	m		23.1	1.43/1.25	
13		44.1		C					
14	1.155	58.9	1	CH	m		38.02, 26.5, 14.5	1.65, 1.44	
15	1.145/1.65	26.5	2	CH2	m				1.145 ppm with 0.64 ppm
16	1.875/1.47	29.3	2	CH2	m		29.3	1.36, 1.145	1.47 ppm with 0.64 ppm
17	1.36	60.5	1	CH	m		14.5, 44.1, 74.2	1.47	
18	0.64	14.5	3	CH3	s		41.9, 44.1, 58.9, 60.5		1.875
19	0.94	25.4	3	CH3	s		36.4, 37.5, 42.9, 44.7		1.88, 1.425
20	3.7	74.2	1	CH	m		29.3, 60.5	1.22, 1.36	0.64, 1.47
21	1.22	25.1	3	CH3	d	6.3	60.5, 74.1	3.7	0.64
1'	4.59	102.7	1	CH	d	8.1	83	3.27	3.83, 1.77
2'	3.27	76.01	1	CH	t	8.7	78.6	3.50 , 4.59	
3'	3.5	78.8	1	CH	m		74.8	3.27	
4'	3.51	74.8	1	CH	m		78.4		
5'	3.69	79.2	1	CH	d	9.8	74.8	3.51	
6'		178.4		COOH					

singlet (s); doublet (d); triplet (t); multiplet (m); triplet of doublets (td); doublet of triplets (dt)

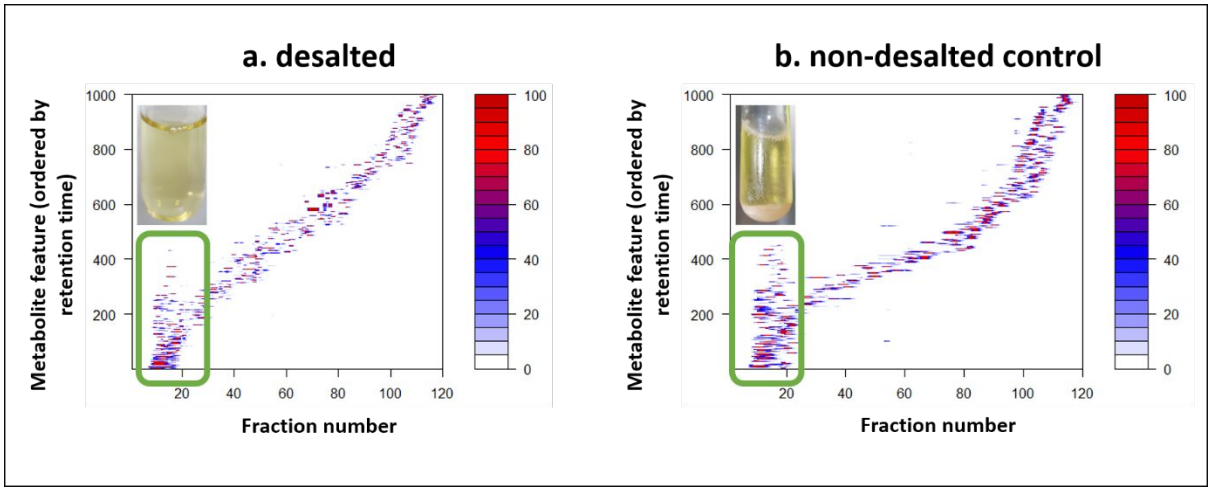

Figure S1. Photographs and metabolite feature heatmaps of desalted urine (a) and non-desalted control urine (b) after 10-fold pre-concentration. Significant salt precipitation is seen in the control concentrated urine. Heatmaps representing the distribution of the top 1000 features (ranked by MS intensity, x-axis) across the 120 collected fractions (y-axis) after 100 repeat fraction collections from 1 mL injections of urine.

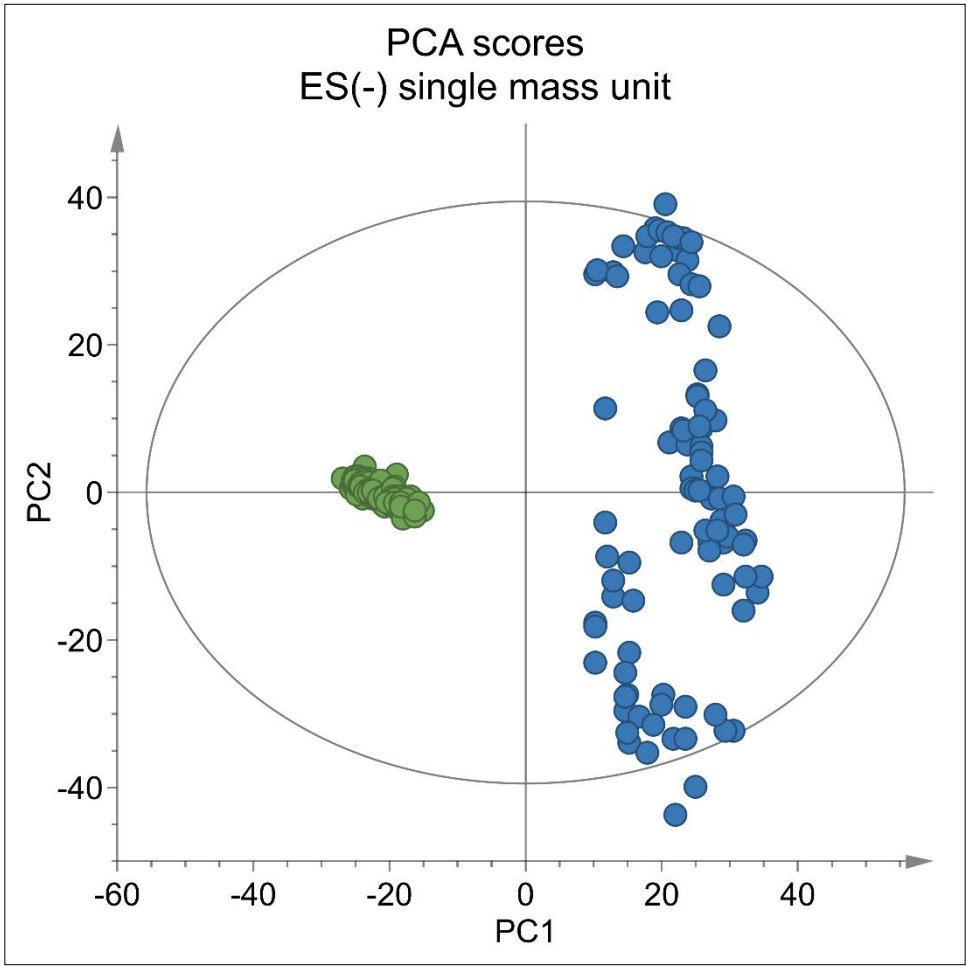

Figure S2. Principal component analysis scores plot comparing LC-MS profiles (unit mass resolution) of 100 repeat 1mL injections of desalted 10x (green, left) and untreated 10x urine (blue, right).

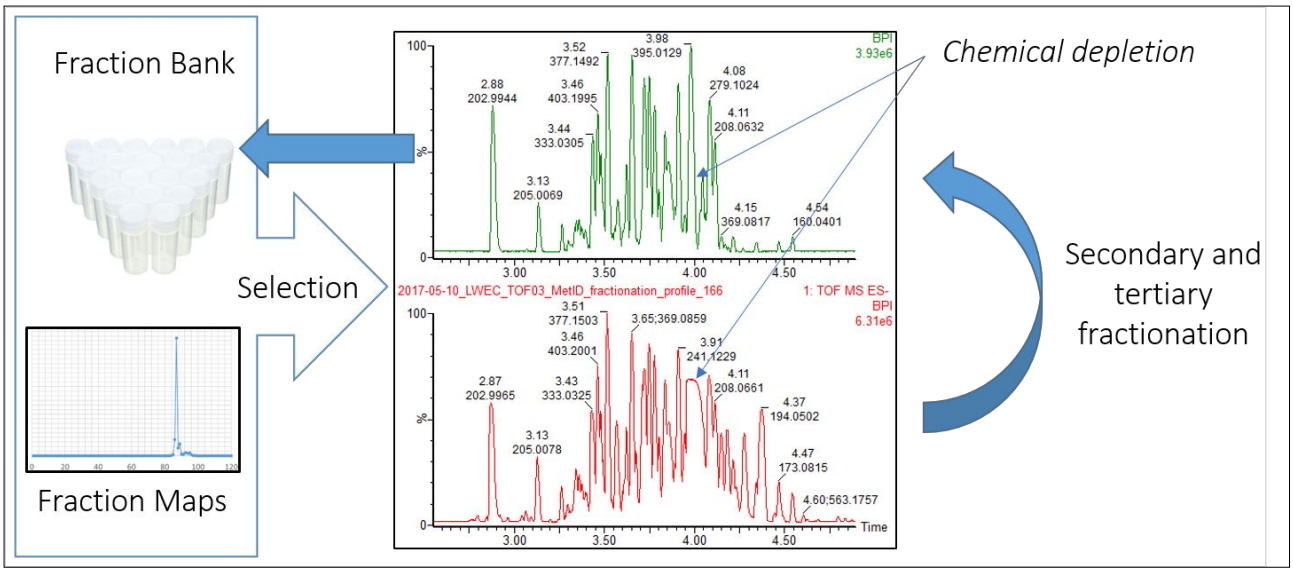

Figure S3. Demonstration of the cyclic fraction bank. Using a metabolite map of the 120 fractions, a fraction with the highest concentration of the feature of intensity can be removed for downstream secondary isocratic purification. All eluent is collected. Secondary fractions with the feature of interest progress to tertiary isocratic purification. The remaining eluent is dried, re-suspended in water and returned to the fraction bank. This is repeated following tertiary isocratic purification. The example in the figure shows the removal of para-cresol sulphate at the secondary stage

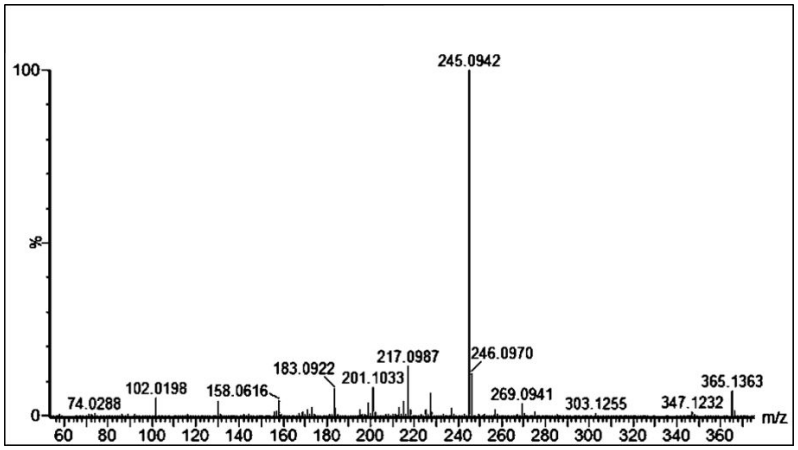

Figure S4. MS/MS spectrum of the m/z 365.134 (ES-) – Feature A- obtained by applying 20V of collision energy.

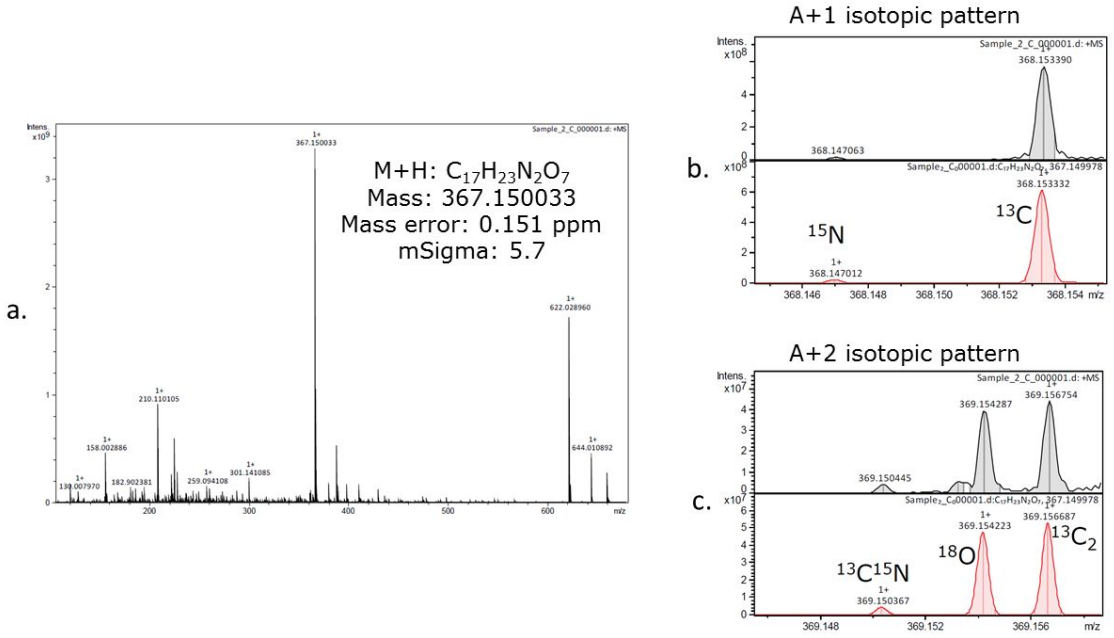

Figure S5. ES+ MRMS with isotopic fine structure confirmation of elemental composition ($C_{17}H_{23}N_2O_7$) of purified Feature A (M+H). **a)** broad band MS spectrum, the mSigma value is a measure for the goodness of fit of the measured vs. simulated isotopic pattern calculated by the Bruker DataAnalysis software. The lower the value the better the fit (with a range between 0-1000). **b)** plot showing A+1 pattern of measured (black) and simulated (red) pattern of $C_{17}H_{23}N_2O_7$, **c)** plot showing A+2 pattern of measured (black) and simulated (red) pattern of $C_{17}H_{23}N_2O_7$

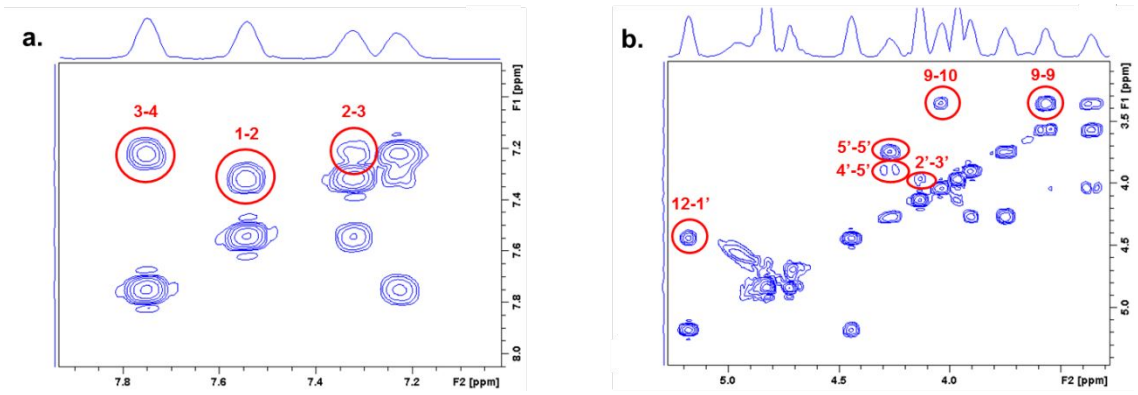

Figure S6. Plots showing two details of the ^{1}H-^{1}H COSY spectrum a) 7-8 ppm, and b) 3-5.5 ppm regions with encircled key correlations corresponding to the structure and assignment of Feature A as presented in Table S2. Labels match the chemical groups as labelled in Figure 2.

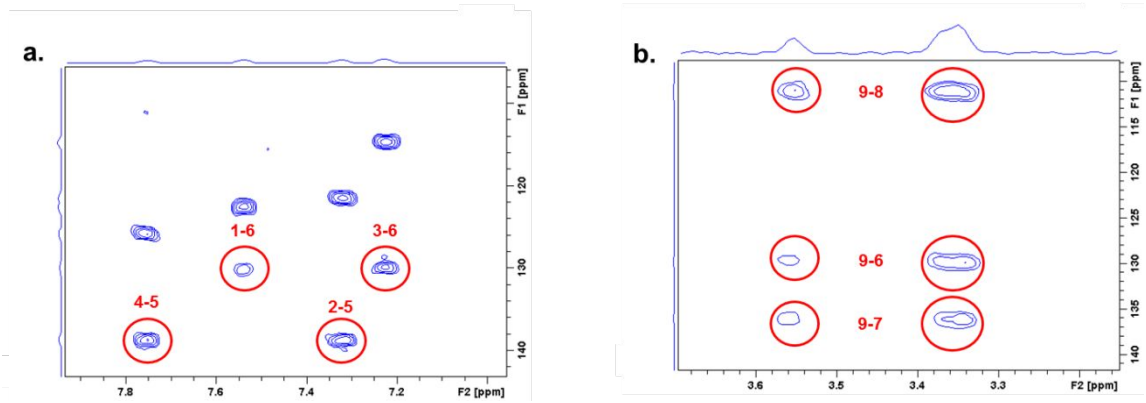

Figure S7. Plots presenting two details of the ^{1}H-^{13}C HMBC spectrum a) 7-8 ppm, and b) 3.1-3.7 ppm regions with encircled key correlations corresponding to the structure and assignment of Feature A as presented in Table S2. Labels match the chemical groups as labelled in Figure 2.

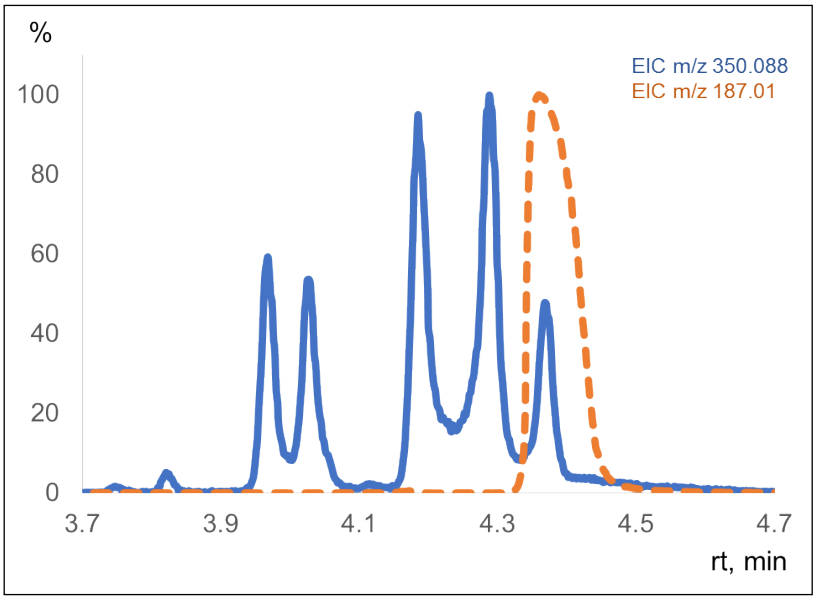

Figure S8. Plot representing the extracted ion chromatograms (EIC) of the m/z 350.088 (blue solid line) and co-eluting feature m/z 187.007 (orange dashed line) in the analytical-scale ES- profiling of the urine pool.

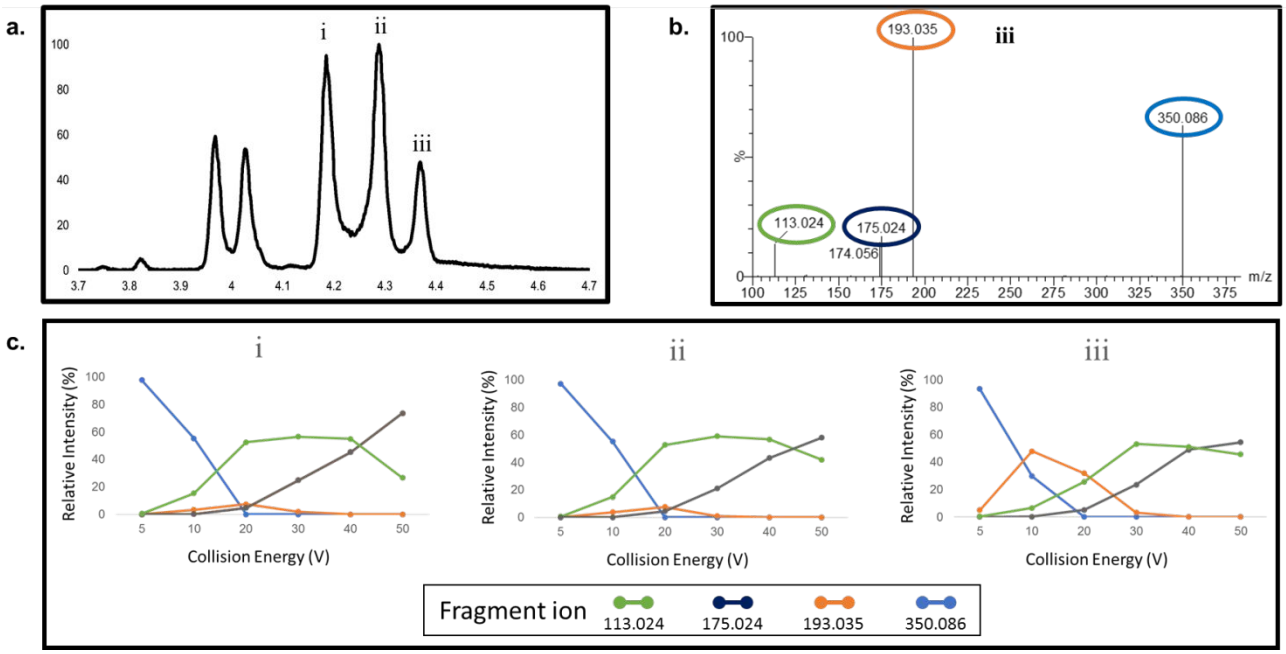

Figure S9. Extracted ion chromatograms (EIC) of the m/z 350.088 (ES-) features, peak iii is the metabolite of interest, peaks i and ii are its isomers (a); MS/MS spectrum obtained in ES- at 10 V of chromatographic peak of interest (iii) showing characteristic fragments of glucuronic acid in ES- (b); Fragment intensity profiles obtained in MS/MS experiments run at varying collision energies for the features i, ii, and iii (c). The profiles are indicative of different structural arrangements that provide differing bond energies and hence fragment ion ratios

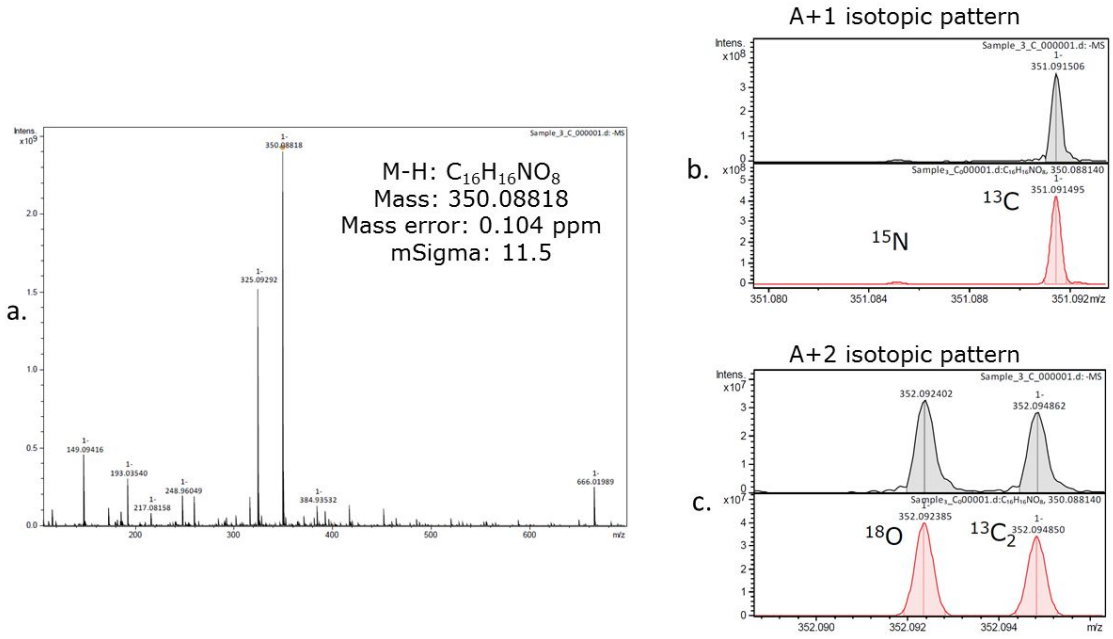

Figure S10. ES- MRMS with isotopic fine structure confirmation of elemental composition ($C_{16}H_{16}NO_8$) of purified Feature B (M-H). **a)** broad band MS spectrum, the mSigma value is a measure for the goodness of fit of the measured vs. simulated isotopic pattern calculated by the Bruker DataAnalysis software. The lower the value the better the fit (with a range between 0-1000). **b)** plot showing A+1 pattern of measured (black) and simulated (red) pattern of $C_{16}H_{16}NO_8$, **c)** plot showing A+2 pattern of measured (black) and simulated (red) pattern of $C_{16}H_{16}NO_8$

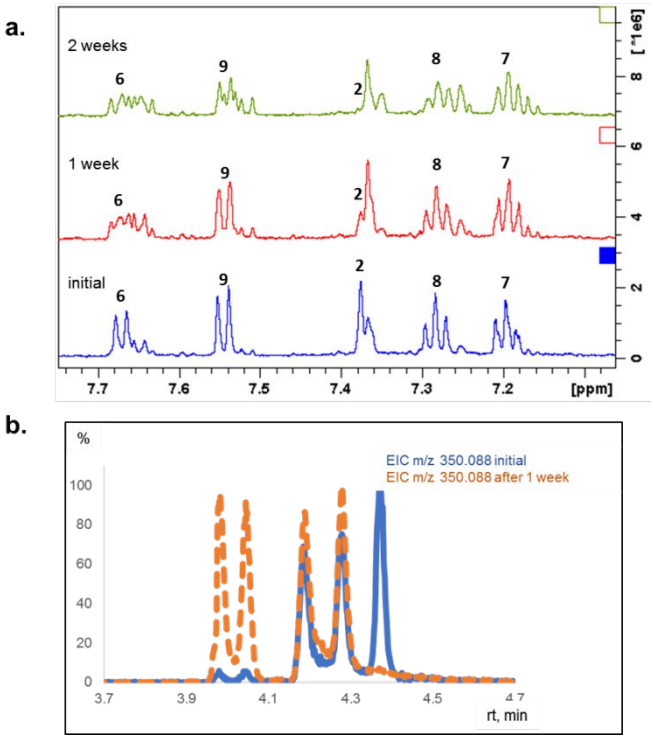

Figure S11. Evidence of the degradation of purified Feature B when stored in phosphate buffer. Plot representing an expansion of the 7-8 ppm region of the ^{1}H NMR spectrum of the purified feature B of interest (iii) acquired in diluted phosphate buffer immediately after preparation, 1 week and 2 weeks later (a), the NMR signals are labelled according to the Table S3 and Figure 2; plots of the extracted ion chromatograms (EIC) of the m/z 350.088 (ES-) profiling - initial solution (blue) and after one week being stored on a phosphate buffer (orange dashed line) (b).

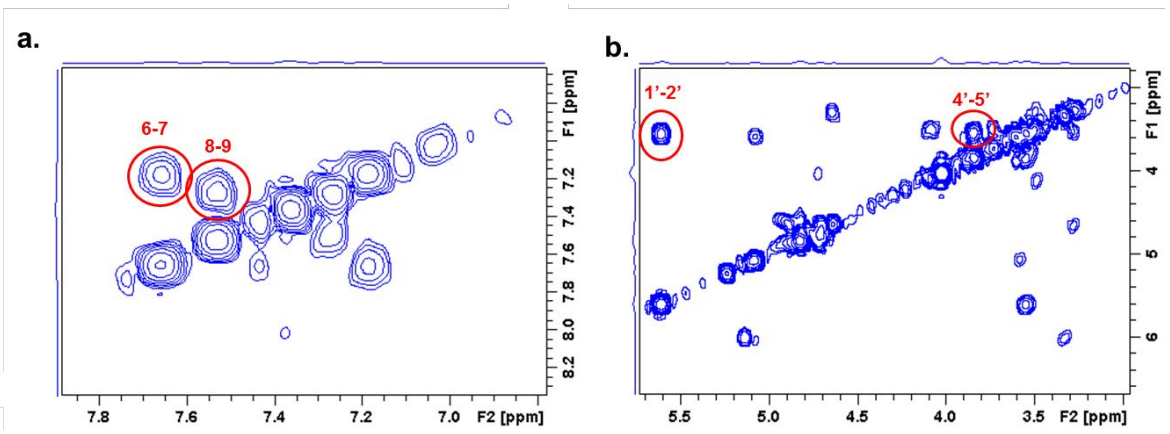

Figure S12. Plots showing two details of the ^{1}H-^{1}H COSY spectrum a) 6.8-8 ppm, and b) 3-5.5 ppm regions with encircled key correlations corresponding to the structure and assignment of Feature B as presented in Table S3 in the main text. Labels match the chemical groups as labelled in Figure 2.

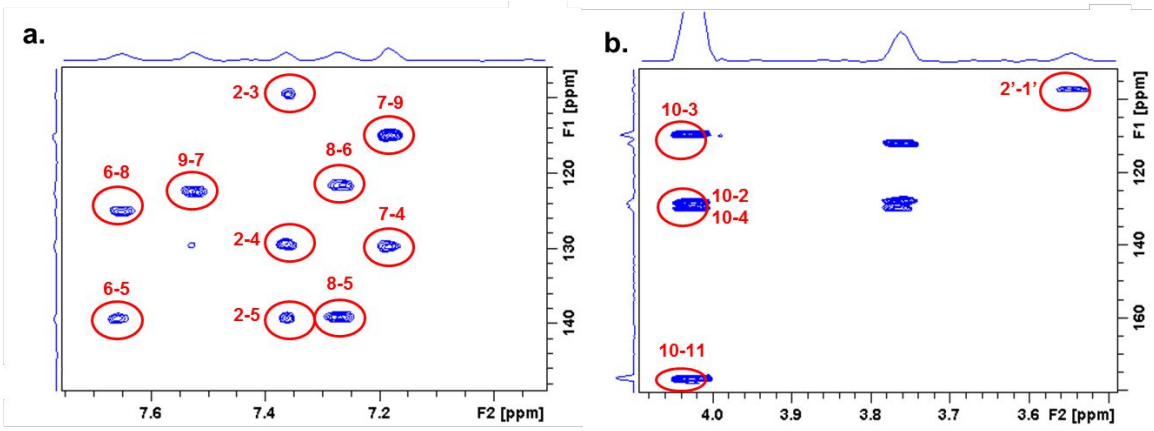

Figure S13. Plots presenting two details of the ^{1}H-^{13}C HMBC spectrum a) 7 -7.8 ppm, and b) 3.5–4.1 ppm regions with encircled key correlations corresponding to the structure and assignment of Feature B as presented in Table S3 in the main text. Labels match the chemical groups as labelled in Figure 2.

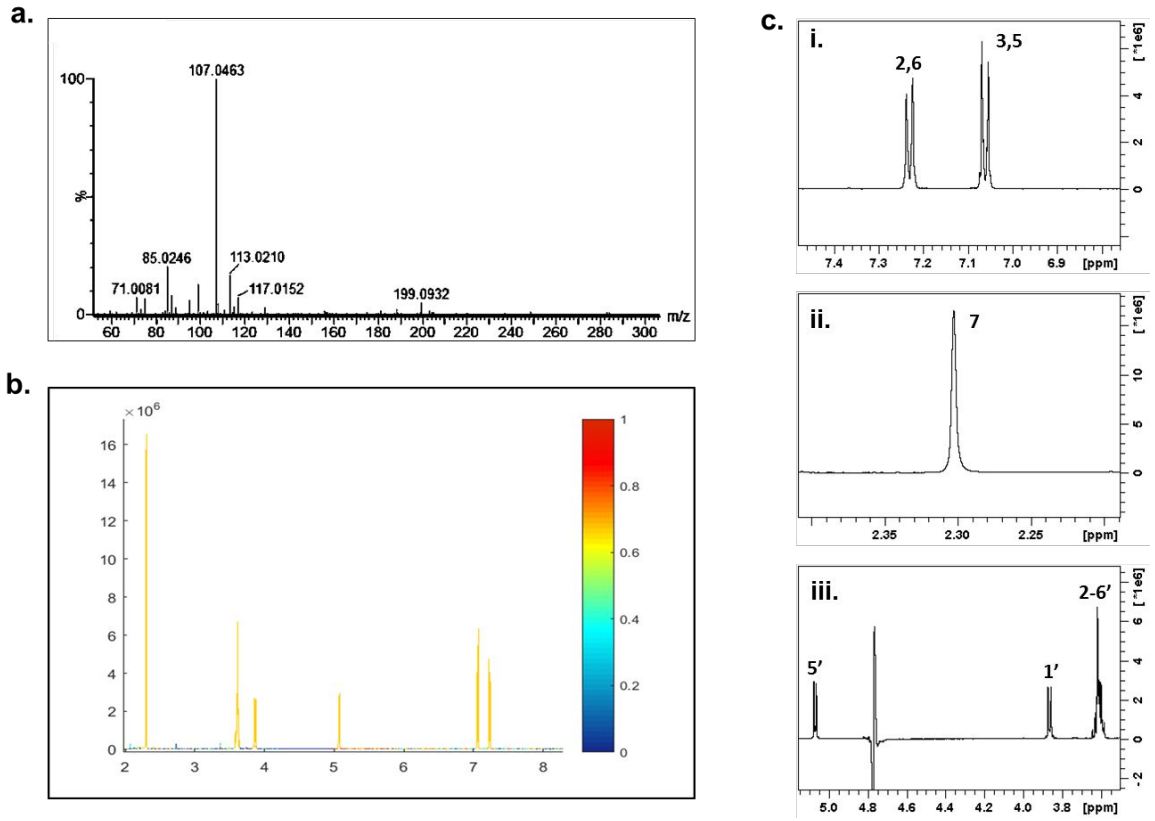

Figure S14. ES- MS/MS spectrum of feature C (m/z 283.082) (a); statistical heterospectroscopy analysis of feature C mapped on the ^{1}H NMR spectrum (b). This allows the isolation of the signals which belong to the compound of interest without further purification. The NMR spectra is colored by Spearman correlation with the LC-MS signal of the fraction containing the co-eluting unknown metabolite; ^{1}H NMR spectrum in 6.8-7.5 ppm region (i), 2.2-2.4 ppm region (ii) and 3.6-5.2 ppm region (iii) of Feature C (c); the NMR signals are labelled according to the Figure 2.

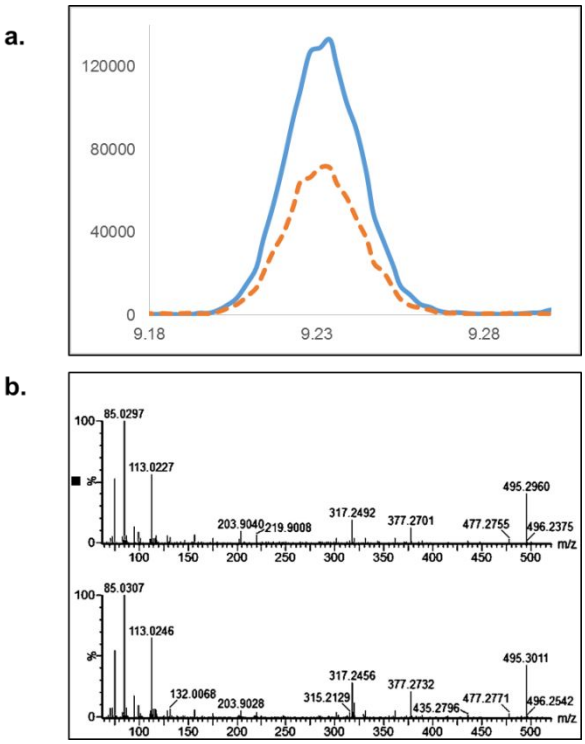

Figure S15. Analytical-scale ES- profiling of the urine pool and 20 ng/mL solution of the standard of pregnanediol-3-glucuronide. Extracted ion chromatogram (EIC) of the m/z 495.297 (blue solid line) in urine and in the standard solution (orange dashed line) (a). MS/MS spectra of the m/z 495.297 (ES-) obtained by applying the collision energy of 40V in the standard solution of pregnanediol-3-glucuronide (top spectrum) and in urine (bottom spectrum) (b).

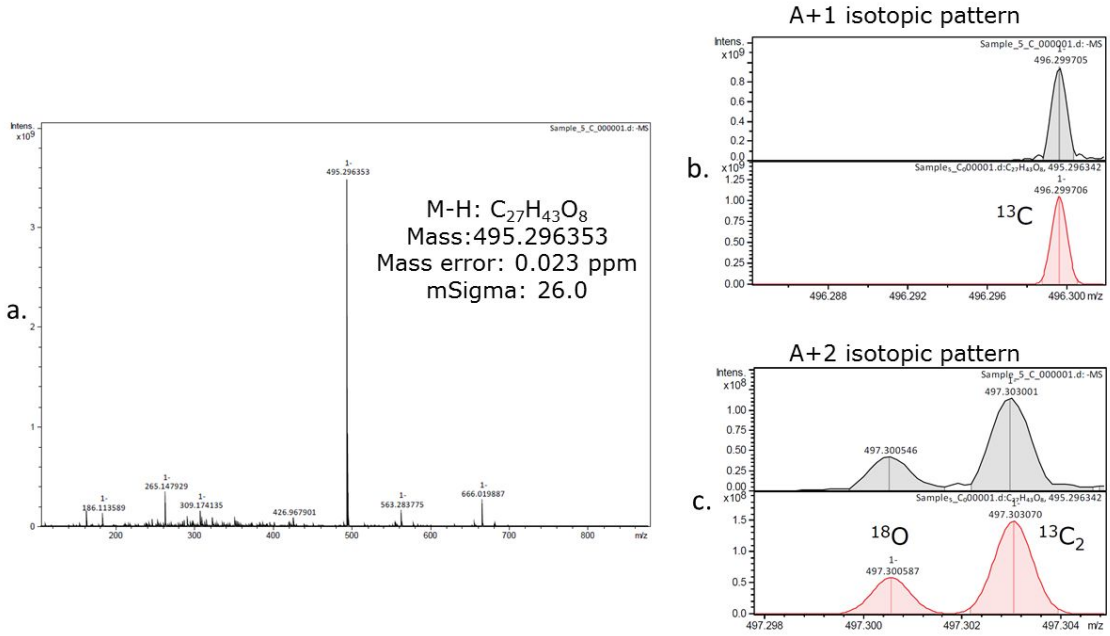

Figure S16. ES- MRMS with isotopic fine structure confirmation of elemental composition ($C_{27}H_{43}O_8$) of purified Feature C (M-H). **a)** broad band MS spectrum, the mSigma value is a measure for the goodness of fit of the measured vs. simulated isotopic pattern calculated by the Bruker DataAnalysis software. The lower the value the better the fit (with a range between 0-1000). **b)** plot showing A+1 pattern of measured and simulated pattern of $C_{27}H_{43}O_8$, **c)** plot showing A+2 pattern of measured and simulated pattern of $C_{27}H_{43}O_8$

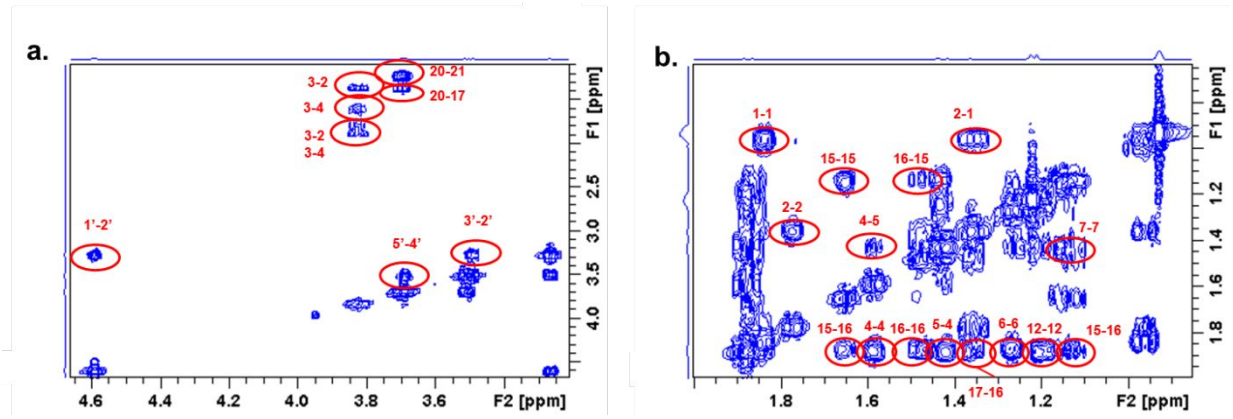

Figure S17. Plots showing two expansions (a) and (b) of the ^{1}H-^{1}H COSY spectrum with encircled key correlations corresponding to the structure and assignment of the Feature D as presented in Table S4. Labels match the chemical groups as labelled in Figure 2.

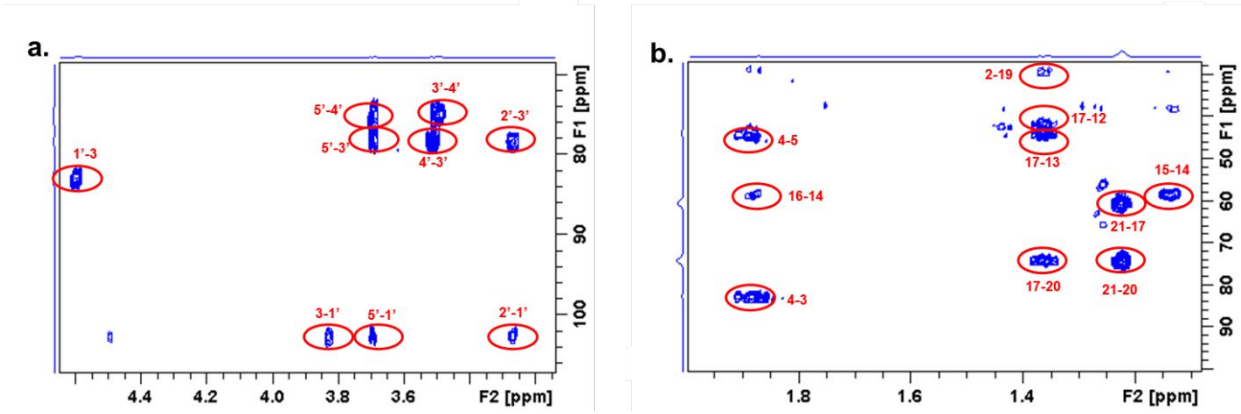

Figure S18. Plots presenting two expansions (a) and (b) of the ^{1}H-^{13}C HMBC spectrum with encircled key correlations corresponding to the structure and assignment of Feature D as presented in Table S4. Labels match the chemical groups as labelled in Figure 2.