SUPPORTING INFORMATION FOR:

PRIORITIZING UNKNOWN TRANSFORMATION PRODUCTS FROM BIOLOGICALLY-TREATED WASTEWATER USING HIGH-RESOLUTION MASS SPECTROMETRY, MULTIVARIATE STATISTICS, AND METABOLIC LOGIC

Jennifer E. Schollée,[1,2] Emma L. Schymanski,[1] Sven E. Avak,[1,3] Martin Loos,[1,2] and Juliane Hollender*[,1,2]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

[2]Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092 Zürich, Switzerland

[3]Department of Chemistry, University of Zürich, 8057 Zürich, Switzerland

*juliane.hollender@eawag.ch. phone: +41-58-765-5493. fax: +41-58-765-5893

LIST OF TABLES:

LIST OF FIGURES:

Table S-1. List of internal standards used for quantification and matrix correction factor calculations.

Internal Standard Name	Formula	Monoisotopic mass	Retention time	$[M+H]^+$
Benzotriazol-D4	C6H1D4N3	123.0729	4.49	124.0807
Metformin-D6	C4H5N5D6	135.1391	0.98	136.1464
5-Methyl-Benzotriazol-D6	C7H1D6N3	139.1011	5.82	140.1089
Amphetamine-D6	C9H7D6N	141.1425	3.79	142.1497
Chloridazon-desphenyl-15N2	C4H4ClN[15]N2O	146.9984	1.25	148.0056
Methamphetamine-D5	C10H10D5N	154.1518	3.91	155.1591
3-Acetamidophenol (Paracetamol-D4)	C8H5D4N1O2	155.0879	2.36	156.0957
Chloridazon-methyl-desphenyl-D3	C5H3D3ClN3O	162.0387	2.18	163.046
Ephedrine-D3	C10H12D3NO	168.1342	3.16	169.1415
Metronidazole-D4	C6H5D4N3O3	175.0895	2.32	176.0968
Gabapentin-D4	C9H13NO2D4	175.151	3.69	176.1583
Atrazine-desisopropyl-D5	C5H3D5ClN5	178.07766	4.34	179.0855
Desethylatrazine-15N3	C6H10ClN2[15]N3	190.0536	5.67	191.0609
Phenazone-D3 (Antipyrine-D3)	C11H9D3N2O	191.1143	4.68	192.1211
2,6-Dichlorbenzamide-3,4,5-D3	C7H2D3Cl2NO	191.9932	3.92	193.0009
Carbendazim-D4	C9H5D4N3O2	195.09513	3.63	196.10186
N,N-diethyl-3-methylbenzamide-D10 (DEET-D10)	C12H7D10NO	201.1932	8.57	202.2011
Atrazine-2-Hydroxy-D5	C8H10D5N5O	202.15959	4.52	203.16632
Caffeine-D9	C8H1N4O2D9	203.13632	3.97	204.14414
Simazine-D5	C7H7D5ClN5	206.109	6.96	207.1168
Isoproturon-D6	C12H12D6N2O	212.179	8.78	213.1869
Chlorotoluron-D6	C10H7D6ClN2O	218.1088	8.23	219.1166
Atrazine-D5	C8H9D5ClN5	220.1246	8.39	221.1324
Primidone-D5	C12H9D5N2O2	223.1375	5.36	224.1442
Chloridazon-D5	C10H3D5ClN3O	226.067	5.24	227.0742
Ritalinic acid-D10	C13H7N1O2D10	229.1887	2.58	230.196
Octilinone-D17	C11H2D17NOS	230.226	11.64	231.2327
Naproxen-D3	C14H11O3D3	233.1131	10.15	234.1204
Terbuthylazine-D5	C9H11D5ClN5	234.1403	10.22	235.1481
Dimethoate-D6	C5H6D6NO3PS2	235.038	5.17	236.0446
Propazine-D6	C9H10D6ClN5	235.1465	9.9	236.1544
Tebutam-D4	C15H19D4NO	237.2025	11.78	238.2103
Diuron-D6	C9H4D6Cl2N2O	238.0541	9.01	239.062
Mefenamic acid-D3	C15H12D3N1O2	244.1297	14.55	245.1364
Carbamazepine-D8	C15H4D8N2O	244.1446	7.71	245.1525
Pirimicarb-D6	C11H12D6N4O2	244.1808	5.01	245.1881
Bisphenol-A-D16	C15D16O2	244.2149	18.63	245.2227
Lidocaine-D10	C14H12N2O1D10	244.236	4.36	245.2433
Terbutryn-D5	C10H14D5N5S1	246.1681	8.71	247.1748
Chlothianidin-D3	C6H5D3ClN5O2S	252.0275	4.85	253.0348

Internal Standard Name	Formula	Monoisotopic mass	Retention time	$[M+H]^+$
N,O-Didesmethylvenlaflaxine-D3	C15H20D3NO2	252.1917	5.01	253.199
Sulfapyridine-D4	C11H7D4N3O2S1	253.0818	3.06	254.0896
Sulfadiazine-D4	C10H6D4N4O2S	254.077	2.43	255.0848
Carbamazepine-10,11-Epoxide-13C1,D2	C14H10D2N2O2[13]C	255.1058	6.3	256.1131
Sulfamethoxazole-D4	C10H7D4N3O3S	257.0767	4.6	258.0845
Methiocarb-D3	C11H12D3NO2S	258.0471	10.31	229.1084
Atomoxetine-D3	C17H18D3NO	258.1811	8.27	259.1884
Sulfathiazole-D4	C9H5D4N3O2S2	259.0382	2.88	260.046
Imidacloprid-D4	C9H6D4ClN5O2	259.0775	4.76	260.0848
Lamotrigine-13C3,D3	C6[13]C3H4D3Cl2N5	261.0367	5.47	262.044
Irgarol-D9	C11H10D9N5S1	262.1932	9.21	263.1999
Cyclophosphamide-D4	C7H11Cl2N2O2PD4	264.0499	6.27	265.0572
Gemcitabine-13C,15N2	C8H11F2NO4[13]C[15]N2	266.0692	1.27	267.0765
Propranolol-D7	C16H14D7NO2	266.2006	6.76	267.2084
N-Desmethylvenlaflaxine-D3	C16H22D3NO2	266.2074	6.47	267.2146
2',2'-Difluoro-2-deoxyuridine-13C,15N2	C8H10F2O5[13]C[15]N2	267.0532	1.86	268.0605
Tramadol-D6	C16H19N1O2D6	269.2262	4.92	270.2335
O-Desmethylvenlaflaxine-D6	C16H19D6NO2	269.2262	4.94	270.2335
Valsartan-acid-D4	C14H6D4N4O2	270.1055	6.83	271.1128
Atenolol-acid-D5	C14H16N1O4D5	272.1784	3.9	273.1857
Atenolol-D7	C14H15D7N2O3	273.207	2.39	274.2143
Metoprolol-D7	C15H18D7NO3	274.2268	5.17	275.2347
Dimethenamid-D3	C12H15D3ClNO2S	278.093	10.12	279.1008
Sotalol-D6	C12H14D6N2O3S	278.1566	1.98	279.1644
EDDP-D3	C20H20D3N	280.2019	6.78	281.2092
Venlafaxine-D6	C17H21N1O2D6	283.2418	6.35	284.2491
Sulfamethazine-13C6	C6H14N4O2S-13C6	284.1001	3.95	285.1112
Morphine-D3	C17H16D3NO3	288.1553	1.43	289.1626
Diazepam-D5	C16H8Cl1N2O1D5	289.103	10.09	290.1103
Metolachlor-D6	C15H16D6ClNO2	289.1716	11.74	290.1788
Oxazepam-D5	C15H6ClD5N2O2	291.0823	9.02	292.0896
Benzoylecgonin-D3	C16H16D3NO4	292.1502	4.69	293.1575
Thiamethoxame-D3	C8H7D3ClN5O3S	294.0381	4.01	295.0454
Climbazole-D4	C15H13D4ClN2O2	296.1224	8.27	297.1302
Diclofenac-D4	C14H7D4Cl2NO2	299.0412	12.77	300.04907
N4-Acetylsulfamethoxazole-D5	C12H9D4N3O4S	299.0878	5.52	300.0951
Aspartame-D5	C14H13D5N2O5	299.1524	2.91	300.1602
Trimethoprim-D9	C14H9D9N4O3	299.1949	3.74	300.2016
N4-Acetyl-Sulfathiazole-D4	C11H7D4N3O3S2	301.0498	4.16	302.0566
Codein-13C,D3	C17H18D3NO3[13]C	303.1743	2.91	304.1816
Cocaine-D3	C17H18D3NO4	306.1659	5.21	307.1732
Fluconazole-D4	C13H8F2N6O1D4	310.1292	5.36	311.1364

Internal Standard Name	Formula	Monoisotopic mass	Retention time	$[M+H]^+$
Clopidogrel-(+/-)-D4	C15H10D4ClNO2S	311.0685	5.68	312.0758
Tebuconazole-D6	C16H16D6ClN3O	313.1824	12.98	314.1901
Sulfadimethoxine-D4	C12H10D4N4O4S	314.0981	5.77	315.106
Diazinon-D10	C12H11D10N2O3PS	314.1633	12.9	315.1711
Fluoxetine-D5	C17H13F3N1O1D5	314.1654	9.62	315.1727
Methadone-D9	C21H18D9NO	318.2658	8.68	319.273
Ranitidine-D6	C13H16N4O3SD6	320.1789	2.38	321.1862
Chlorpyrifos-methyl-D6	C7HD6Cl3NO3PS	326.9328	13.93	327.9401
Citalopram-D6	C20H15D6FN2O	330.2015	6.87	331.2087
Sulcotrione-D3	C14H10D3ClO5S	331.0366	6.92	332.0433
Clozapine-D8	C18H11D8ClN4	334.18	6.54	335.1873
Mesotrione-D3	C14H10D3NO7S	342.0607	6.21	343.0674
Propiconazole-D5	C15H12D5Cl2N3O2	346.1006	13.25	347.1084
Chlorpyrifos-D10	C9HD10Cl3NO3PS	358.9885	15.87	359.9963
Indomethacin-D4	C19H12Cl1N1O4D4	361.1019	12.8	362.1092
Bezafibrate-D4	C19H16D4ClNO4	365.13264	10.53	366.14047
Fenofibrate-D6	C20H15D6ClO4	366.1511	15.53	367.1578
Amisulpride-D5	C17H22D5N3O4S	374.2036	3.94	375.2109
Methylprednisolol-D4	C22H26O5D4	378.2344	4.3	379.2417
Neotame-D3	C20H27D3N2O5	381.2343	9.14	382.2416
Prochloraz-D7	C15H9D7Cl3N3O2	382.0742	12.27	383.082
Metsulfuron-methyl-D3	C14H12D3N5O6S	384.0926	7.18	385.1004
Cetirizine-D8	C21H17D8ClN2O3	396.2056	9.49	397.2129
Diflufenican-D3	C19H8D3F5N2O2	397.0924	14.5	398.1002
Eprosartan-D3	C23H21D3N2O4S	427.1645	6.33	428.1718
Irbesartan-D3	C25H25D3N6O	431.2513	9.34	432.2586
Valsartan-15N,13C5	[13]C5C19H29[15]N1N4O3	441.2403	11.06	442.2481
Candesartan-D5	C24H15D5N6O3	445.1905	9.41	446.1983
Verapamil-D6	C27H32N2O4D6	460.3208	7.54	461.3281
Nelfinavir-D3	C32H42D3N3O4S	570.3319	10.09	571.3392
Ritonavir-D6	C37H42N6O5S2D6	726.3504	13.51	727.3577
Erythromycin-13C2	C35H67NO13[13]C2	735.468	8.49	736.4752
Clarithromycin-D3	C38H66D3NO13	750.4952	10.06	751.503
Azithromycin-D3	C38H69D3N2O12	751.5268	6.22	752.53463

DATA-DEPENDENT MS/MS MEASUREMENTS

For data-dependent MS/MS measurements a scan resolution of 17,500 and scan range of *m/z* 200 to 2000 were used. A "top 5, pick others" experiment was executed, where first masses on the inclusion list were selected, then the other (*i.e.*, most intense) masses were selected and fragmented. Five cycles of MS/MS measurement were included for each cycle of full scan MS. A dynamic exclusion window of 8 seconds was set, meaning that once triggered, a mass would be placed on a temporary exclusion list for 8 seconds to ensure that one mass would not dominate all MS/MS scans. The automatic gain control (AGC) target was set at 200,000 and maximum injection time was 100 milliseconds. For fragmentation, normalized collision energies (NCE) were used for the nontarget peaks on the inclusion list. NCE was determined based on the mass of the nontarget compound. See Table S-2 below for the distribution.

Table S-2. Summary of the higher collision dissociation (HCD) energies used for nontarget masses on the inclusion list for MS/MS measurement.

m/z Range	HCD energy (NCE)
<124	90
124 – 148	85
148 – 172	80
172 – 197	75
197 – 224	70
224 – 251	65
251 – 277	60
277 – 304	55
304 – 329	50
329 – 355	45
355 – 380	40
380 – 407	35
407 – 434	30
434 – 460	25
460 – 478	20
>478	15

Table S-3. Settings used for peak picking in the enviPick software.

Parameter	Value
dmzgap	10
dmzdens	5
ppm	TRUE
drtgap	1000
drtsmall	20
drtdens	250
drtfill	10
drttotal	200
minpeak	4
recurs	5
weight	2
SB	3
SN	2
minint	1.00E+05
maxint	1.00E+08
ended	2
progbar	TRUE

Table S-4. Settings used for feature profiling in enviMass software. Two separate functions were used for the profile building, carried out sequentially.

Parameter	Value
agglomer function	
dmass	5
ppm	TRUE
dret	50
partcluster function	
dmass	4
ppm	TRUE
dret	50
from	FALSE
to	FALSE
progbar	FALSE
plotit	FALSE

Table S-5. Results from targeted screening of the validation pairs, including PC1 loading in principal component analysis (PCA), and additional statistical tools applied. Fold change is ratio of mean intensity of influent samples to mean intensity of effluent samples. Wilcox U-test and indicator species analysis were calculated using the intensities of the features in each sample. "Related to" column is the majority vote of the evidence from each of these tools and to which group (if any) each compound was associated with based on the evidence. The reason for non-association is provided in last column.

Compound	CAS number	PCA PC1 loading	PLS-DA PC1 loading	Fold change	Wilcox p-value	Influent indicator	Effluent indicator	Related to	Reasons
Atrazine	1912-24-9	-0.159	-0.162	16	9.77E-04	F	F	None	Indicator analysis
Carbamazepine	298-46-4	0.088	-0.156	1.5	9.77E-04	F	F	None	PCA, fold change, & indicator analysis
Diuron	330-54-1	-0.151	-0.158	4.5	9.77E-04	F	F	None	Fold change & indicator analysis
Ethofumesate	26225-79-6	-0.158	-0.160	4.8E7	3.52E-03	T	F	Influent	
Irgarol	28159-98-0	0.152	0.147	0.12	3.68E-03	F	T	Effluent	
Isoproturon	34123-59-6	-0.160	-0.161	39	9.77E-04	F	F	None	Indicator analysis
Metamitron	41394-05-2	-0.159	-0.161	2.0E8	3.52E-03	T	F	Influent	
Metribuzin	21087-64-9	-0.157	-0.158	85	9.77E-04	T	F	Influent	
Oseltamivir	196618-13-0	-0.143	-0.144	73	1.43E-02	T	F	Influent	
Ranitidine	66357-35-5	-0.125	-0.151	2.4	9.77E-04	F	F	None	Fold change & indicator analysis
Sulfadiazine	68-35-9	-0.158	-0.159	3.4E7	3.52E-03	T	F	Influent	
Sulfadimethoxine	122-11-2	-0.152	-0.153	1.5E8	3.52E-03	T	F	Influent	
Sulfamethazine	57-68-1	-0.158	-0.160	48	3.84E-03	T	F	Influent	
Sulfamethoxazole	723-46-6	-0.154	-0.156	4.7	9.77E-04	F	F	None	Fold change & indicator analysis
Sulfathiazole	72-14-0	-0.158	-0.159	3.4E7	3.52E-03	T	F	Influent	
Terbuthylazine	5915-41-3	-0.157	-0.160	24	9.77E-04	F	F	None	Indicator analysis
Venlafaxine	93413-69-5	0.158	-0.129	1.1	9.77E-04	F	F	None	PCA, fold change, & indicator analysis
Carbamazepine-10,11-dihydro-10,11-dihydroxy	58955-93-4	0.156	0.135	0.74	9.77E-04	F	F	None	Fold change & indicator analysis
Carbamazepine-10,11-epoxide	36507-30-9	0.160	0.160	0.23	9.77E-04	F	F	None	Fold change & indicator analysis
Desethylatrazine	6190-65-4	0.160	0.160	0.25	9.77E-04	F	F	None	Fold change & indicator analysis
Desisoproylatrazine	1007-28-9	0.159	0.159	0.13	9.77E-04	F	F	None	Fold change & indicator analysis

Compound	CAS number	PCA PC1 loading	PLS-DA PC1 loading	Fold change	Wilcox p-value	Influent indicator	Effluent indicator	Related to	Reasons
Diuron-desdimethyl	2327-02-8	0.160	0.160	0.00	3.27E-03	F	T	Effluent	
Diuron-desmonomethyl	3567-62-2	0.157	0.157	0.00	3.27E-03	F	T	Effluent	
Ethofumesate-2-keto	26244-33-7	0.109	0.109	0.00	2.95E-03	F	T	Effluent	
Irgarol-descyclopropyl	30125-65-6	0.160	0.160	0.01	3.52E-03	F	T	Effluent	
Isoproturon-didemethyl	56046-17-4	0.160	0.161	0.00	3.52E-03	F	T	Effluent	
Isoproturon-monomethyl	34123-57-4	0.160	0.161	0.00	3.27E-03	F	T	Effluent	
Metamitron-deamino	36993-94-9	0.161	0.160	0.16	9.77E-04	F	F	None	Fold change & indicator analysis
Metribuzin-desamino	35045-02-4	0.161	0.161	0.00	3.27E-03	F	T	Effluent	
N,N-Didesmethylvenlafaxin	93413-77-5	0.160	0.159	0.40	9.77E-04	F	F	None	Fold change & indicator analysis
O,N-Didesmethylvenlafaxine	135308-74-6	0.160	0.159	0.40	9.77E-04	F	F	None	Fold change & indicator analysis
N4-Acetyl-Sulfadiazine	127-74-2	0.160	0.161	0.00	3.27E-03	F	T	Effluent	
N4-Acetyl-Sulfadimethoxine	24341-30-8	0.160	0.161	0.02	3.52E-03	F	T	Effluent	
N4-Acetyl-Sulfamethazine	100-90-3	0.160	0.161	0.00	3.27E-03	F	T	Effluent	
N4-Acetyl-Sulfamethoxazole	21312-10-7	-0.150	-0.153	4.3	9.77E-04	F	F	None	Fold change & indicator analysis
N4-Acetyl-Sulfathiazole	127-76-4	0.158	0.158	0.00	3.27E-03	F	T	Effluent	
N-Desmethylvenlafaxine	149289-30-5	0.155	0.139	0.76	9.77E-04	F	F	None	Fold change & indicator analysis
O-Desmethylvenlafaxine	93413-62-8	0.155	0.139	0.76	9.77E-04	F	F	None	Fold change & indicator analysis
Oseltamivir-carboxylate	187227-45-8	0.160	0.160	0.15	9.77E-04	F	F	None	Fold change & indicator analysis
Ranitidine-N-oxide	73857-20-2	0.160	0.160	0.03	3.79E-03	F	T	Effluent	
Ranitidine-S-oxide	73851-70-4	0.160	0.160	0.03	3.79E-03	F	T	Effluent	
Terbuthylazine-2-hydroxy	66753-07-9	0.145	0.116	0.59	9.77E-04	F	F	None	Fold change & indicator analysis

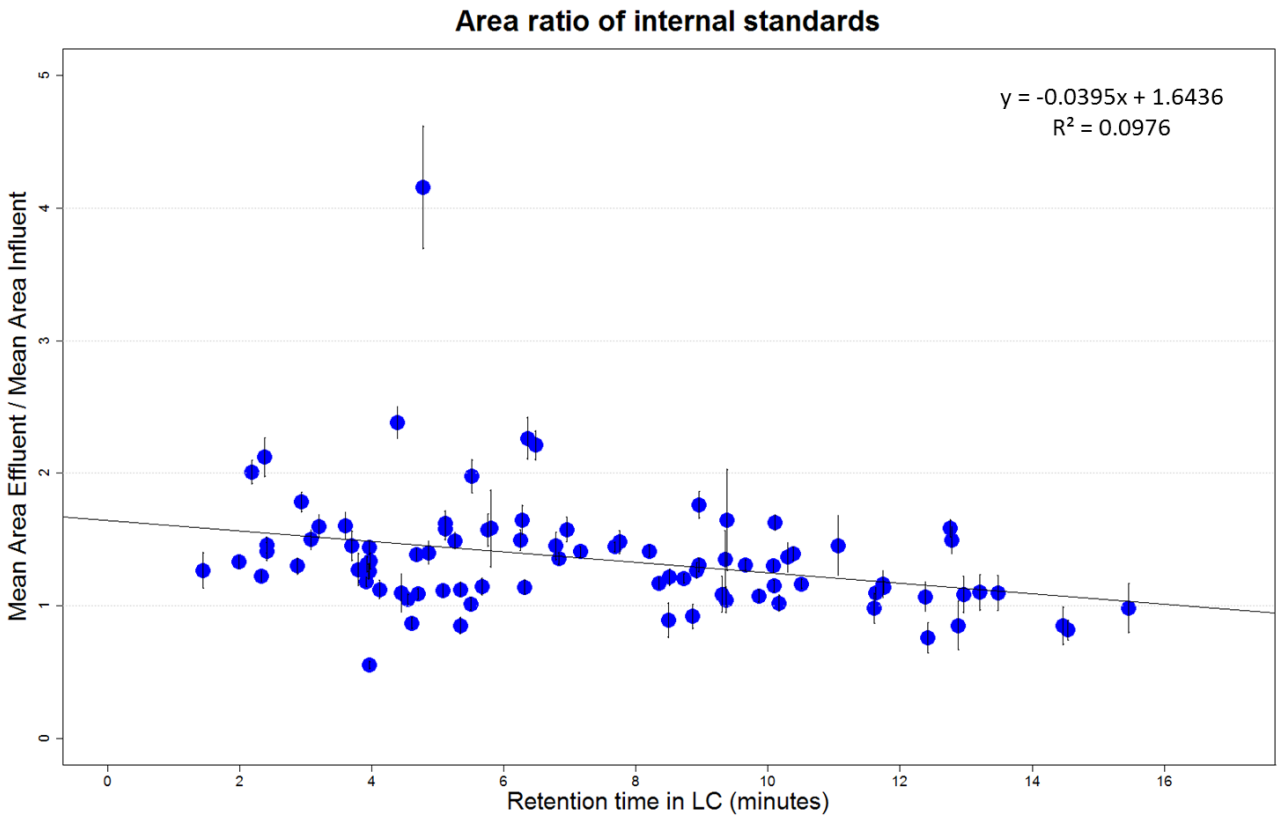

Figure S-1. Ratio of areas of each internal standard for matrix factor correction, measured in positive ionization. While a slight downward trend could be observed, the correlation was not strong enough to justify a blanket matrix factor correction of all nontarget peaks. The outlier at 4.78 minutes is the compound Ritaline.

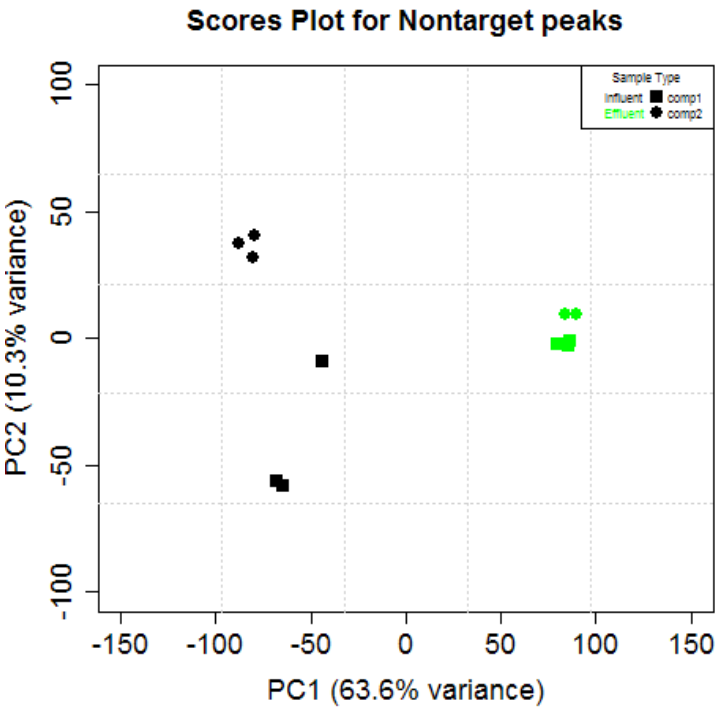

Figure S-2a. PC1 vs. PC2 scores plot for all peaks in validation samples, measured in positive ionization. Here the first two PC are displayed, with PC1 on the x-axis and PC2 along the y-axis. Both of the major variances in the data are capture here; for example influent and effluent sample types are well explained by PC1, while composites are differentiated in PC2.

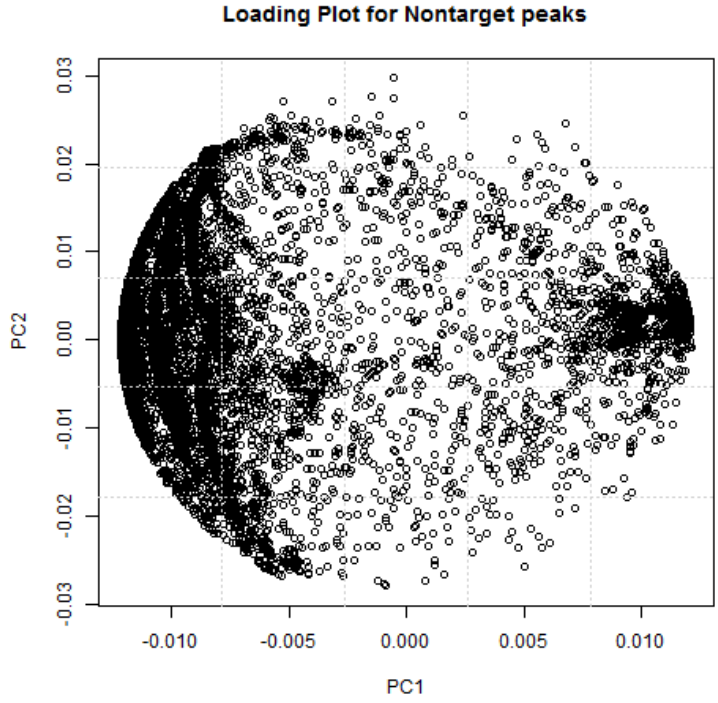

Figure S-2b. PC1 vs. PC2 loading plot for all peaks in validation samples, measured in positive ionization mode. Here the first two PC are displayed, with PC1 on the x-axis and PC2 along the y-axis. Each circle represents a unique *m/z* and RT feature detected during the peak picking and feature building. Classification of the peaks was done by comparing the feature loading to the sample scores plot.

Table S-6. Summary of the validation pairs used for the PCA and difference analysis in the nontarget workflow. Parent and the corresponding transformation product are shown; some parent compounds are related to multiple transformation products. The biotransformation that occurs is also indicated, as well as the compound class of the parent compound. The last column summarizes the result of the workflow for each pair, with a "yes" if the two compounds were linked together or a "no" and the reason when the match was not found.

	Parent compound	Compound class	Transformation	Transformation Product	Found when workflow applied to spiked samples?
1	Atrazine	Herbicide	deethylation	Desethylatrazine	Yes
2			desisopropyl	Desisopropylatrazine	Yes
3	Diuron	Herbicide	ethene loss	Diuron-desdimethyl	Yes
4			demethylation	Diuron-desmonomethyl (DCPMU)	Yes
5	Ethofumesate	Herbicide	deethylation	Ethofumesate-2-keto	Yes
6	Isoproturon	algicide, herbicide	ethene loss	Isoproturon-didemethyl	Yes
7			demethylation	Isoproturon-monodemethyl	Yes
8	Metamitron	Herbicide	amino loss	Metamitron-Desamino	Yes
9	Metribuzin	Herbicide	amino loss	Metribuzin-Desamino	Yes
10	Terbutylazine	Herbicide	hydrolysis	Terbutylazine-2-hydroxy	No, TP in influent
11	Oseltamivir	pharmaceutical	ethene loss	Oseltamivir-carboxylate	Yes
12	Ranitidine	pharmaceutical	oxidation	Ranitidine-S-oxide	No, parent is not well degraded so not associated with either influent or effluent
13			oxidation	Ranitidine-N-oxide	No, parent is not well degraded so not associated with either influent or effluent
14	Sulfadiazine	pharmaceutical	acetylation	N4-Acetyl-Sulfadiazine	Yes
15	Sulfadimethoxine	pharmaceutical	acetylation	N4-Acetyl-Sulfadimethoxine	Yes
16	Sulfamethazine	pharmaceutical	acetylation	N4-Acetyl-Sulfamethazine	Yes
17	Sulfathiazole	pharmaceutical	acetylation	N4-Acetyl-Sulfathiazole	No, poor peak shapes so not selected by the peak picker.

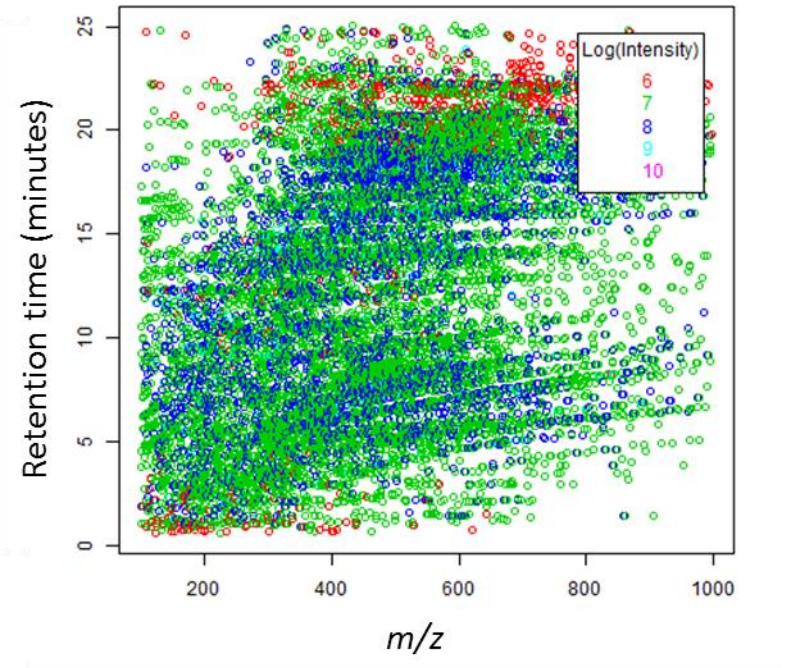

Figure S-3a. Distribution of detected features (n=14268) in the influent samples in positive electrospray. Each circle represents the mean *m/z* and retention time in influent samples of this feature. Color of the circle indicates the mean log10 intensity of the feature in all influent samples.

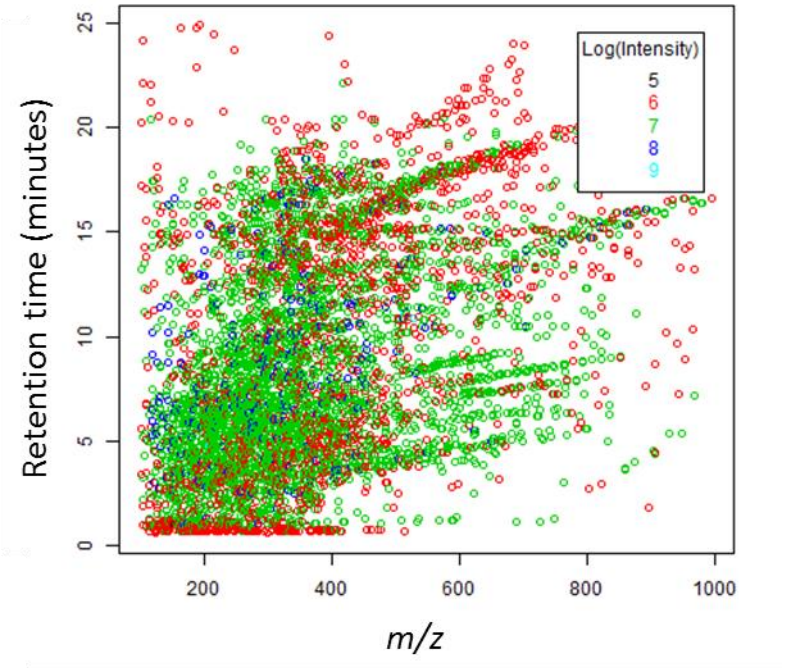

Figure S-3b. Distribution of detected features (n=5964) in the effluent samples in positive electrospray. Each circle represents the mean *m/z* and retention time in effluent samples of this feature. Color of the circle indicates the mean log10 intensity of the feature in all effluent samples. In comparison to Figure S-3a, a large decrease of features can be observed, as well as a general trend towards lower retention times, lower masses, and lower intensities.

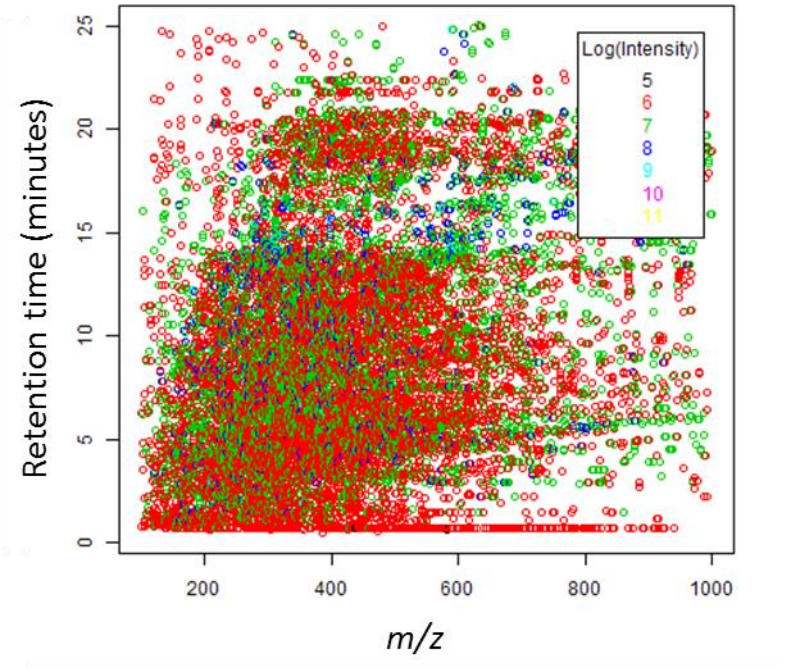

Figure S-4a.Distribution of detected features (n=16483) in the influent samples in negative electrospray. Each circle represents the mean *m/z* and retention time in influent samples of this feature. Color of the circle indicates the mean log10 intensity of the feature in all influent samples.

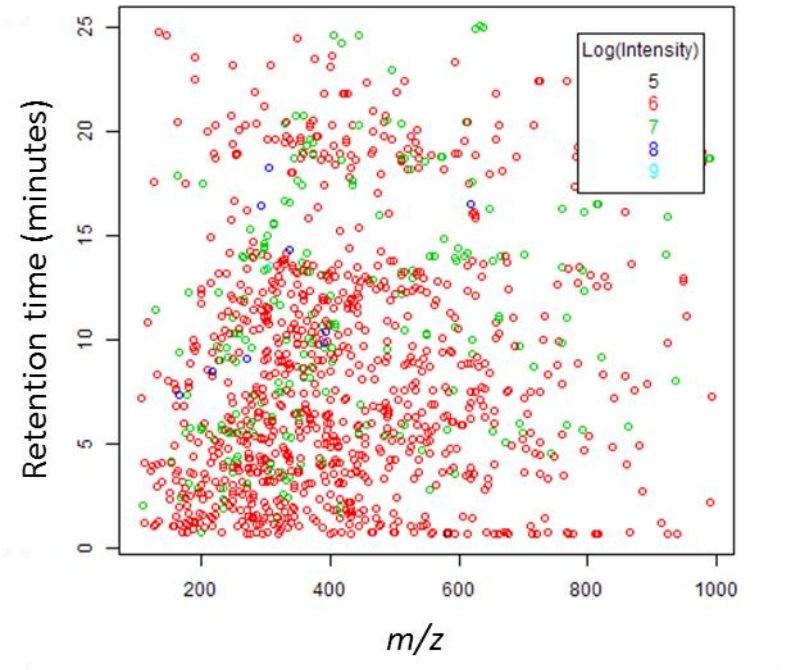

Figure S-4b. Distribution of detected features (n=1172) in the effluent samples in negative electrospray. Each circle represents the mean *m/z* and retention time in effluent samples of this feature. Color of the circle indicates the mean log10 intensity of the feature in all effluent samples. In comparison to Figure S-3a, a large decrease of features can be observed, as well as a general trend towards lower retention times, lower masses, and lower intensities.

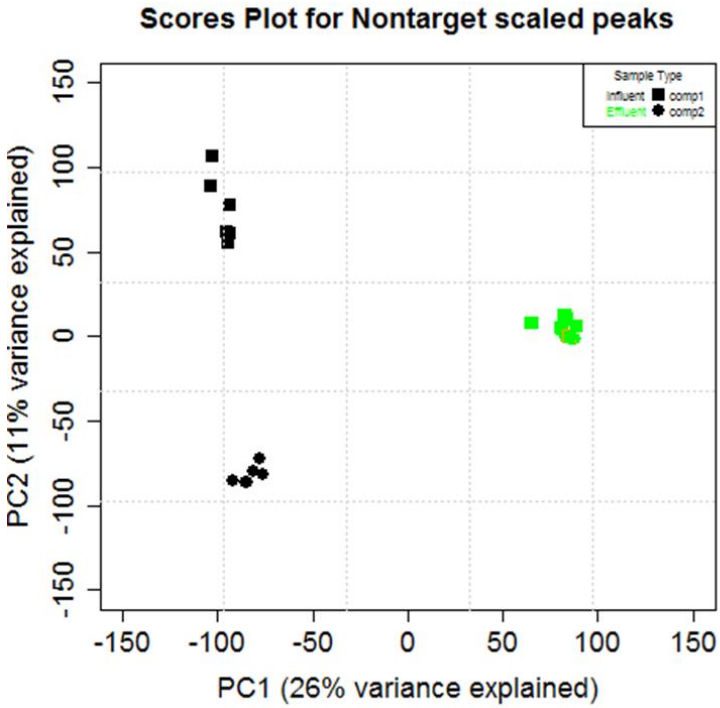

Figure S-5a. PC1 vs. PC2 scores plot for all peaks in the unspiked samples, measured in positive ionization mode. The first two PC are displayed, with PC1 along the x-axis and PC2 on the y-axis. Both of the major variances in the data are captured here; for example influent and effluent sample types are well separated by PC1, while composites are differentiated in PC2. It appears that variance in the influent samples is greater than that in the effluent samples, which are here not separated at all in the 2nd dimension. A quality control effluent sample without spiked internal standards is displayed in orange and was correctly grouped with the other effluent samples.

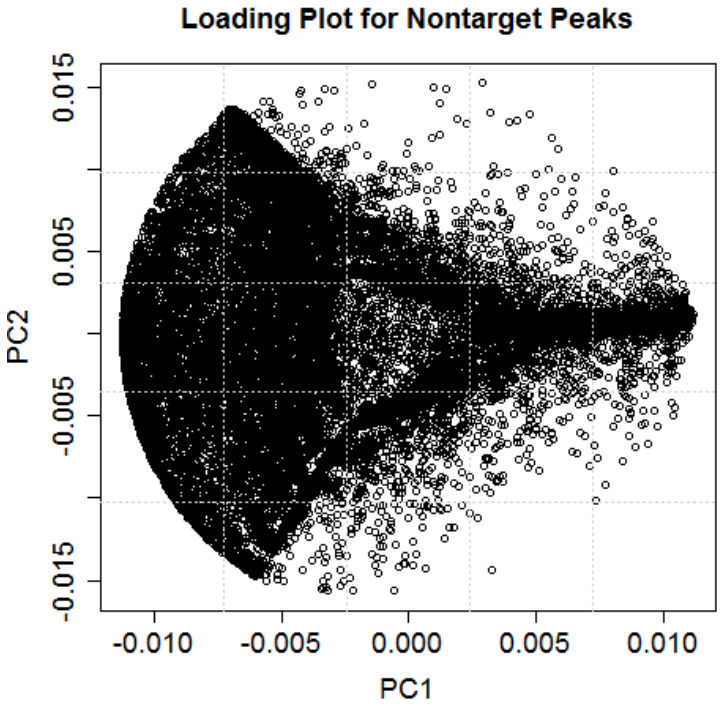

Figure S-5b. PC1 vs. PC2 loading plot for all peaks in the unspiked samples, measured in positive ionization mode. The first two PC are displayed, with PC1 along the x-axis and PC2 on the y-axis. Each circle represents a unique *m/z* and RT feature detected during the peak picking and feature building. Classification of the peaks was done by comparing the feature loading to the sample scores plot in Figure S-5a.

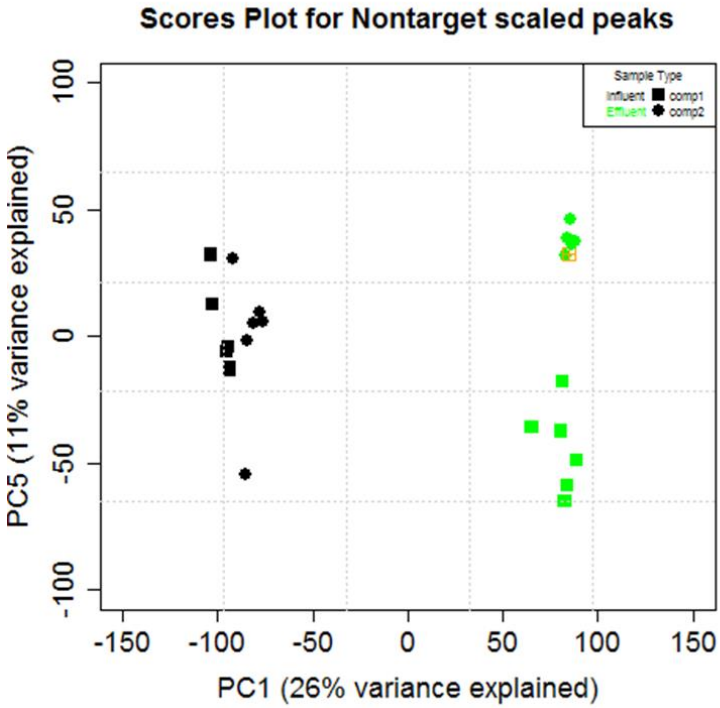

Figure S-6. PC1 vs. PC5 scores plot for all peaks in the unspiked samples, measured in positive ionization mode. Here the first and fifth PCs are displayed, with PC1 along the x-axis and PC5 on the y-axis. The first PC shows the separation of the influent and effluent sample types, while the fifth PC appears to capture the variance resulting from the different effluent composites. But compared to variance contributed to the influent composites (11%, Figure S-5a), this effect appears to be quite minor. A quality control effluent sample without spiked internal standards is displayed in orange and was correctly grouped with the other effluent samples.

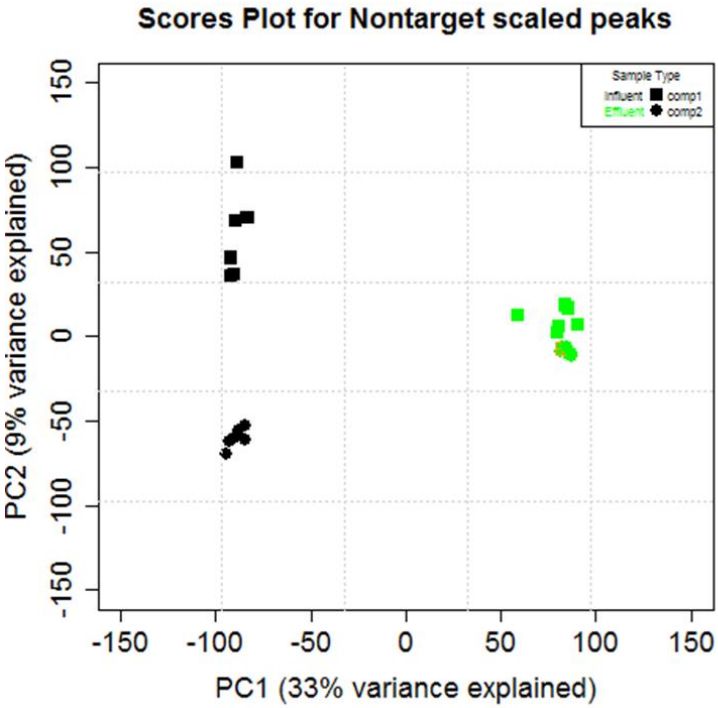

Figure S-7a. PC1 vs. PC2 scores plot for all peaks in the unspiked samples, measured in negative ionization mode. The first two PC are displayed, with PC1 along the x-axis and PC2 on the y-axis. Both of the major variances in the data are captured here; for example influent and effluent sample types are well separated by PC1, while composites are differentiated in PC2. It appears that variance in the influent samples is greater than that in the effluent samples, which are here not separated at all in the 2nd dimension. A quality control effluent sample without spiked internal standards is displayed in orange and was correctly grouped with the other effluent samples. The patterns observed here were very similar to those found in analysis of the positive ionization data (Figure S-5a).

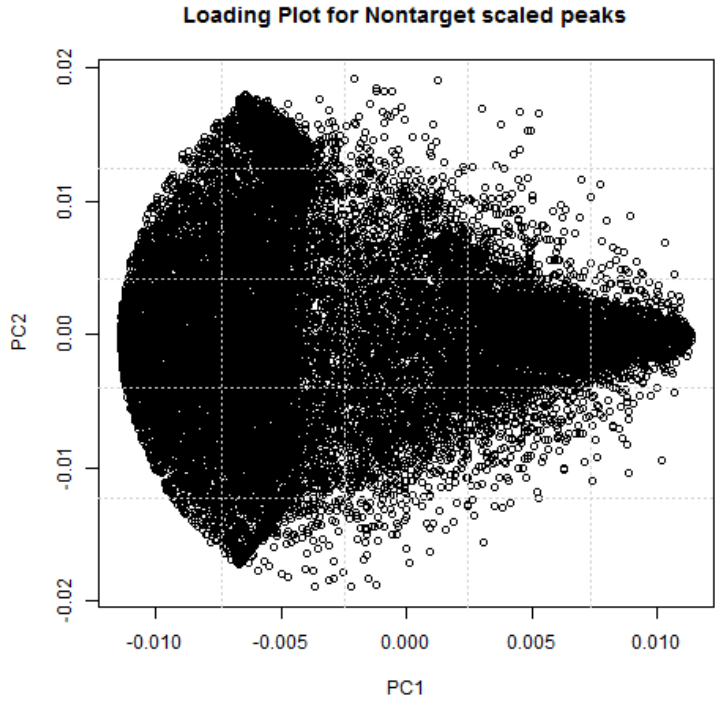

Figure S-7b. PC1 vs. PC2 loading plot for all peaks in the unspiked samples, measured in negative ionization mode. The first two PC are displayed, with PC1 along the x-axis and PC2 on the y-axis. Each circle represents a unique *m/z* and RT feature detected during the peak picking and feature building. Classification of the peaks was done by comparing the feature loading to the sample scores plot in Figure S-a.

SUMMARY OF NEGATIVE MODE DATA

Principal component analysis of the negative mode data showed very similar trends as were observed in the positive mode. Influent and effluent samples could be very well separated in the first PC. From this information, features could be attributed to either influent or effluent samples. Using the same PC1 cutoffs as in the positive mode (*i.e.*, +/- 0.005) 10,881 features were classified with the influent, while 4,080 features were classified with the effluent. These figures are similar to what was observed in the positive mode, where 9,758 influent peaks and 3,011 effluent peaks were selected.

Table S-7. Biotransformations included in the difference analysis, with mass difference and atomic loss or gain. Inventory of links and the detected isotopes calculated using the R 'nontarget' package. In some cases Cl/Br and S isotopes may be overlapping and therefore reported twice. Similarly N isotopes may also overlap with ^{13}C isotopes. Data presented here is from nontarget analysis performed on negative ionization mode data.

Transformation		Atomic loss / gain	Mass difference (Da)	# of links before RT filter	# of links after RT filter	# of links with		
						Cl/Br isotopes	S isotopes	N isotopes
1	Hydroxylation	+ O	+15.9949	2242	1173	205	127	79
2	Demethylation	- CH_2	-14.0157	2143	1107	189	129	84
3	Deethylation	- C_2H_4	-28.0313	2105	1168	220	115	75
4	Dehydrogenation	+ H_2	-2.0157	2191	1118	192	111	64
5	Hydrogenation	- H_2	+2.0157	2012	776	169	105	73
6	Dehydration	- H_2O	-18.0106	2140	957	173	129	76
7	Chlorine Reduction	- Cl / + H	-33.9611	1114	313	64	35	18
8	Acetylation	+ C_2H_2O	+42.0106	2177	NA	386	236	151
9	Deacetylation	- C_2H_2O	-42.0106	1909	NA	365	193	130
10	Glucuronidation	+ $C_6H_8O_6$	+176.0320	1118	NA	210	118	62
11	Deglucuronidation	- $C_8H_8O_6$	-176.0320	851	NA	148	117	86
12	Sulfonation	+ SO_3	+79.9568	1537	NA	292	194	107
13	Desulfonation	- SO_3	-79.9568	1083	NA	198	137	112
	Total			22,622	15,287	2,811	1,746	1,117

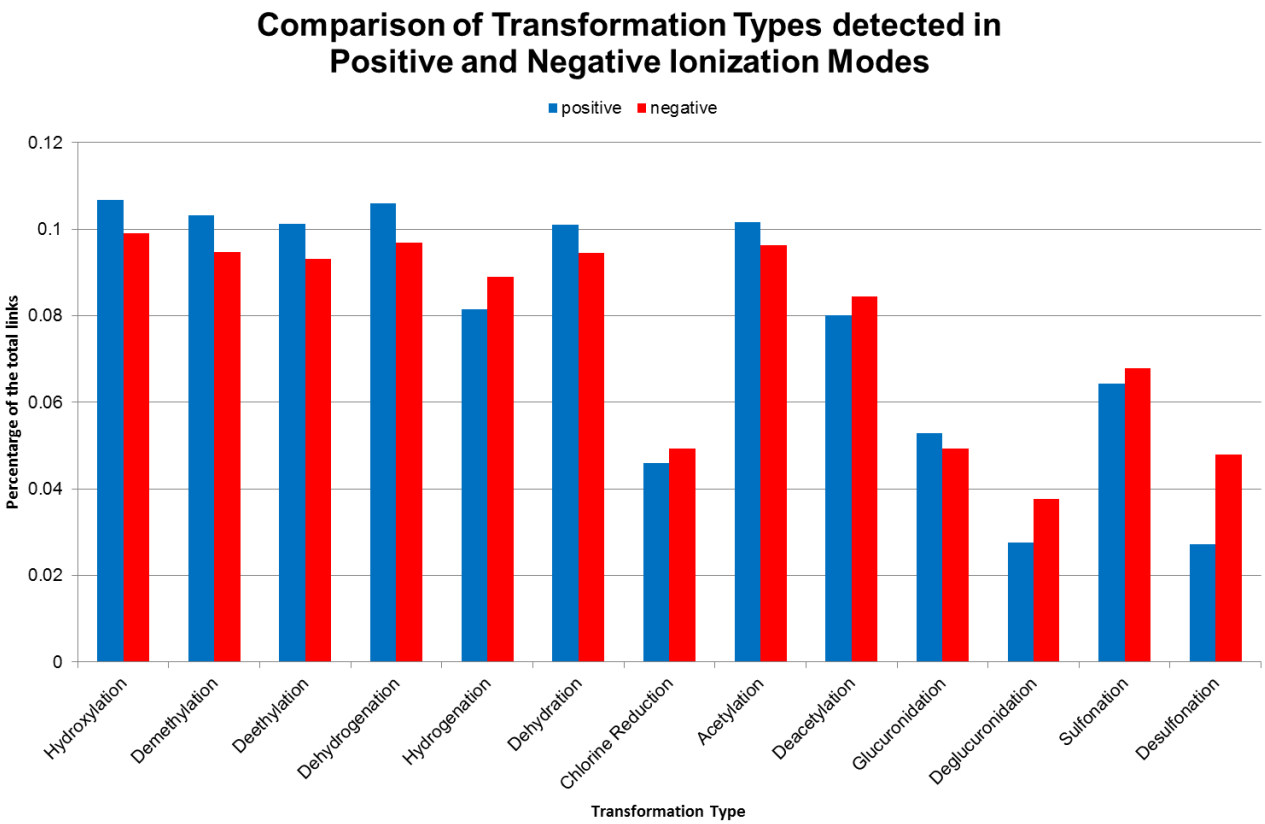

Figure S-8. Comparison of the different transformation types detected in the linkage analysis in both positive (in blue) and negative (in red) ionization modes, in the percentage of the total number of links detected. Overall the distribution is very similar between the two ionization modes.

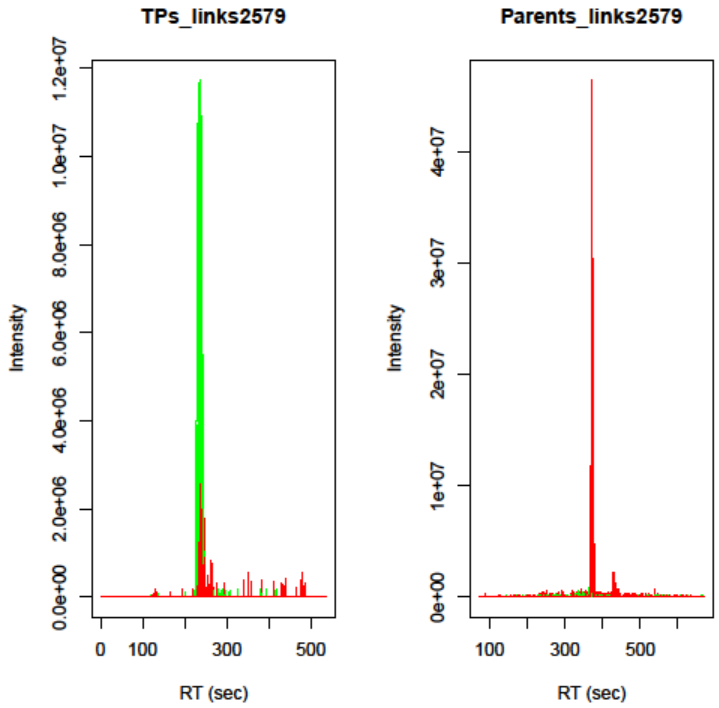

Figure S-9. Example of the trend visualization for selecting links for targeted MS/MS measurements. Extracted ion chromatographs (EICs) were plotted for both potential parent and potential TP masses of the top 300 most significant TP peaks. Here parent EICs are shown on the right, while TP EICs are shown on the left. The EICs are also colored according to the sample type, with influent samples in red and effluent samples in green. This example illustrates the ideal trend, with no parent peak detected in the effluent samples (*i.e.*, no green peak visible on the right) and only a very small TP peak detected in the influent samples (*i.e.*, small red peak on the left).

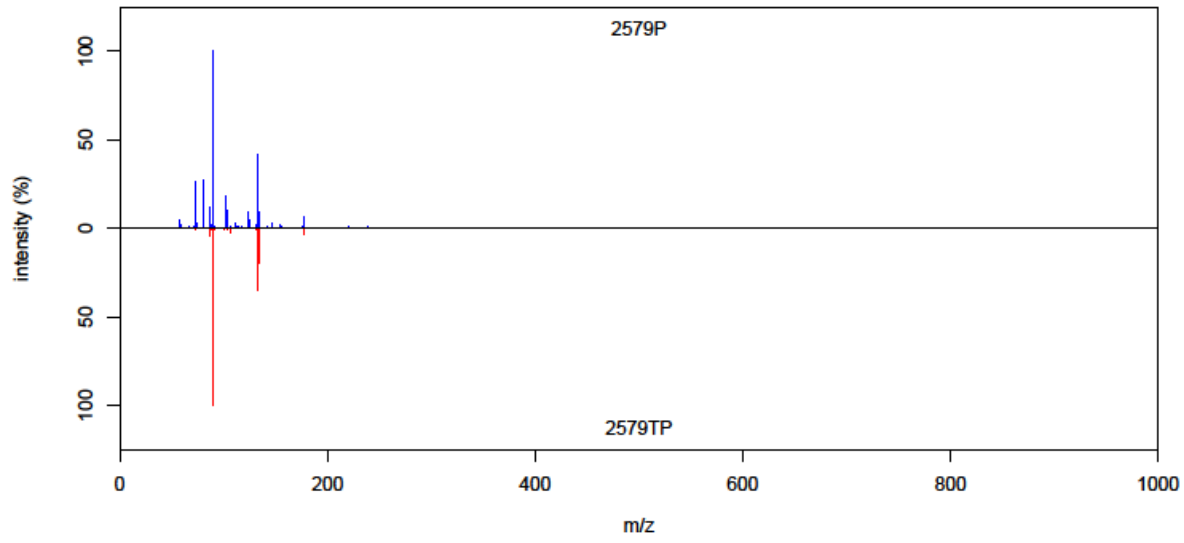

Figure S-10. Example of head to tail plot for visual comparison of the MS/MS similarity of each link. These plots are automatically generated as an output from the 'OrgMassSpecR' package. Parent compound MS/MS is plotted in blue in the upper half, while the MS/MS of the TP is shown in red in the lower half of the figure. The x-axis shows *m/z* values and the y-axis is relative intensities of the fragment peaks.

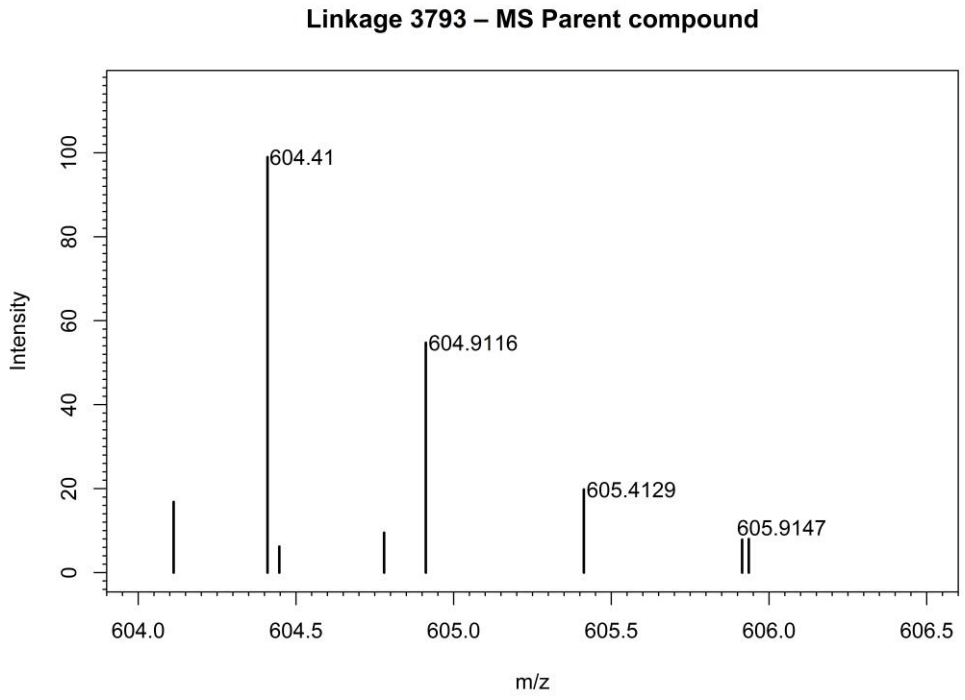

Figure S-11. Isotopic pattern of a parent compound identified to likely be a polypeptide. The x-axis is the *m/z* values and the y-axis is the relative intensity of the peaks. Peaks are annotated with the measured exact mass. Mass difference of approximately 0.5 Da between the isotope peaks indicates a charge of +2 on this compound.

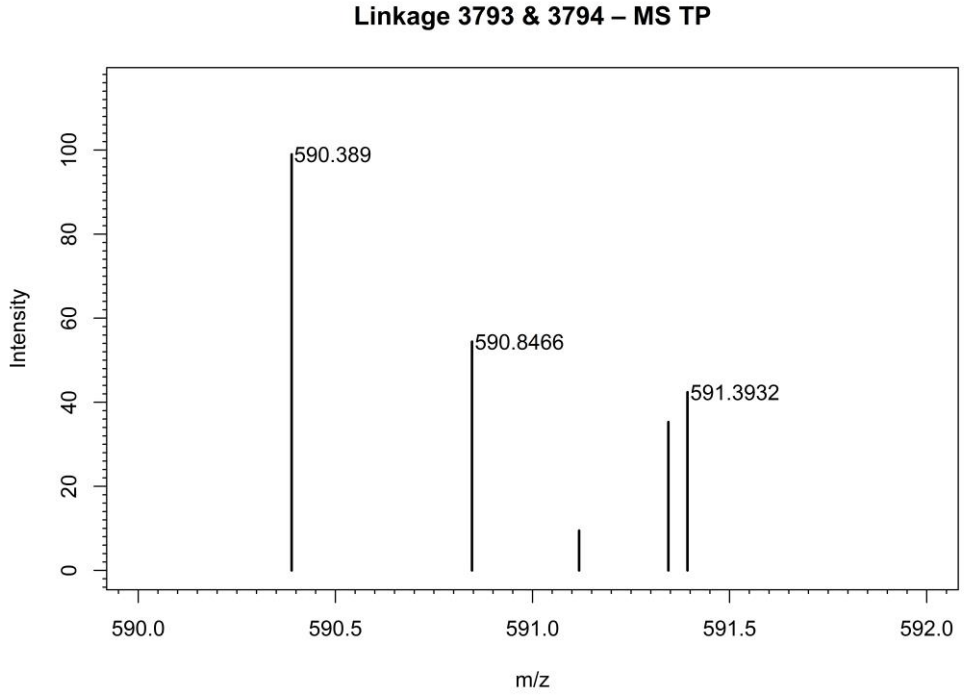

Figure S-12. Isotopic pattern of a transformation product identified as likely to be resulting from a polypeptide. The x-axis is the *m/z* values and the y-axis is the relative intensity of the peaks. Peaks are annotated with the measured exact mass. Mass difference of approximately 0.5 Da between the isotope peaks indicates a charge of +2 on this compound.

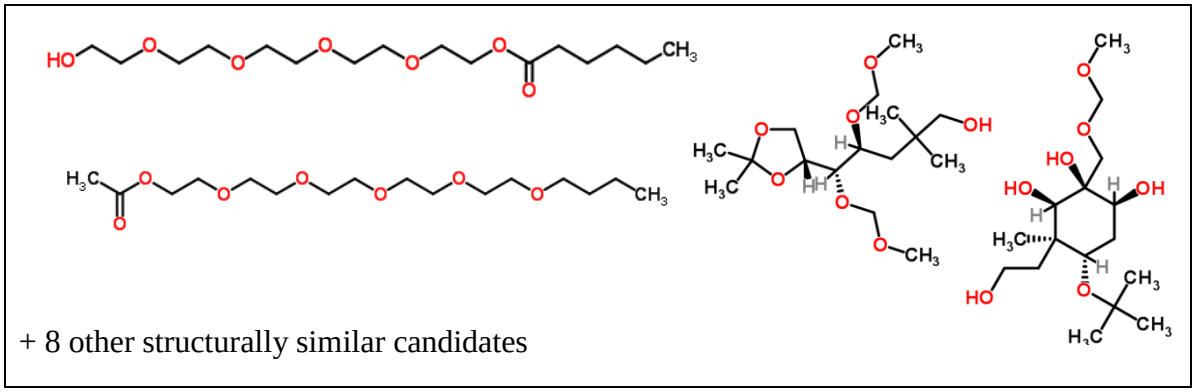

Figure S-13. Proposed structural candidates from MetFrag for possibly ethoxylated parent compound structure

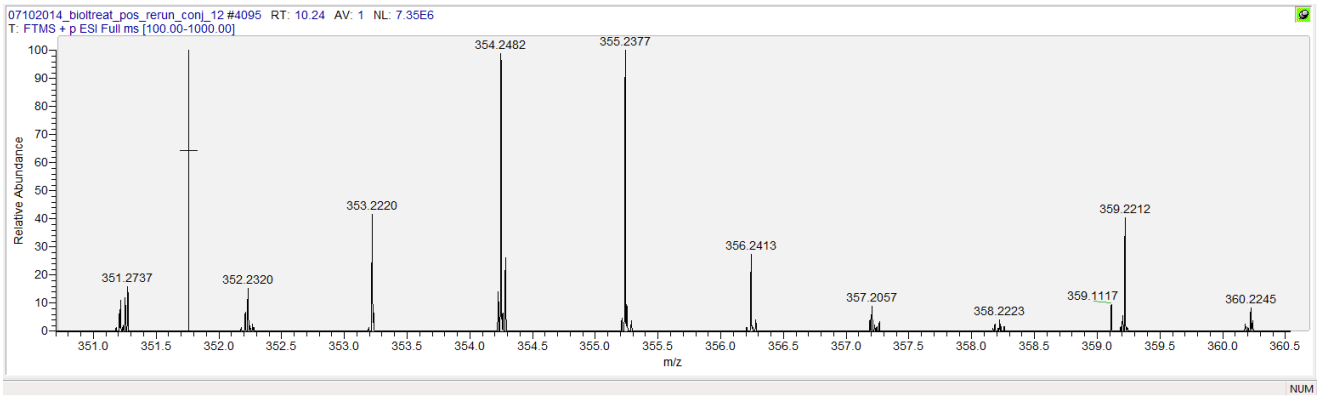

Figure S-14. Example of interferences in a nontarget peak. Exact mass of interest for the parent compound is 354.2486. A mass of equal/higher intensity is eluting simultaneously at 355.2377, which would obscure the $^{13}C/^{15}N$ isotope peaks. Additionally, the mass difference of these peaks is less than 1.0 Da, so both would be collected and fragmented for MS/MS, meaning that any fragments collected could not be attributed to a specific parent mass.

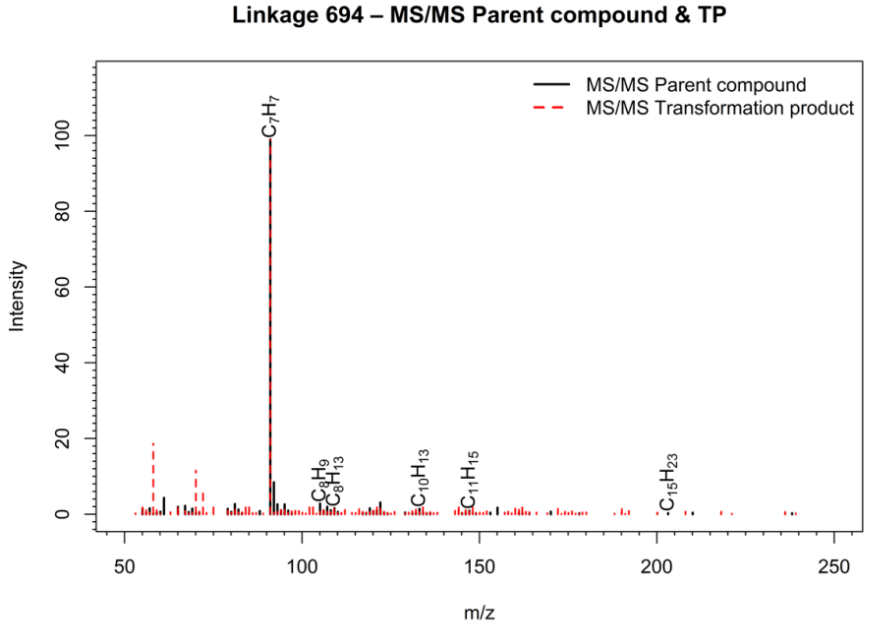

Figure S-15. MS/MS of the parent and transformation product of a hydrogenation link with highest MS/MS similarity score. The x-axis is the *m/z* values and the y-axis is the relative intensity of the fragments. Fragments from the parent compound are shown in black, while fragments in the transformation product are shown in red hatched. Fragments were annotated with formulas explained by GenForm.

Table S-8. Summary of the MS/MS similarity score calculations with the validation pairs. MS/MS spectra were collected for known parent/TP pairs in wastewater samples and in reference standards and then the similarity score was calculated with the 'OrgMassSpecR' package. Displayed here is the transformation product and the calculated score. "NoTP" indicates that no MS/MS spectra was collected for the transformation product in wastewater samples, while "NoPAR" means no spectra was collected for the parent compound in wastewater samples. "NA" indicates that no MS/MS spectra was collected for this reference standard.

	Transformation Product	Parent	Score in wastewater samples	Score in reference standards
1	Desethylatrazine	Atrazine	0	0.708081
2	1-Hydroxy-Benzotriazole	Benzotriazole	0	0
3	4-Hydroxy-Benzotriazole	Benzotriazole	0	0
4	Carbamazepine-10,11-epoxide	Carbamazepine	0	0.002348
5	Carbamazepine-10,11-dihydroxy-10,11-dihydro	Carbamazepine	0	0.002127
6	Diuron-desdimethyl	Diuron	NoTP	0.010848
7	Diuron-desmonomethyl	Diuron	NoTP	0.00395
8	Ethofumesate-2-keto	Ethofumesate	NoPAR	0.19484
9	Irgarol-descyclopropyl	Irgarol	NoPAR	0.024471
10	Isoproturon-didemethyl	Isoproturon	NoTP	0.007296
11	Isoproturon-monodemethyl	Isoproturon	0	0.005474
12	Metamitron-Desamino	Metamitron	NoTP	0.543897
13	Metribuzin-Desamino	Metribuzin	NoPAR	0.155758
14	Oseltamivir-carboxylate	Oseltamivir	NoPAR	0.094223
15	Ranitidine S-Oxide	Ranitidine	NoPAR	0.072668
16	Ranitidine N-Oxide	Ranitidine	NoPAR	0.986723
17	N4-Acetylsulfadiazine	Sulfadiazine	NoPAR	0.333396
18	N4-Acetyl-Sulfadimethoxine	Sulfadimethoxine	NoPAR	0
19	N4-Acetyl-Sulfamethazine	Sulfamethazine	NoPAR	0.13762
20	N4-Acetyl-Sulfamethoxazole	Sulfamethoxazole	0.000963	0.319804
21	N4-Acetyl-Sulfathiazole	Sulfathiazol	NoPAR	0.25032
22	Terbuthylazine-desethyl	Terbuthylazine	0	0.160193
23	N-Desmethylvenlafaxine	Venlafaxine	0.002377	0.234515
24	O-Desmethylvenlafaxine	Venlafaxine	0.002377	0.931976
25	O,N-Didesmethylvenlafaxine	Venlafaxine	0	0.017823
26	N,N-Didesmethylvenlafaxine	Venlafaxine	0	0.252916
27	5-Methyl-benzotriazole	Benzotriazole	NoTP	NA
28	Desisopropylatrazine	Atrazine	0.002574	0.312438