

SUPPORTING INFORMATION

Multi-criteria approach to select polyaromatic river mutagen candidates

*Christine M.J. Gallampois^{1†}, Emma L. Schymanski², Martin Krauss¹, Nadin Ulrich³,
Mahmoud Bataineh⁴, Werner Brack^{1*}*

¹ UFZ - Helmholtz Centre for Environmental Research, Department of Effect-Directed
Analysis, Permoserstr. 15, D-04318 Leipzig, Germany

² Eawag – Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133,
CH-8600 Dübendorf, Switzerland

³ Institute for Prevention and Occupational Medicine of the German Social Accident
Insurance, Institute of the Ruhr-Universität Bochum (IPA), Bürkle-de-la-Camp-Platz 1, 44789
Bochum, Germany

⁴ Abu Dhabi Men's College, P.O.Box 25035, Abu Dhabi, United Arab Emirates

* Corresponding author email address: werner.brack@ufz.de, Telephone: +49 341 235 1531,
Fax: +49 341 235 2401

† Current affiliation: Department of Chemistry, Umeå University, 90187 Umeå, Sweden

This SI document contains 33 pages, including the front page

List of contents:

List of Tables:

- Table S1: Reagents, solvents and reference standards (page S3)
- Table S2: Chromatographic information for the compounds to identify (page S4)
- Table S3: Reference standards behavior under the different ionization modes (page S6)
- Table S4: Peak candidate selection (pages S12 – S26)
- Table S5: Retention and mutagenicity predictions (page S27 – S29)

List of Figures:

- Figure S1: Confirmation of the benzyl(diphenyl)phosphine presence in the fraction (page S30)
- Figure S2a: Ames test results with TA98 without S9 activation (page S31)
- Figure S2b: Ames test results with TA98 with S9 activation (page S32)
- Figure S3: Candidate structures for Peaks 2, 4, 12 and 14 (page S33)

List of Text Description:

- SI S1: Description of the liquid chromatographic and mass spectrometric systems (page S5)
- SI S2: Experiment description to confirm the presence of the benzyl(diphenyl) phosphine oxide in fraction N2-8 and mutagenicity assay results (pages S31 – S32)

Table S1: Lists of solvents, reagents and standards used in the study

List of solvents/ reagents	Suppliers
HPLC grade methanol; acetic acid water was bi-distilled in the laboratory	Merck
List of standards	Suppliers
benzotriazole [95-14-7]; benzophenone [119-61-9]; 1,3-dinitrobenzene [99-65-0]; 1-naphthol [90-15-3]; 1,4 naphthalene-dione [130-15-4]; 4-nitroquinoline 1-oxid [56-57-5]; metazachlor [67129-08-2]	Fluka
2-amino-3-methylimidazo[4,5-f]quinoline [2-09-9]; 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine [105650-23-5]; nitrofen [1836-75-5];	ABCR
2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline [2-10-9]	Toronto Research Chemicals
2-amino-9H-pyrido[2,3-b]indole [2-09-8]	MP Biomedicals LCC
1-methyl-9H-pyrido[3,4-b]indole [486-84-0]; N,N'-diethyl-N,N'-diphenylurea [85-98-3]; carbazole [86-74-8]; 1,10-phenanthroline [66-71-7]; 9-fluorenone [486-25-9]; acridine [260-94-6]; benzo(h)quinoline [230-27-3]; 2-Naphthalenamine, N-phenyl- [135-88-6]; benzanthrone [82-05-3]; 1,8,9-trihydroxyanthracene [480-22-8]; 2-nitro-9-fluorenone [3096-52-4]; 1,5-dihydroxyanthraquinone [117-12-4]; 1-nitropyrene [5522-43-0]; 1,8-dinitropyrene [42397-65-9]	Aldrich
3,3'-dichlorobenzidine [91-94-1]; 1H-isoindole-1,3(2H)-dione [85-41-6]; 2-nitrotoluene [88-72-2]	Riedel
benz(c)acridine [225-51-4]; trinitrotoluene [118-96-7]; 2,4-dinitrotoluene [121-14-2]; 1-chloro-2,4-dinitrobenzene [97-00-7]; carbamazepine [298-46-4]	Sigma Aldrich
benalaxyl [71626-11-4]	Ehrenstorfer
4-methyl-2-nitrophenol [119-33-5]; 2-naphthoic acid [93-09-4];	Alfa Aesar
4H-cyclopenta[def]phenanthren-4-one [5737-13-3]; 3-nitrobenzanthrone [17117-34-9]; 1,4-chrysenquinone; 5,6-chrysenquinone; 1,8-diaminopyrene; 1-hydroxypyrene [5315-79-7]; 3-hydroxybenzo(a)pyrene [98-17-9]	Chiron
9,10-phenanthrenedione [84-11-7]; atrazine [1912-24-9]	Promochem
1-hydroxyanthraquinone [129-43-1]; 2-hydroxyanthraquinone [605-32-3]	Tokyo Kasei
caffeine [58-08-2]	Merck

Table S2: Chromatographic Information for the 17 compounds of interest present in fraction N2-8

Peak Identity	m/z	Retention time (min)
Peak 1	204.1012	13.2
Peak 2	276.1020	13.5
Peak 3	302.0804	14.5
Peak 4	259.0528	16.2
Peak 5	250.0855	16.5
Peak 6	185.1066	16.6
Peak 7	326.0802	17.0
Peak 8	330.1438	17.2
Peak 9	360.1545	18.4
Peak 10	233.0963	18.5
Peak 11	273.1226	19.4
Peak 12	293.1084	20.0
Peak 13	377.1598	20.2
Peak 14	241.1329	20.4
Peak 15	225.0540	20.5
Peak 16	168.0801	24.2
Peak 17	323.0694	25.3

SI S1: LC and MS systems for reference standards analysis and chemical screening of active fractions

- *Liquid chromatography system:* The LC system used consisted of an Agilent series 1200 HPLC system including a degasser, a high-pressure binary SL pump, an autosampler, a column oven and a diode array detector operated from 210 to 250 nm (Agilent Technologies, Santa Clara, CA, USA), controlled by the ChemStation software.

- *Liquid chromatography gradient for the separation of standards and active fraction N2-8:* Standards and fraction N2-8 were separated on an analytical C18 reversed-phase column (LC-PAH, 250 x 2.1 mm, 5 μ m particle size, Supelco, CA, USA). A gradient elution was used, with water and methanol for negative ionization or an ammonium acetate buffer (10 mM, pH 4) in water and methanol for positive ionization. Elution of analytes was at a flow of 0.2 mL/min with the following conditions: 0-50 min, 50-95% of methanol; 50-65 min, 95% of methanol, then re-equilibration of the phase.

- *Liquid chromatography gradient for the LSER experiments:* A linear gradient of ammonium acetate buffer (10 mM/L, pH 7.3) and 0.1% formic acid in methanol was used with a flow of 0.2 mL/min with the following conditions: 0-3.2 min, 10% of methanol; 3.2-17.8 min, 10-95% of methanol; 17.8-37.8 min, 95% of methanol, then re-equilibration of the phase. The column was placed in an oven at 22°C

- *Mass spectrometry system:* MS spectra in the mass range of m/z 140-1400 were acquired in the Orbitrap at a resolving power of 100 000 (full width at half maximum, at m/z 400). The fraction was run in data-dependent acquisition mode including tandem MS experiments using collision-induced dissociation (CID) (MS^2 and MS^3) and the higher energy collision dissociation (HCD) cell (MS^2). The fragmentation of the standards and fraction were acquired using CID (collision energy 35 %). For the fraction, further experiments were run using the HCD cell to obtain complementary information regarding fragmentation of the unknown compounds. The energy was set at 90%. In the case of Peaks 2, 3, 4, 5, 6 and 15, fragments from CID and HCD were combined manually for the MetFrag search. The fragmentation pattern of the standards (CID fragmentation) was used to derive some substructure fragments association. The MS^n products were transferred to the Orbitrap for detection, such that high resolution fragments were obtained. Resolving power used was 15000 (m/z 400). Calibration of the Orbitrap was performed prior to sample analysis limiting the deviation of the calibration masses from theoretical values to the range -1 to 5 ppm.

Table S3: Ionization of standard compounds in APCI and ESI, limits of detection (DL), and neutral losses observed during CID fragmentation. nd = non determined (no signal or > 1000 pg). (UAXXXXXX): reference number for the available spectra on NORMAN MassBank Database (www.massbank.eu)

Name	Mass error in ppm (-/+)	APCI DL pg (-/+)	ESI DL pg (-/+)	Neutral losses recorded in APCI positive mode	Neutral losses recorded in APCI negative mode	Neutral losses recorded in ESI positive mode	Neutral losses recorded in ESI negative mode
Azaarenes							
acridine	-/-2.4	nd/50	nd/10	M+H (UA000301) no fragments	no signal	M+H (UA000301) CH ₂ N	no signal
2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline	-2.6/-2.0	nd/25	nd/5	M+H (UA000601) C ₃ H ₄ N ₂ , CH ₂ N ₂ , C ₂ H ₃ N, CHN, H ₃ N, CH ₃	M-H (UA000651) CH ₃	M+H (UA000601) C ₂ H ₃ N ₃ , C ₃ H ₄ N ₂ , CH ₂ N ₂ , C ₂ H ₃ N, CHN, H ₃ N, H ₂ N, CH ₃	M-H (UA000651) no fragments
2-amino-3-methylimidazo[4,5-f]quinoline	-3.4/-2.2	nd/25	nd/5	M+H (UA000501) C ₃ H ₄ N ₂ , CH ₂ N ₂ , C ₂ H ₃ N, CHN, H ₃ N, CH ₃	M-H (UA000551) CH ₃	M+H (UA000501) C ₂ H ₃ N ₃ , C ₃ H ₄ N ₂ , CH ₂ N ₂ , C ₂ H ₃ N, CHN, H ₃ N, H ₂ N, CH ₃	M-H (UA000551) no fragments
2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine	-2.8/-1.6	nd/50	nd/5	M+H (UA000901) CHN, H ₃ N,	M-H (UA000951) C ₂ H ₃ N	M+H (UA000901) CHN, H ₃ N	no signal
2-amino-9H-pyrido[2,3-b]indole	-2.1/-1.9	nd/25	nd/5	M+H (UA000801) C ₃ H ₄ N ₂ , C ₂ H ₃ N, CHN, H ₃ N, CH ₄ , CH ₃	M-H (UA000851) no fragments	M+H (UA000801) C ₃ H ₄ N ₂ , C ₂ H ₃ N, CHN, H ₃ N, CH ₄ , CH ₃	M-H (UA000851) no fragments
benz(c)acridine	-/-1.3	nd/25	nd/5	M+H (UA001001) C ₃ H ₇ N ₃ , C ₃ H ₅ N ₃ , C ₃ H ₄ N ₃ , C ₆ H ₆ , C ₂ H ₅ N ₃ , C ₃ H ₄ N ₃ , C ₂ H ₆ N ₂ , C ₂ H ₅ N ₂ , CH ₄ N ₂ , CH ₂ N ₂ , C ₂ H ₃ N, CHN, H ₃ N, CH ₃	M-H CH ₃	M+H (UA001001) C ₃ H ₇ N ₃ , C ₆ H ₆ , C ₃ H ₄ N ₂ , C ₂ H ₆ N ₂ , CH ₄ N ₂ , CH ₂ N ₂ , C ₂ H ₃ N, CHN, H ₃ N, CH ₃	no signal
benzo(h)quinoline	-/-2.3	nd/50	nd/10	M+H (UA000401) CHN	no signal	M+H (UA000401) CH ₃ N, CH ₂ N, CHN	no signal
benzotriazole	-0.9/-4.1	50/50	50/50	M+H (UA000201) N ₂	M-H (UA000251) no fragments	M+H (UA000201) CHN ₃ , N ₂	M-H (UA000251) no fragments
carbazole	-2.5/-2.4	50/50	nd/500	M+H (UA000101) no fragments	M-H no fragments	M+H (UA000101) no fragments	no signal

Name	Mass error in ppm (-/+)	APCI DL pg (-/+)	ESI DL pg (-/+)	Neutral losses recorded in APCI positive mode	Neutral losses recorded in APCI negative mode	Neutral losses recorded in ESI positive mode	Neutral losses recorded in ESI negative mode
1-methyl-9H-pyrido[3,4-b]indole	-2.8/-2.4	nd/25	nd/5	M+H (UA000701) C ₃ H ₄ N ₂ , C ₂ H ₆ N ₂ , C ₂ H ₂ N ₂ , C ₃ H ₃ N, CH ₄ N ₂ , CH ₂ N ₂ , C ₂ H ₃ N, CH ₃ N, CHN, C ₂ H ₂ , H ₃ N, H ₂ N, CH ₃	M-H (UA000751) CH ₃	M+H (UA000701) C ₃ H ₄ N ₂ , C ₂ H ₆ N ₂ , C ₂ H ₂ N ₂ , CH ₂ N ₂ , C ₂ H ₃ N, CH ₃ N, CHN, H ₃ N, H ₂ N, CH ₃	M-H (UA000751) no fragments
Keto-PAHs							
benzanthrone	-3.5/-1.4	50/50	nd/10	M+H (UA001201) CHO, CO	M- no fragments	M+H (UA001201) CHO, CO	no signal
4H-cyclopenta [def] phenanthrene-4-one	-3.4/-2.2	50/50	nd/250	M+H (UA001401) C ₃ H ₂ O, CHO, CO	M- no fragments	M+H (UA001401) CO	no signal
9-fluorenone	-2.8/-1.5	50/50	nd/100	M+H (UA001101) CHO, CO	M- no fragments	M+H (UA001101) CHO, CO	no signal
Ketone							
benzophenone	-/-1.9	nd/25	nd/25	M+H (UA001501) C ₇ H ₆ O, C ₆ H ₆	no signal	M+H (UA001501) C ₇ H ₆ O, C ₆ H ₆	no signal
Nitro-keto-PAHs							
2-nitro-9-fluorenone	-3.46/-2.2	75/nd	nd/nd	M+H (UA001701) CNO ₂ , NO ₂ , NO, OH	M- (UA001783) NO	no signal	no signal
3-nitrobenzanthrone	-4.0/-0.7	50/25	nd/500	M+H (UA001601) C ₃ H ₅ O ₃ , C ₂ HNO ₃ , CNO ₂ , C ₂ O ₂ , NO ₂ , NO, HO, N	M- (UA001683) NO, CHO, CO	(no M+H signal) (UA001601) CNO ₂ , NO ₂ , NO, HO	no signal
Quinones							
1,4-chrysenequinone	-3.9/-0.9	50/400	nd/1000	M+H (UA002101) No fragments	M- (UA002183) no fragments	(no M+H signal) (UA002101) C ₃ H ₂ O ₂ , C ₂ O ₂ , CO ₂ , C ₂ H ₂ O, CO, H ₂ O	no signal
5,6-chrysenequinone	-3.9/-1.1	50/500	nd/1000	M+H (UA002201) C ₂ HO ₂ , CO	M- no fragments	M+H (UA002201) C ₂ HO ₂ , CO	no signal
1H-isoindole-1,3(2H)dione	-2.1/-3.2	50/nd	500/nd	M+H (UA002001) H ₂ O	M-H (UA002051) no fragments	(no M+H signal) (UA002001) H ₂ O	M-H (UA002051) no fragments

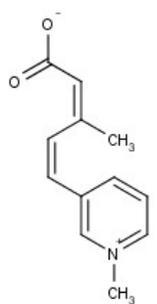
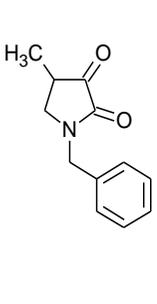
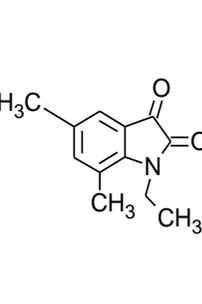
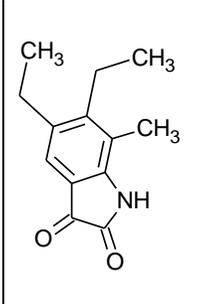
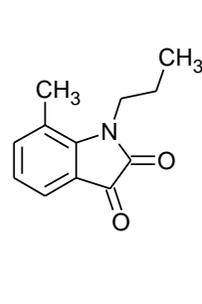
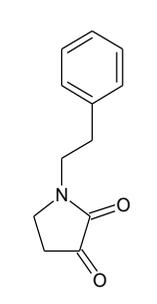
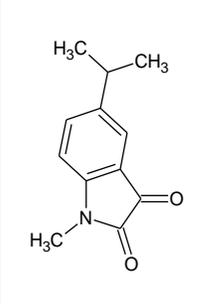
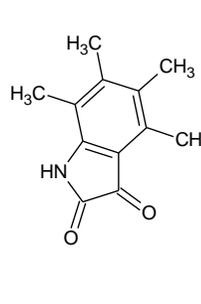
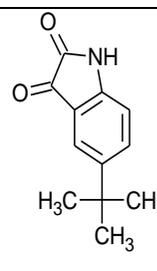
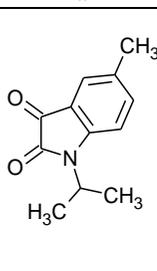
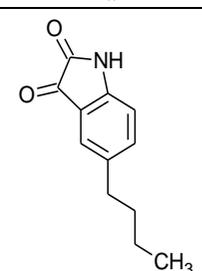
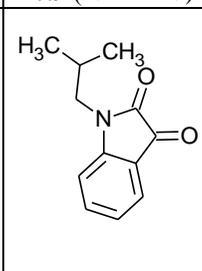
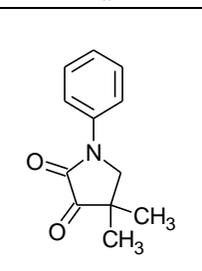
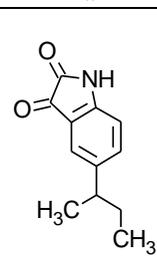
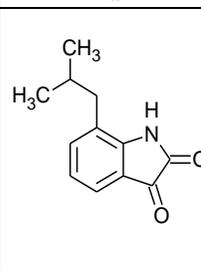
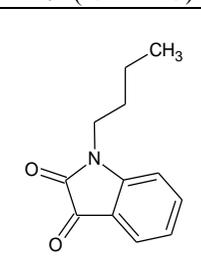
Name	Mass error in ppm (-/+)	APCI DL pg (-/+)	ESI DL pg (-/+)	Neutral losses recorded in APCI positive mode	Neutral losses recorded in APCI negative mode	Neutral losses recorded in ESI positive mode	Neutral losses recorded in ESI negative mode
naphthalene-1,4-dione	-1.9/-	100/nd	nd/nd	M+H (UA001801) CO	M- (UA001883) no fragments	no signal	no signal
phenanthrene-9,10-dione	-3.4/-2.0	50/50	nd/nd	M+H (UA001901) C ₄ H ₂ O ₂ , C ₂ HO ₂ , C ₂ O ₂ , CO	M- no fragments	no signal	no signal
OH-quinones							
1,5-dihydroxyanthraquinone	-3.75/-	50/nd	nd/nd	M+H (UA002301) C ₇ H ₄ O ₂ , C ₃ O ₄ , C ₂ O ₄ , C ₃ O ₃ , C ₂ H ₂ O ₃ , C ₂ O ₃ , C ₂ O ₂ , CH ₂ O ₂ , CO ₂ , C ₂ H ₂ O, CO, H ₂ O	M- no fragments	M+H (UA002301) C ₈ H ₄ O ₄ , C ₇ H ₄ O ₂ , C ₃ O ₄ , C ₃ O ₃ , C ₂ O ₃ , C ₂ O ₂ , CH ₂ O ₂ , CO ₂ , C ₂ H ₂ O, CO, H ₂ O	no signal
1-hydroxyanthraquinone	-4.0/-1.9	50/50	nd/50	M+H (UA002501) C ₈ H ₄ O ₃ , C ₇ H ₄ O ₂ , C ₃ O ₃ , C ₂ H ₂ O ₃ , C ₂ HO ₃ , C ₂ O ₃ , C ₂ O ₂ , CH ₂ O ₂ , CO ₂ , C ₂ H ₂ O, CO, H ₂ O	M- (UA002583) no fragments	M+H (UA002501) C ₇ H ₄ O ₂ , C ₃ O ₃ , C ₂ HO ₃ , C ₂ O ₃ , C ₂ O ₂ , CO ₂ , C ₂ H ₂ O, CO, H ₂ O	M-H (UA002551) CO
2-hydroxyanthraquinone	-3.6/-1.6	50/75	nd/50	M+H (UA002401) C ₈ H ₄ O ₃ , C ₇ H ₄ O ₂ , C ₃ O ₃ , C ₂ H ₂ O ₃ , C ₂ HO ₃ , C ₂ O ₃ , C ₂ O ₂ , CH ₂ O ₂ , CO ₂ , C ₂ H ₂ O, CO, H ₂ O	M- no fragments	M+H (UA002401) C ₇ H ₄ O ₂ , C ₃ O ₃ , C ₂ O ₃ , C ₂ O ₂ , C ₂ H ₂ O, CO	M-H (UA002451) CO
Amino-compounds							
atrazine	-/-0.6	nd/25	nd/5	M+H (UA002901) C ₄ H ₉ N ₂ Cl, C ₆ H ₁₂ N ₂ , C ₄ H ₈ N ₂ , CH ₃ N ₂ Cl, C ₃ H ₆ N ₂ , C ₃ H ₆ , HCl, C ₂ H ₄	no signal	M+H (UA002901) C ₆ H ₁₃ N ₂ Cl, C ₇ H ₁₁ N ₃ , C ₄ H ₉ N ₂ Cl, C ₆ H ₁₂ N ₂ , C ₄ H ₈ N ₂ , CH ₃ N ₂ Cl, C ₅ H ₁₀ , C ₃ H ₆ N ₂ , C ₃ H ₆ , HCl, C ₂ H ₄	no signal
3,3'-dichlorobenzidine	-/-0.7	nd/50	nd/500	M+ (UA004901) HCl ₂ , Cl ₂ , CHNCl, H ₃ NCl, H ₂ NCl, Cl, H ₂ N	no signal	M+H (UA004901) H ₂ Cl ₂ , HCl ₂ , HCl, Cl, H	no signal

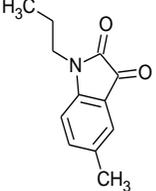
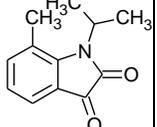
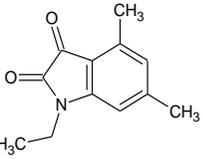
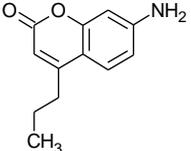
Name	Mass error in ppm (-/+)	APCI DL pg (-/+)	ESI DL pg (-/+)	Neutral losses recorded in APCI positive mode	Neutral losses recorded in APCI negative mode	Neutral losses recorded in ESI positive mode	Neutral losses recorded in ESI negative mode
1,8-diaminopyrene	-/-1.3	nd/25	nd/10	M+ (UA002701) CH ₃ N ₂ , HN ₂ , CH ₂ N, CHN, H ₃ N, H ₂ N, HN, H	no signal	M+H (UA002701) CH ₂ N, CHN, H ₃ N	no signal
N-phenylnaphthalen-2-amine	-/-0.9	nd/10	nd/5	M+H (UA002801) C ₁₀ H ₈ , C ₉ H ₈ , C ₇ H ₇ N, C ₇ H ₅ N, C ₆ H ₆ N, C ₆ H ₆ , C ₆ H ₅ , C ₂ H ₄ N, C ₂ H ₃ N, C ₂ H ₅ , CHN, H ₄ N, H ₃ N, CH ₄ , CH ₃ , H ₃ ,H ₂	no signal	M+H (UA002801) C ₁₀ H ₈ , C ₉ H ₈ , C ₇ H ₇ N, C ₇ H ₅ N, C ₆ H ₆ N, C ₆ H ₆ , C ₆ H ₅ , C ₂ H ₄ N, C ₂ H ₃ N, C ₂ H ₅ , CHN, H ₄ N, H ₃ N, CH ₄ , CH ₃ , H ₃ ,H ₂	no signal
Amides							
benalaxyl	-/-0.6	nd/25	nd/5	M+H (UA003201) C ₁₀ H ₁₀ O ₃ , C ₈ H ₆ O, C ₂ H ₄ O ₂ , CH ₄ O	no signal	no signal	no signal
carbamazepine	-/-1.7	nd/50	nd/5	M+H (UA003001) CH ₃ NO, CHNO, H ₃ N	no signal	(no M+H signal) (UA003001) CH ₃ NO, CH ₂ NO, CHNO, H ₃ N	no signal
N,N'-dimethyl-N,N'-diphenylurea	-/-1.4	nd/10	nd/5	M+H (UA003301) C ₈ H ₉ NO, C ₇ H ₉ N	no signal	M+H (UA003301) C ₉ H ₁₂ N ₂ O, C ₉ H ₁₀ N ₂ O, C ₈ H ₉ NO, C ₈ H ₁₂ N, C ₇ H ₉ N	no signal
metazachlor	-/-1.25	nd/50	nd/5	no signal	no signal	(no M+H signal) (UA003101) C ₅ H ₅ N ₂ OCl, C ₃ H ₄ N ₂	no signal
NO₂-compounds							
1-chloro-2,4-dinitrobenzene	-3.5/-	50/nd	nd/nd	no signal	M- (UA004183) NO ₂ , NO	no signal	no signal
1,3-dinitrobenzene	-3.0/-	50/nd	nd/nd	no signal	M- (UA003983) NO	no signal	no signal

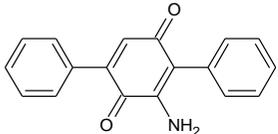
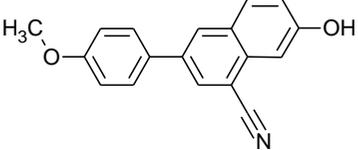
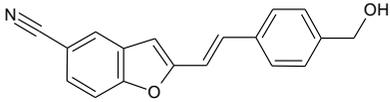
Name	Mass error in ppm (-/+)	APCI DL pg (-/+)	ESI DL pg (-/+)	Neutral losses recorded in APCI positive mode	Neutral losses recorded in APCI negative mode	Neutral losses recorded in ESI positive mode	Neutral losses recorded in ESI negative mode
1,8-dinitropyrene	-/-1.1	50/nd	nd/nd	M+H (UA003501) C ₂ NO ₄ , HN ₂ O ₄ , HN ₂ O ₃ , N ₂ O ₃ , CNO ₃ , HNO ₃ , N ₂ O ₂ , CNO ₂ , HNO ₂ , NO ₂ , NO, HO	M- (UA003583) NO ₂ , NO	no signal	no signal
2,4-dinitrotoluene	-3.3/-	50/nd	nd/nd	no signal	M- (UA004083) CH ₂ NO ₃ , CNO ₂ , H ₂ NO ₂ , CH ₃ O ₂ , HNO ₂ , CH ₂ O ₂ , CHO ₂ , NO, CHO, H ₂ O, HO	no signal	no signal
4-methyl-2-nitrophenol	-2.0/-	50/nd	nd/nd	no signal	M- (UA003683) no fragments	no signal	no signal
nitrofen	-3.9/-	50/nd	nd/nd	no signal	M- (UA004383) C ₆ H ₂ Cl ₂ , C ₆ H ₃ Cl, Cl, NO	no signal	no signal
4-nitroquinoline-1-oxide	-3.2/-2.0	50/50	nd/75	M+H (UA003701) C ₃ H ₂ N ₂ O ₃ , C ₂ HN ₂ O ₃ , C ₃ H ₂ NO ₃ , C ₂ HNO ₃ , C ₂ NO ₃ , CHNO ₃ , CNO ₃ , CHNO ₂ , CNO ₂ , HNO ₂ , NO ₂ , HNO, NO, HO	M- (UA003783) CNO ₃ , C ₂ H ₂ NO ₂ , CNO ₂ , H ₂ NO ₂ , NO ₂ , NO, HO	M+H (UA003701) C ₃ H ₂ N ₂ O ₃ , C ₂ HN ₂ O ₃ , C ₃ H ₂ NO ₃ , C ₂ HNO ₃ , C ₂ NO ₃ , CHNO ₃ , CNO ₃ , CHNO ₂ , CNO ₂ , O ₃ , HNO ₂ , NO ₂ , HNO, NO	no signal
1-nitropyrene	-3.6/-1.3	25/25	nd/nd	M+H (UA003401) HNO ₂ , NO ₂ , NO, OH	M- (UA003483) NO	no signal	no signal
2-nitrotoluene	-1.5/-	50/nd	nd/nd	no signal	M- no fragments	no signal	no signal

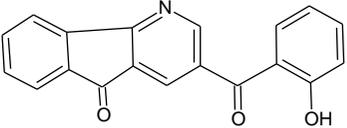
Name	Mass error in ppm (-/+)	APCI DL pg (-/+)	ESI DL pg (-/+)	Neutral losses recorded in APCI positive mode	Neutral losses recorded in APCI negative mode	Neutral losses recorded in ESI positive mode	Neutral losses recorded in ESI negative mode
trinitrotoluene	-4.0/-	125/nd	nd/nd	no signal	M- (UA004283) C ₂ H ₂ NO ₄ , CHNO ₄ , N ₃ O ₃ , CH ₂ NO ₃ , CHNO ₃ , H ₂ NO ₃ , CH ₃ O ₃ , CH ₂ NO ₂ , N ₂ O ₂ , CHNO ₂ , CNO ₂ , H ₂ NO ₂ , HNO ₂ , CH ₂ O ₂ , CHO ₂ , CH ₂ O, NO, CHO, HO	no signal	no signal
OH-PAHs							
3-hydroxybenzo (a)pyrene	-3.2/-2.1	50/100	nd/750	M+ (UA004501) CHO, CO, H ₂ O, HO, O, H	M-H (UA004551) no fragments	M+ & M+H (UA004501) CHO, CO, HO, O	M-H (UA004551) no fragments
1-hydroxypyrene	-3.7/-1.5	125/75	50/nd	M+H (UA004401) C ₁₀ H ₆ O, CHO, CO, H ₂ O, HO, H	M-H no fragments	no signal	M-H (& M-) (UA004451) CO
1-naphthol	-2.1/-3.4	50/500	50/nd	M+H (UA004601) CO, H ₂ O	M-H (UA004651) no fragments	no signal	M-H no fragments
COOH-PAHs							
2-naphthoic acid	-2.9/-2.5	100/-	50/-	M+H (UA004801) H ₂ O, CO, CO ₂	M-H (UA004851) CO ₂	M+H (UA004801) H ₂ O, CO, CO ₂	M-H (UA004851) CO, CO ₂
Purine-dione							
caffeine	-/-2.3	nd/25	nd/50	M+H (UA005001) C ₅ H ₆ N ₂ O ₂ , C ₃ H ₃ NO ₂ , C ₂ H ₃ NO ₂ , C ₂ HNO ₂ , CH ₃ NO ₂ , C ₂ H ₃ NO, CO ₂ , CH ₄ O, CH ₃ , CH ₂	no signal	M+H (UA005001) C ₅ H ₆ N ₂ O ₂ , C ₄ H ₄ N ₂ O ₂ , C ₃ H ₃ NO ₂ , C ₃ H ₄ N ₂ O, C ₂ H ₃ NO, CO ₂ , CH ₄ O	no signal

Table S4: Peak candidate selection. na: not applicable. nd: not predicted

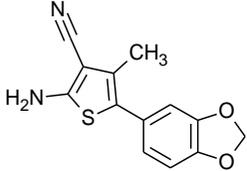
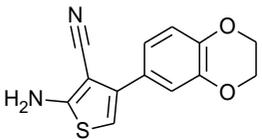
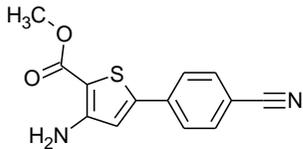
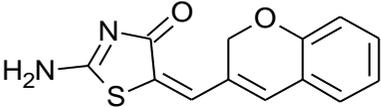
Peak 1								
Information	<p>This peak of formula C₁₂H₁₃NO₂ (m/z 204.1013), eluting at 13.2 min was detected in APCI + and ESI +. The APCI + MS² spectrum of the unknown showed only one fragment (m/z 148.1117), corresponding to the loss of C₂O₂ (or two carbonyl groups). Using ChemSpider, 1345 chemicals with C₁₂H₁₃NO₂ as molecular formula were proposed. MetFrag identified twenty chemicals explaining the fragment observed with a score greater than 0.89. No signal in negative mode to support the presence of candidate 1 (carboxylic acid better observed in negative mode), which can be removed from the list. Expectation of loss of benzyl group for structures 1, 5 and 12, but not seen, thus can be removed. Expectation of loss of alkyl group on the nitrogen atom for structures 2, 4, 6, 11, 12, 15, 16, 17 and 18, but not seen, thus can be removed. After matching the pK_a and log K_{ow}, the best candidates remaining on the list are compounds 3, 7, 8, 10, 13, 14 and 19</p>							
Candidate structures								
ChemSpider ID	1) 1293508	2) 219476	3) 20523428	4) 13494962	5) 2407695	6) 24798308	7) 24021760	8) 19461480
MetFrag score (number of predicted peak)	1(1)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)
Log K_{ow} ACD/EPI SUITE	2.14/2.74	1.0/0.2	2.0/1.9	2.5/2.5	2.1/1.8	1.8/nd	1.9/nd	2.4/2.6
pK_a		na	na	10.9 (NH ↔ N ⁻)	na	na	na	11.4 (NH ↔ N ⁻)
Candidate structures								
ChemSpider ID	9) 517156	10) 1606080	11) 2339601	12) 1783307	13) 9238673	14) 16812332	15) 513041	16) 1283920

MetFrag score (number of predicted peak)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)	0.945 (1)
Log K_{ow} ACD/EPI SUITE	2.2/2.4	1.0/1.7	2.6/2.5	2.0/1.7	1.6/0.4	2.4/1.6	2.4/2.4	2.2/1.8
p<i>K_a</i>	10.6 (NH ↔ N ⁻)	na	10.6 (NH ↔ N ⁻)	na	na	10.6 (NH ↔ N ⁻)	10.6 NH ↔ N ⁻	na
Candidate structures								
ChemSpider ID	17) 2378726	18) 3347751	19) 24209284	20) 2012579				
MetFrag score (number of predicted peak)	0.945 (1)	0.945 (1)	0.945 (1)	0.89 (1)				
Log K_{ow} ACD/EPI SUITE	2.1/1.8	1.9/1.7	2.1/nd	2.2/2.1				
p<i>K_a</i>	na	na	na	2.2 (NH ₃ ⁺ ↔ NH ₂)				

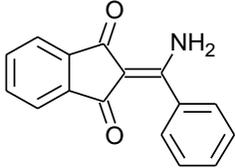
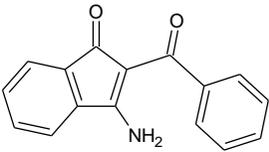
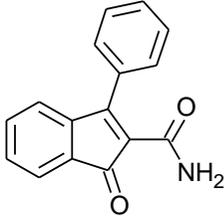
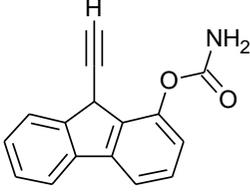
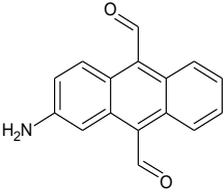
Peak 2			
Information	<p>This peak of formula C₁₈H₁₃NO₂ (m/z 276.1021), eluting at 13.5 min was detected in APCI +/- and ESI+. The APCI + MS² spectrum of the unknown showed five fragments in total (259.0758, 258.0915, 234.0916, 231.0801 and 169.1906). Four of the five neutral losses were identified, and corresponded to NH₃, H₂O, C₂H₂O, and CH₃NO. Using ChemSpider, 417 chemicals with the formula C₁₈H₁₃NO₂ were proposed and Metfrag identified seven of them with a high score (≥ 0.749), explaining two or three fragments. After matching the log <i>K</i>_{ow} range and the p<i>K</i>_a values only three of them remained as candidates. NH₃ losses mainly connected to the presence of an amino group in PAHs skeleton, or present for some azaarenes and these losses associated to a strong signal. Candidate 1 can produce this NH₃ loss, while candidate 2 and 3 not. Candidate 1 could also be able to produce the fragment at 231 Da (loss of CH₃NO) by the breakage of two C-C bonds to release a carbonyl group and the release of the already observed of NH₃. Candidates 2 and 3 are less likely. Furthermore candidates 2 and 3 have a hydroxyl group in their structure, but not detected in ESI-. Thus these two compounds can be removed.</p>		
Candidate structures			
ChemSpider ID	1) 12251096	2) 9429393	3) 13815809
MetFrag score (number of predicted peak)	1 (2)	0.927 (3)	0.758 (3)
Log <i>K</i>_{ow} ACD/EPI SUITE	3.3/3.1	4.4/4.1	3.5/3.7
p<i>K</i>_a	12.1 (NH ₂ ↔ NH ⁻)	8.2 (OH ↔ O ⁻)	14.8 (OH ↔ O ⁻)

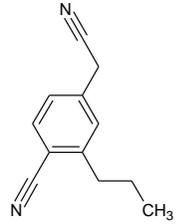
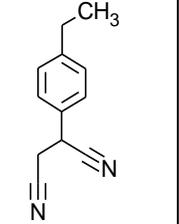
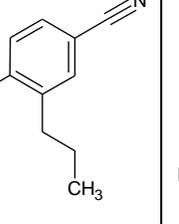
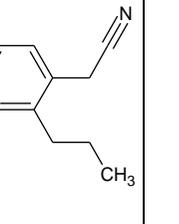
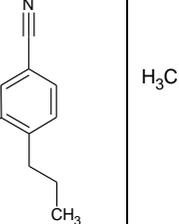
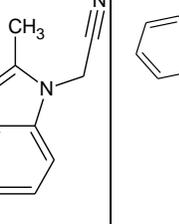
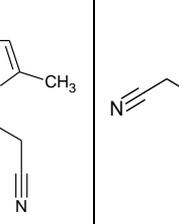
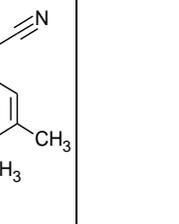
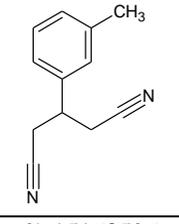
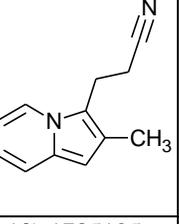
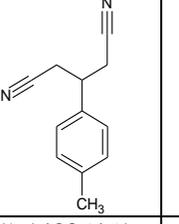
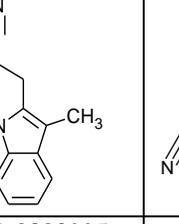
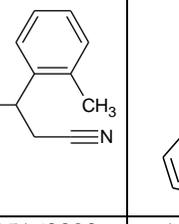
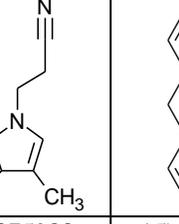
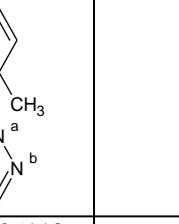
Peak 3	
Information	This peak of formula C ₁₉ H ₁₁ NO ₃ (m/z 302.0804), eluting at 14.5 min was detected in APCI +/- and ESI +/- . The APCI + MS ² spectra of the unknown were obtained from CID and HCD fragmentation, and four fragments were observed (284.0705, 274.0862, 256.0756 and 227.0728). MS ³ (CID) of fragments 274 and 284 Da were performed and three fragments were obtained (256.0756, 246.0912 and 228.0806). The corresponding losses were H ₂ O, CO from 302, CH ₂ O ₂ from 302, CO from 284, H ₂ O from 274, CO from 274, C ₂ H ₂ O ₂ from 284) and C ₂ H ₃ O ₃ from 302. Using ChemSpider, 21 chemicals with the formula C ₁₉ H ₁₁ NO ₃ were proposed. MetFrag identified three chemicals with a high score (0.925) and the three most intense fragments could be explained. However, after matching the different selection criteria, only one of these candidates remained on the list. For the other candidates over the 21 proposed, the score was lower than 0.291 so we removed them from the candidates list. Regarding the MS ² prediction, the most three intense fragments are explained. The fragment 246 could also be possible by the loss of the carbonyl group present in the 5 rings of the molecule skeleton, by breakage of two C-C bonds. We saw that keto-PAHs can release such fragments (e.g. 9-fluorenone, benzanthrone, 4H-cyclopenta-[def]phenanthren-4-one). Thus, we would be able to explain four of the six fragments. The two fragments which cannot be explained with the presented structure are very low in intensity and this structure enabled to explain the most intense fragments. However, a loss of a phenyl group could be expected from parent ion or the fragment 284, as seen for the benzophenone. Furthermore, the presence of signal in ESI +/- supports this candidate.
Candidate structures	
ChemSpider ID	1) 10514029
MetFrag score (number of predicted peak)	0.925 (3)
Log K_{ow} ACD/EPI SUITE	3.8/3.8
pK_a	10.0 (OH ↔ O ⁻)

Peak 4

Information	<p>This peak of formula C₁₃H₁₀O₂N₂S (m/z 259.0538 Da), eluting at 16.2 min was detected in APCI + and ESI +. The APCI + MS² spectrum of the unknown showed that eight fragments in total were found (244.0304, 231.0808, 216.0236, 215.0160, 202.0776, 188.0292, 187.0211 and 171.0257). The neutral losses identified by MOLGEN-MSMS were: CH₄N₂, C₂H₄N₂O and C₂H₄N₂O₂. Using the ChemSpider database, 293 chemicals having C₁₃H₁₀N₂O₂S as molecular formula. This number of candidates was decreased to four using MetFrag. For the remaining 289 compounds, only one or two fragments were predicted and they had a very low score (< 0.263). Candidate 1 is better ranked as it has higher number of predicted fragments (higher number of substituents on skeleton structure). The most intense fragment (216) predicted for all candidates. The presence of a amino functional group is supported by the detection only in positive mode. These four compounds had a log <i>K_{ow}</i> values matching fractions properties. The p<i>K_a</i> value for candidates 1, 2 and 3 allow them to elute completely present in neutral fraction, while the p<i>K_a</i> value for candidate 4 allows it to elute mainly in the basic fraction and this peak was not present in there, so candidate 4 can be removed from the list.</p>			
Candidate structures				
ChemSpider ID	1) 291445	2) 4204525	3) 10167241	4) 3386684
MetFrag score (number of predicted peak)	1 (5)	0.939 (4)	0.81(4)	0.778 (3)
Log <i>K_{ow}</i> ACD/EPI SUITE	1.8/1.4	1.8/1.4	3.2/2.0	1.9/1.7
p<i>K_a</i>	na	na	na	3.9 (N _{Ar} H ⁺ ↔ N _{Ar})

Peak 5

Information	<p>This peak of formula C₁₆H₁₁NO₂ (m/z 250.0866), eluting at 16.5 min was detected in APCI +/- and ESI +. The APCI + MS² spectrum of the unknown showed two fragments (233.0604 and 205.0652), corresponding to the loss of NH₃ and CH₃NO. The signal resulting from a loss of NH₃ was the most intense and the molecule probably contains an amino group. The second loss, CH₃NO (NH₃ + CO), indicates the possible presence of a carbonyl group, although the corresponding [M+H-CO] signal was not detected. Using ChemSpider, 346 chemicals with the formula C₁₆H₁₁NO₂ were proposed. MetFrag identified five chemicals with a high score (≥ 0.879). These chemicals have an amino group in their structure, and can explain one or two of the fragments. It is expected that structures 2 and 3 will lose a benzyl group, but this fragment is not observed, thus these 2 compounds can be removed. The unknown was also present in the acidic fraction A-2-8 and compound 5 cannot elute in this fraction, thus it can be removed from the list. Structures 1 and 4 are the best candidates</p>				
Candidate structures					
ChemSpider ID	1) 4373802	2) 10520884	3) 216520	4) 2315078	
MetFrag score (number of predicted peak)	1 (1)	1 (1)	1 (0.879)	1 (0.879)	1 (0.879)
Log K_{ow} ACD/EPI SUITE	2.5/2.0	1.0/1.7	2.4/2.2	2.9/2.9	2.2/2.9
pK_a	na	12.8 (NH ₂ ↔ NH ⁻)	na	12.2 (NH ₂ ↔ NH ⁻)	2.1 (NH ₃ ⁺ ↔ NH ₂)

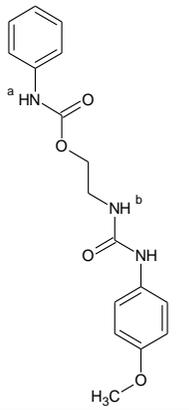
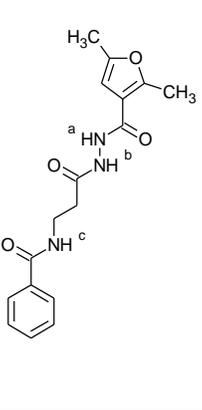
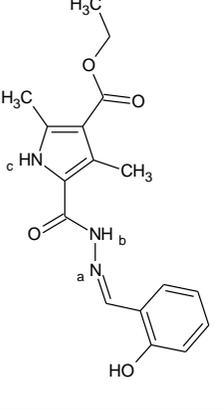
Peak 6								
Information	This peak of formula C ₁₂ H ₁₂ N ₂ (m/z 185.1066), eluting at 16.6 min was detected in APCI +, only. MS ² fragments were obtained (m/z 170.0833, 169.0756, 167.0721, 158.0961, 157.0754, 145.0880 and 144.0802). The losses corresponding to the fragments were assigned using MOLGEN-MSMS and were: CH ₃ , CH ₄ , CHN, N ₂ , C ₂ H ₄ , C ₂ H ₂ N and C ₂ H ₃ N. Using ChemSpider, 500 chemicals with the formula C ₁₂ H ₁₂ N ₂ were proposed and Metfrag identified 59 candidates with a high score (≥ 0.802 including 15 with a score ≥ 0.916 with six assigned fragments) and explaining five or six of the seven observed fragments. Amongst these candidates twelve of them had an amino functional group and were removed from the list as amino-compounds are better ionized in ESI + and should have a greater signal than in APCI +. This unknown was not detected in ESI +, supporting the absence of amino group. After matching the log <i>K_{ow}</i> and p <i>K_a</i> values, 15 candidates remained within the score range of 0.916 to 1.							
Candidate structures								
ChemSpider ID	1) 15554727	2) 25620509	3) 15555002	4) 15555691	5) 15555015	6) 15425225	7) 15350264	8) 269821
MetFrag score (number of predicted peak)	1 (6)	1 (6)	1 (6)	1 (6)	1 (6)	0.953 (6)	0.945 (6)	0.941 (6)
Log <i>K_{ow}</i> ACD/EPI SUITE	2.4/2.6	1.6/nd	2.4/2.6	2.4/2.6	2.4/2.6	2.7/2.7	2.5/2.7	1.8/2.2
p<i>K_a</i>	na	na	na	na	na	na	na	na
Candidate structures								
ChemSpider ID	9) 15160596	10) 1735125	11) 14306461	12) 2820995	13) 15160320	14) 2075988	15) 15426110	
MetFrag score (number of predicted peak)	0.941 (6)	0.941 (6)	0.941 (6)	0.941 (6)	0.941 (6)	0.941 (6)	0.916 (7)	
Log <i>K_{ow}</i> ACD/EPI SUITE	1.4/2.0	2.5/3.2	1.4/2.0	2.1/2.7	1.4/2.0	1.4/2.0	1.7/nd	
p<i>K_a</i>	na	na	na	na	na	na	2.8 (N ^a H ⁺ ↔ N ^a) 2.8 (N ^b H ⁺ ↔ N ^b)	

Peak 7

Information	This peak of formula C ₂₁ H ₁₁ NO ₃ (m/z 326.0815), eluting at 17.0 min was detected in APCI +/- and ESI +. The APCI + MS ² and MS ³ spectra of the unknown showed six fragments in total (298.0868 (-CO), 283.0759, 282.0919 (-CO ₂), 254.0961, 241.0501 (no assignment), 223.0394 (-C ₇ H ₅ NO), 195.0436 (-C ₈ H ₅ NO from 326, -CO from 223) and 139.0537 (-C ₁₀ H ₅ NO ₃). Using ChemSpider, eight chemicals with the formula C ₂₁ H ₁₁ NO ₃ were proposed and Metfrag identified only one with a score of 1, 2-phenylanthra[2,3-d][1,3]oxazole-5,10-dione, explaining only two of six fragments, not including the most intense fragments. Furthermore its log <i>K_{ow}</i> did not fit the range (5.2 and 4.8 predicted by ACD and EPISUITE respectively). All the other candidates had a score of 0. For this peak, no candidate can be provided.
--------------------	--

Peak 8

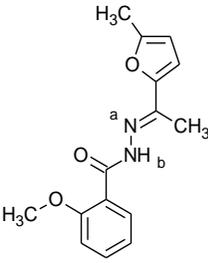
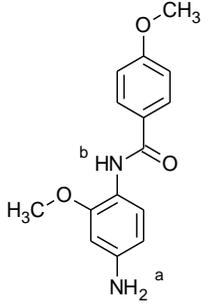
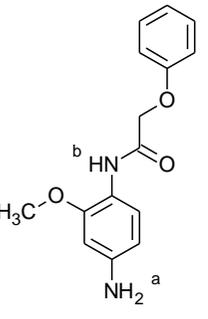
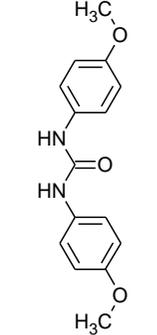
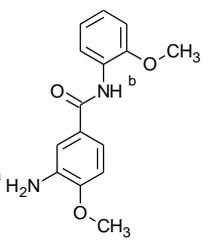
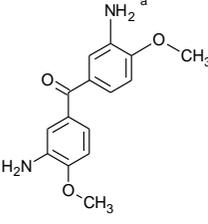
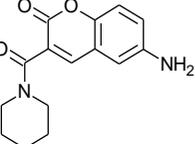
Information	This peak of formula C ₁₇ H ₁₉ N ₃ O ₄ (m/z 330.1443), eluting at 17.2 min was detected in APCI +/- and ESI +. The APCI + MS ² spectrum of the unknown was obtained from CID fragmentation, and two fragments were observed (211.1076 and 165.0657), corresponding to the loss of C ₇ H ₅ NO and C ₉ H ₁₁ NO ₂ . MS ³ (CID) of the 211 Da fragment was performed, and two fragments were obtained (196.0844 and 183.0762 Da) with the corresponding losses of CH ₃ O and C ₂ H ₄ . Using ChemSpider, 2257 chemicals with the formula C ₁₇ H ₁₉ N ₃ O ₄ were proposed. Metfrag identified ten chemicals with a score greater than 0.853 and explaining at least two fragments. Based on the explained fragments, candidate 4 would be the best as the four fragments can be predicted from the structure, candidate 10 has a hydroxyl group in its structure, the unknown is not present in ESI-, which does not support the presence of this candidate that can be removed. Candidates 2, 3, 5, 6, 7 and 8 only enabled the prediction of two fragments and thus are less good candidates. Candidate 9 could not explain the second most intense fragment and thus can be removed. The best candidates are structure 1,4 and 10 and match the log <i>K_{ow}</i> range and p <i>K_a</i> values.						
Candidate structures							
ChemSpider ID	1) 16867074	2) 832693	3) 7177168	4) 9338397	5) 20907755	6) 7724394	7) 22444805
MetFrag score (number of predicted peak)	1 (3)	0.923 (2)	0.923 (2)	0.892 (4)	0.859 (2)	0.859 (2)	0.859 (2)
Log <i>K_{ow}</i> ACD/EPI SUITE	1.5/1.9	2.9/2.5	1.9/2.4	0.2/2.4	1.0/2.0	0.8/1.8	0.9/nd
p<i>K_a</i>	11.1 (N ^a H ↔ N ^{a-})	9.0 (N ^a H ↔ N ^{a-}) 9.1 (N ^b H ↔ N ^{b-}) 9.1 (N ^c H ↔ N ^{c-})	0.7 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 13.1 (N ^b H ↔ N ^{b-}) 13.5 (N ^c H ↔ N ^{c-})	0.8 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 0.7 (N ^b H ⁺ ↔ N ^b)	0.5 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 10.1 (N ^b H ↔ N ^{b-})	0.8 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 13.3 (N ^b H ↔ N ^{b-}) 12.9 (N ^c H ↔ N ^{c-})	0.6 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 12.3 (N ^b H ↔ N ^{b-})

Candidate structures			
ChemSpider ID	8) 24782508	9) 11702249	10) 18392286
MetFrag score (number of predicted peak)	0.859 (2)	0.792 (3)	0.792 (3)
Log <i>K_{ow}</i> ACD/EPI SUITE	2.6/nd	0.1/1.6	4.0/3.3
p<i>K_a</i>	10.5 (N ^a H ↔ N ^{a-}) 12.8 (N ^b H ↔ N ^{b-})	9.5 (N ^a H ↔ N ^{a-}) 9.3 (N ^b H ↔ N ^{b-}) 13.9 (N ^c H ↔ N ^{c-})	3.3 (N ^a H ⁺ ↔ N ^a) 11.1 (N ^b H ↔ N ^{b-}) 13.5 (N ^c H ↔ N ^{c-})

Peak 9					
Information	This peak of formula C ₁₈ H ₂₁ N ₃ O ₅ (m/z 360.1557), eluting at 18.4 min was detected in APCI + and ESI +. The APCI + MS ² and MS ³ spectra of the unknown showed six fragments in total (211.1080 (-C ₈ H ₇ NO ₂), 196.0838 (-CH ₃ from 211), 183.0760 (-C ₂ H ₄ from 211), 165.0655 (-C ₂ H ₆ O from 211), 124.0755 (-C ₁₁ H ₁₂ N ₂ O ₂) and 109.0521 (-C ₁₂ H ₁₅ N ₂ O ₂)). Using ChemSpider, 1520 chemicals with the formula C ₂₁ H ₁₁ NO ₃ were proposed and Metfrag identified 8 of them with a high score (> 0.833), explaining at least three fragments of the six observed. After matching all selection criteria, three candidates were removed. Most intense ion (211) predicted for the 5 candidates remaining. Candidate 1 was available as standard reference, but did not match MS ² fragmentation pattern.				
Candidate structures					
ChemSpider ID	1) 16896706	2) 9385631	3) 4991426	4) 21985955	5) 15224742
MetFrag score (number of predicted peak)	1 (4)	0.906 (5)	0.833 (3)	0.833 (3)	0.833 (3)
Log K_{ow} ACD/EPI SUITE	1.5/2.0	0.3/2.5	2.3/1.8	2.5/nd	2.1/nd
pK_a	11.3 (N ^a H ↔ N ^{a-})	1.2 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 0.9 (N ^b H ⁺ ↔ N ^b)	11.9 (N ^a H ↔ N ^{a-}) 12.8 (N ^b H ↔ N ^{b-}) 13.5 (N ^c H ↔ N ^{c-})	13.5 (N ^a H ↔ N ^{a-}) 13.4 (N ^b H ↔ N ^{b-}) 10.9 (N ^c H ↔ N ^{c-})	13.5 (N ^a H ↔ N ^{a-}) 13.5 (N ^b H ↔ N ^{b-}) 11.4 (N ^c H ↔ N ^{c-})

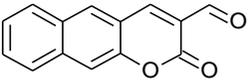
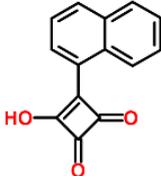
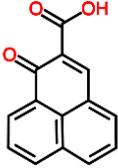
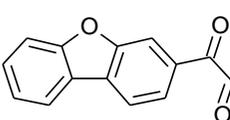
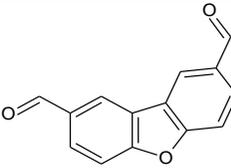
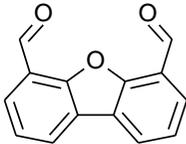
Peak 10	
Information	This peak of formula C ₁₇ H ₁₂ O (m/z 233.0963), eluting at 18.5 min was detected in APCI + only. The APCI + MS ² spectrum of the unknown showed six fragments in total (218.0730, 217.1589, 215.0860, 205.1016 and 203.0857, 188.0924). The neutral losses were assigned to four of these fragments: CH ₃ , H ₂ O, CO and CHO. Using ChemSpider, 113 chemicals with C ₁₇ H ₁₂ O as formula were found. MetFrag identified five chemicals explaining all four fragments observed (score from 1 to 0.862), but none fitted the log K _{ow} range (> 3.2) or the pK _a value allowing an elution in the neutral fraction. Other structures fitting these properties were found, with a maximum of two fragments predicted. Their score ranked from 0 to 0.687, with only one molecule having a score higher than 0.5. Thus, the best candidate found was 7,9-dimethyl-8H-cyclopenta[a]acenaphthylen-8-one. The most intense fragment was predicted and corresponded to the loss of the methyl group. The loss of the carbonyl group was also predicted. Retention time of chemicals increases with increases of number of aromatic rings. The 9-fluorenone (EPI SUITE log K _{ow} 3.6, 3 fused rings eluted at 23 minutes), benzanthrone, (EPI SUITE log K _{ow} 4.8 eluted at 32 minutes), this candidate should elute after the 9-fluorenone and cannot be the candidate for Peak 40. No candidate can be provided for this peak.

Peak 11

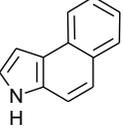
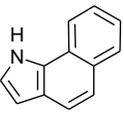
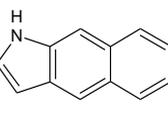
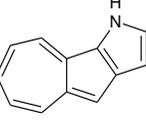
Information	<p>This peak of formula C₁₅H₁₆N₂O₃ (m/z 273.1236), eluting at 19.4 min was detected in APCI + and ESI +. The APCI + MS² spectrum of the unknown showed two fragments (124.0755 and 109.0521) with the corresponding loss of C₈H₇NO₂ and C₉H₁₀NO₂ (or loss of CH₃ from the fragment at 124 Da). Using ChemSpider database more than 2900 chemicals with the formula C₁₂H₁₃NO₂ were proposed. MetFrag identified 18 chemicals that explained one or two of the observed fragments with a score greater than 0.873. Three candidates had a hydroxyl group in their structure, but no signal in ESI-, which does not support the presence of such compounds. After matching the pK_a and log K_{ow} seven remained. The presence of an amino group in candidate structures is supported by the presence of intense signal in ESI+.</p>						
Candidate structures							
ChemSpider ID	1) 757961	2) 833606	3) 833624	4) 120696	5) 8811261 (+ 2 isomers 16805321 and 13317570)	6) 660964	7) 26898
MetFrag score (number of predicted peak)	1 (1)	0.987 (1)	0.987 (1)	0.987 (1)	0.987 (1)	0.987 (1)	0.873 (2)
Log K_{ow} ACD/EPI SUITE	2.3/3.4	0.9/1.4	1.6/1.4	2.8/3.1	2.1/1.4	2.5/1.5	0.4/0.7
pK_a	3.8 (N ^a H ⁺ ↔ N ^a) 11.7 (N ^b H ↔ N ^{b-}) 1.8 (N ^b H ₂ ⁺ ↔ N ^b H)	3.4 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 13.6 (N ^b H ↔ N ^{b-}) 0.07(N ^b H ₂ ⁺ ↔ N ^b H)	3.4 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 13.6 (N ^b H ↔ N ^{b-}) 0.07(N ^b H ₂ ⁺ ↔ N ^b H)	13.6 (NH ↔ N ⁻)	2.7 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 13.6 (N ^b H ↔ N ^{b-}) 0.07(N ^b H ₂ ⁺ ↔ N ^b H)	2.7 (N ^a H ₃ ⁺ ↔ N ^a H ₂)	3.6 (NH ₃ ⁺ ↔ NH ₂)

Peak 13				
Information	<p>This peak of formula C₂₁H₂₀N₄O₃ (m/z 377.1598), eluting at 20.2 min was detected in APCI +/- and ESI +. The MS² spectra showed two fragments corresponded to the two halves of the molecule (254.0921 corresponding to C₁₄H₁₂N₃O₂ and 124.0756 to C₇H₁₀NO). Thus, the molecule fragmented easily in two fragments as observed for two standards, benzophenone and 1,3-dimethyl-1,3-diphenylurea, for which the cleavage appeared at the carbonyl group. MS³ fragmentation of the most intense ion, 254.0921 Da was performed. The fragments obtained were 169.0488 (-C₈H₅O), 135.0549 (-C₇H₅NO), 107.0599 (-CO from 135 ion) and 94.0647 (-C₈H₄N₂O₂). In total six fragments were used to search the ChemSpider database with MetFrag. There were more than 1600 hits, but only seven of them showed a very high score (≥ 0.856) and could explain four or five fragments of the six observed fragments. After verification of the log <i>K_{ow}</i> and p<i>K_a</i> values, three were excluded as their p<i>K_a</i> values would have led to elution the basic fraction. APCI + signal is similar to ESI + signal, which do not support the presence of amino group in the structure of the unknown. Candidates 2, 3 and 4 are less likely.</p>			
Candidate structures				
ChemSpider ID	1) 16999411	2) 11536340	3) 11536336	4) 8197974
MetFrag score (number of predicted peak)	1 (4)	0.945 (5)	0.945 (5)	0.945 (5)
Log <i>K_{ow}</i> ACD/EPI SUITE	3.9/3.5	2.0/1.9	2.5/2.5	3.8/1.4
p<i>K_a</i>	0.5 (N ^a H ⁺ ↔ N ^a) 11.8 (N ^b H ↔ N ^{b-}) 13.5 (N ^c H ↔ N ^{c-}) 14.0 (N ^d H ↔ N ^{d-})	2.8 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 13.2 (N ^b H ↔ N ^{b-}) 11.3 (N ^c H ↔ N ^{c-}) 3.8 (N ^d H ↔ N ^{d-})	2.8 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 13.2 (N ^b H ↔ N ^{b-}) 11.4 (N ^c H ↔ N ^{c-}) 3.3 (N ^d H ↔ N ^{d-})	2.8 (N ^a H ₃ ⁺ ↔ N ^a H ₂) 13.2 (N ^b H ↔ N ^{b-}) 11.4 (N ^c H ↔ N ^{c-}) 4.4 (N ^d H ↔ N ^{d-})

Peak 15

Information	<p>This peak of formula C₁₄H₈O₃ (m/z 225.0543), eluting at 20.5 min was detected in APCI +/- and ESI -. The APCI + MS² spectrum of the unknown showed three fragments (197.0596 (-CO), 141.0695 (-C₃O₃) and 100.1116 (no assignment)). Using ChemSpider, 44 chemicals with the formula C₁₄H₈O₃ were proposed. MetFrag identified three chemicals with a score greater than 0.956, explaining the two assigned fragments, and also identified three other chemicals with a score between 0.77 and 0.956 explaining only one of the assigned fragments. No ESI – signal, does not support OH presence. No negative mode signal, this does not support the carboxylic acid functional group presence. Thus, candidates 2 and 3 can be removed from the list. Based on the structures of candidates 4, 5 and 6 it is impossible that they lose a C₃O₃ fragment without splitting the whole molecule in two. Thus, candidate 1 is the best candidate that we can propose as it is the only one explaining the two assigned fragments.</p>					
Candidate structures						
ChemSpider ID	1) 9416609	2) 8189431	3) 321437	4) 4436917	5) 13920496	6) 10129130
MetFrag score (number of predicted peak)	0.996 (2)	0.996 (2)	0.996 (2)	0.77 (2)	0.77(1)	0.77 (1)
Log K_{ow} ACD/EPI SUITE	2.6/1.8	1.9/2.0	3.2/2.7	2.9/3.0	3.0/3.5	2.8/3.5
pK_a	na			na	na	na

Peak 16

Information	<p>This peak of formula C₁₂H₉N (m/z 168.0801 Da), eluting at 24.1 min was detected in APCI + only. Its MS² spectrum obtained by APCI + revealed only one fragment, corresponding to the loss of a hydrogen atom (m/z 167.0727), similar to carbazole (m/z 168.0808, empirical formula C₁₂H₉N and retention time of 28.1 min). The search performed with ChemSpider provided 103 hits, with only 58 fitting the log <i>K_{ow}</i> range (0.5 – 4.2 including the error range). As only the parent ion was present in the MS² spectrum, MetFrag could not be used to reduce the number of candidates. The 58 hits presented two types of structure, one type with substituents on aromatic rings, for which fragment(s) were likely and one without substituents on the aromatic rings, for which fragment(s) were less likely. As, substituents on aromatic rings are easily lost during fragmentation (e.g. methyl group: 2-amino-3-methylimidazo[4,5-f]quinoline or 1-methyl-9H-pyrido[3,4-b]indole, amino group: N-phenylnaphthalen-2-amine or 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine or 1,8-diaminopyrene), such compounds could be removed from the list. Furthermore, this unknown was only observed in APCI + and not in ESI mode, excluding the presence of an amino group. Further decrease was possible due to the match of log <i>K_{ow}</i> and p<i>K_a</i> value.</p>			
Candidate structures				
ChemSpider ID	1) 9519149	2) 89061	3) 10496645	4) 14524926
MetFrag score (number of predicted peak)	na	na	na	na
Log <i>K_{ow}</i> ACD/EPI SUITE	3.4/3.2	3.4/3.2	3.7/3.2	3.7/3.2
p<i>K_a</i>	na	na	na	na

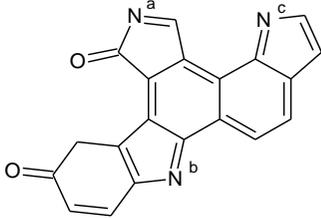
Peak 17	
Information	<p>This peak (m/z 323.0706), eluting at 25.3 min was detected in APCI + and ESI +. The APCI + MS² spectrum of the unknown showed one fragment at 295.0763, corresponding to the loss of a carbonyl group.</p> <p>Two formulas were assigned.</p> <p>The first one was C₂₂H₁₀O₃ Using ChemSpider, five chemicals with the formula C₂₂H₁₀O₃ were proposed, but none of them matched the log <i>K_{ow}</i> range (> 4.5).</p> <p>The second formula to search for a structure was C₂₀H₉N₃O₂. Using ChemSpider, two hits were found but only one chemical fitted the log <i>K_{ow}</i> range. Most of this compound should elute in the basic fraction, due to the p<i>K_a</i> values. However, a small part is still expected in the neutral fraction. The loss of CO can be predicted. However from the structure more fragments would be expected (e.g. another loss of a carbonyl group as seen for the keto-PAHs and hydroxyquinone and the loss of a CHN group as seen for most of the azaarenes). There is a large difference between the log <i>k_{ow}</i> predicted value, and based on chromatographic retention time, this candidate is expected to elute later than 25 min.</p> <p>No candidate can be provided for this peak</p>
Candidate structures	
ChemSpider ID	10705285
MetFrag score (number of predicted peak)	1 (1)
Log <i>K_{ow}</i> ACD/EPI SUITE	2.9/4.5
p<i>K_a</i>	1.7 (N ^a H ⁺ ↔ N ^a) 0.3 (N ^b H ⁺ ↔ N ^b) 2.5 (N ^c H ⁺ ↔ N ^c)

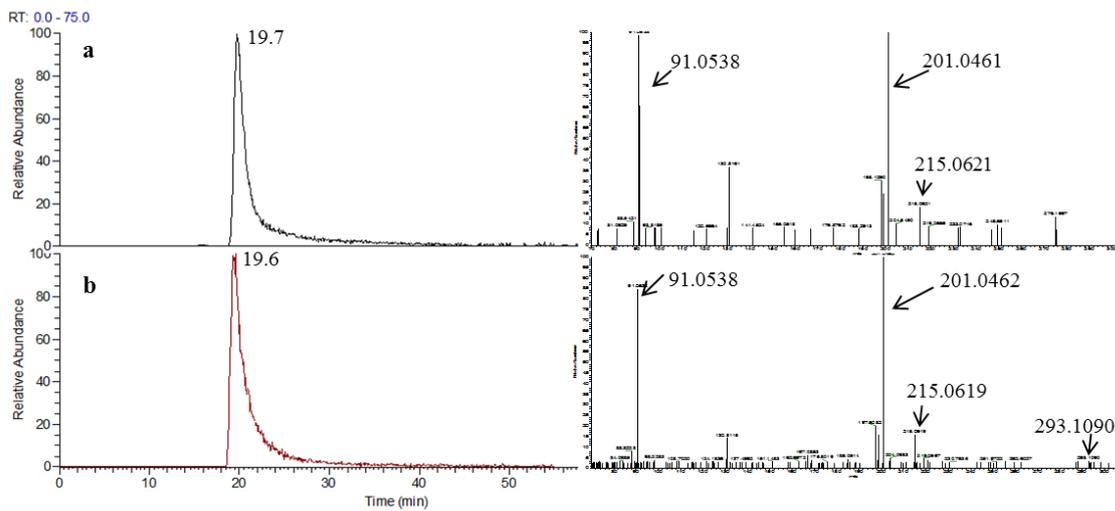
Table S5: Retention and mutagenicity prediction for the unknown compounds and their candidates, with +: candidate retained on the list and -: candidate eliminated from the list, unknown: prediction model not applicable to these compounds, Ames+: high probability of being mutagenic, Ames -: high probability of being non mutagenic.

Name (ChemSpider ID)	Candidates CHI range	Unknown CHI	Retention prediction	$\Delta\Delta E$ (kcal/mol)	Mutagenicity prediction	Candidates of interest
Unknown Peak 1		80.01				
5,6-diethyl-1H-indole-2,3-dione (13494962)	66.12 – 78.34		-	na	unknown	no
4,5,6,7-tetramethyl-1H-indole-2,3-dione (19461480)	67.67 – 79.90		-	na	unknown	no
5-tert-butyl-1H-indole-2,3-dione (517156)	64.68 – 76.90		-	na	unknown	no
5-butyl-1H-indole-2,3-dione (2339601)	65.40 – 77.61		-	na	unknown	no
5-(1-methylpropyl)-1H-indole-2,3-dione (16812332)	64.58 – 76.80		-	na	unknown	no
7-(2-methylpropyl)-1H-indole-2,3-dione (513041)	64.58 – 76.80		-	na	unknown	no
7-amino-4-propyl-2H-chromen-2-one (2012579)	71.76 – 83.95		+	5.48	Ames -	no
Unknown Peak 2		76.46				
3-amino-2,5-diphenylcyclohexa-2,5-diene-1,4-dione (PubChem ID 12251096)	71.92 – 84.87		+	4.57	inconclusive	yes
Unknown Peak 3		77.73				
3-(2-hydroxybenzoyl)-5H-indeno[1,2-b]pyridin-5-one (10514029)	73.54 – 86.05		+	na	unknown	no
Unknown Peak 4		78.23				
2-amino-5-(1,3-benzodioxol-5-yl)-4-methylthiophene-3-carbonitrile (291445)	65.82 – 78.16		-	- 432.34	Ames +	no
2-amino-4-(2,3-dihydro-1,4-benzodioxin-6-yl) thiophene-3-carbonitrile (4204525)	59.56 – 72.14		-	- 49.01	Ames +	no
methyl 3-amino-5-(4-cyanophenyl)thiophene-2-carboxylate (10167241)	72.72 – 85.00		+	0.48	inconclusive	yes
Unknown Peak 5		79.12				
2-[amino(phenyl)methylidene]-1H-indene-1,3(2H)-dione (8197974)	57.99 – 70.49		-	na	unknown	no
9-ethynyl-9H-fluoren-1-yl carbamate (2315078)	76.72 – 88.99		+	na	unknown	no
Unknown Peak 6		49.08				
4-(cyanomethyl)-2-propylbenzotrile (15554727)	83.87 – 96.23		-	na	unknown	no
2-(4-ethylphenyl) butanedinitrile (25620509)	79.28 – 91.60		-	na	unknown	no
4-(cyanomethyl)-3-propylbenzotrile (15555002)	83.87 – 96.23		-	na	unknown	no
3-(cyanomethyl)-2-propylbenzotrile (15555691)	83.87 – 96.23		-	na	unknown	no
3-(cyanomethyl)-4-propylbenzotrile (15555015)	83.87 – 96.23		-	na	unknown	no
(2,3-dimethyl-1H-indol-1-yl)acetonitrile (15425225)	82.31 – 94.50		-	na	unknown	no
3-(2-methyl-1H-indol-1-yl) propanenitrile (15350264)	81.37 – 93.56		-	na	unknown	no
2,2'-(4,5-dimethyl benzene-1,2-diyl) diacetonitrile (269821)	84.13 – 96.44		-	na	unknown	no
3-(3-methylphenyl) pentanedinitrile (15160596)	81.71 – 94.07		-	na	unknown	no
3-(2-methylindolizin-3-yl) propanenitrile (1735125)	80.73 – 92.92		-	na	unknown	no
3-(4-methylphenyl) pentanedinitrile (14306461)	81.71 – 94.07		-	na	unknown	no
3-(3-methyl-1H-indol-2-yl) propanenitrile (15160320)	80.03 – 92.22		-	na	unknown	no
3-(2-methylphenyl) pentanedinitrile (2075988)	81.71 – 94.07		-	na	unknown	no
3-(3-methyl-1H-indol-1-yl) propanenitrile (15426110)	81.37 – 93.56		-	na	unknown	no

Name (ChemSpider ID)	Candidates CHI range	Unknown CHI	Retention prediction	$\Delta\Delta E$ (kcal/mol)	Mutagenicity prediction	Candidates of interest
Unknown Peak 8		80.08				
1-methyl-2-oxo-2-(phenylamino)ethyl 6-oxo-1-propyl-1,6-dihydropyridazine-3-carboxylate (16867074)	54.22 – 67.06		-	na	unknown	no
ethyl 3-[(2-aminophenyl)carbonyl]-1-(cyclopropyl carbonyl)-4,5-dihydro-1H-pyrazole-5-carboxylate (9338397)	56.31 – 69.17		-	- 44.50	Ames +	no
ethyl 5-(((2Z)-2-[(2-hydroxyphenyl) methylidene] hydrazine)carbonyl)-2,4-dimethyl-1H-pyrrole-3-carboxylate (18392286)	61.32 – 73.86		-	na	unknown	no
Unknown Peak 9		80.72				
ethyl 3-(2-amino-5-methoxybenzoyl)-1-(cyclopropyl-carbonyl)-4,5-dihydro-1H-pyrazole-5-carboxylate (9385631)	55.32 – 68.38		-	- 1.25	inconclusive	no
N-(2,4-dimethoxy-5-[(2-methoxy phenyl) carbamoyl] amino)phenyl)acetamide (4991426)	58.66 – 71.44		-	na	unknown	no
N-(2,5-dimethoxy phenyl)-2-[(3-methoxy phenyl) carbamoylamino] acetamide (21985955)	52.86 – 65.72		-	na	unknown	no
N-(2,4-dimethoxy phenyl)-2-[(3-methoxy phenyl) carbamoylamino]acetamide (15224742)	52.96 – 65.99		-	na	unknown	no
Unknown Peak 11		81.07				
2-methoxy-N'-[1-(5-methylfuran-2-yl) ethylidene] benzohydrazide (757961)	74.01 – 86.36		+	na	unknown	no
benzamide, N-(4-amino-2-methoxyphenyl)-4-methoxy- (833606)	56.14 – 68.61		-	44.05	Ames -	no
acetamide, N-(4-amino-2-methoxyphenyl)-2-phenoxy- (833624)	55.17 – 67.66		-	153.34	Ames -	no
1,3-bis(4-methoxyphenyl)urea (120696)	65.47 – 77.82		-	na	unknown	no
1,3-bis(2-methoxyphenyl)urea (260536)	68.93 – 81.25		+	na	unknown	no
1-(2-methoxyphenyl)-3-(4-methoxyphenyl)urea (260538)	67.23 – 79.56		-	na	unknown	no
1-(3-methoxyphenyl)-3-(4-methoxyphenyl)urea (761644)	67.53 – 79.86		-	na	unknown	no
1,3-bis(3-methoxyphenyl)urea (641853)	69.66 – 81.98		+	na	unknown	no
1-(2-methoxyphenyl)-3-(3-methoxyphenyl)urea (260537)	69.28 – 81.60		+	na	unknown	no
3-amino-4-methoxy-N-(2-methoxyphenyl)benzamide (8811261)	55.78 – 68.25		-	38.54	Ames -	no
3-amino-4-methoxy-N-(3-methoxyphenyl)benzamide (16805321)	56.16 – 68.63		-	-12.40	Ames +	no
3-amino-4-methoxy-N-(4-methoxyphenyl)benzamide (13317570)	54.02 – 66.52		-	44.68	Ames -	no
bis(3-amino-4-methoxyphenyl) methanone (660964)	53.62 – 66.18		-	150.54	Ames -	no
6-amino-3-(piperidin-1-ylcarbonyl)-2H-chromen-2-one (26898)	56.08 – 68.67		-	- 1.49	inconclusive	no
Unknown Peak 13		82.03				
2-ethoxy-N-(3-[(phenylcarbamoyl) amino]phenyl) pyridine-3-carboxamide (16999411)	56.96 – 69.94		-	na	unknown	no
Unknown Peak 14		81.71				
(6Z)-6-[1-[2-(2-methylphenyl)hydrazino]ethylidene] cyclohexa-2,4-dien-1-one (5201826)	63.12 – 75.60		-	na	unknown	no
(6Z)-6-[1-[2-(3-methylphenyl)hydrazino]ethylidene] cyclohexa-2,4-dien-1-one (5247178)	63.12 – 75.60		-	na	unknown	no

Name (ChemSpider ID)	Candidates CHI range	Unknown CHI	Retention prediction	$\Delta\Delta E$ (kcal/mol)	Mutagenicity prediction	Candidates of interest
6-[1-[2-(4-methylphenyl)hydrazino]ethylidene] cyclohexa-2,4-dien-1-one (5253314)	63.12 – 75.60		-	na	unknown	no
4-[1-[2-(4-methylphenyl)hydrazinyl]ethylidene] cyclohexa-2,5-dien-1-one (7349810)	63.12 – 75.60		-	na	unknown	no
4-[1-[2-(3-methylphenyl)hydrazinyl]ethylidene] cyclohexa-2,5-dien-1-one (7350957)	63.12 – 75.60		-	na	unknown	no
4-[1-[2-(2-methylphenyl)hydrazino]ethylidene] cyclohexa-2,5-dien-1-one (7305321)	63.12 – 75.60		-	na	unknown	no
3-amino-4-methyl-N-(4-methylphenyl)benzamide (15117243)	65.89 – 78.20		-	- 0.99	inconclusive	no
4-amino-3-methyl-N-(2-methylphenyl)benzamide (14815425)	65.89 – 78.20		-	2.49	inconclusive	no
3-amino-4-methyl-N-(3-methylphenyl)benzamide (16801713)	65.89 – 78.20		-	0.11	inconclusive	no
4-amino-3-methyl-N-(3-methylphenyl)benzamide (16808230)	65.89 – 78.20		-	2.47	inconclusive	no
3-amino-4-methyl-N-(2-methylphenyl)benzamide (16817934)	65.89 – 78.20		-	- 0.19	inconclusive	no
3-amino-2-methyl-N-(2-methylphenyl)benzamide (16817998)	65.89 – 78.20		-	- 0.16	inconclusive	no
3-amino-2-methyl-N-(3-methylphenyl)benzamide (16802496)	65.89 – 78.20		-	- 0.12	inconclusive	no
3-amino-2-methyl-N-(4-methylphenyl)benzamide (16810443)	65.89 – 78.20		-	- 0.16	inconclusive	no
4-amino-3-methyl-N-(4-methylphenyl)benzamide (16810500)	65.89 – 78.20		-	2.46	inconclusive	no
2-amino-3-methyl-N-(2-methylphenyl)benzamide (16814789)	70.33 – 82.60		+	2.92	inconclusive	yes
2-amino-3-methyl-N-(3-methylphenyl)benzamide (16810019)	70.33 – 82.60		+	2.58	inconclusive	yes
2-amino-3-methyl-N-(4-methylphenyl)benzamide (16801544)	70.33 – 82.60		+	2.59	inconclusive	yes
2-amino-6-methyl-N-(2-methylphenyl)benzamide (16824275)	70.33 – 82.60		+	2.56	inconclusive	yes
2-amino-6-methyl-N-(3-methylphenyl)benzamide (16805125)	70.33 – 82.60		+	2.72	inconclusive	yes
1-[2-amino-5-(benzylamino)phenyl]ethanone (9086908)	70.67 – 82.95		+	- 18.12	Ames +	yes
1-[4-amino-3-(benzylamino)phenyl]ethanone (19441053)	67.97 – 80.27		-	- 21.36	Ames +	no
bis(4-amino-3-methylphenyl)methanone (11255017)	63.92 – 76.26		-	103.69	Ames -	no
3,3'-diamino-4,4'-dimethylbenzophenone (508004)	63.92 – 76.26		-	111.26	Ames -	no
(4-amino-2-methylphenyl)(4-amino-3-methylphenyl) methanone (15990795)	63.92 – 76.26		-	112.10	Ames -	no
bis(4-amino-2-methylphenyl)methanone (15990812)	63.92 – 76.26		-	113.19	Ames -	no
(3-amino-5-methylphenyl)(4-amino-3-methylphenyl) methanone (15990839)	63.92 – 76.26		-	108.50	Ames -	no
bis(3-amino-5-methylphenyl)methanone (15990847)	63.92 – 76.26		-	106.73	Ames -	no
(3-amino-5-methylphenyl)(4-amino-2-methylphenyl) methanone (15990856)	63.92 – 76.26		-	109.68	Ames -	no
4-amino-N-benzyl-3-methylbenzamide (16491734)	65.91 – 78.26		-	3.64	inconclusive	no
3-amino-N-benzyl-2-methylbenzamide (16815979)	65.91 – 78.26		-	- 0.60	inconclusive	no
3-amino-N-benzyl-4-methylbenzamide (16822378)	65.91 – 78.26		-	-0.70	inconclusive	no
2-amino-N-benzyl-3-methylbenzamide (16821380)	70.28 – 82.55		+	2.80	inconclusive	yes
2-amino-N-benzyl-6-methylbenzamide (16818487)	70.28 – 82.55		+	3.17	inconclusive	yes
2-amino-N-benzyl-5-methylbenzamide (497152)	70.28 – 82.55		+	0.88	inconclusive	yes
Unknown Peak 15		86.58				
2-oxo-2H-benzo[g]chromene-3-carbaldehyde (9416609)	68.15 – 80.48		-	na	unknown	no
Unknown Peak 16		87.57				
3H-benz[e]indole (9519149)	78.40 – 90.62		+	na	unknown	no
1H-benz[f]indole (10496645)	78.40 – 90.62		+	na	unknown	no
1H-azuleno[1,2-b]pyrrole (14524926)	65.52 – 78.06		-	na	unknown	no

Figure S1: Comparison of retention time and MS² spectra of unknown Peak 36 (a) with the reference standard benzyl(diphenyl)phosphine oxide (b).



SI S2: Experiment to confirm the presence of the benzyl(diphenyl) phosphine oxide in fraction N2-8 and mutagenicity assay

A standard solution at a concentration of 1 mg/L was injected on the LC-MS system described above (Supplementary Information S2) with the same chromatographic gradient and MS system used for the analysis of fraction N2-8.

The Ames fluctuation assay (Ames II test) using the *Salmonella typhimurium* strain TA98 with and without external metabolic activation was used to screen for mutagenic compounds, as described by Perez and co-workers (S. Pérez, G. Reifferscheid, P. Eichhorn, D. Barceló, *Environmental Toxicology Chemistry*, 22 (2003) 2576) with slight variations as stated below. Samples were tested with three replicates per sample, four replicates for the positive controls and at least eight replicates for the negative controls (DMSO only). The positive controls were 2-nitrofluorene (20 ng/well) for TA98 without metabolic activation and 2-aminoanthracene (5ng/well) for TA98 with metabolic activation by S9 mix. The significance of mutagenic effects was tested with the Wilcoxon-Mann-Whitney test with $\alpha=0.5$. The benzyl(diphenyl) phosphine oxide was tested from 0.0029 to 11.73 mg/L (maximum solubility). The results are presented in the two following Figure S6a and Figure S6b

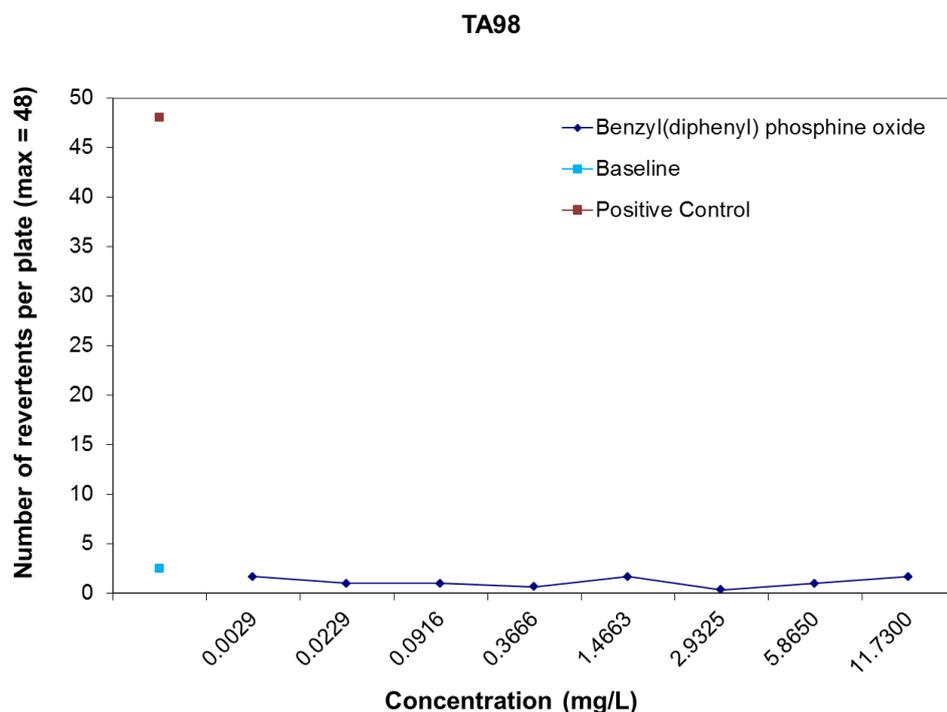


Figure S2a: Ames test results with TA98 without S9 activation

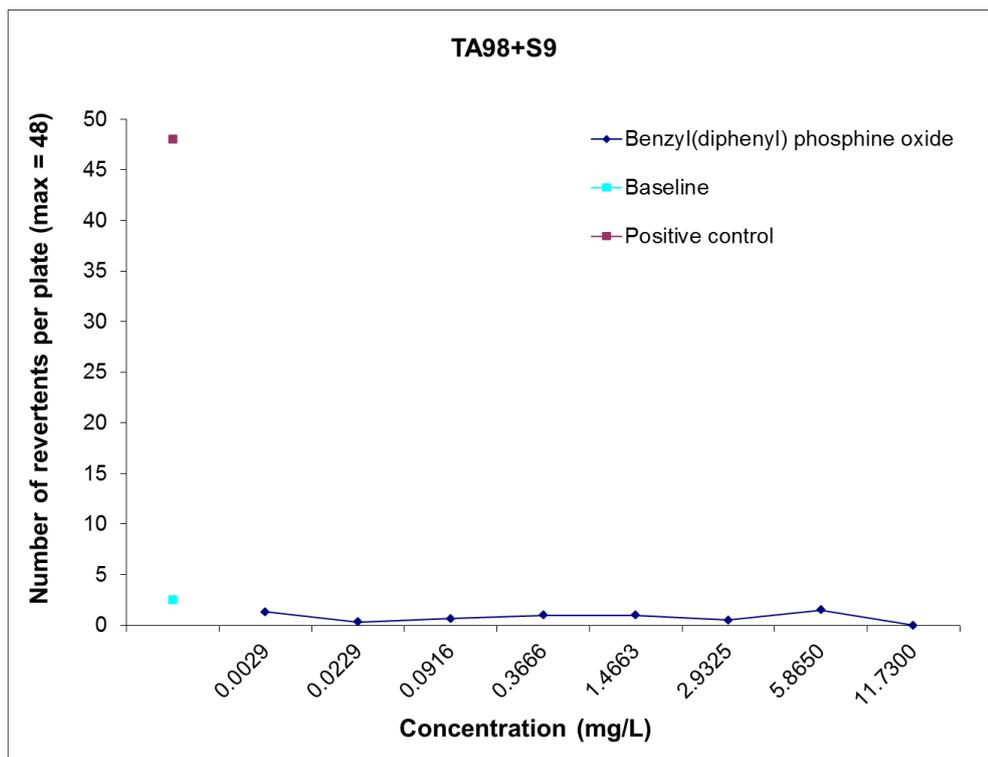
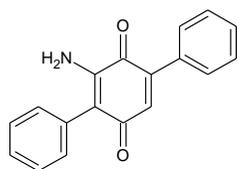
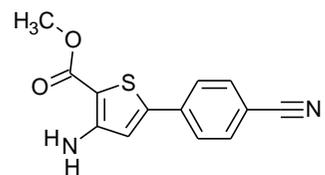


Figure S2b: Ames test results with TA98 with S9 activation

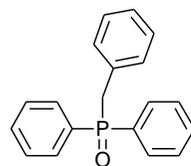
Figure S3: Candidate structures for Peaks 2, 4, 12 and 14



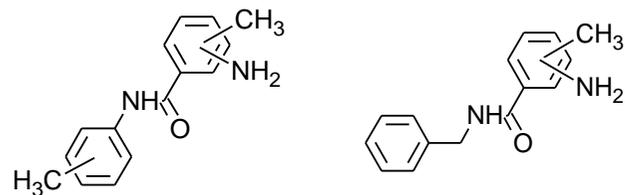
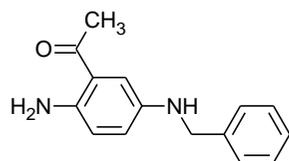
Peak 2



Peak 4



Peak 12



Peak 14