Supporting information for:

Characterization of advanced wastewater treatment with ozone and activated carbon using LC-HRMS based non-target screening with automated trend assignment

Jennifer E. Schollée[a], Juliane Hollender[a,b], Christa S. McArdell[a]

[a] Eawag: Swiss Federal Institute of Aquatic Science and Technology, 8600 Duebendorf, Switzerland
[b] ETH Zurich, Institute of Biopollutant Dynamics, 8092 Zurich, Switzerland

Table of Contents:

Section S1. Sampling Locations 6
Section S2. Analytical procedure 12
Section S3. Data processing methods 14
Method Evaluation 18
Step (3): Profile grouping function 20
Step (4): Replicate filter 21
Step (5): Intensity normalization 21
Step (6): profile QC filter 23
Step (7): Isotope, Adduct, Homologue Filter 23
Step (8): Dilution factor correction 25
Step (9): Missing value imputation 25
Automated trend analysis 26
Qualtitative Target and Suspect Screening 30
Linkage analysis 31
Section S4: Non-target characterization 33
General Characterization of Influent Samples 34
Identification Information for Potential Industrial Compounds 37
Section S5. Sensitivity analysis of automated trend assignment 67
Section S6. Results of Trend Analysis 69
Ozonation Transformation Products 71
Section S7: Linkage Analysis 74
Section S8: References 77

List of Figures:

Figure S 1. Schematic from WWTP Glarnerland (GL). Sampling locations in the influent (INF), after biological treatment (BIO), after ozonation (OZO), and after post-treatment with a GAC filter in the effluent (EFF) are indicated. Two advanced treatment setups were investigated, one (top) which treated wastewater with only GAC filtration and one (bottom) with pre-ozonation followed by GAC filtration. ... 6

Figure S 2. Schematic from WWTP Altenrhein (AR). Sampling locations in the influent (INF), after biological treatment (BIO), after ozonation (OZO), and after post-treatment with a GAC filter in the effluent (EFF) are indicated. ... 7

Figure S 3. Schematic from WWTP ProRheno (PR). Sampling locations in the influent (INF), after biological treatment (BIO), after ozonation (OZO), and after post-treatment with powder activated carbon (PAC) dosed on a sand filter in the effluent (EFF) are indicated. Two influent streams were sampled, PR-Comm, which is collected from mainly municipal sources, and PR-Chem, which is collected from the pharmaceutical industries in the area. These two streams were then mixed in the WWTP in a reaction of 90:10 PR-Comm:PR-Chem prior to treatment. ... 8

Figure S 4. Figure of non-target screening pre-processing workflow. New workflow steps are dashed. ... 18

Figure S 5. Example of broad, tailing peak, for which split profiles were detected ... 20

Figure S 6. Comparison of internal standard intensities (a; top) before intensity normalization and (b; bottom) after intensity normalization. Plotted on the x-axis is the sequence order and on the y-axis is the log10 of the intensity of the internal standards. The boxes indicate the interquantile range (IQR), with the lower edge of the box at the 25th percentile (*i.e.*, Q1) and the upper edge of the box at the 75th percentile (*i.e.*, Q3). The lower and upper whisker give Q1-1.5xIQR and Q3+1.5xIQR, respectively. Points outside of this range are plotted as open circles. Colors indicate sample matrix type. Red = EFF; blue = OZO; black = BIO; green = INF; cyan = calibration standards. First are samples from WWTP Altenrhein; then from WWTP Glarnerland; finally from WWTP ProRheno. ... 22

Figure S 7. *enviMass* output of the detected isotopes. On the x-axis is the isotope and charge z (directly under the bar), while on the y-axis is the absolute frequency, *i.e.*, the number of detections. Pattern groups are the detected isotopologues based on m/z difference and intensity pattern prior to profiling. ... 23

Figure S 8. *enviMass* output of the detected adducts. On the x-axis are the adduct types. In the top portion of the figure the absolute frequency of each adduct is shown; on the bottom portion box-whisker plots of the log10 of peak intensities of each adduct type. ... 24

Figure S 9. Venn diagram of the isotope, adduct, and homologue filtering. In black are the features that were annotated as an isotope (*i.e.*, not a monoisotopic peak), in red are the features not annotated as an [M+H]+ adduct (*i.e.*, either not annotated or annotated as another adduct), and in blue are the features that belonged to at least 1 homologue series. Only features in the intersecting portion of the diagram (10,231 features) were filtered out of the final dataset. ... 25

Figure S 10. Example of missing value imputation for 400 randomly selected non-target features. Colors indicate the matrix type of the sample. On the x-axis is the profileID of the non-target features; on the y-axis is the log10 of the intensity (measured and gap-filled). ... 26

Figure S 11. Visualization of cutoff domains for profile barcoding. On the x-axis are the four sampling location along the wastewater treatment train. On the y-axis is the profile intensity, normalized to the maximum profile intensity. Shown are the 20% and 60% cutoffs used for cutoff_01 scenario. INF = in the WWTP influent; BIO = after biological treatment; OZO = after ozonation; EFF = after post-treatment (*i.e.*, GAC filtration or PAC+SF), in the WWTP effluent. Note that the “20% cutoff” refers to a 80% removal of the feature. ... 27

Figure S 12. All 81 possible minor trends defined for the profile barcoding. Of these, 16 were not present in the profile barcoding analysis because none of the four point was equal to 1, which was a requirement due to the normalization procedure applied on each intensity profile. ... 28

Figure S 13. Defined major trends for pattern recognition. On the x-axis are the four sampling location at each WWTP; INF – influent, BIO – after biological treatment, OZO – after ozonation, EFF – after post-treatment with activated carbon, in the effluent. On the y-axis is the normalized profile intensity. Profiles

are normalized to the maximum profile intensity. Each of the major trend correspond to an expected trend in wastewater treatment. 29
Figure S 14. The number of features detected in each wastewater treatment plant (WWTP) and each matrix, in the individual samples as well as in the two sets of pooled samples. Unspiked pooled samples were a mix of samples from one matrix (*i.e.*, influent, biological, ozonation, or effluent) in one WWTP (*i.e.*, Altenrhein, Glarnerland, ProRheno). Spiked pooled samples were the same mix as in the unspiked samples but spiked with a set of target compounds (1000 ng/L in influent samples, 250 ng/L in all other matrix types). Shown at the base of each bar are the number of samples measured in the respective category. 33
Figure S 15. Principal component analysis (PCA) of all wastewater samples, including spiked and unspiked pooled samples. In (a) first principal component (PC1) vs. second principal component (PC2). PC1 explains 19.2% of variance and PC2 explained 7.7%. In (b) PC1–PC5 are compared in a pairs plot matrix. U: unspiked pooled samples, S: spiked pooled samples 34
Figure S 16. Principal component analysis (PCA) of only wastewater samples (*i.e.*, spiked and unspiked pooled samples removed). In (a) scores plot of first principal component (PC1) vs. second principal component (PC2). In (b) the loading plot of PC1 vs. PC2. PC1 (13.7% of variance) and PC2 (7.1% of variance) 34
Figure S 17. Characterization of influent samples. In (top), a 4-group Venn diagram to visualize the presence of non-target features in the influent samples of different inputs. Assigned class are indicated in a box in each Venn quadrant. In (bottom), the first principal component (PC1) vs. second principal component (PC2). Influents of the different wastewater treatment plants are color-coded and labeled with the sampling date. Black: AR – Altenrhein; red: GL – Glarnerland; green: PR-Chem – ProRheno industrial wastewater; blue: PR-Comm – ProRheno domestic wastewater. 35
Figure S 18. Characterization of influent non-target features as possibly originating from industrial sources for (a) Glarnerland, (b) Altenrhein, (c) ProRheno municipal wastewater, and (d) ProRheno domestic wastewater. On the x-axis is the intensity spread for each profile, defined as the ratio of the 95^{th} and 5^{th} percentiles, and on the y-axis is the maximum intensity for each profile. Non-target features were classified as industrial if the intensity spread is >1E4; cutoff is indicated with a vertical dotted grey line. 36
Figure S 19. Tornado plot, visualizing the difference in number of nontarget features in each major trend with different cutoff values in the automated trend assignment algorithm. On the x-axis is percent change relative to cutoff_01 (20%, 60%), while on the y-axis the 12 major trends are listed..... 67
Figure S 20. Barchart of non-target features in each major trend with different cutoff values in automated trend assignment algorithm. Standard deviation are calculated across all 24 sampling dates. On x-axis are the 12 major trends, on the y-axis is the percent of non-target features. In the inset is a close up of Trends 2-12. 68
Figure S 21. Distribution of non-target features in trends across 27 sampling dates. On the x-axis is the sample, indicated with the two letter abbreviation for the WWTP, applied ozone dose and GAC or PAC dose. On the y-axis is the number of non-target features and the assigned major trend is colored as indicated in the legend. Due to the influence of rain in the first sampling of $AR_{0.3O3\_GAC44000}$, which is also clearly visible due to the lower number of features detected, this sample was removed from the data set. 69
Figure S 22. Cumulative intensity of non-target features at each step of the WWTP (INF: influent; BIO: after biological treatment; OZO: after ozonation; EFF: after post-treatment, effluent). The average for each sample setting is shown; the sample settings are shown in the title of each graph. GL, granular activated carbon (GAC) filtration without pre-ozonation (top row). GL and AR, pre-ozonation followed by GAC filtration (middle row, ordered by increasing GAC bed volumes). PR, pre-ozonation followed by powdered activated carbon (PAC) dosed onto sand filter (bottom row). Colors indicate the fate of features in each treatment step as shown in the legend. Number of samples at each setting shown in the respective title. 70
Figure S 23. Comparison of *m/z* values for well-removed OTPs (in blue) and for stable OTPs (in orange) for each sampling date. In the upper right panel is a scatter plot of the median *m/z* of stable OTPs vs. the median *m/z* of well-removed OTPs for each sampling date, as indicated by the letter and the corresponding plot. Signficant differences are indicated by a star (t-test, $p<0.05$). 72

Figure S 24. Comparison of retention times (RT) of well-removed OTPs (in blue) and for stable OTPs (in orange) for each sampling date. In the upper right panel is a scatter plot of the median retention time of stable OTPs vs. the median retention time of well-removed OTPs for each sampling date, as indicated by the letter and the corresponding plot. Signficant differences are indicated by a star (t-test, p<0.05). 73
Figure S 25. Distribution of detected reaction types with the linkage analysis. Reaction types are shown along the perimeter of the circle. Detected reactions for each sampling date are shown separately and indicated by color, as shown in the legend..... 75
Figure S 26. Removal of potential OTPs of various reactions types in the different post-treatments. In (a, on the left/top), OTP removal in the post-treatment with the freshest GAC ($GL_{0.2O3\_GAC8000}$) is always plotted on the x-axis. On the y-axis is the OTP removal during other GAC post-treatments ($GL_{0.2O3\_GAC18000}$, $GL_{0.2O3\_GAC33000}$, $AR_{0.1O3\_GAC44000}$, $AR_{0.3O3\_GAC48000}$). In (b, on the left/bottom), OTP removal in the PR-MK1 is plotted ($PR_{0.2O3\_PAC7.7}$), versus OTP removal in the three other PAC doses on the y-axis ($PR_{0.2O3\_PAC13.3}$, $PR_{0.2O3\_PAC13.4}$, $PR_{0.3O3\_noPAC}$). In c, a heatmap of the linkage reaction types based on log2 of the ratio between well-removed and stable OTPs at each sampling date. Five clusters were defined and are indicated by color along the left side of the plot and by number and color and number on the right side of the plot. Fromation of clusters appears to be driven by the removal in fresh GAC, as can be seen in the correlation the removal in the fresh GAC (x-axis in (a)) with the defined clusters. 76

List of Tables:

Table S 1. Summary of Samples, including waterwater treatment plant (WWTP), location in the treatment train, sampling date, sample name, and corresponding pooled sample..... 9
Table S 2. List of 147 isotopically labeled internal standards spiked and used for quality control of non-target screening workflow (available in separate Excel). 12
Table S 3. Summary of R packages used, including version and source. 15
Table S 4. Summary of the total number of features and the number of internal standard features detected after each step of the non-target screening workflow. The total number of spiked internal standards was 147. 18
Table S 5. List of minor trends and the associated major trend (available in separate Excel file)..... 29
Table S 6. Validation of automated trend assignment with quantified target compounds. (available in separate Excel)..... 30
Table S 7. List of 66 wastewater relevant organic micropollutants detected with qualitative target screening (MP66 List) (available in separate Excel) 30
Table S 8. Suspect list of 999 known ozonation transformation products (OTPs) compiled from literature sources (available in separate Excel)..... 30
Table S 9. Summary of transformation reactions considered during linkage analysis. Listed are the reaction type, reaction abbreviation, mass difference between parent and transformation product, formula change from parent to transformation product, expected oxidant during ozonation and example reactions from literature. Adapted from Schollée *et al*. 2018. 31
Table S 10. Results of identification of likely nontarget features originating from industrial sources, including measured accurate mass (m/z), retention time, intensity spread, msPurity, compound name and compound class (available in separate Excel). 36
Table S 11. Removal of potential industrial non-target features detected in the two influent stream in PR. Potential industrial features were selected based on intensity spread over the sampling dates in each influent stream. The detected trend for each feature on the day of maximum intensity (*i.e.*, assumed to be the day of maximum discharge) is shown. In the first column is the absolute number of non-target features with this trend, while in the second column is the percentage. 36
Table S 12. Summary of cutoffs considered in sensitivity analysis of pattern recognition algorithm.... 67
Table S 13. Distribution of non-target features in major trends across 27 sampling dates (available in separate Excel)..... 69

Table S 14. Summary of non-target features in each step of the WWTP. Reported are number of features, cumulative feature intensity, mean m/z of non-target features and mean retention time (rt) in minutes of non-target features (available in separate Excel). 69
Table S 15. Binning of major trends in each treatment step. The columns are the four WWTP sampling points (INF – in the influent; BIO – after biological treatment; OZO – after ozonation; EFF – after post-treatment, in the effluent). Rows are the seven defined classes used for visualization (see Figure S22). 70
Table S 16. Number of suspect and non-target OTPs detected on each of the 20 sampling dates were ozonation was applied. Sampling campaign is listed, along with the applied ozone does, the granular activated carbon (GAC) bed volumes (BVs), and the powdered activated carbon (PAC) dose in mg/L. The absolute number of detections and the percent of features that were removed >80% (Trend 5) in the respective post-treatment are given. 71

Section S1. Sampling Locations

WWTP Glarnerland

WWTP Glarnerland (*GL*) treats an annual volume of 7.1 million m^3, with 70,000 PE and 45,000 connected inhabitants. It has a calculated industrial contribution of approximately 40%; the various connected industries include manufacturers of pharmaceuticals, paper goods, cement products, and a textile refinery. The full-scale plant includes first mechanical treatment, then followed by a biological treatment using the S::Select® technology with hydrocyclones used to separate the excess sludge.

The advanced treatment pilot plant received a portion of flow after biological treatment (Figure S 1). This stream was split, with part of the wastewater being treated with pre-ozonation and granular activated carbon (GAC) filtration, while the other part of the wastewater being treated with GAC filtration directly. Both GAC filters used Pool W 1-3 (Carbotech), particle size sieved to 0.85-2.0 mm (10x20 mesh). Inflow over the filters was dosed flow proportionally during dry weather and reached an empty bed contact time (EBCT) of approximately 24 minutes. More information about the WWTP and the pilot plant can be found in Oltramare et al. in prep and McArdell et al. 2020.

Three sample campaigns were carried out at the plant, on June 8, 2017, January 16, 2018, and September 5, 2018. Additional sampling campaigns were conducted and measured at GL but ultimately not used in this analysis due to the lack of influent samples (Table S 1). GAC BVs were 6,700–33,000 and ozone doses were 0.18–0.22 gO3/gDOC. Samples were collected as 24-h composites in borosilicate glass bottles at the influent (INF) of the WWTP, after biological treatment, *i.e.*, at the inflow of the pilot plant (BIO), after ozonation (OZO), and after GAC filtration, *i.e.*, at the effluent of the pilot plant (EFF). Influent samples were collected one day prior, to account for the hydraulic retention time (HRT) in the WWTP. For samples without pre-ozonation, in the automated trend assignment, the effluent sample was used for both the OZO and EFF time point.

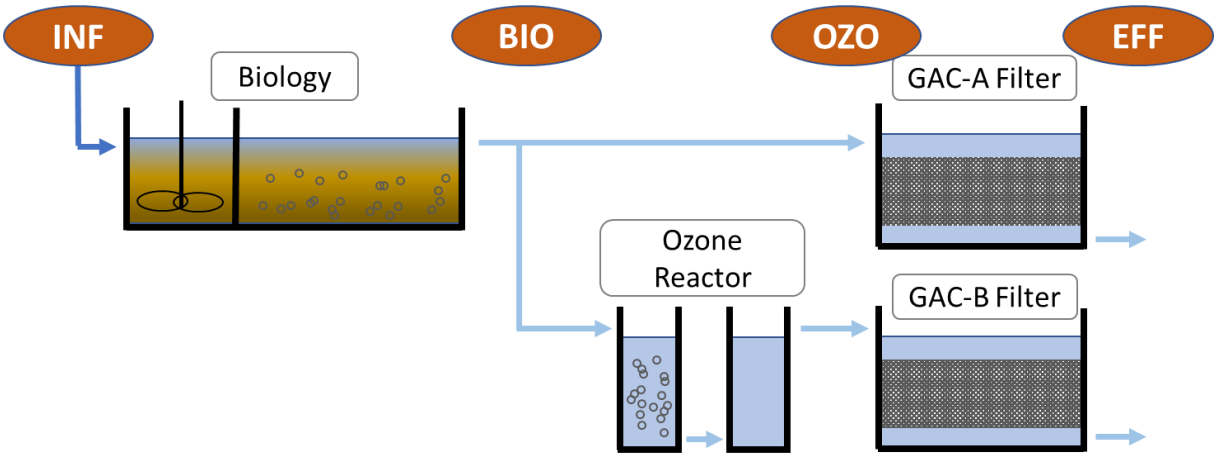


Figure S 1. Schematic from WWTP Glarnerland (GL). Sampling locations in the influent (INF), after biological treatment (BIO), after ozonation (OZO), and after post-treatment with a GAC filter in the effluent (EFF) are indicated. Two advanced treatment setups were investigated, one (top) which treated wastewater with only GAC filtration and one (bottom) with pre-ozonation followed by GAC filtration.

WWTP Altenrhein

WWTP Altenrhein (*AR*) currently treats the wastewater of 82,000 population equivalents (PEs). In 2017 the average inflow volume was approximately 9 million m^3, and the calculated industrial contribution was 26%. The conventional treatment includes mechanical treatment, followed by biological treatment, which consists of activated sludge and fixed bed processes operating in parallel), and lastly a sand filter.

A portion of the conventionally treated wastewater was redirected to a pilot plant onsite, equipped with pre-ozonation and GAC filtration (Figure S 2). The GAC filter contained the Cyclecarb 401 (Virgin) from Chemviron Carbon and and was run time-proportional with an EBCT of 20 minutes. Cyclecarb

specification were a particle size of 0.425-2.36 mm (8x40 mesh), with a maximum 10% (weight) > 2.36mm (=8 US mesh) and a minimum 5% (w) < 0.425 mm (=40 US mesh).

Two sampling campaigns were carried out at this location and for each, samples were collected during the week for 3 consecutive days (Table S 1). During the first campaign, July 16-18, 2018, a pre-ozonation dose of approximately 0.15±0.03 gO3/gDOC was applied, while during the second campaign, September 3-5, 2018, a pre-ozonation dose of 0.33±0.04 gO3/gDOC was applied. GAC BVs during the sampling campaigns were around 44,000 and 48,000, respectively. Samples were collected as 24-h composites in borosilicate glass bottles at four sampling points along the treatment train, namely after mechanical filtration (INF), after biological treatment (BIO), after ozonation (OZO), and after GAC filtration, ie., at the effluent (EFF). Influent samples were collected one day prior to account for the HRT in the WWTP.

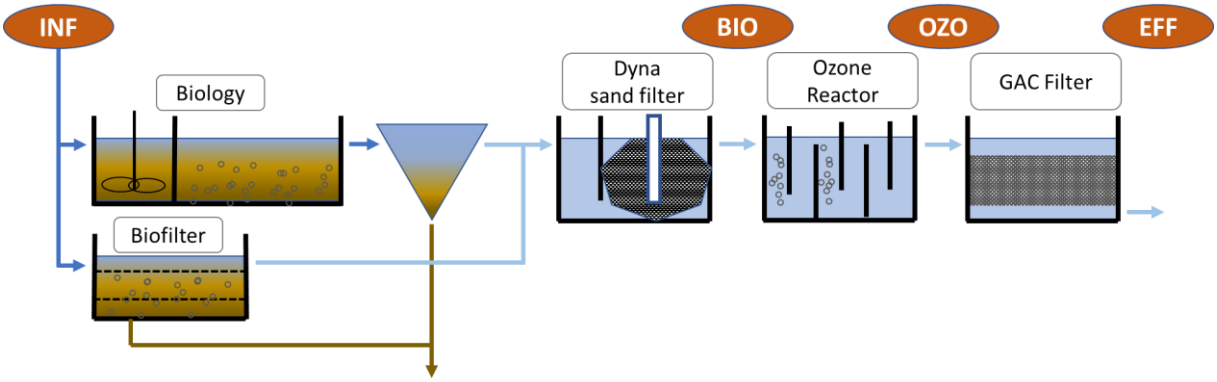


Figure S 2. Schematic from WWTP Altenrhein (AR). Sampling locations in the influent (INF), after biological treatment (BIO), after ozonation (OZO), and after post-treatment with a GAC filter in the effluent (EFF) are indicated.

WWTP ProRheno

WWTP ProRheno (*PR*) treats 28.8 million m^3, with a PE of 470,000. Wastewater was collected from two separate input streams; one for the mainly municipal wastewater from 270,000 connected inhabitants (*PR-Comm*; 20-25% industrial wastewater) and the other from local pharmaceutical industries (*PR-Chem*; 100% industrial wastewater). The two wastewater streams were manually mixed in ratio 90:10 (v:v, PR-Comm:PR-Chem), which corresponds to a DOC ratio of 80:20 (Krahnstöver *et al*. , 2018). A pilot scale treatment plant consisted of a sequencing batch reactor (SBR; approximate volume 380 L), followed by a buffer tank (approximate volume 400L; cycle time 6.5 hours). The flow was then directed to an advanced treatment train at a constant rate of 20-25 L/h.

The advanced treatment consisted of an ozonation reaction chamber (two-column; residence time approximately 30 minutes), followed by a stirred reactor where PAC and iron (III) cloride (residence time approximately 13 minutes) was dosed (Figure S 3). Finally, the wastewater containing the PAC was directed over a sand filter (two-layer; upper layer 120 cm clay, bottom layer 60 cm sand), with an approximate residence time of 12.5 minutes plus 3.5-10 minutes residence time in the supernatant on top of the filter bed. Sludge water with PAC was recirculated to the SBR. More information on the WWTP and the pilot plant can be found in Krahnstöver et al. 2018.

Five sampling campaigns were conducted at this WWTP, with varying ozone doses and PAC doses (Table S 1). The first sampling campaign was conducted on February 21, 2017, March 2, 2017, and March 9, 2017, where an ozone dose of 0.23±0.04 gO3/gDOC and 7.7±1.6 mg/L PAC was applied. This sampling was meant to use the same ozone dose as the second sampling campaign (defined at the reference setting), but with a lower PAC dose. Sampling for the second campaign with higher PAC dose was conducted one month later on April 6, 2017, April 12, 2017, and April 20, 2017. During this campaign, which represented the reference setting, an ozone dose of 0.20±0.01 gO3/gDOC and 12.1±2.0 mg/L PAC was applied. For the third sampling campaign, samples were collected on April 27, 2017, May 4, 2017, and May 11, 2017. This sample campaign used lower ozone dose (0.09±0.01 gO3/gDOC)

compared to the reference setting, with the same PAC dose (12.5±3.6 mg/L PAC). In a fourth sampling campaign, conducted on June 6, 2017, June 15, 2017, and June 22, 2017, no ozone was dosed and wastewater was treated only with PAC (13.6±2 mg/L PAC) dosed onto the sand filter. For the fifth sampling campaign, samples were collected again a month later to ensure PAC flushing out of the sand filter on July 20, 2017, July 27, 2017, and August 10, 2017. During this campaign, an ozone dose similar to the reference setting was applied (0.26±0.04 gO3/gDOC), but no PAC was dosed, so only the sand filter served as a post-treatment. During these sampling events, samples were collected in borosilicate glass bottles as 48-h composites at the following five sampling points: at the influent of the municipal wastewater stream (PR-Comm); at the influent of the industrial wastewater stream (PR-Chem); after biological treatment (BIO); after ozonation (OZO); and after the sand filter, at the effluent (EFF). For the INF sample in the trend analysis, 90:10 mixture of PR-Comm and PR-Chem was taken into account to calculate the intensities of the individual features.

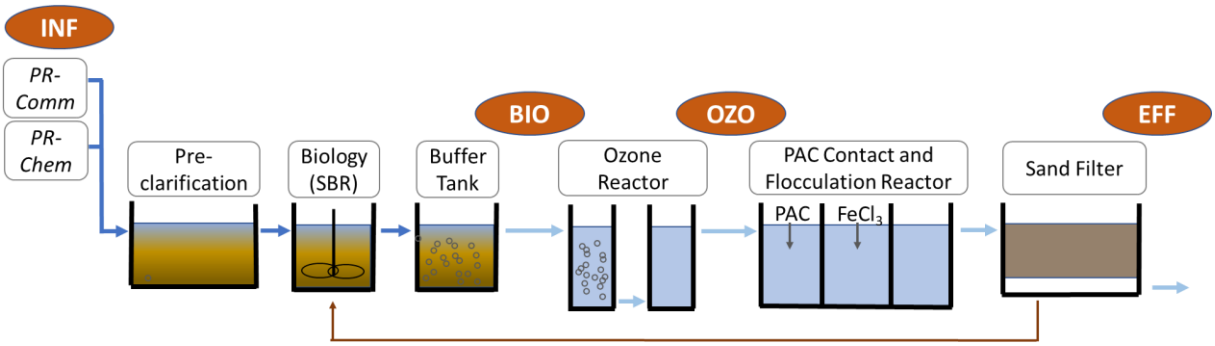


Figure S 3. Schematic from WWTP ProRheno (PR). Sampling locations in the influent (INF), after biological treatment (BIO), after ozonation (OZO), and after post-treatment with powder activated carbon (PAC) dosed on a sand filter in the effluent (EFF) are indicated. Two influent streams were sampled, PR-Comm, which is collected from mainly municipal sources, and PR-Chem, which is collected from the pharmaceutical industries in the area. These two streams were then mixed in the WWTP in a reaction of 90:10 PR-Comm:PR-Chem prior to treatment.

A complete list of samples collected can be found in Table S 1.

Table S 1. Summary of Samples, including waterwater treatment plant (WWTP), location in the treatment train, sampling date, sample name, and corresponding pooled sample.

WWTP	Sampling location	Sampling date (YYMMDD)	Sample Name	Pooled Sample
Altenrhein	in influent	180716	AR-INF-180716	AR-INF-pool
Altenrhein	in influent	180717	AR-INF-180717	AR-INF-pool
Altenrhein	in influent	180718	AR-INF-180718	AR-INF-pool
Altenrhein	in influent	180903	AR-INF-180903	AR-INF-pool
Altenrhein	in influent	180904	AR-INF-180904	AR-INF-pool
Altenrhein	in influent	180905	AR-INF-180905	AR-INF-pool
Altenrhein	after biological treatment	180716	AR-BIO-180716	AR-BIO-pool
Altenrhein	after biological treatment	180717	AR-BIO-180717	AR-BIO-pool
Altenrhein	after biological treatment	180718	AR-BIO-180718	AR-BIO-pool
Altenrhein	after biological treatment	180903	AR-BIO-180903	AR-BIO-pool
Altenrhein	after biological treatment	180904	AR-BIO-180904	AR-BIO-pool
Altenrhein	after biological treatment	180905	AR-BIO-180905	AR-BIO-pool
Altenrhein	after ozonation	180716	AR-OZO-180716	AR-OZO-pool
Altenrhein	after ozonation	180717	AR-OZO-180717	AR-OZO-pool
Altenrhein	after ozonation	180718	AR-OZO-180718	AR-OZO-pool
Altenrhein	after ozonation	180903	AR-OZO-180903	AR-OZO-pool
Altenrhein	after ozonation	180904	AR-OZO-180904	AR-OZO-pool
Altenrhein	after ozonation	180905	AR-OZO-180905	AR-OZO-pool
Altenrhein	after post-treatment, in effluent	180716	AR-EFF-180716	AR-EFF-pool
Altenrhein	after post-treatment, in effluent	180717	AR-EFF-180717	AR-EFF-pool
Altenrhein	after post-treatment, in effluent	180718	AR-EFF-180718	AR-EFF-pool
Altenrhein	after post-treatment, in effluent	180903	AR-EFF-180903	AR-EFF-pool
Altenrhein	after post-treatment, in effluent	180904	AR-EFF-180904	AR-EFF-pool
Altenrhein	after post-treatment, in effluent	180905	AR-EFF-180905	AR-EFF-pool
Glarnerland	in influent	170608	GL-INF-170608	GL-INF-pool
Glarnerland	in influent	180116	GL-INF-180116	GL-INF-pool
Glarnerland	in influent	180905	GL-INF-180905	GL-INF-pool
Glarnerland	after biological treatment	170214[1]	GL-BIO-170214	GL-BIO-pool
Glarnerland	after biological treatment	170516[1]	GL-BIO-170516	GL-BIO-pool
Glarnerland	after biological treatment	170518[1]	GL-BIO-170518	GL-BIO-pool
Glarnerland	after biological treatment	170523[1]	GL-BIO-170523	GL-BIO-pool
Glarnerland	after biological treatment	170601[1]	GL-BIO-170601	GL-BIO-pool
Glarnerland	after biological treatment	170608	GL-BIO-170608	GL-BIO-pool
Glarnerland	after biological treatment	170926[1]	GL-BIO-170926	GL-BIO-pool
Glarnerland	after biological treatment	171121[1]	GL-BIO-171121	GL-BIO-pool
Glarnerland	after biological treatment	180116	GL-BIO-180116	GL-BIO-pool
Glarnerland	after biological treatment	180418[1]	GL-BIO-180418	GL-BIO-pool
Glarnerland	after biological treatment	180905	GL-BIO-180905	GL-BIO-pool
Glarnerland	after ozonation	170214[1]	GL-OZO-170214	GL-OZO-pool
Glarnerland	after ozonation	170516[1]	GL-OZO-170516	GL-OZO-pool
Glarnerland	after ozonation	170518[1]	GL-OZO-170518	GL-OZO-pool
Glarnerland	after ozonation	170523[1]	GL-OZO-170523	GL-OZO-pool
Glarnerland	after ozonation	170601[1]	GL-OZO-170601	GL-OZO-pool
Glarnerland	after ozonation	170608	GL-OZO-170608	GL-OZO-pool
Glarnerland	after ozonation	170926[1]	GL-OZO-170926	GL-OZO-pool
Glarnerland	after ozonation	171121[1]	GL-OZO-171121	GL-OZO-pool
Glarnerland	after ozonation	180116	GL-OZO-180116	GL-OZO-pool
Glarnerland	after ozonation	180418[1]	GL-OZO-180418	GL-OZO-pool
Glarnerland	after ozonation	180905	GL-OZO-180905	GL-OZO-pool
Glarnerland	after post-treatment, in effluent	170608	GL-EFFa-170608[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	170926[1]	GL-EFFa-170926[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	171121[1]	GL-EFFa-171121[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	180116	GL-EFFa-180116[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	180418[1]	GL-EFFa-180418[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	180905	GL-EFFa-180905[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	170214[1]	GL-EFFn-170214[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	170516[1]	GL-EFFn-170516[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	170518[1]	GL-EFFn-170518[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	170523[1]	GL-EFFn-170523[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	170601[1]	GL-EFFn-170601[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	170608	GL-EFFn-170608[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	170926[1]	GL-EFFn-170926[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	171121[1]	GL-EFFn-171121[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	180116	GL-EFFn-180116[2]	GL-EFF-pool

WWTP	Sampling location	Sampling date (YYMMDD)	Sample Name	Pooled Sample
Glarnerland	after post-treatment, in effluent	180418[1]	GL-EFFn-180418[2]	GL-EFF-pool
Glarnerland	after post-treatment, in effluent	180905	GL-EFFn-180905[2]	GL-EFF-pool
ProRheno	in influent	170221	PR-InfChem-170221[3]	PR-INF-pool
ProRheno	in influent	170302	PR-InfChem-170302[3]	PR-INF-pool
ProRheno	in influent	170309	PR-InfChem-170309[3]	PR-INF-pool
ProRheno	in influent	170406	PR-InfChem-170406[3]	PR-INF-pool
ProRheno	in influent	170412	PR-InfChem-170412[3]	PR-INF-pool
ProRheno	in influent	170420	PR-InfChem-170420[3]	PR-INF-pool
ProRheno	in influent	170427	PR-InfChem-170427[3]	PR-INF-pool
ProRheno	in influent	170504	PR-InfChem-170504[3]	PR-INF-pool
ProRheno	in influent	170511	PR-InfChem-170511[3]	PR-INF-pool
ProRheno	in influent	170606	PR-InfChem-170606[3]	PR-INF-pool
ProRheno	in influent	170615	PR-InfChem-170615[3]	PR-INF-pool
ProRheno	in influent	170622	PR-InfChem-170622[3]	PR-INF-pool
ProRheno	in influent	170720	PR-InfChem-170720[3]	PR-INF-pool
ProRheno	in influent	170810	PR-InfChem-170810[3]	PR-INF-pool
ProRheno	in influent	170221	PR-InfComm-170221[3]	PR-INF-pool
ProRheno	in influent	170302	PR-InfComm-170302[3]	PR-INF-pool
ProRheno	in influent	170309	PR-InfComm-170309[3]	PR-INF-pool
ProRheno	in influent	170406	PR-InfComm-170406[3]	PR-INF-pool
ProRheno	in influent	170412	PR-InfComm-170412[3]	PR-INF-pool
ProRheno	in influent	170420	PR-InfComm-170420[3]	PR-INF-pool
ProRheno	in influent	170427	PR-InfComm-170427[3]	PR-INF-pool
ProRheno	in influent	170504	PR-InfComm-170504[3]	PR-INF-pool
ProRheno	in influent	170511	PR-InfComm-170511[3]	PR-INF-pool
ProRheno	in influent	170606	PR-InfComm-170606[3]	PR-INF-pool
ProRheno	in influent	170615	PR-InfComm-170615[3]	PR-INF-pool
ProRheno	in influent	170622	PR-InfComm-170622[3]	PR-INF-pool
ProRheno	in influent	170720	PR-InfComm-170720[3]	PR-INF-pool
ProRheno	in influent	170727	PR-InfComm-170727[3]	PR-INF-pool
ProRheno	in influent	170810	PR-InfComm-170810[3]	PR-INF-pool
ProRheno	after biological treatment	170221	PR-BIO-170221	PR-BIO-pool
ProRheno	after biological treatment	170302	PR-BIO-170302	PR-BIO-pool
ProRheno	after biological treatment	170309	PR-BIO-170309	PR-BIO-pool
ProRheno	after biological treatment	170406	PR-BIO-170406	PR-BIO-pool
ProRheno	after biological treatment	170412	PR-BIO-170412	PR-BIO-pool
ProRheno	after biological treatment	170420	PR-BIO-170420	PR-BIO-pool
ProRheno	after biological treatment	170427	PR-BIO-170427	PR-BIO-pool
ProRheno	after biological treatment	170504	PR-BIO-170504	PR-BIO-pool
ProRheno	after biological treatment	170511	PR-BIO-170511	PR-BIO-pool
ProRheno	after biological treatment	170606	PR-BIO-170606	PR-BIO-pool
ProRheno	after biological treatment	170615	PR-BIO-170615	PR-BIO-pool
ProRheno	after biological treatment	170622	PR-BIO-170622	PR-BIO-pool
ProRheno	after biological treatment	170720	PR-BIO-170720	PR-BIO-pool
ProRheno	after biological treatment	170727	PR-BIO-170727	PR-BIO-pool
ProRheno	after biological treatment	170810	PR-BIO-170810	PR-BIO-pool
ProRheno	after ozonation	170221	PR-OZO-170221	PR-OZO-pool
ProRheno	after ozonation	170302	PR-OZO-170302	PR-OZO-pool
ProRheno	after ozonation	170309	PR-OZO-170309	PR-OZO-pool
ProRheno	after ozonation	170406	PR-OZO-170406	PR-OZO-pool
ProRheno	after ozonation	170412	PR-OZO-170412	PR-OZO-pool
ProRheno	after ozonation	170420	PR-OZO-170420	PR-OZO-pool
ProRheno	after ozonation	170427	PR-OZO-170427	PR-OZO-pool
ProRheno	after ozonation	170504	PR-OZO-170504	PR-OZO-pool
ProRheno	after ozonation	170511	PR-OZO-170511	PR-OZO-pool
ProRheno	after ozonation	170606	PR-OZO-170606	PR-OZO-pool
ProRheno	after ozonation	170615	PR-OZO-170615	PR-OZO-pool
ProRheno	after ozonation	170622	PR-OZO-170622	PR-OZO-pool
ProRheno	after ozonation	170720	PR-OZO-170720	PR-OZO-pool
ProRheno	after ozonation	170727	PR-OZO-170727	PR-OZO-pool
ProRheno	after ozonation	170810	PR-OZO-170810	PR-OZO-pool
ProRheno	after post-treatment, in effluent	170221	PR-EFF-170221	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170302	PR-EFF-170302	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170309	PR-EFF-170309	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170406	PR-EFF-170406	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170412	PR-EFF-170412	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170420	PR-EFF-170420	PR-EFF-pool

WWTP	Sampling location	Sampling date (YYMMDD)	Sample Name	Pooled Sample
ProRheno	after post-treatment, in effluent	170427	PR-EFF-170427	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170504	PR-EFF-170504	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170511	PR-EFF-170511	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170606	PR-EFF-170606	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170615	PR-EFF-170615	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170622	PR-EFF-170622	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170720	PR-EFF-170720	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170727	PR-EFF-170727	PR-EFF-pool
ProRheno	after post-treatment, in effluent	170810	PR-EFF-170810	PR-EFF-pool

[1]Measured but not analyzed, on account of missing influent sample

[2] Effn = effluent after ozonation and GAC filtration; Effa = Effluent after only GAC filtration

[3] InfComm = PR-Comm; InfChem = PR-Chem

Section S2. Analytical procedure

Chemicals

Reference compounds (*i.e.*, target compounds) and isotopically labeled internal standards (ISs) were purchased from various distributors, including Sigma Aldrich (Switzerland), ReseaChem (Switzerland), Lipmed (Switzerland), Novartis (Switzerland), Dr. Ehrenstorfer (Germany), HPC Standards (Germany), CDN Isotopes (Germany), TCI Europe (Belgium), LGC (UK), Toronto Research Chemicals (Canada), TRC Canada (Canada), Cambridge Isotope Laboratories (USA), and Cerilliant (USA). Organic solvents were obtained from Fisher Scientific (Switzerland), Sigma Aldrich (Switzerland), and Merck (Germany) in HPLC grade. Formic acid was also obtained from Merck (Germany) at >- 98% purity and nanopure water was produced onsite with a purification system (Barnstead Nanopure, Thermo Scientific, USA).

Sample preparation

Samples were first thawed overnight and then filtered through a two glass fiber filters (Whatman GF/D 2.7 µm pore size on top; Whatman GF/F 0.7 µm pore size on bottom) under vacuum in a borosilicate vacuum manifold. After filtration, all samples were diluted 1:1 w:w with nanopure water, except influent samples, which were diluted 1:3 w:w to account for possible higher matrix effects. Each sample was also spiked with a mixture of 147 isotopically labeled internal standards (ISs; each 200 ng/L; SI, Table S 2), which covered a broad mass (124.0806–844.5748) and retention time (9–24 minutes) range and were used as quality controls in the data evaluation. After filtration and spiking, samples were stored for a maximum of 8 days at 4°C prior to measurement.

In addition to individual samples, pooled samples were produced for each matrix at each WWTP. For example, equal aliquots of all BIO samples from WWTP Altenrhein were mixed to generate an *AR-BIO-pool* (for a complete summary of which individual samples were used for each pooled sample, see Table S1). These pooled samples were also split into two sets, one that was only spiked with ISs, and one set that was spiked with ISs, as well as a mixture of reference micropollutants (*i.e.*, target compounds). Target compounds were spiked at a level of 250 ng/L in BIO, OZO, and EFF samples, and at 1000 ng/L in INF samples. These target compounds were used to calculate relative recoveries and for target quantification in a very limited number of samples (Sideris 2019). For the non-target screening, the pooled samples were injected in triplicate and used for data cleaning through the implementation of a replicate filter (discussed in Section S3).

Table S 2. List of 147 isotopically labeled internal standards spiked and used for quality control of non-target screening workflow (available in separate Excel).

Sample measurement

For measurement, a method established for a large group of micropollutants (Bourgin et al. 2018, Huntscha et al. 2014) was used. First, samples were enriched with an automated two-phase online solid-phase extraction (SPE) with a PAL autosampler (CTC Analytics) as developed by Stoob *et al*. 2005. The SPE cartridge consisted of 9 mg Oasis HLB Sorbent (Waters) and 9 mg of a mixture of Strata-X-AW, Strata-X-CW (both Phenomenex), and Env+ (Biotage) in a ratio of 1:1:1.5 (w:w:w). The sorbents were manually packed into an aluminum cylinder and closed with stainless steel frits and an PTFE sealing-ring. The 20 mL sample was then injected and enriched at a flow rate of 1.27 mL/min and finally eluted in backflush mode with methanol with 0.1% formic acid at a flow rate of 40 µL/min.

Reverse-phase liquid chromatography was done with an Atlantis T3 column (Waters, 3 µm particle size, 3.0x150 mm, 100 Å inner diameter) fitted with a pre-column of the same material and precolumn filter on an Ultimate 3000 RS pump (Thermo Scientific). Separation was done with nanopure water (eluent A) and methanol (eluent B), both modified with 0.1% (v) formic acid. The starting ratio was 90:10 A:B for four minutes (which was sent to the waste), after which the ratio of B increased linearly to 95% over 16 minutes, for a hold of 9 minutes, prior to reequilibrium at 90:10 for 6 minutes. The column temperature was kept at 30°C with a column oven (Portmann Instruments).

The LC was coupled to a quadropule Orbitrap high-resolution mass spectrometer (HRMS; Q Exactive Plus, Thermo Scientific) through a heated electrospray ion source (ESI) operated in positive ionization

mode. In each cycle, full scan spectra (MS1; 140,000 R @ 200 *m/z* FHMW) were acquired in profile mode, followed by 5 fragmentation spectra (MS2; 17,500 R @ 200 *m/z* FHMW). The mass range was 100-1000 mass to charge ratio (*m/z*) and the mass accuracy was <5 ppm. Chemically induced dissociation (CID) fragmentation was achieved with nitrogen in a collision cell. MS2 scans were triggered based on an inclusion list containing the target compounds, with a dynamic exclusion window of 8 seconds. The 'pick other' option was enabled, meaning MS2 scans were secondarily triggered for the most intense *m/z*'s in the corresponding full scan, if none of the masses on the inclusion list were present.

Section S3. Data processing methods

Conversion RAW files

After measurement, acquired RAW files were converted to .mzXML centriod data with ProteoWizard (v. 3.0.11781). Unless otherwise noted, data analysis was performed in R (v.3.5.0 and v.3.5.3) and RStudio (v.1.1.453 and v.1.2.1335).

Summary of R software, packages and versions

setting	value
version	R version 3.5.3 (2019-03-11)
os	Windows Server 2012 R2 x64
system	x86_64, mingw32
ui	RStudio
language	(EN)
collate	English_United States.1252
ctype	English_United States.1252
tz	Europe/Berlin
date	2020-03-03

Table S 3. Summary of R packages used, including version and source.

package	version	date	source	Citation
affy	1.60.0	10/30/18	Bioconductor	Gautier et al. 2004
affyio	1.52.0	10/30/18	Bioconductor	Bolstad 2018b
assertthat	0.2.1	3/21/19	CRAN (R 3.5.3)	Wickham 2019a
backports	1.1.5	10/2/19	CRAN (R 3.5.3)	Lang and Team 2019c
base64enc	0.1-3	7/28/15	CRAN (R 3.5.2)	Urbanek 2015
Biobase	2.42.0	10/30/18	Bioconductor	Huber et al. 2015
BiocGenerics	0.28.0	10/30/18	Bioconductor	Huber et al. 2015
BiocManager	1.30.10	11/16/19	CRAN (R 3.5.3)	Morgan 2019
BiocParallel	1.16.6	2/10/19	Bioconductor	Morgan et al. 2019
bit	1.1-15.1	1/14/20	CRAN (R 3.5.3)	Oehlschlägel 2018
bitops	1.0-6	8/17/13	CRAN (R 3.5.2)	Dutky 2013
broom	0.5.4	1/27/20	CRAN (R 3.5.3)	Robinson and Hayes 2019
callr	3.4.1	1/24/20	CRAN (R 3.5.3)	Csárdi and Chang 2019
caTools	1.17.1.2	3/6/19	CRAN (R 3.5.3)	Tuszynski 2019
cellranger	1.1.0	7/27/16	CRAN (R 3.5.3)	Bryan 2016
class	7.3-15	1/1/19	CRAN (R 3.5.3)	Venables and Ripley 2002
cli	2.0.1	1/8/20	CRAN (R 3.5.3)	Csárdi 2019
codetools	0.2-16	12/24/18	CRAN (R 3.5.3)	Tierney 2018
colorspace	1.4-1	3/18/19	CRAN (R 3.5.3)	Zeileis et al. 2019
crayon	1.3.4	9/16/17	CRAN (R 3.5.3)	Csárdi 2017
DBI	1.1.0	12/15/19	CRAN (R 3.5.3)	(R-SIG-DB) et al. 2018
dbplyr	1.4.2	6/17/19	CRAN (R 3.5.3)	Wickham and Ruiz 2019
depRec	* 1.0	1/23/20	Github (blosloos/depRec@b4b148b)	
desc	1.2.0	5/1/18	CRAN (R 3.5.3)	Csárdi et al. 2018
devtools	* 2.2.1	9/24/19	CRAN (R 3.5.3)	Wickham et al. 2019b
digest	0.6.23	11/23/19	CRAN (R 3.5.3)	Eddelbuettel et al. 2019
doParallel	1.0.15	8/2/19	CRAN (R 3.5.3)	Corporation and Weston 2018
dplyr	* 0.8.3	7/4/19	CRAN (R 3.5.3)	Wickham et al. 2019a
dynamicTreeCut	1.63-1	3/11/16	CRAN (R 3.5.2)	Langfelder et al. 2016
e1071	1.7-3	11/26/19	CRAN (R 3.5.3)	Meyer et al. 2019
ellipsis	0.3.0	9/20/19	CRAN (R 3.5.3)	Wickham 2019b
enviPat	* 2.4	4/7/19	CRAN (R 3.5.3)	Loos et al. 2015
enviPick	* 1.5	6/6/16	CRAN (R 3.5.3)	Loos 2016
fansi	0.4.1	1/8/20	CRAN (R 3.5.3)	Gaslam 2018
fastcluster	1.1.25	6/7/18	CRAN (R 3.5.2)	Müllner 2013
fastmap	1.0.1	10/8/19	CRAN (R 3.5.3)	Chang 2019a
ff	2.2-14	5/15/18	CRAN (R 3.5.3)	Adler et al. 2018
fingerprint	3.5.7	1/7/18	CRAN (R 3.5.3)	Guha 2018
forcats	* 0.4.0	2/17/19	CRAN (R 3.5.3)	Wickham 2019c
foreach	1.5.1	11/26/18	R-Forge (R 3.5.1)	Microsoft and Weston 2017
fs	1.3.1	5/6/19	CRAN (R 3.5.3)	Hester and Wickham 2019
gdata	2.18.0	6/6/17	CRAN (R 3.5.3)	Warnes et al. 2017
generics	0.0.2	11/29/18	CRAN (R 3.5.3)	Kuhn et al. 2018
ggplot2	* 3.2.1	8/10/19	CRAN (R 3.5.3)	Wickham 2016
glue	1.3.1	3/12/19	CRAN (R 3.5.3)	Hester 2019
gplots	3.0.1.2	1/11/20	CRAN (R 3.5.3)	Warnes et al. 2019
gtable	0.3.0	3/25/19	CRAN (R 3.5.3)	Wickham and Pedersen 2019
gtools	3.8.1	6/26/18	CRAN (R 3.5.2)	Warnes et al. 2018
haven	2.2.0	11/8/19	CRAN (R 3.5.3)	Wickham and Miller 2019
hms	0.5.3	1/8/20	CRAN (R 3.5.3)	Müller 2018a
htmltools	0.4.0	10/4/19	CRAN (R 3.5.3)	Inc. 2017
httpuv	1.5.2	9/11/19	CRAN (R 3.5.3)	Cheng et al. 2019
httr	1.4.1	8/5/19	CRAN (R 3.5.3)	Wickham 2018b
impute	1.56.0	10/30/18	Bioconductor	Hastie et al. 2018
InterpretMSSpectrum	1.2	5/3/18	CRAN (R 3.5.3)	Lisec 2018
IRanges	2.16.0	10/30/18	Bioconductor	Lawrence et al. 2013
iterators	1.0.12	7/26/19	CRAN (R 3.5.3)	Analytics and Weston 2018
itertools	0.1-3	3/12/14	CRAN (R 3.5.3)	Weston and Wickham 2014
jsonlite	1.6	12/7/18	CRAN (R 3.5.3)	Ooms 2014
KernSmooth	2.23-15	6/29/15	CRAN (R 3.5.3)	Wand 2015
later	1.0.0	10/4/19	CRAN (R 3.5.3)	Cheng and Chang 2019
lattice	* 0.20-38	11/4/18	CRAN (R 3.5.3)	Sarkar 2008
lazyeval	0.2.2	3/15/19	CRAN (R 3.5.3)	Wickham 2019d
lifecycle	0.1.0	8/1/19	CRAN (R 3.5.3)	Henry 2020
limma	3.38.3	12/2/18	Bioconductor	Ritchie et al. 2015

package	version	date	source	Citation
lubridate	1.7.4	4/11/18	CRAN (R 3.5.3)	Grolemund and Wickham 2011
magrittr	1.5	11/22/14	CRAN (R 3.5.3)	Bache and Wickham 2014
MALDIquant	1.19.3	5/12/19	CRAN (R 3.5.3)	Gibb and Strimmer 2012
MASS	7.3-51.1	11/1/18	CRAN (R 3.5.3)	Venables and Ripley 2002
Matrix	1.2-15	11/1/18	CRAN (R 3.5.3)	Bates and Maechler 2019
memoise	1.1.0	4/21/17	CRAN (R 3.5.3)	Wickham et al. 2017
mgcv	* 1.8-27	2/6/19	CRAN (R 3.5.3)	Wood 2011
mime	0.8	12/19/19	CRAN (R 3.5.3)	Xie 2018
modelr	0.1.5	8/8/19	CRAN (R 3.5.3)	Wickham 2019e
MSnbase	2.8.3	1/5/19	Bioconductor	Gatto and Lilley 2011
munsell	0.5.0	6/12/18	CRAN (R 3.5.3)	Wickham 2018a
mzID	1.20.1	1/4/19	Bioconductor	Pedersen et al. 2019
mzR	2.14.0	5/1/18	Bioconductor	Chambers et al. 2012
nlme	* 3.1-137	4/7/18	CRAN (R 3.5.3)	Pinheiro et al. 2018
nontarget	* 1.9	9/27/16	CRAN (R 3.5.3)	Loos 2015
nontargetData	* 1.1	7/22/14	CRAN (R 3.5.2)	Loos and Corona 2014
pcaMethods	1.74.0	10/30/18	Bioconductor	Stacklies et al. 2007
pillar	1.4.3	12/20/19	CRAN (R 3.5.3)	Müller and Wickham 2018
pkgbuild	1.0.6	10/9/19	CRAN (R 3.5.3)	Wickham and Hester 2019
pkgconfig	2.0.3	9/22/19	CRAN (R 3.5.3)	Csárdi 2018
pkgload	1.0.2	10/29/18	CRAN (R 3.5.3)	Wickham et al. 2018a
plyr	* 1.8.5	12/10/19	CRAN (R 3.5.3)	Wickham 2011a
png	0.1-7	12/3/13	CRAN (R 3.5.2)	Urbanek 2013
preprocessCore	1.44.0	10/30/18	Bioconductor	Bolstad 2018a
prettyunits	1.1.1	1/24/20	CRAN (R 3.5.3)	Csardi 2018
processx	3.4.1	7/18/19	CRAN (R 3.5.3)	Csardi and Chang 2019
promises	1.1.0	10/4/19	CRAN (R 3.5.3)	Cheng 2018
ProtGenerics	1.14.0	10/30/18	Bioconductor	Gatto 2018
ps	1.3.0	12/21/18	CRAN (R 3.5.3)	Loden et al. 2018
purrr	* 0.3.3	10/18/19	CRAN (R 3.5.3)	Henry and Wickham 2019
R.methodsS3	* 1.7.1	2/16/16	CRAN (R 3.5.2)	Bengtsson 2003
R.oo	* 1.23.0	11/3/19	CRAN (R 3.5.3)	Bengtsson 2003
R.utils	* 2.9.2	12/8/19	CRAN (R 3.5.3)	Bengtsson 2019
R6	2.4.1	11/12/19	CRAN (R 3.5.3)	Chang 2019b
RAMClustR	* 1.0.9	9/20/19	CRAN (R 3.5.3)	Broeckling et al. 2014
rcdk	3.4.7.2	12/21/19	CRAN (R 3.5.3)	Guha 2007
rcdklibs	2	6/11/17	CRAN (R 3.5.3)	Guha 2017
Rcpp	1.0.3	11/8/19	CRAN (R 3.5.3)	Eddelbuettel 2013, Eddelbuettel and Balamuta 2017, Eddelbuettel and Francois 2011
RCurl	1.98-1.1	1/19/20	CRAN (R 3.5.3)	Lang and team 2019a
Rdisop	1.40.0	5/1/18	Bioconductor	Bocker et al. 2006, 2009, Bocker and Liptak 2007, Bocker et al. 2008
readMzXmlData	* 2.8.1	9/16/15	CRAN (R 3.5.3)	Gibb 2015
readr	* 1.3.1	12/21/18	CRAN (R 3.5.3)	Wickham et al. 2018b
readxl	1.3.1	3/13/19	CRAN (R 3.5.3)	Wickham and Bryan 2019a
remotes	2.1.0	6/24/19	CRAN (R 3.5.3)	Csardi et al. 2019
reprex	0.3.0	5/16/19	CRAN (R 3.5.3)	Bryan et al. 2019
reshape	* 0.8.8	10/23/18	CRAN (R 3.5.3)	Wickham 2007
reshape2	* 1.4.3	12/11/17	CRAN (R 3.5.3)	Wickham 2007
D rJava	0.9-11	3/29/19	CRAN (R 3.5.3)	Urbanek 2019
rjson	0.2.20	6/8/18	CRAN (R 3.5.2)	Couture-Beil 2018
rlang	0.4.3	1/24/20	CRAN (R 3.5.3)	Henry and Wickham 2020
RMassBank	2.99.0	7/12/19	Github (MassBank/RMassBank@275f843)	Stravs et al. 2013
RMassScreening	* 0.2	1/23/20	Github (meowcat/RMassScreening@c59edab)	Stravs 2015
rprojroot	1.3-2	1/3/18	CRAN (R 3.5.3)	Müller 2018b
rstudioapi	0.1	3/19/19	CRAN (R 3.5.3)	Ushey et al. 2019
rvest	0.3.5	11/8/19	CRAN (R 3.5.3)	Wickham 2019f
S4Vectors	0.20.1	11/9/18	Bioconductor	Pagès et al. 2018
scales	1.1.0	11/18/19	CRAN (R 3.5.3)	Wickham 2018c
sessioninfo	1.1.1	11/5/18	CRAN (R 3.5.3)	Csardi et al. 2018
shiny	* 1.4.0	10/10/19	CRAN (R 3.5.3)	Chang et al. 2019
shinyAce	* 0.4.1	9/24/19	CRAN (R 3.5.3)	Nijs et al. 2019
stringi	1.4.5	1/11/20	CRAN (R 3.5.3)	Gagolewski 2019
stringr	* 1.4.0	2/10/19	CRAN (R 3.5.3)	Wickham 2019g
testthat	2.3.1	12/1/19	CRAN (R 3.5.3)	Wickham 2011b
tibble	* 2.1.3	6/6/19	CRAN (R 3.5.3)	Müller and Wickham 2019

package	version	date	source	Citation
tidyr	* 1.0.2	1/24/20	CRAN (R 3.5.3)	Wickham and Henry 2019
tidyselect	0.2.5	10/11/18	CRAN (R 3.5.3)	Henry and Wickham 2018
tidyverse	* 1.3.0	11/21/19	CRAN (R 3.5.3)	Wickham 2017
usethis	* 1.5.1	7/4/19	CRAN (R 3.5.3)	Wickham and Bryan 2019b
vctrs	0.2.2	1/24/20	CRAN (R 3.5.3)	Wickham 2018d
vsn	3.50.0	10/30/18	Bioconductor	Huber et al. 2002
withr	2.1.2	3/15/18	CRAN (R 3.5.3)	Hester et al. 2018
XLConnect	* 0.2-15	4/5/18	CRAN (R 3.5.3)	GmbH 2018a
XLConnectJars	* 0.2-15	4/5/18	CRAN (R 3.5.3)	GmbH 2018b
XML	3.99-0.3	1/20/20	CRAN (R 3.5.3)	Lang and team 2019b
xml2	1.2.2	8/9/19	CRAN (R 3.5.3)	Wickham et al. 2018c
xtable	1.8-4	4/21/19	CRAN (R 3.5.3)	Dahl et al. 2018
yaml	2.2.0	7/25/18	CRAN (R 3.5.3)	Stephens et al. 2018
zlibbioc	1.28.0	10/30/18	Bioconductor	Morgan 2018

Method Evaluation

A number of new pre-processing steps were developed and/or incorporated into a previously developed non-target screening workflow to increase the quality of the data. A set of 147 internal standards were used to evaluate the efficacy and effect of each step (Table S 2). These results are outlined in the following section. The final workflow is shown in Figure S 4.

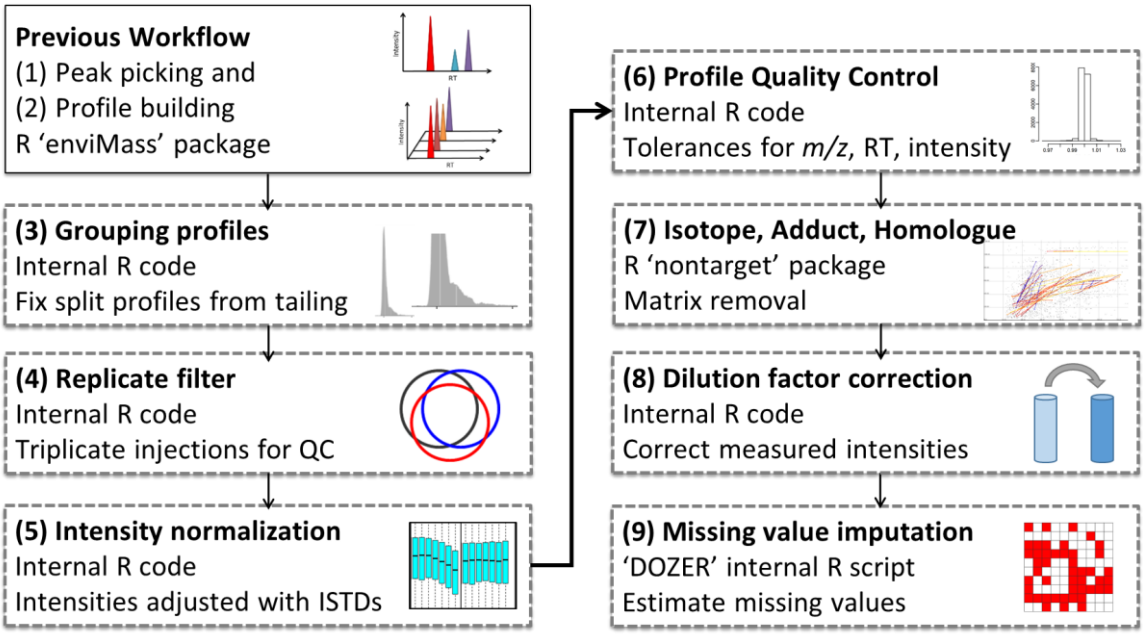


Figure S 4. Figure of non-target screening pre-processing workflow. New workflow steps are dashed.

Table S 4. Summary of the total number of features and the number of internal standard features detected after each step of the non-target screening workflow. The total number of spiked internal standards was 147.

	After (1) peak picking and (2) profiling	After (3) profile grouping	After (4) replicate filter	After (5) intensity normalization	After (6) profile quality control filter	After (7) isotope, adduct, and homologue filtering	After (8) dilution factor correction	After (9) missing value imputation
non-targets	459867	395408	67994	67994	60953	50722	50722	50722
IS false positives	989	782	14	14	11	9	9	9
IS true positives	147	147	147	147	147	147	147	147

Step (1) and (2): Peak picking and profiling with enviMass functions

```
Input to enviPick function:
  dmzgap: 10
  ppm: TRUE
  drtgap: 800
  minpeak: 5
  maxint: 1.0E+20
  dmzdens: 3
  drtdens: 450
  drtsmall: 40
  drtfill: 20
  drttotal: 400
  recurs: 3
  weight: 1
  SB: 3
  SN: 2
  minint: 1.0E+5
  ended: 2
  scantypes:
    "1-pos-*":
      maxint: 1.0E+20
    "1-neg-*":
      maxint: 3.0E+20
    "2-pos-*":
      maxint: 5.0E+20
      minint: 5.0e+3
    "2-neg-*":
      maxint: 5.e+20
      minint: 5.0e+3
screenProfiles:
  ppmLimit: 3
  rtLimit: NULL
viewer:
  hitsLimit: 2000
```

Step (3): Profile grouping function

It was observed that the peak picking would select both the main chromatographic peak and the tailing portion of the peak. This problem was seen especially for compounds present at high intensities (for an example of such a peak, see Figure S 5). These so-called 'split peaks' led to the formation of false positive features, because multiple non-target features were then associated with 1 compound. To remove these features, an algorithm was developed, *profileGroup.upwards*, to combine these features into one feature. The feature list was sorted from highest to lowest intensity, and then features suspected of belonging to the same compound were found using an *m/z* window of 1 ppm and a RT window of 20 seconds. Once groups were established, the intensities of grouped features were summed and assigned to the profileID of the more intense feature (since this is likely to be the main chromatographic peak and therefore also closer to the correct RT).

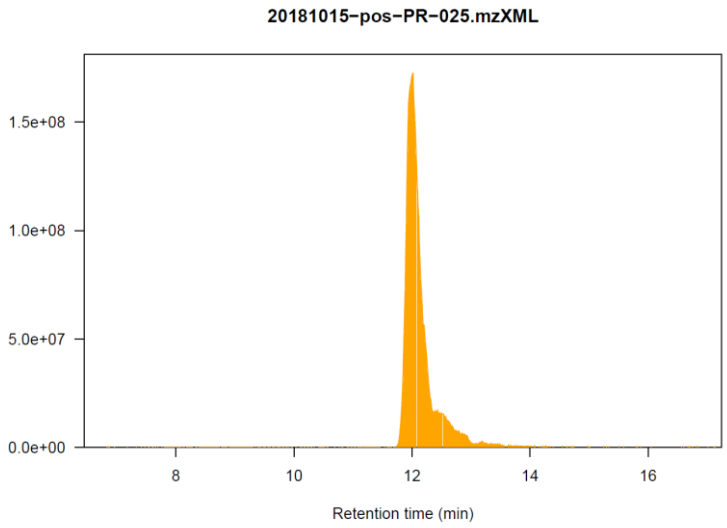


Figure S 5. Example of broad, tailing peak, for which split profiles were detected

Step (4): Replicate filter

As shown previously (Bader et al. 2016), including a replicate filter in the pre-processing of LC-HRMS data is recommended to remove noise. The use of pooled samples for quality control is standard in metabolomics analysis, where all samples are mixed together to generate a pooled sample, which is measured repeatedly throughout the measurement sequence. For environmental analysis, there is higher variability in the presence of compounds in the samples, because compounds can differ, depending on the sample location. Additionally, many environmental compounds are present at low concentrations and may fall below the limit of detection (LOD) when mixing many samples together. Therefore, it was decided to generate a pooled sample for each matrix in each WWTP (Table S 1), in order to retain the quality control aspect, without potentially diluting low intensity peaks so much so as to make them undetectable. Each pooled sample was injected in triplicate and non-target features were only retained if they were detected in all three replicates of a specific pooled sample type; all other non-target features were removed from the dataset.

Step (5): Intensity normalization

Intensity normalization is applied to correct for potential differences between matrices, leading to matrix suppression and/or enhancement. This is especially relevant for samples from wastewater influents, which have a high amount of matrix present (green samples in Figure S 6a). The algorithm applied is equivalent to the one implemented in *enviMass* and described in Albergamo et al. 2019. In short, using ISs with a detection frequency > 80%, a median intensity per internal standard is calculated over all samples. In each sample, the deviation of each IS $d_{i,j}$ to the median intensity is then calculated. Subsequently, the median deviation is calculated for each sample, which is then indicative of the matrix suppression in that sample and is then applied as a normalization factor to all intensities in that sample.

For a *n* x *m* matrix $W = \log_{10} V$, where *n* is the number of internal standards, *m* the number of samples, and element $w_{i,j}$ the log intensity of internal standard *i* in sample *j:*

$\tilde{w}_{i\in 1..n} = median_{j\in 1..m}(w_{i,j})$ *(median log intensity per IS)*

$d_{i,j} = w_{i,j} - \tilde{w}_i \mid i \in 1..n, j \in 1..m$ *(log intensity deviation for each IS in each sample)*

$\tilde{d}_{j\in 1..m} = median_{i\in 1..n}(d_{i,j})$ *(median log intensity deviation per sample)*

The correction factor *per sample* $\tilde{d}_{j\in 1..m}$ is then applied to the log sample intensity matrix $U = \log_{10} X$ (with *k* rows and *m* columns, where *k* is the number of features) with elements $u_{i,j}$, to yield the corrected matrix U_{corr}:

$u_{\text{corr}\,i,j} = u_{i,j} - d_j \mid i \in 1..k, j \in 1..m$ *(corrected log intensities per feature and sample)*

$X_{\text{corr}} = 10^{U_{\text{corr}}}$ *(corrected intensities per feature and sample)*

The effect of intensity normalization on the ISs (n=141) is shown below in Figure S 6 (top is uncorrected intensities, bottom is corrected intensities). Samples are shown in the order of measurement, colors indicate sample type (red=effluent, blue=ozonation, black=biological, green=influent, cyan=standards). Samples 84 to 89 had a very high mass deviation (>10ppm) and therefore no ISs were detected and not used in the intensity normalization or in the subsequent data analysis.

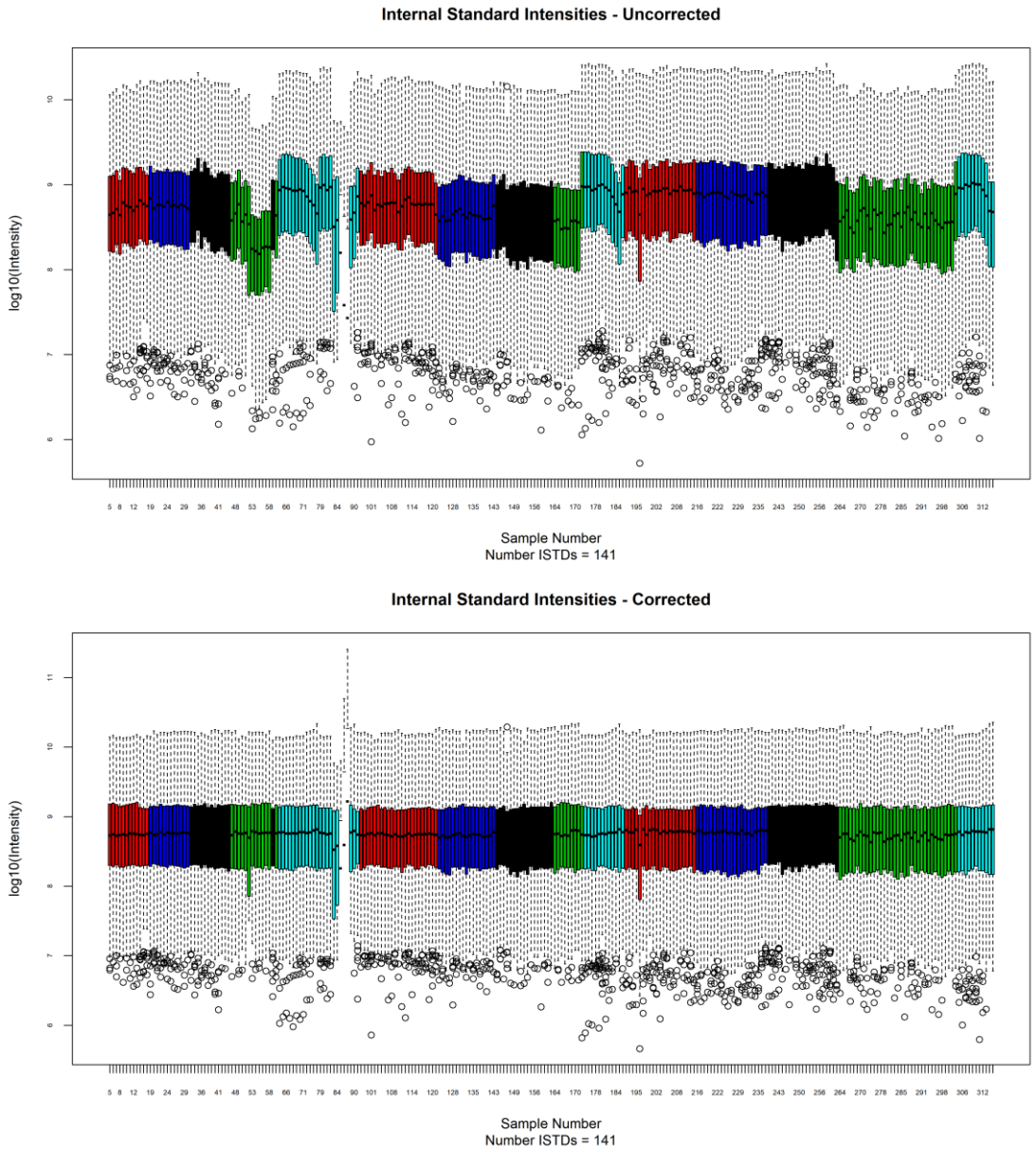


Figure S 6. Comparison of internal standard intensities (a; top) before intensity normalization and (b; bottom) after intensity normalization. Plotted on the x-axis is the sequence order and on the y-axis is the log10 of the intensity of the internal standards. The boxes indicate the interquantile range (IQR), with the lower edge of the box at the 25th percentile (*i.e.*, Q1) and the upper edge of the box at the 75th percentile (*i.e.*, Q3). The lower and upper whisker give Q1-1.5xIQR and Q3+1.5xIQR, respectively. Points outside of this range are plotted as open circles. Colors indicate sample matrix type. Red = EFF; blue = OZO; black = BIO; green = INF; cyan = calibration standards. First are samples from WWTP Altenrhein; then from WWTP Glarnerland; finally from WWTP ProRheno.

Step (6): profile QC filter

A new profile quality control filter was included in this workflow to address the cases, where peaks originating from different compounds were incorrectly profiled together. To detect these cases, first, the ratio of the median and the mean were calculated for *m/z* and RT in each detected feature assigned to a profile (called mzDist or rtDist). Then, the interquartile range (IQR; *i.e.*, the middle 50%) was calculated for both mzDist and rtDist across all features. Next, the lower bound of expected variability was set as the first quartile - 1.5xIQR, while the upper bound was set at the third quartile + 1.5xIQR. Profiles where the mzDist or rtDist were outside of the lower and upper bounds were labeled as outliers and were removed from the dataset.

Step (7): Isotope, Adduct, Homologue Filter

Isotopes, adducts, and homologues were detected with the *nontarget* package (v.1.9) and the relevant input parameters are provided below. In general, the detection of isotopes, adducts, and homologues is done based on expected mass differences of the centroid HRAM peaks. For isotope and adduct assignment, a RT criterion is also included to ensure correct assignment and for the isotope assignment, intensity ratios are used to further filter potential matches (https://cran.r-project.org/web/packages/nontarget/nontarget.pdf).

Settings for nontarget package

Input to *pattern.search2* function
  quantiz: OrbitrapXL_VelosPro_R60000at400_q
  isotopes: isotopes
  use_charges: c(1,2)
Input to *adduct.search* function:
  adducts: adducts
  rttol: 0.05
  ion_mode: positive
  use_adducts: M+K, M+H, M+Na, M+NH4
Input to *my.homol* function:
  isotopes: isotopes
  elements: C, H, O
  use_C: TRUE
  rttol: 0.5
  mztol: 3.5
  minlength: 6
  vec_size: 5e7

For the isotope detection the function *pattern.search2* was used; selected isotopes were C, O, N, S, Cl, and Br and charges were c(1,2). All non-targets features identified as a possible isotope were removed from the dataset. The results of the isotope detection are shown in Figure S 7.

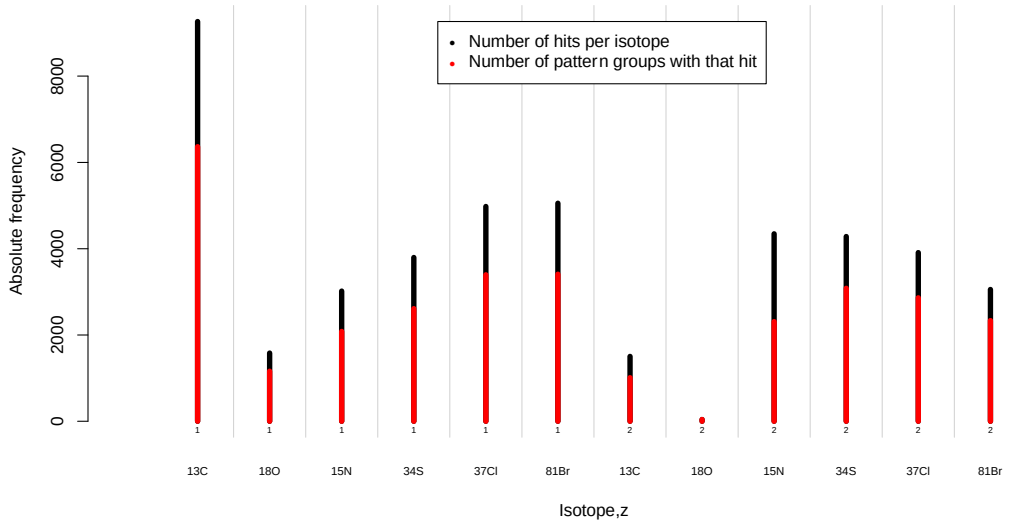


Figure S 7. *enviMass* output of the detected isotopes. On the x-axis is the isotope and charge z (directly under the bar), while on the y-axis is the absolute frequency, *i.e.*, the number of detections. Pattern groups are the detected isotopologues based on m/z difference and intensity pattern prior to profiling.

Adduct detection was performed with the function *adduct.search*. Selected adducts were M+H, M+Na, M+NH4, and M+K. The intensities and number of detections of these adducts is shown in Figure S 8. Any non-target features identified as possible adducts were removed from the dataset.

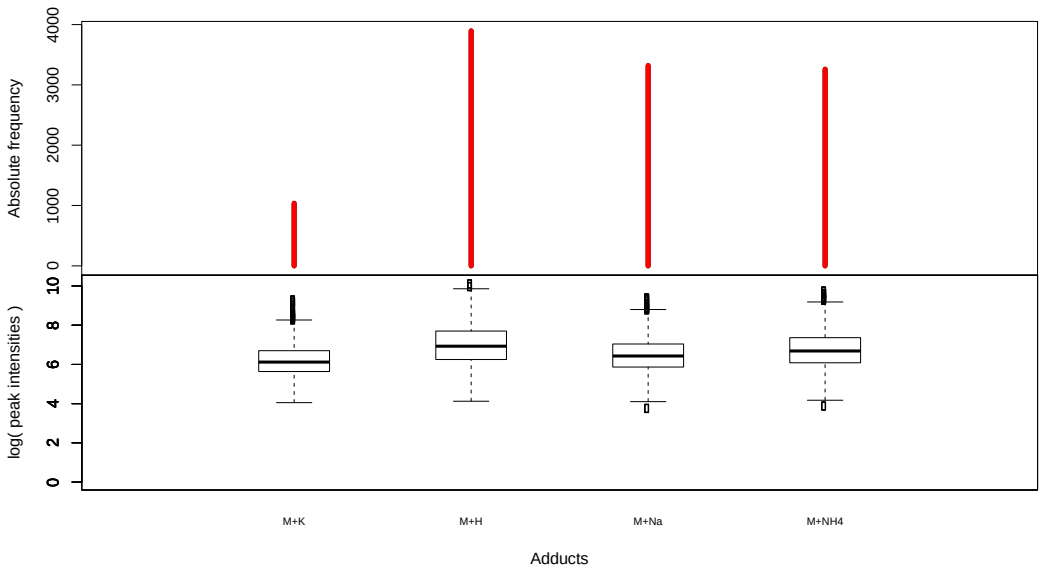


Figure S 8. *enviMass* output of the detected adducts. On the x-axis are the adduct types. In the top portion of the figure the absolute frequency of each adduct is shown; on the bottom portion box-whisker plots of the log10 of peak intensities of each adduct type.

The homologue search was executed with a slightly modified version of the *homol.search* function from the *nontarget* package. The elements C, H, and O were considered, use_C TRUE, RT tolerance was 0.5 minutes, *m/z* tolerance for finding homologues was 3.5 ppm, and a minimum series length of 6 were used. Despite these relatively strict tolerances, 48,941 of the 60,953 non-target features were grouped into homologue series. Removing all of these features was deemed to be too extreme, as only 12,012 non-target features would be remaining. Therefore, only the annotated homologes that were overlapping with the results of the isotope and adduct searching were removed (Figure S 9), resulting in the removal of 10,231 non-target features and a final dataframe with 50,722 non-target features.

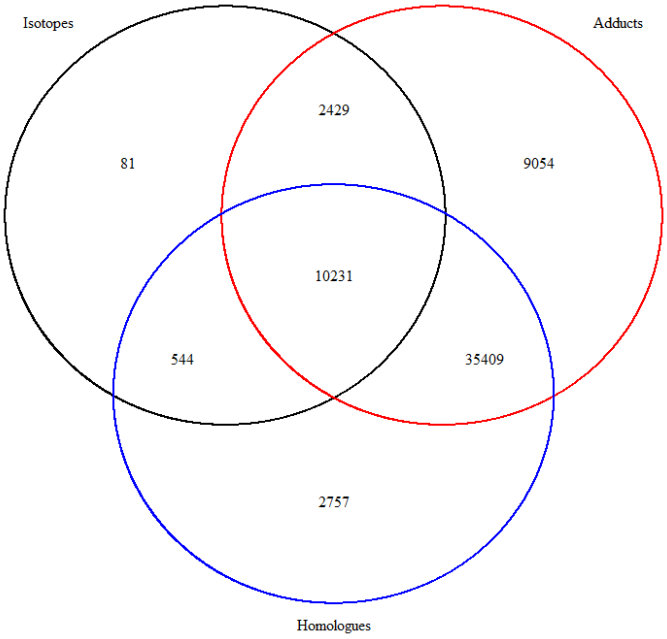


Figure S 9. Venn diagram of the isotope, adduct, and homologue filtering. In black are the features that were annotated as an isotope (*i.e.*, not a monoisotopic peak), in red are the features not annotated as an [M+H]+ adduct (*i.e.*, either not annotated or annotated as another adduct), and in blue are the features that belonged to at least 1 homologue series. Only features in the intersecting portion of the diagram (10,231 features) were filtered out of the final dataset.

Step (8): Dilution factor correction

Measured intensities of all features were corrected for the dilution factor applied during sample preparation, *i.e.*, x4 for INF sample and x2 for BIO, OZO, and EFF samples.

Step (9): Missing value imputation

It is known that in non-target screening, a high proportion of values are non-detects. This scarcity of data can influence data analysis and therefore it is recommended to include missing value imputation of the dataset. This newly developed algorithm addressed the problem of missing value imputation in multiple steps. First, the global instrument intensity limit of quantification (LOQ), based on the minimum, median, and variance of all detected intensities. Second, the minimum, median and variance was calculated per feature. The feature LOQ was then randomly selected from between the global minimum intensity and the local feature minimum. The missing values for this features were then sampled from between the estimated feature LOQ and the global LOQ. A visualization of imputed values is shown in Figure S 10, where values below 1 have been imputated. While this method functions sufficiently for not missing at random errors, it does not deal with missing at random errors.

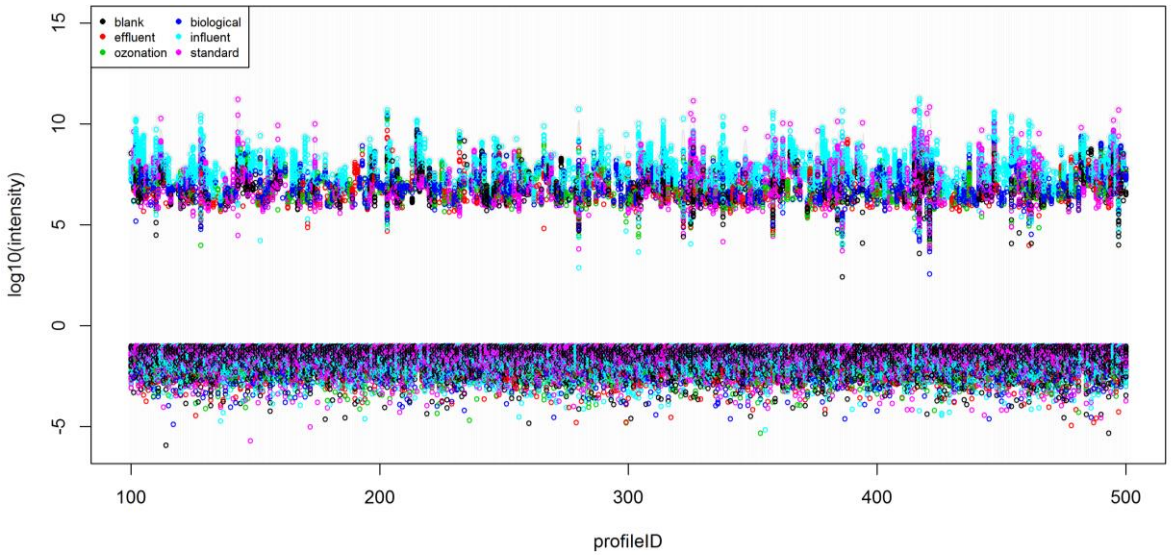


Figure S 10. Example of missing value imputation for 400 randomly selected non-target features. Colors indicate the matrix type of the sample. On the x-axis is the profileID of the non-target features; on the y-axis is the log10 of the intensity (measured and gap-filled).

Automated trend analysis

Previous non-target screening that incorporated trend analysis applied hierarchical cluster analysis to group together features with similar patterns (Albergamo et al. 2019, Chiaia-Hernández et al. 2017, Schollée et al. 2018). Trend were then assigned manually by visual inspection of each cluster. This method enabled a broad assessment of observed trends, distribution of features across different trends, and, most importantly, was used for the priorization of interesting non-target features for subsequent structure elucidation and identification. The method was useful for the identiciation of compounds in sediment and in groundwater. Nevertheless the method also presented a number of shortcomings. First, much time and effort was spent to select the correct number of clusters, in order to simulanteously remain specific enough to be useful, but general enough to be efficient. Second, even with cluster refinement, clusters were composed of a mixture of feature patterns that could not always be easily assigned to one trend. Finally, trends had to be assigned manually, which was only feasible for a small number of clusters and was subject to interpretation. To address these issues, a new method of automated trend assignment of individual non-target features was developed.

For this study, it was desirable to assign trends for each sampling date and each non-target feature individually, because different wastewater treatment settings were applied and the goal was to compare and constrast the different trends. On each sampling date, samples were collected at the four previously described locations: INF, BIO, OZO, and EFF. To assign a trend, each non-target intensity profile across these four locations was normalized to the maximum intensity of that profile. The intensities at each sampling point were then binned into one of the following three intensity domains: high (100-60%), middle (60-20%), and low (20-0%). These domains were assigned values of 1, 0, and -1, respectively; as such each intensity profile was converted to a four digit ‘barcode’. This barcode was then automatically converted to one of 65 possible trends (Figure S 12). To simplify interpretation, twelve major trends were defined (Figure S 13), which corresponded to known/expected trends in wastewater treatment and each minor trend was assigned to one major trend (Table S 5).

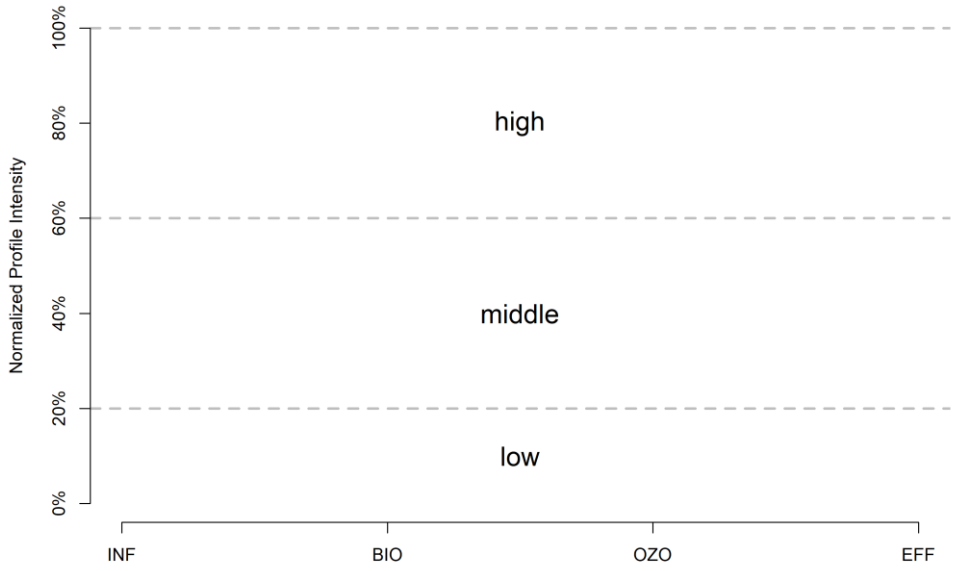


Figure S 11. Visualization of cutoff domains for profile barcoding. On the x-axis are the four sampling location along the wastewater treatment train. On the y-axis is the profile intensity, normalized to the maximum profile intensity. Shown are the 20% and 60% cutoffs used for cutoff_01 scenario. INF = in the WWTP influent; BIO = after biological treatment; OZO = after ozonation; EFF = after post-treatment (*i.e.*, GAC filtration or PAC+SF), in the WWTP effluent. Note that the "20% cutoff" refers to a 80% removal of the feature.

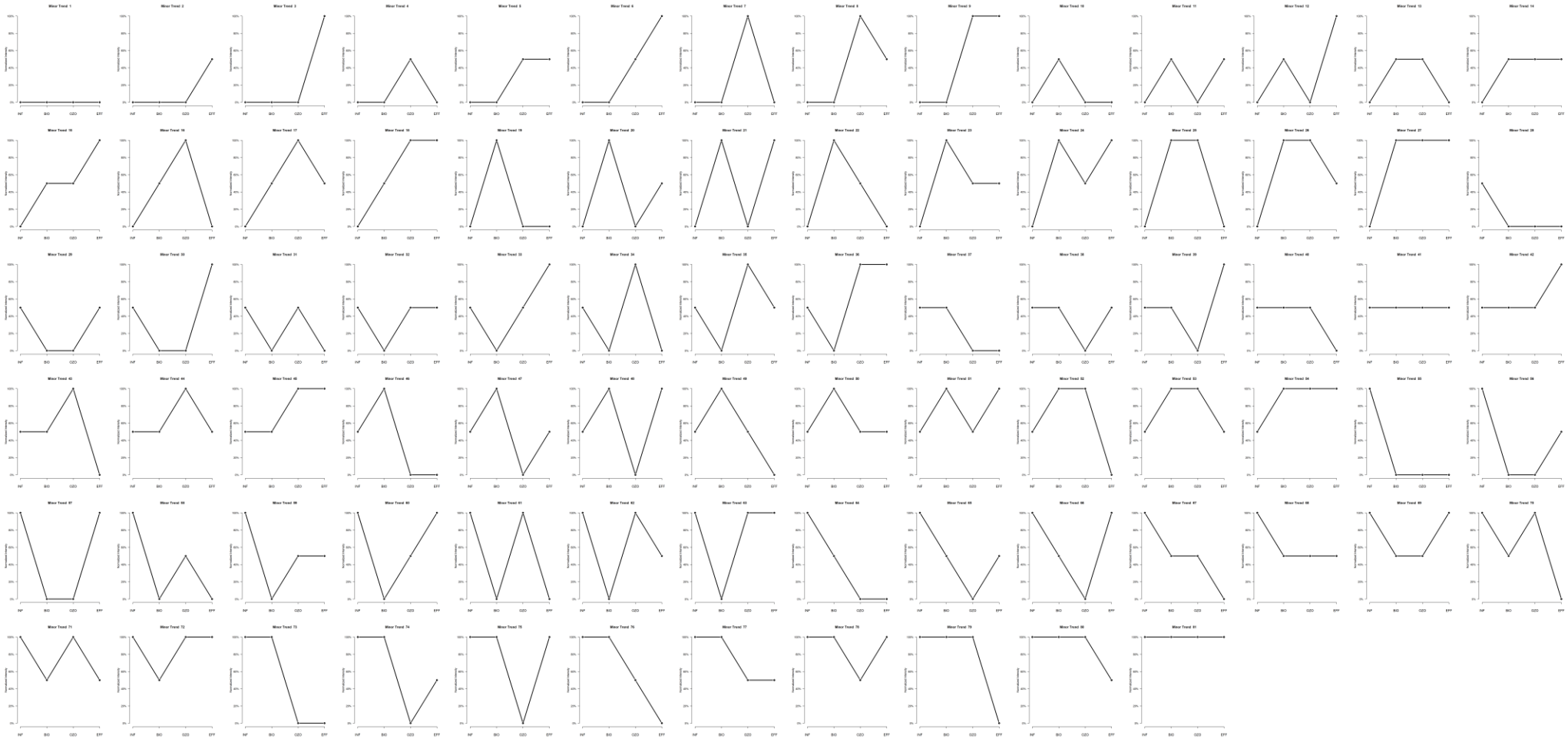


Figure S 12. All 81 possible minor trends defined for the profile barcoding. Of these, 16 were not present in the profile barcoding analysis because none of the four point was equal to 1, which was a requirement due to the normalization procedure applied on each intensity profile.

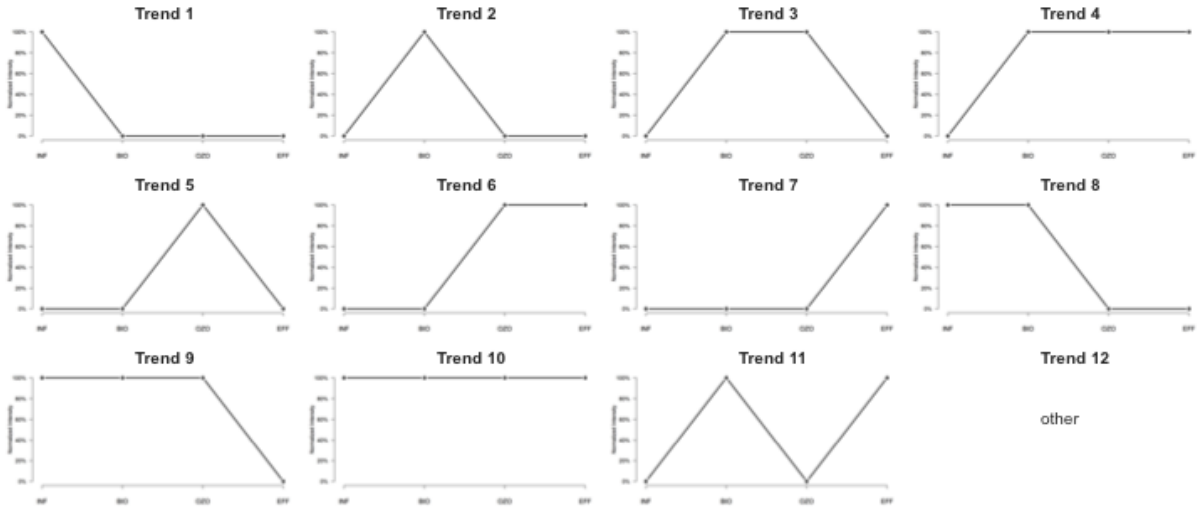


Figure S 13. Defined major trends for pattern recognition. On the x-axis are the four sampling location at each WWTP; INF – influent, BIO – after biological treatment, OZO – after ozonation, EFF – after post-treatment with activated carbon, in the effluent. On the y-axis is the normalized profile intensity. Profiles are normalized to the maximum profile intensity. Each of the major trend correspond to an expected trend in wastewater treatment.

Each of the 12 major trends defined corresponded to an expected trend in wastewater treatment, as detailed in the following. The description of the trend generally included (1) in which treatment step a feature originated and (2) in which treatment step a feature was removed >80%. Trend 1 was 'influent feature, removed in biological treatment'. Trend 2 was 'biological transformation product, removed in ozonation'. Trend 3 was ' biological transformation product, removed in post-treatment'. Trend 4 was 'stable biological transformation product'. Trend 5 was 'ozonation transformation product, removed in post-treatment', while Trend 6 was 'stable ozonation transformation product'. Trend 7 was 'transformation product from post-treatment'. Trend 8 was 'influent feature, removed in ozonation' and Trend 9 was 'influent feature, removed in post-treatment'. Trend 10 was for 'persistent influent features', while Trend 11 was for 'biological TPs formed in both biological treatment and post-treatment'. Trend 12 was for trends that did not correspond easily to one of the 12 defined major trends. With these defined major trends, comparisons and conclusions could be made based on the likely origin of the feature and in which treatment step >80% removal was achieved (if any).

There are of course many cases where a non-target feature is partically removed in one treatment step and then further removed in a subsequent step. A number of micropollutants for example are partially removed during biological treatment (by <80%) and then further removed during ozonation. The minor trend associated with this pattern was binned in with the 'removed in ozonation' major trend (Trend 8) because ultimately, >80% removal was only reached when ozonation was applied. Depending on the study question, associated minor with major trends can be adjusted, or eliminated altogether. A complete list of minor trends and the associated major trends in is Figure S 5

Table S 5. List of minor trends and the associated major trend (available in separate Excel file).

The automated trend assignment method was validated with a set of target compounds quantified in the samples (GL: McArdell et al. 2020, Oltramare et al. in prep; AR: Bogler 2019; PR: Krahnstöver et al. 2018). Target compounds were amisulpride, benzotriazol, candesartan, carbamazepine, citalopram, diclofenac, metoprolol, sulfamethoxazole, and venlafaxine and belong to the set of 12 Swiss "indicator substances" used to measure the abatement of micropollutants in advanced wastewater treatment (Götz et al. 2015). Trends were assigned to the quantified target compounds in the same manner as for the nontarget features, using concentrations in this case instead of feature intensities. First, for each measurement date and compound, concentrations were normalized to the maximum concentration. The normalized concentrations were then assigned to either high (100-60%), middle (60-20%) or low (20-

0%) domain, which was then converted to the associated minor and finally, major trend. Overall 8-9 target compounds were quantified on 25 sampling days, providing 206 trend assignments for comparison. From these 206 trend assignments, 172 were correct between the two measurement types. On average 84±17% trends were correct per sampling day.

Some of the discrepancies are maybe also due to measurement differences and are not a reflection on the trend assignment. For example, the data in Bogler 2019 were quantified with the exact same RAW (measurement) files and the overlap was 96%±7% (only 2/48 trends are different). In Glarnerland (Oltramare et al. in prep) the exact same sample bottles were used, but were measured with different analytical methods; here the overlap was 82±11% (10/54 measurements were different). Finally in PR, where both the samples and the methods were different, the overlap was 79±19% (22/104 were different).

Table S 6. Validation of automated trend assignment with quantified target compounds. (available in separate Excel)

Qualtitative Target and Suspect Screening

Annotation of known and/or suspected compounds was done with (1) a list of 427 organic micropollutants and (2) a list of 999 known ozonation transformation products. For the first, nontarget features were annotated as targets by comparison to expected exact mass (±5 ppm) and measured retention time (±30 seconds) of the respective reference standard. This method is considered a qualtitative target screening because comparison to reference standards was done for exact mass and retention time but not for MSMS fragmentation. 269 target compounds were detected in at least one sample (data not shown). A subset of these compounds was then selected to compare micropollutant elimination across the diverse sample set and different treatment setting. The detected target compounds were first ranked in order of highest mean intensity and the top 100 most intense were selected. Next, only targets previously detected in Swiss wastewater (Bourgin et al. 2018) were retained and finally, only target compounds detected on all measurement days were kept. Based on these filter criteria, a list of 66 wastewater relevant micropollutants (MP66 list) was generated.

Table S 7. List of 66 wastewater relevant organic micropollutants detected with qualitative target screening (MP66 List) (available in separate Excel)

In addition to the qualitative target screening, a suspect screening was conducted for known OTPs. A list of 999 known OTPs was compiled from the literature, including both laboratory experiments in wastewater matrix and TPs detected in wastewater. In total, OTPs from 84 parent organic micropollutants were included and OTPs covered wide *m/z* (86.0598-764.4791) and RT (5.1-24.7) ranges. The complete list of literature sources used is in the following and the list of 999 OTPs with names and exact mass are in Table S 8.

Literature sources: Abellán et al. 2008, Acero et al. 2000, Badawy et al. 2011, Barron et al. 2006, Benitez et al. 2015, Benner and Ternes 2009, Bianchini et al. 2011, Bollmann et al. 2016, Borowska et al. 2016, Boule et al. 2002, Calza et al. 2011, Calza et al. 2013, Chen et al. 2012, Christophoridis et al. 2016, Coelho et al. 2009, Dantas et al. 2011, Dantas et al. 2007, Diehle et al. 2019, Dodd et al. 2010, Favier et al. 2015, Feng et al. 2008, Gómez-Ramos et al. 2011, Gulde et al. submitted, Hörsing et al. 2012, Hübner et al. 2014, Keen et al. 2014, Kuang et al. 2013, Lajeunesse et al. 2013, Lange et al. 2006 Lester et al. 2013, Madhavan et al. 2010, Marotta et al. 2013, Mawhinney et al. 2012, McDowell et al. 2005, Mehta et al. 2010, Miao et al. 2015, Müller et al. 2012, Radjenović et al. 2009, Rodayan et al. 2010, Salgado et al. 2013, Šojić et al. 2012, Szabó et al. 2011, Tay et al. 2011, Tay et al. 2012, Topalov et al. 2000, Vogna et al. 2004, Zimmermann et al. 2012.

Table S 8. Suspect list of 999 known ozonation transformation products (OTPs) compiled from literature sources (available in separate Excel).

Linkage analysis

Table S 9. Summary of transformation reactions considered during linkage analysis. Listed are the reaction type, reaction abbreviation, mass difference between parent and transformation product, formula change from parent to transformation product, expected oxidant during ozonation and example reactions from literature. Adapted from Schollée *et al*. 2018.

	Reaction type	Mass difference	Formula change	Expected oxidant	Example reaction / source of reaction
1[a]	Addition of 3 oxygens	47.9847	+3O	reactions with O3 and OH	benzoquinone plus OH at multiple sites
2[a]	Oxygen addition	31.9898	+2O	reactions with O3	hydroquinone to benzoquinone; also benzotriazole to TP[1]
3[a]	Methyl to carboxylic acid	29.9741	+O2-H2	reactions with O3	second generation product, eg oxicooh plus oh; also methyl to carboxylic acid[2]; also from ozone reactions with aniline[3]
4[a]	Hydration	18.0106	+H2O	reactions with O3	hydrochlorothiazide TP1 to TP3[4]
5[a]	Hydroxylation, N/S-Oxidation, Epoxidation	15.9949	+O	reactions with O3 and OH	n-oxide formation; hydroxylation at heterocyclic aromatic[5]
6[a]	Oxidative deamination	14.9632	+O2-NH3	reactions with O3 and OH	deamination to ketone, followed by oxidation
7[a]	Alcohol to carboxylic acid or primary amine to nitro	13.9792	+O-H2	reactions with O3 and OH	multiple atrazine reactions[6]
8[a]	Hydrogenation	2.0157	+H2	does not occur	
9[a]	Oxidative displacement of amine	1.9918	+OH-NH2	reactions with O3 or OH	O3: chlorothiazide oxidation on sulfate group[4]; OH: amide to carboxylic acid[7]
10[a]	Deamination to ketone	-1.0317	+O-NH3	does not occur	
11[a]	Oxidative displacement of fluorine	-1.9957	+OH-F	reactions with OH	ipso attach by OH; detected for 5 fluoroquinolones[8]
12[a]	Dehydrogenation	-2.0157	-2H	reactions with O3	multiple atrazine reactions[6]
13[a]	Demethylation	-14.0157	-CH2	reactions with O3	diuron to DCPMU[9]
14[a]	Deamination	-15.0109	-NH	only during extensive ozonation	
15[a]	Oxidative displacement of chlorine	-17.9662	+OH-Cl	reactions with OH	diuron substitution at the ring[9]; OH: clobifric acid and bezafibrate[10]
16[a]	Dehydration	-18.0106	-H2O	spontaneous reaction	diclofenac cyclization (unclear if O3 or OH driven)[11]
17[a]	Di-demethy or Deethylation	-28.0313	-C2H4	reactions with OH	des-ethyl-atrazine[6]
18[a]	Dealkylation	-30.0470	-C2H6	reactions with O3 and OH	
19[a]	Reductive displacement of chlorine	-33.9611	+H-Cl	reactions with OH only	clofibric acid TP[10]
20[a]	Descyclopropyl	-40.0313	-C3H4	reactions with O3	Dealkylation of cyclopropyl group[12]

21[a]	De-acetylation	-42.0106	-C2H2O	reactions with OH	multiple steps involving ester and quinone (ciprofloxacin OTPs)[12]
22[a]	Deisopropyl	-42.0470	-C3H6	reactions with OH and O3	multiple atrazine reactions[6]
23[a]	Decarboxylation	-43.9898	-CO2	reactions with OH	[13]
24	Addition of 4 oxygens	63.9796	+O4	reactions with O3	top 20 predicted[14]
25	Addition of 5 oxygens	79.9745	+O5	reactions with O3	top 20 predicted[14]
26	multiple reactions	45.9691	+O3-H2	reactions with O3	top 20 predicted[14]
27	multiple reactions	61.9640	+O4-H2	reactions with O3	top 20 predicted[14]
28	multiple reactions	62.9718	+O4-H	reactions with O3	top 20 predicted[14]
29	multiple reactions	44.9613	+O3-H3	reactions with O3	top 20 predicted[14]
30	multiple reactions	46.9769	+O3-H	reactions with O3	top 20 predicted[14]
31	multiple reactions	33.9691	+O3-CH2	reactions with O3	top 20 predicted[14]
32	multiple reactions	95.9695	+O6	reactions with O3	top 20 predicted[14]
33	multiple reactions	78.9668	+O5-H	reactions with O3	top 20 predicted[14]
34	multiple reactions	81.9902	+O5H2	reactions with O3	top 20 predicted[14]
35	multiple reactions	65.9953	+O4H2	reactions with O3	top 20 predicted[14]
36	multiple reactions	77.9589	+O5-H2	reactions with O3	top 20 predicted[14]
37	multiple reactions	51.9797	+O4-C	reactions with O3	top 20 predicted[14]
38	multiple reactions	50.9718	+O4-CH	reactions with O3	top 20 predicted[14]
39	multiple reactions	28.9664	+O2-H3	reactions with O3	top 20 predicted[14]
40	multiple reactions	-11.0160	+O-CHN	reactions with O3	top 20 predicted[14]
41	multiple reactions	17.9742	+O2-CH2	reactions with O3	top 20 predicted[14]
42	multiple reactions	-10.0207	+O-C2H2	reactions with OH and O3	Reaction reported in the literature (see footnotes)
43	multiple reactions	5.9742	+O2-C2H2	reactions with OH and O3	Reaction reported in the literature (see footnotes)
44	multiple reactions	-81.0215	-C4H3NO	reactions with OH and O3	Reaction reported in the literature (see footnotes)
45	multiple reactions	-95.0371	-C5H5NO	reactions with OH and O3	Reaction reported in the literature (see footnotes)

[a] Reaction previously included in Schollée et al. (2018)

[1] Mawhinney et al. 2012
[2] Müller et al. 2012
[3] von Sonntag and von Gunten 2012
[4] Borowska et al. 2016
[5] Tekle-Röttering et al. 2016
[6] Acero et al. 2000
[7] Song et al. 2008
[8] Santoke et al. 2009
[9] Mestankova et al. 2011
[10] Razavi et al. 2009
[11] Coelho et al. 2009
[12] DeWitte et al. 2008
[13] Andreozzi et al. 2003
[14] Schollée et al. in prep

Section S4: Non-target characterization

The final data set included 50,722 non-target features in 293 samples (including blanks, pooled samples, and calibration samples). A principal component analysis (PCA) was used to visualize variance and patterns in the dataset (Figure S 15; excluding blanks and standards). The first principal component (PC1) explained 19.2% of the variance among the samples and is related to the differences among the influent matrices. The second dimension explained 7.7% of the variance and is related to the target compounds, seen in that the spiked pooled samples have higher loadings in this dimension compared to the unspiked pooled samples and the WWTP samples. In higher PCs, the dominant sources of variation are from the changes in influent composition and, to a minor degree, differences in biological and ozonation sample composition. The effluent samples remain clustered together near the center of the PCA in PC1-5, indicating their increased similarity in composition compared to the wastewater collected at the other steps in the treatment train. If unspiked and spiked pooled samples are removed from the PCA (Figure S 16), PC1 (13.7% of variance) and PC2 (7.1% of variance) are both related to differences among the influent samples, with the samples from the first sampling campaign in AR appearing to be most different, while a tight clustering of the other sample types can be observed.

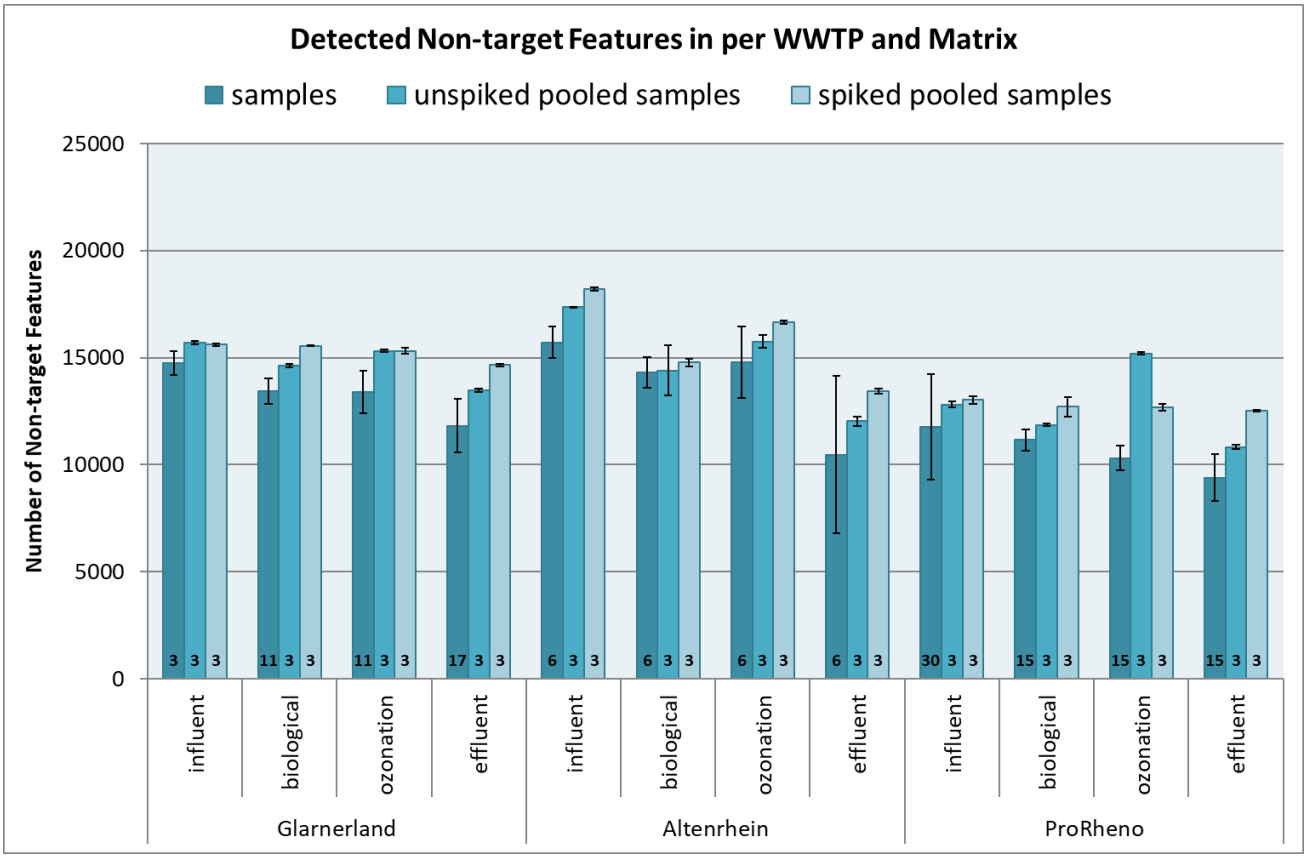


Figure S 14. The number of features detected in each wastewater treatment plant (WWTP) and each matrix, in the individual samples as well as in the two sets of pooled samples. Unspiked pooled samples were a mix of samples from one matrix (*i.e.*, influent, biological, ozonation, or effluent) in one WWTP (*i.e.*, Altenrhein, Glarnerland, ProRheno). Spiked pooled samples were the same mix as in the unspiked samples but spiked with a set of target compounds (1000 ng/L in influent samples, 250 ng/L in all other matrix types). Shown at the base of each bar are the number of samples measured in the respective category.

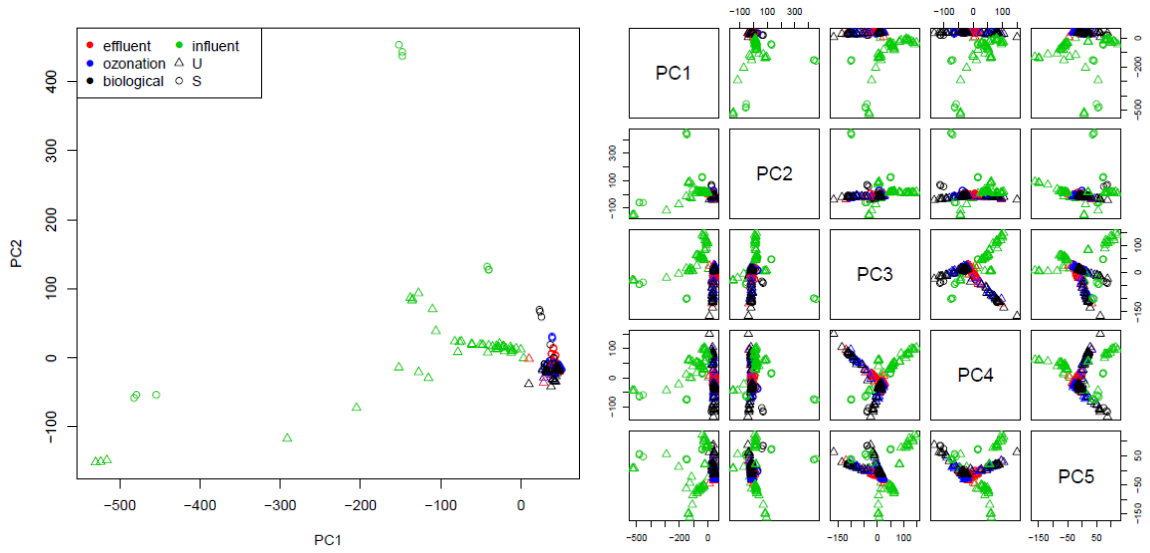


Figure S 15. Principal component analysis (PCA) of all wastewater samples, including spiked and unspiked pooled samples. In (a) first principal component (PC1) vs. second principal component (PC2). PC1 explains 19.2% of variance and PC2 explained 7.7%. In (b) PC1–PC5 are compared in a pairs plot matrix. U: unspiked pooled samples, S: spiked pooled samples

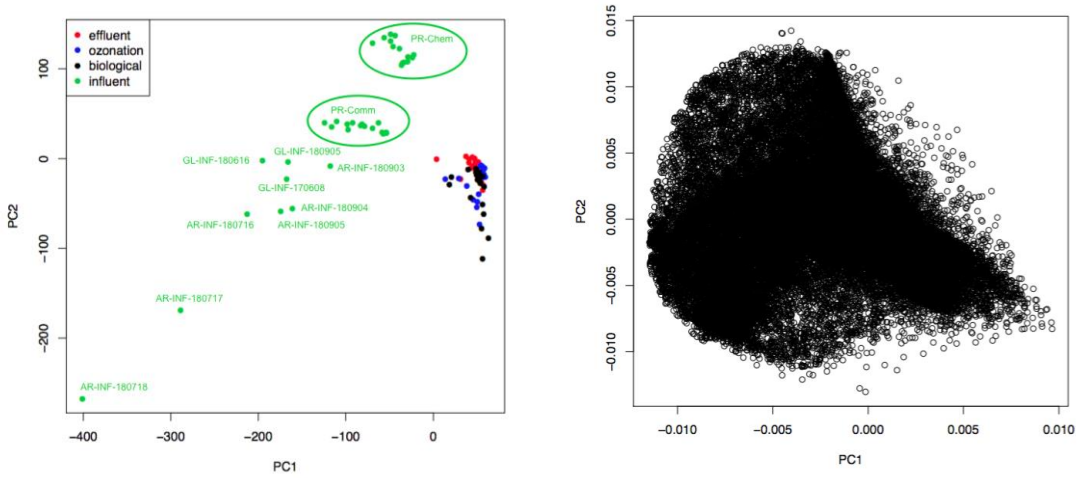


Figure S 16. Principal component analysis (PCA) of only wastewater samples (*i.e.*, spiked and unspiked pooled samples removed). In (a) scores plot of first principal component (PC1) vs. second principal component (PC2). In (b) the loading plot of PC1 vs. PC2. PC1 (13.7% of variance) and PC2 (7.1% of variance)

General Characterization of Influent Samples

In GL, AR, and PR-Comm, the pluralities of non-target features (*i.e.*, 39% (n_{GL}=3), 28% (n_{AR}=5), and 24% ($n_{PR\text{-}Comm}$=15), respectively) were detected on all sampling dates in the respective WWTP. In contrast, only 13% of non-target features were detected on all sampling dates for PR-Chem ($n_{PR\text{-}Chem}$=14). In PR-Chem, the largest portion of features (21%) were detected on only 1 sampling date, reinforcing the notion that inputs from industry are generally short and sporadic. The percentages of non-target features unique to only 1 sampling date were lower in GL (35%, n_{GL}=3), AL (19%, n_{AR}=5), and PR-Comm (15%, $n_{PR\text{-}Comm}$=15) and are generally in line with the expected amount of industrial inputs at each location (GL: 40%; AL: 26%; PR-Comm: 20-25%). Although this method estimates the possible industrial discharges, it only considers presence/absence of a compound. Finally, influent features were assigned a class, based on where they were detected (visualized as Venn, SI, Figure S12a).

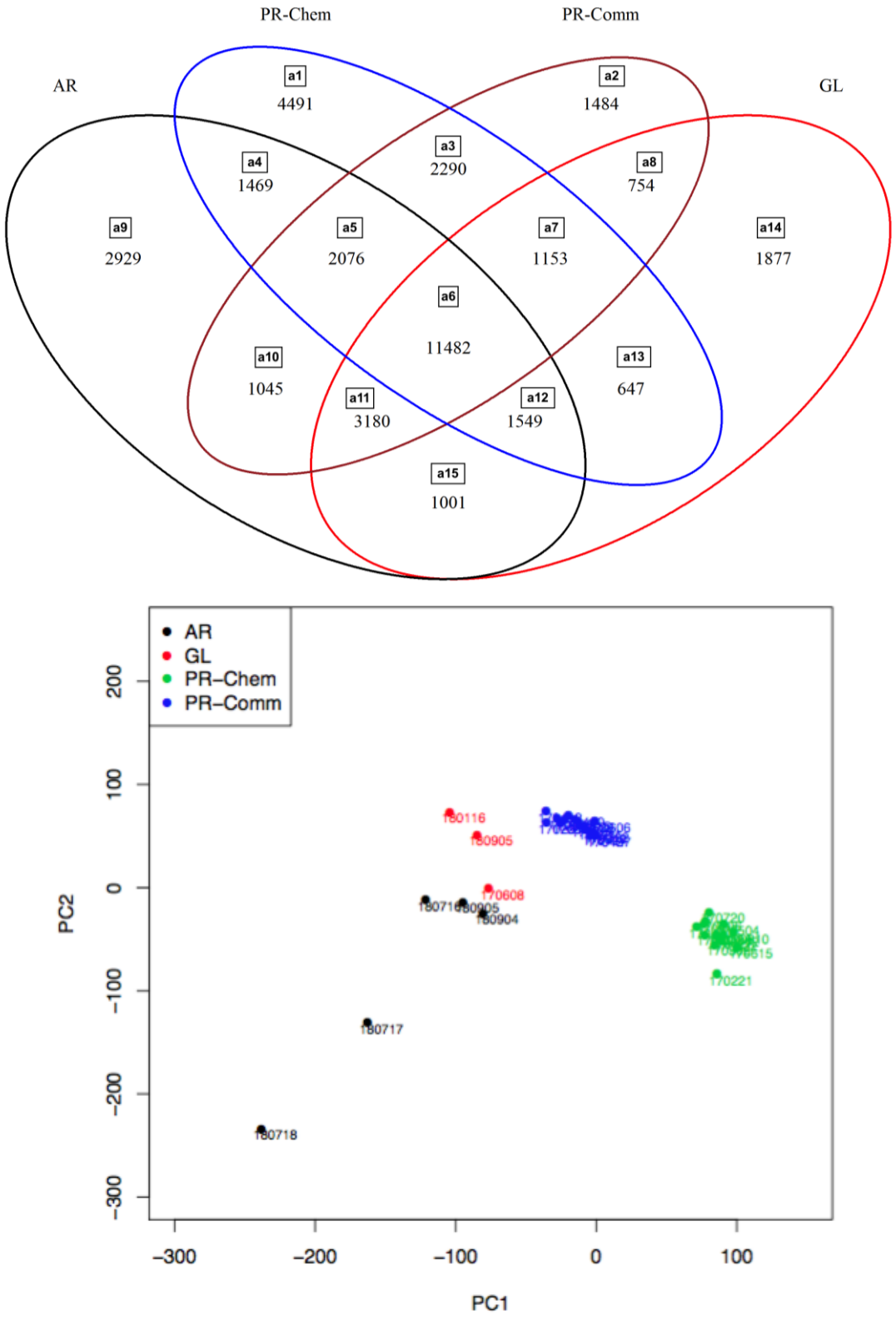


Figure S 17. Characterization of influent samples. In (top), a 4-group Venn diagram to visualize the presence of non-target features in the influent samples of different inputs. Assigned class are indicated in a box in each Venn quadrant. In (bottom), the first principal component (PC1) vs. second principal component (PC2). Influents of the different wastewater treatment plants are color-coded and labeled with the sampling date. Black: AR – Altenrhein; red: GL – Glarnerland; green: PR-Chem – ProRheno industrial wastewater; blue: PR-Comm – ProRheno domestic wastewater.

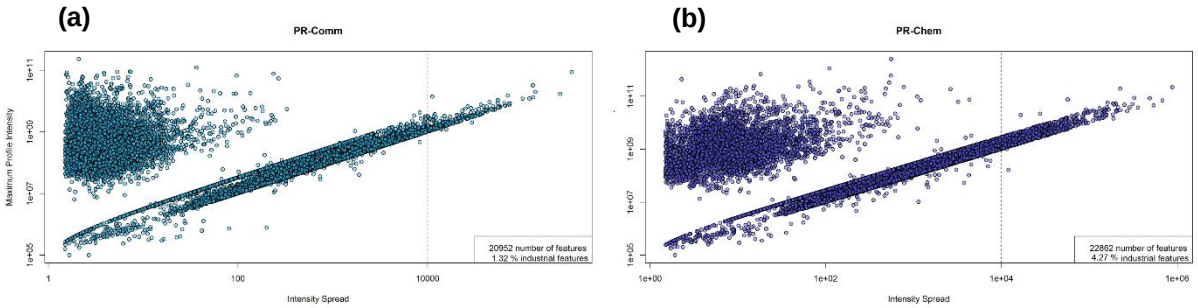


Figure S 18. Characterization of influent non-target features as possibly originating from industrial sources for (a) ProRheno municipal wastewater and (b) ProRheno domestic wastewater. On the x-axis is the intensity spread for each profile, defined as the ratio of the 95th and 5th percentiles, and on the y-axis is the maximum intensity for each profile. Non-target features were classified as industrial if the intensity spread is >1E4; cutoff is indicated with a vertical dotted grey line.

Table S 10. Results of identification of likely nontarget features originating from industrial sources, including measured accurate mass (m/z), retention time, intensity spread, msPurity, compound name and compound class (available in separate Excel).

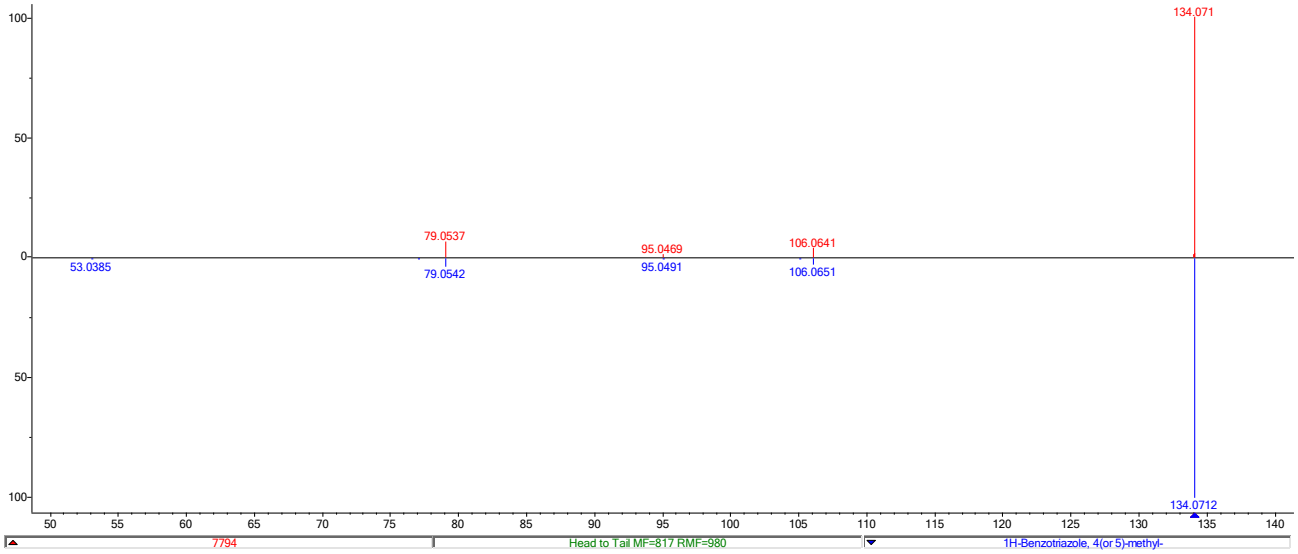
Table S 11. Removal of potential industrial non-target features detected in the two influent stream in PR. Potential industrial features were selected based on intensity spread over the sampling dates in each influent stream. The detected trend for each feature on the day of maximum intensity (*i.e.*, assumed to be the day of maximum discharge) is shown. In the first column is the absolute number of non-target features with this trend, while in the second column is the percentage.

	PR-Comm		PR-Chem	
Number of non-target features classified as potentially of industrial origin	274		977	
Number of tentatively identified non-targets	43		16	
Removed in biological treatment[1]	257	93.8%	813	83.2%
Removed in ozonation[2]	11	4.0%	88	9.0%
Removed in post-treatment[3]	2	0.7%	21	2.1%
Persistent[4]	0	0.0%	29	3.0%
Other trends (*e.g.*, some formation in other treatment steps) [5]	4	1.5%	26	2.7%

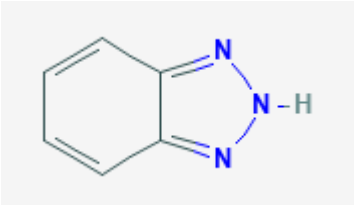
[1] Features with Trend 1; [2] Features with Trend 8; [3] Features with Trend 9; [4] Features with Trend 10; [5] Features with Trend 2, 3, 4, 7, or 11. No industrial features were detected with Trend 5 or 6. The visualization of these Trends can be found in Figure S 13.

In the following section, the evidence is provided for the identification of 54 non-target features found to be likely of industrial origin. For each idenitified non-target feature, first the profile number is provided for reference. Then the head to tail plot for spectral comparison is provided (obtained directly from NIST MSSearch), with *m/z* on the x-axis and relative fragment intensity on the y-axis. In red on the top is the measured MS2 spectrum of the non-target feature; in blue on the bottom is the library spectrum for the proposed structure. The Match Factor (MF) and Reverse Match Factor (RMF) calculated between the spectra by NIST is given in green below the figure. Further information is also provided about the identified non-target, including name of the proposed compound, molecular formula and PubChemID of the proposed compound, and the expected compound class. Also given is the influent stream in which this non-target feature was classified as likely of industrial origin, as well as fate of this non-target feature in wastewater treatment. Finally, on the right, the structure of the proposed compound is provided. All identifications are considered to be of confidence level 2(b), according to the scheme proposed by Schymanski et al. (2014).

Profile 7794

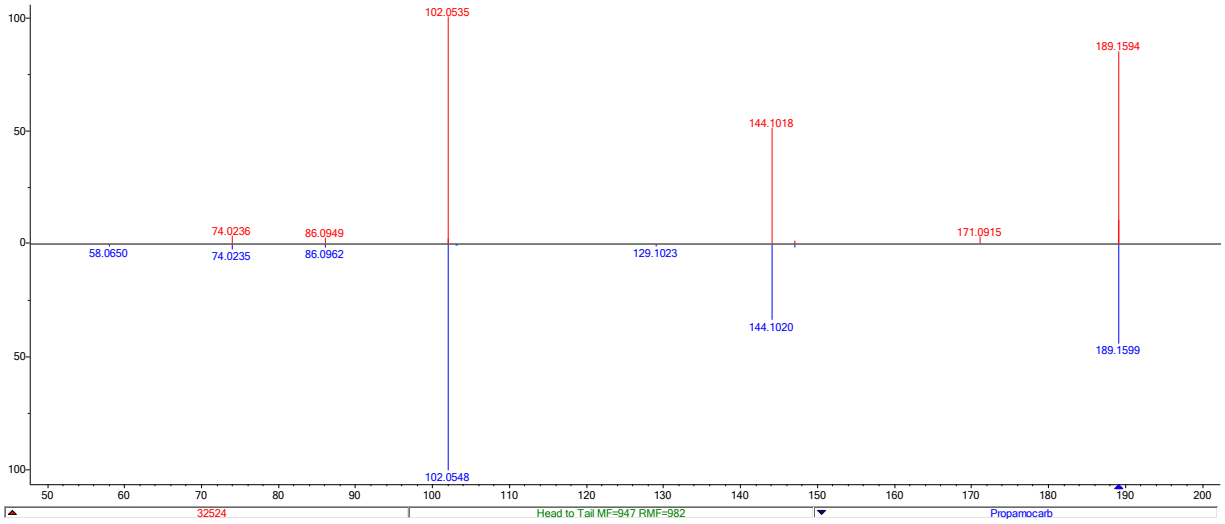


Name: 1H-Benzotriazole
Formula: C7H7N3
PubChemID: 7220
Compound class: corrosion inhibitor

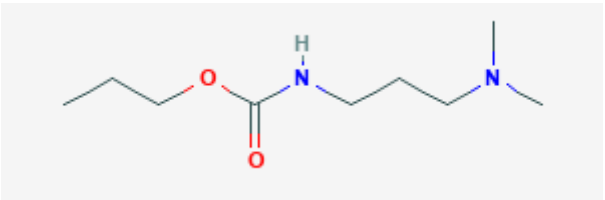


WWTP stream: PR-Chem
Fate: 100% average elimination in PR, mostly in BIO

Profile 32524

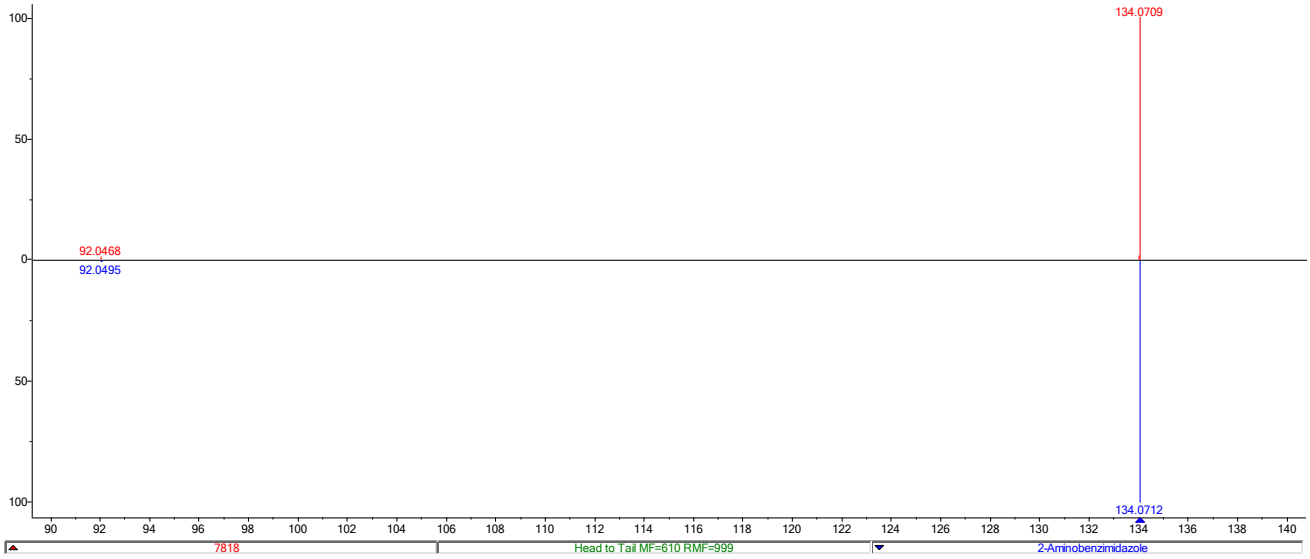


Name: Propamocarb
Formula: C9H20N2O2
PubChemID:32490
Compound class: fungicide

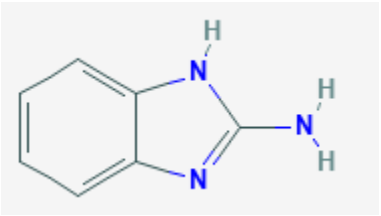


WWTP stream: PR-Chem
Fate: ND in almost all effluents in PR, average eliimation 94%, mainly in BIO

Profile 7818

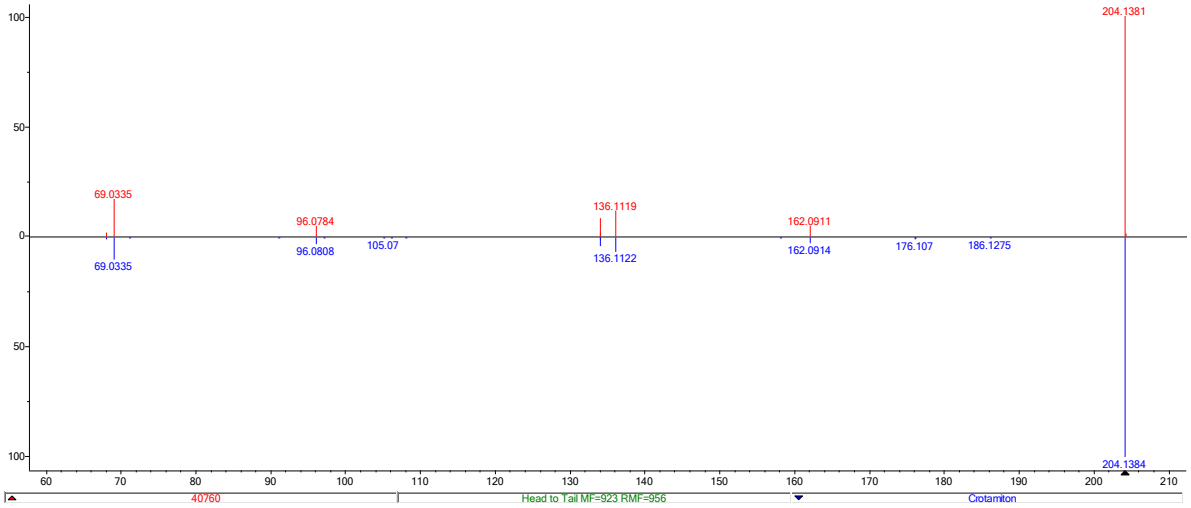


Name: 2-Aminobenzimidazole
Formula: C7H7N3
PubChemID: 13624
Compound class: multiple uses

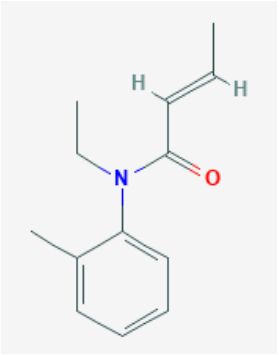


WWTP stream: AR
Fate: ND in all AL effluents. Average elimation was 100%, removed in BIO

Profile 40760

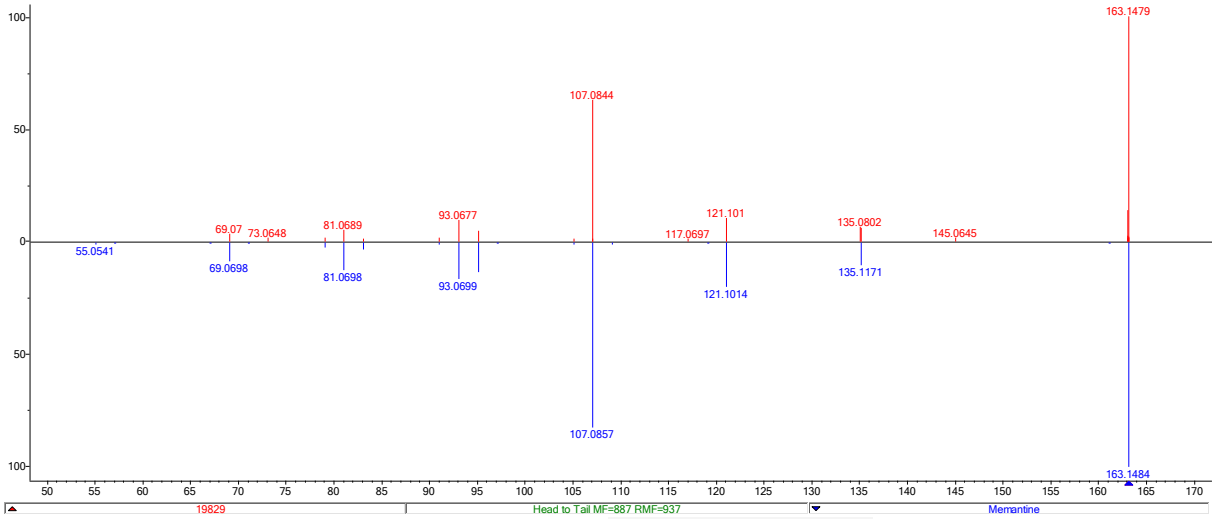


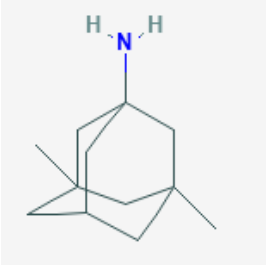
Name: Crotamiton
Formula: C13H17NO
PubChemID: 688020
Compound class: pharmaceutical



WWTP stream: PR-Chem
Fate: Overage removal in PR 92%, mostly removed in OZO

Profile 19829

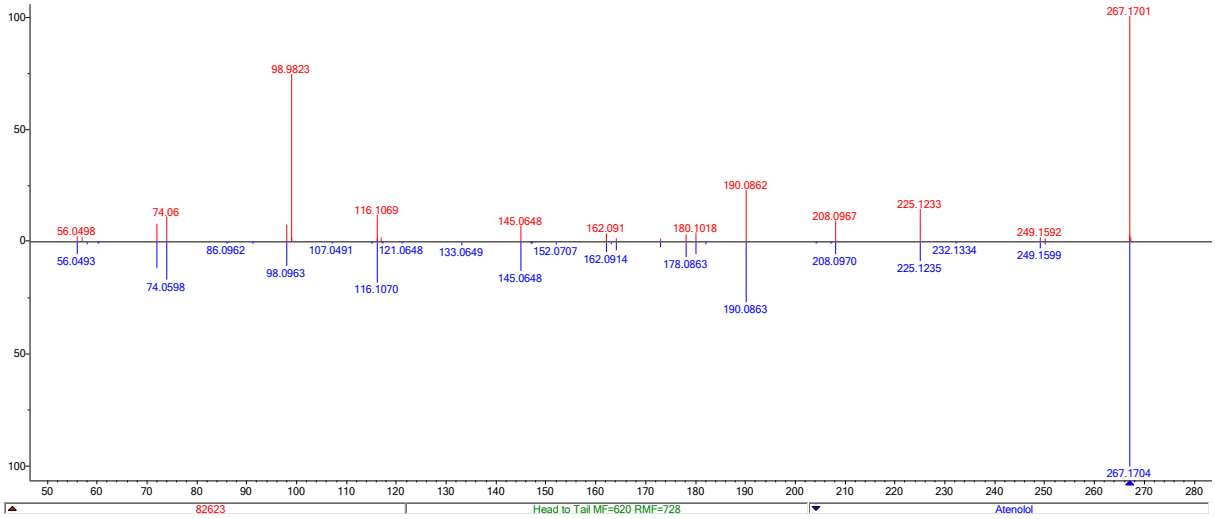


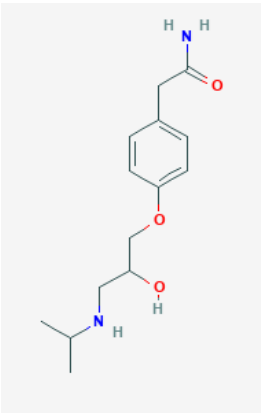


Name: Memantine [M+H-NH3]+
Formula: C12H21N
PubChemID: 4054
Compound class: pharmaceutical

WWTP stream: PR-Chem
Fate: Average removal in PR 85%, mostly in PAC

Profile 82623

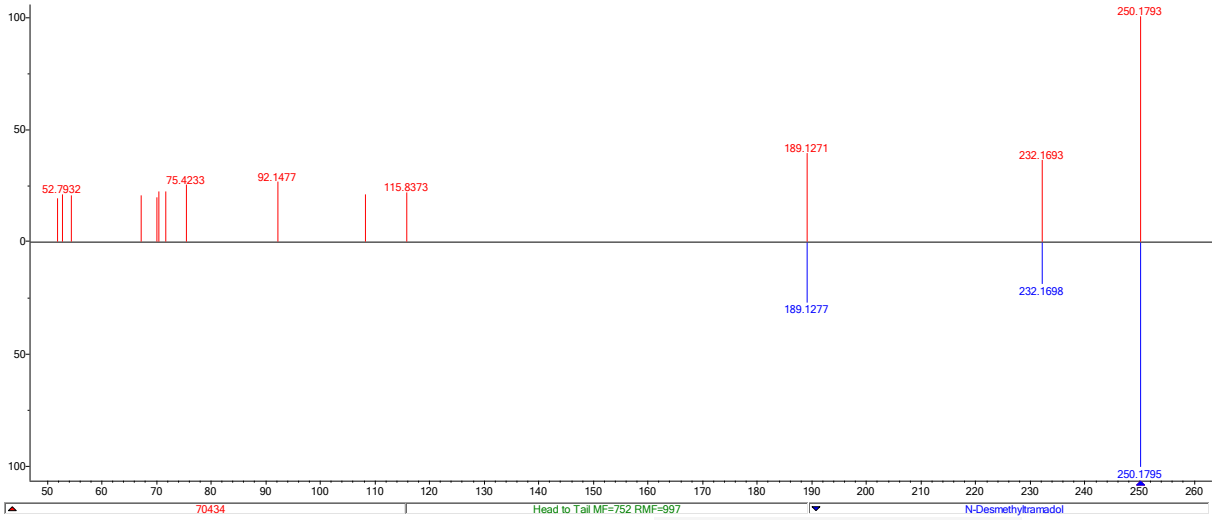




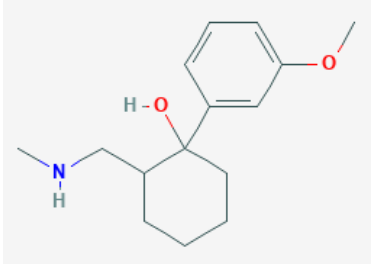
Name: Atenolol
Formula: C14H22N2O3
PubChemID: 2249
Compound class: pharmaceutical

WWTP stream: PR-Comm
Fate: 100% average elimination in PR, mostly in BIO

Profile 70434

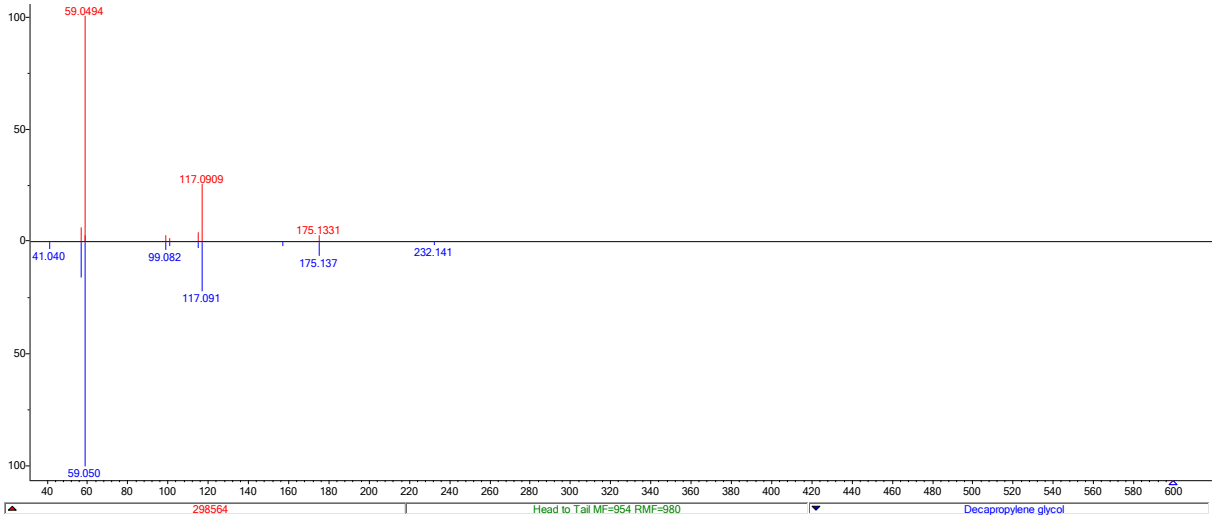


Name: N-Desmethyltramadol
Formula: C15H23NO2
PubChemID: 198555
Compound class: pharmaceutical transformation product

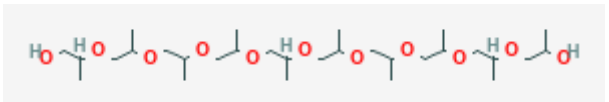


WWTP stream: PR-Chem
Fate: 92% average in PR, mostly in OZO

Profile 298564

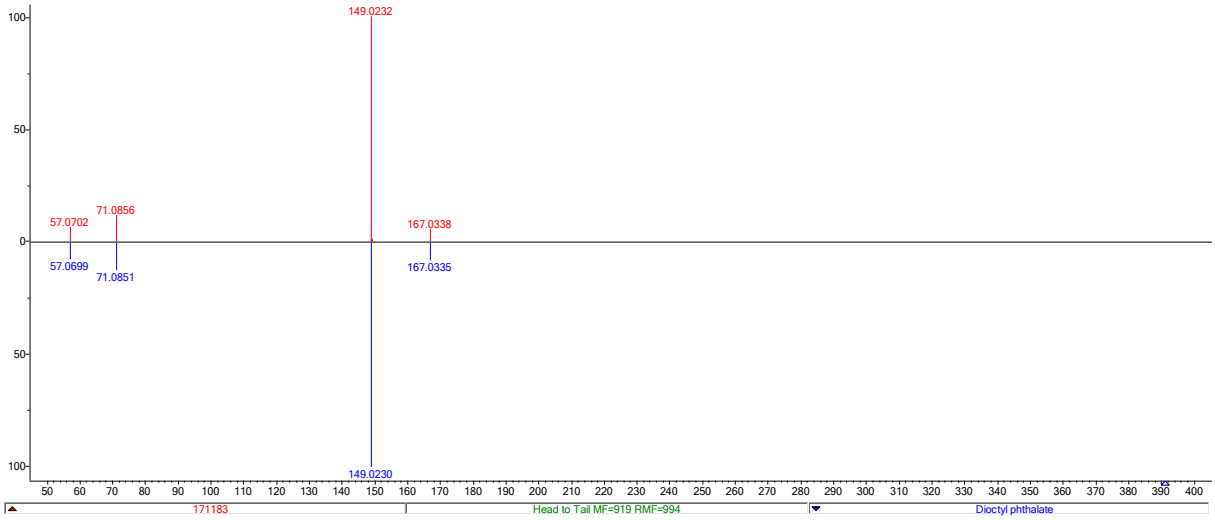


Name: Decapropylene glycol
Formula: C30H62O11
PubChemID: 87390959
Compound class: multiple uses

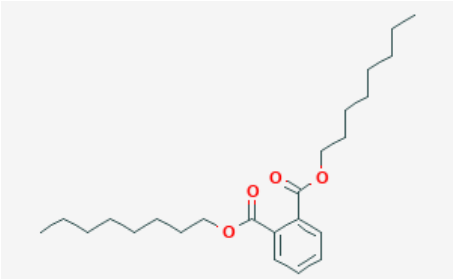


WWTP stream: AR
Fate: 100% average elimination in both AR and PR, mostly in BIO

Profile 171183

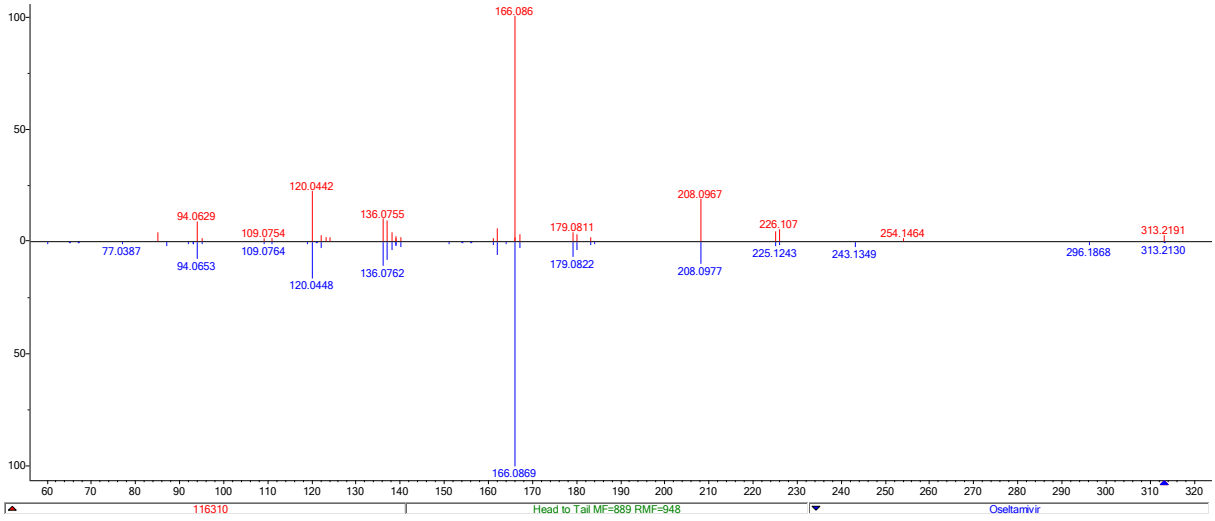


Name: Dioctyl phthalate
Formula: C24H38O4
PubChemID: 8346
Compound class: platicizer

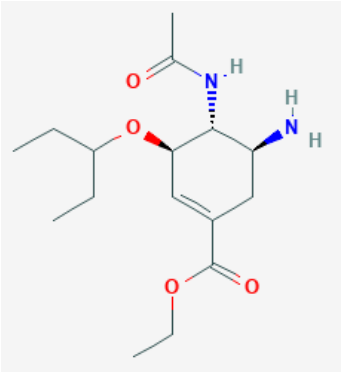


WWTP stream: AR
Fate: 65% average overall removal in AR, in BIO and OZO. In PR formation observed

Profile 116310

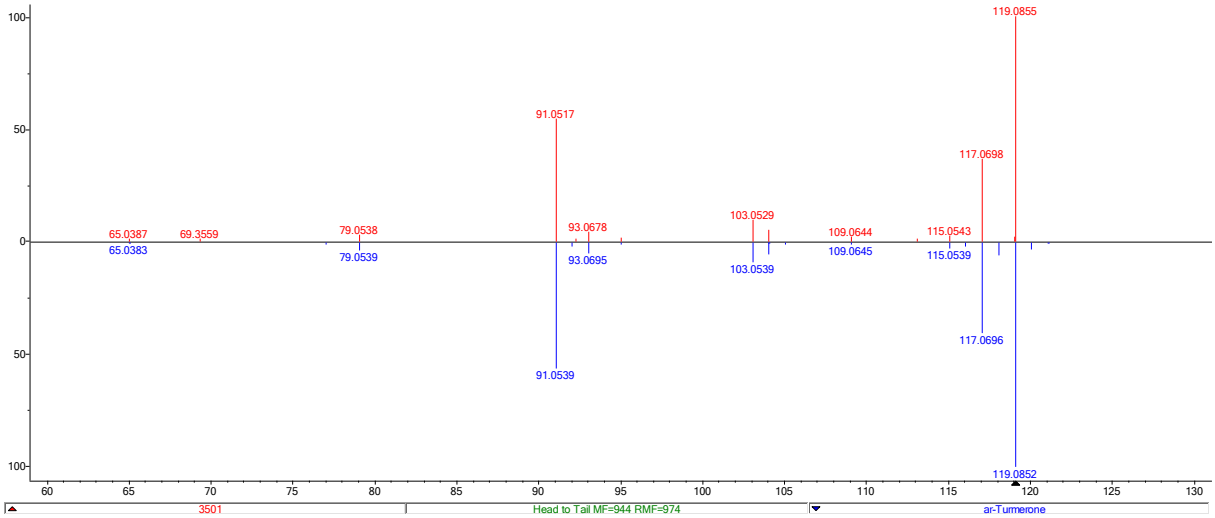


Name: Oseltamivir
Formula: C16H28N2O4
PubChemID: 65028
Compound class: pharmaceutical

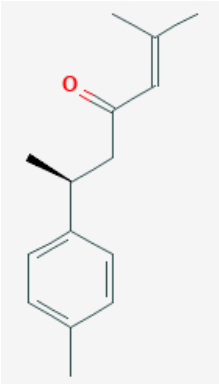


WWTP stream: PR-Comm
Fate: 96% average overall removal in PR, mostly in OZO but also in PAC

Profile 3501

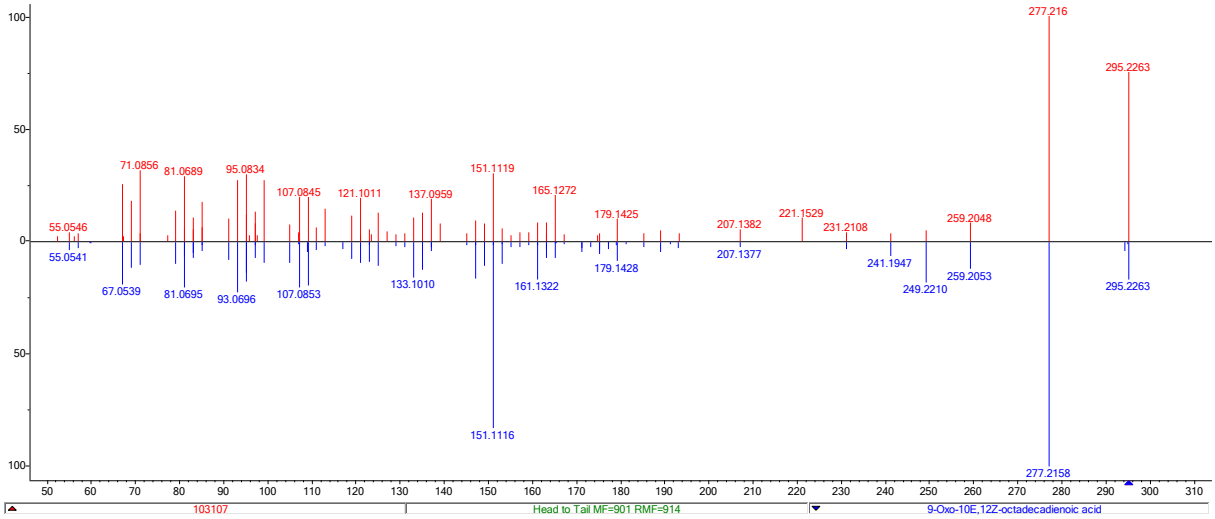


Name: ar_Tumerone
Formula: C15H20O
PubChemID: 160512
Compound class: Natural medicine

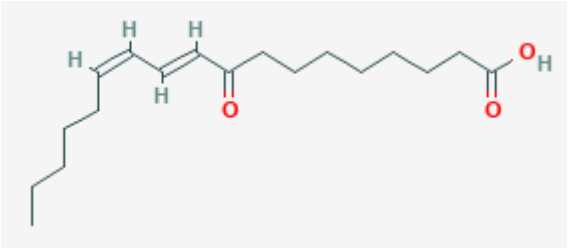


WWTP stream: PR-Comm
Fate: 100% average elimination in PR, in BIO

Profile 103107

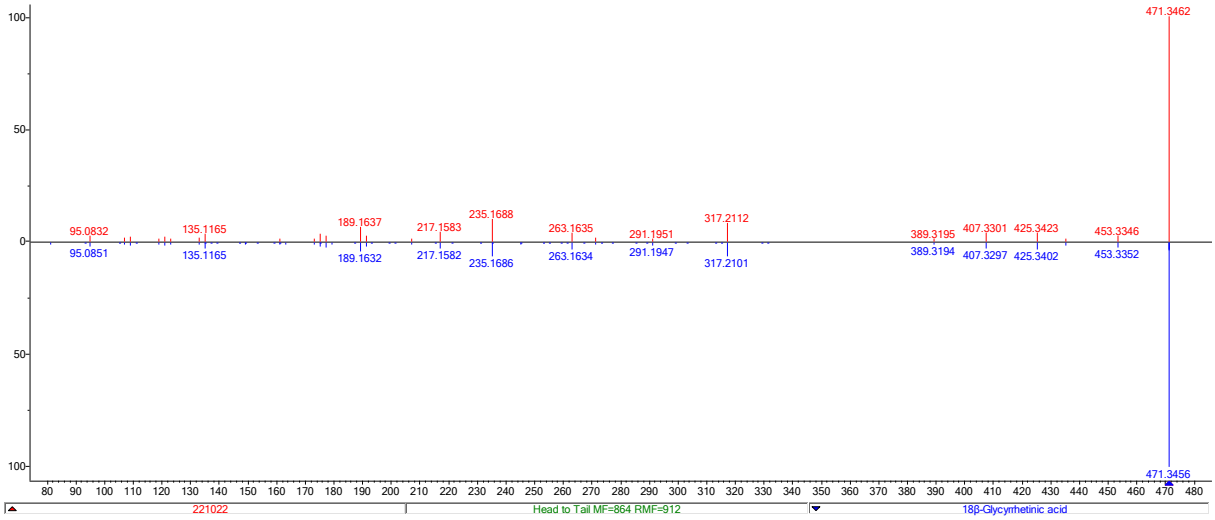


Name: 9-oxo-10E, 12Z-octadecadienoic acid
Formula: C18H30O3
PubChemID: 9839084
Compound class: fatty acid



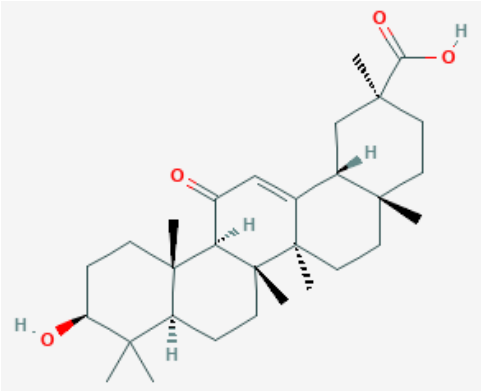
WWTP stream: AR
Fate: detected in effluents of PR. On highest emission day, 76% removal detected. On average compound eliminated 74% in PR, mainly in BIO

Profile 221022

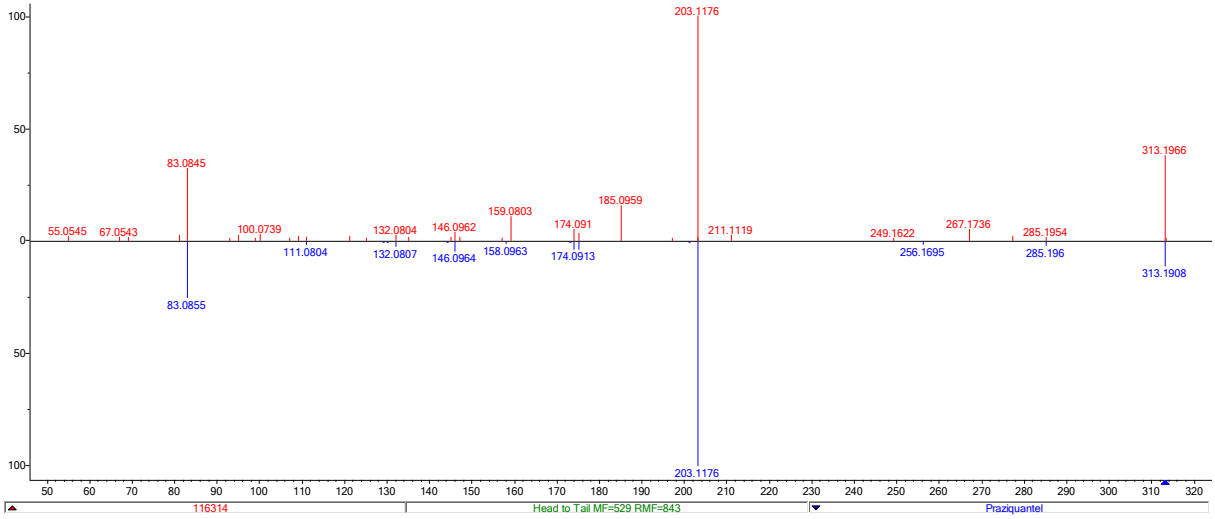


Name: 18β-Glycyrrhetinic acid / Enoxolone / glycyrrhetic acid
Formula: C30H46O4
PubChemID: 10114
Compound class: pharmaceutical / flavoring agent

WWTP stream: PR-Chem
Fate: 100% overall average elimination in PR, all in BIO

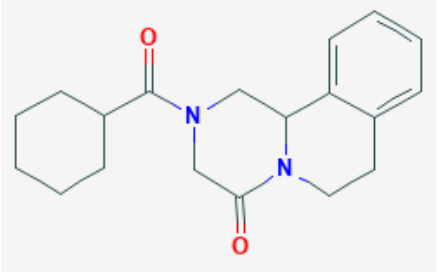


Profile 116314

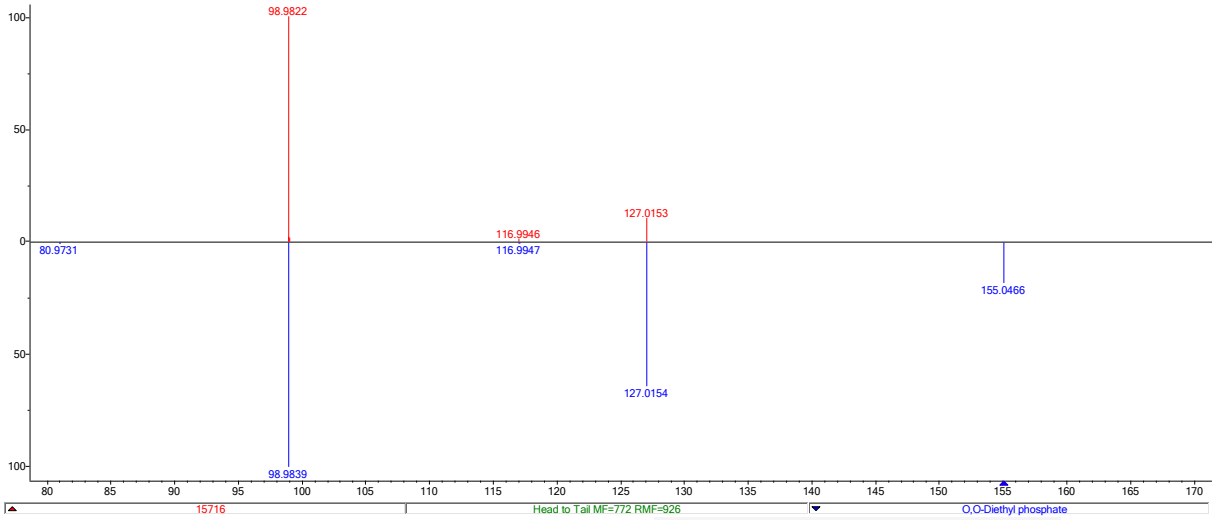


Name: Praziquantel
Formula: C19H24N2O2
PubChemID: 4891
Compound class: pharmaceutical

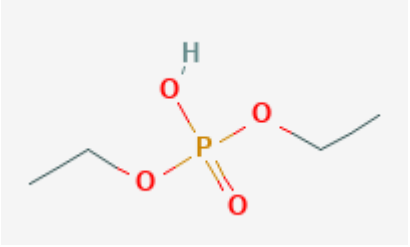
WWTP stream: PR-Chem
Fate: 100% average elimination in PR, some in BIO, most in OZO



Profile 15716

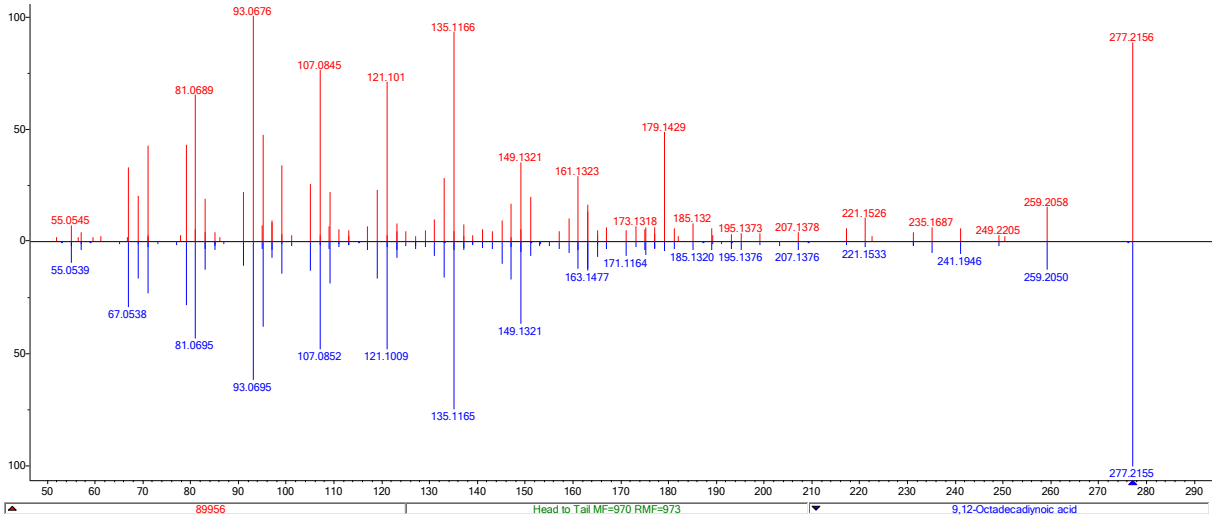


Name: O,O-Diethyl phosphate
Formula: C4H11O4P
PubChemID: 654
Compound class: organophosphate pesticides metabolite

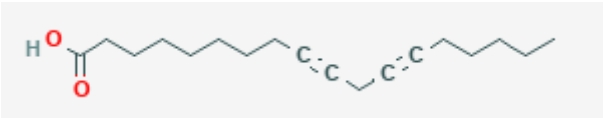


WWTP stream: PR-Chem
Fate: Persistant in PR, wide range of elimination/formation

Profile 89956

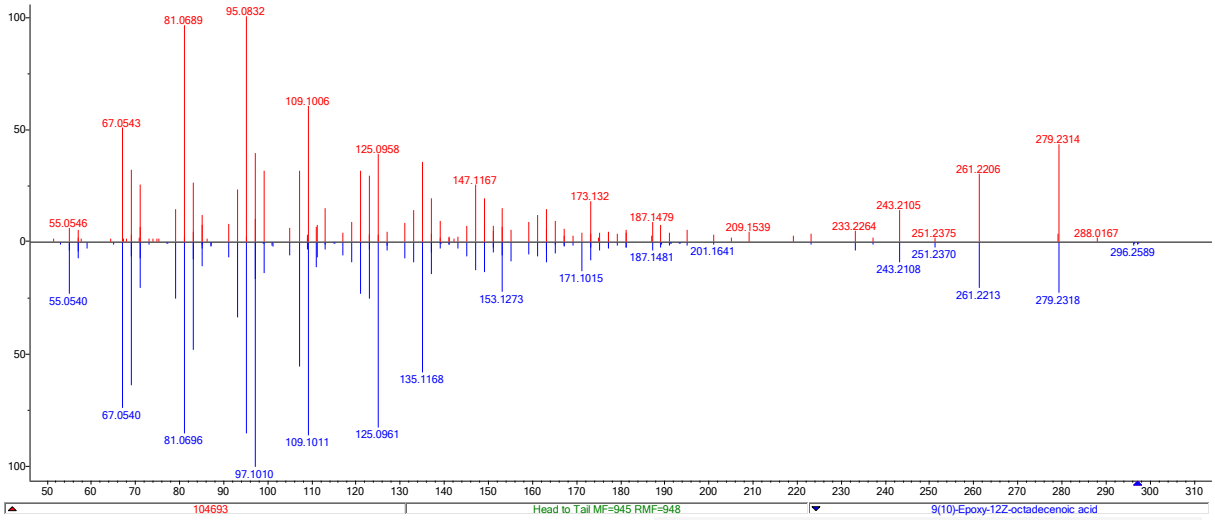


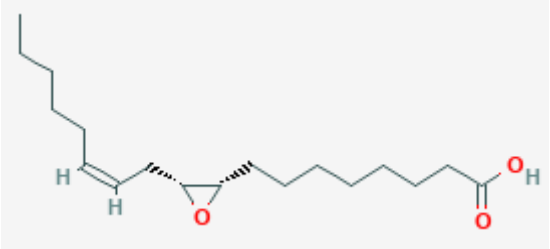
Name: 9,12-Octadecaiynoic acid
Formula: C18H28O2
PubChemID: 1931
Compound class: fatty acid



WWTP stream: PR-Chem
Fate: Average 71% elimination in PR, mostly in BIO

Profile 104693

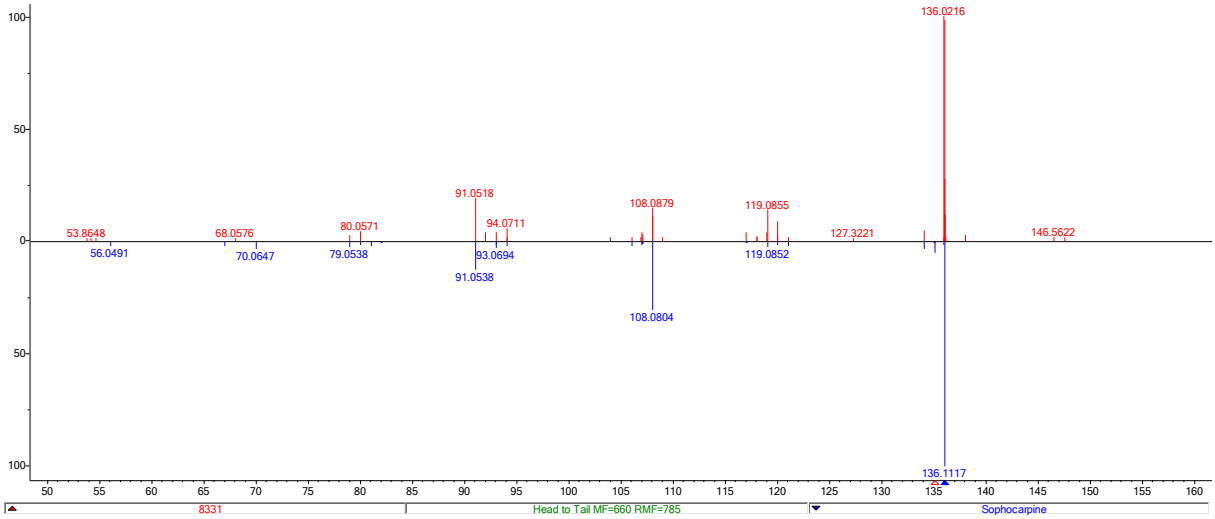


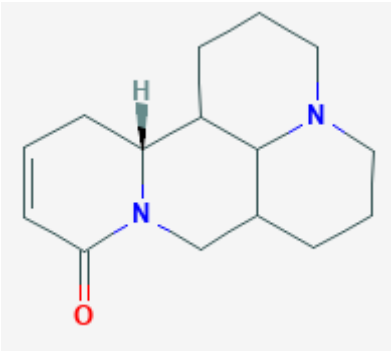


Name: 9(10)-Epoxy-12Z-octadecenoic acid / Coronaric acid
Formula: C18H32O3
PubChemID:12097313
Compound class: fatty acid

WWTP stream: PR-Comm
Fate: Average elimination 89% in PR

Profile 8331

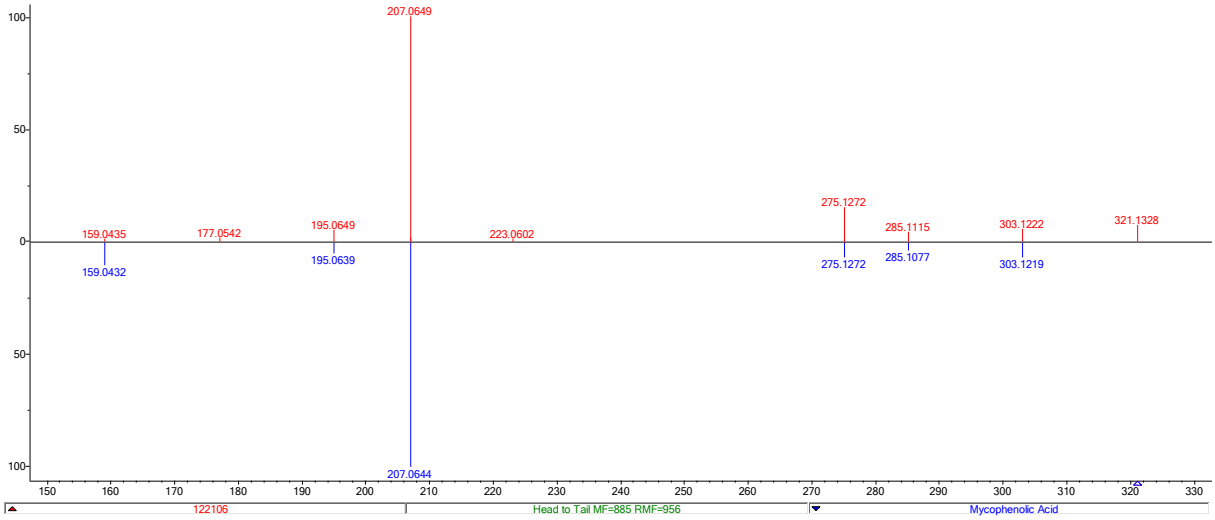




Name: Sophocarpine [M+H-C6H9ON]+
Formula: C15H22N2O
PubChemID: 5271988
Compound class: natural medicine / herbal extract

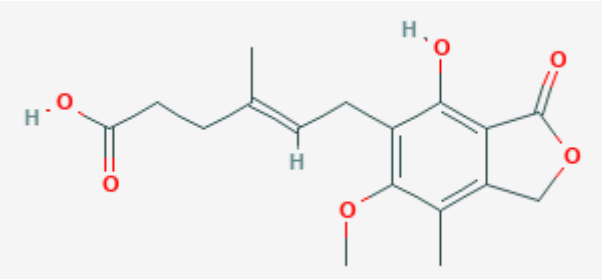
WWTP stream: PR-Chem
Fate: Average 87% elimination in PR on selected dates, mainly in OZO

Profile 122106

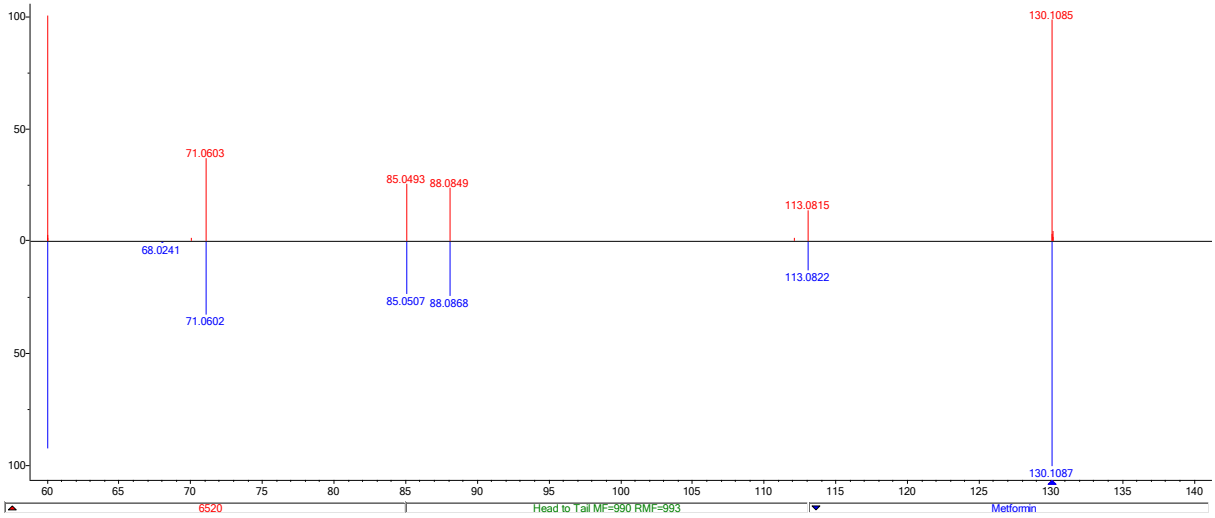


Name: Mycophenolic acid
Formula: C17H20O6
PubChemID: 446541
Compound class: pharmaceutical

WWTP stream: PR-Chem
Fate: 100% average overall elimination, mostly in BIO

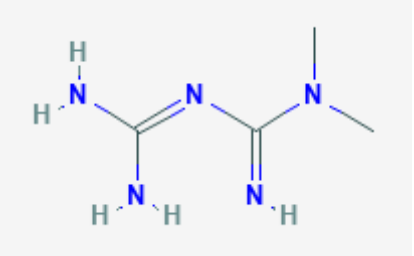


Profile 6520

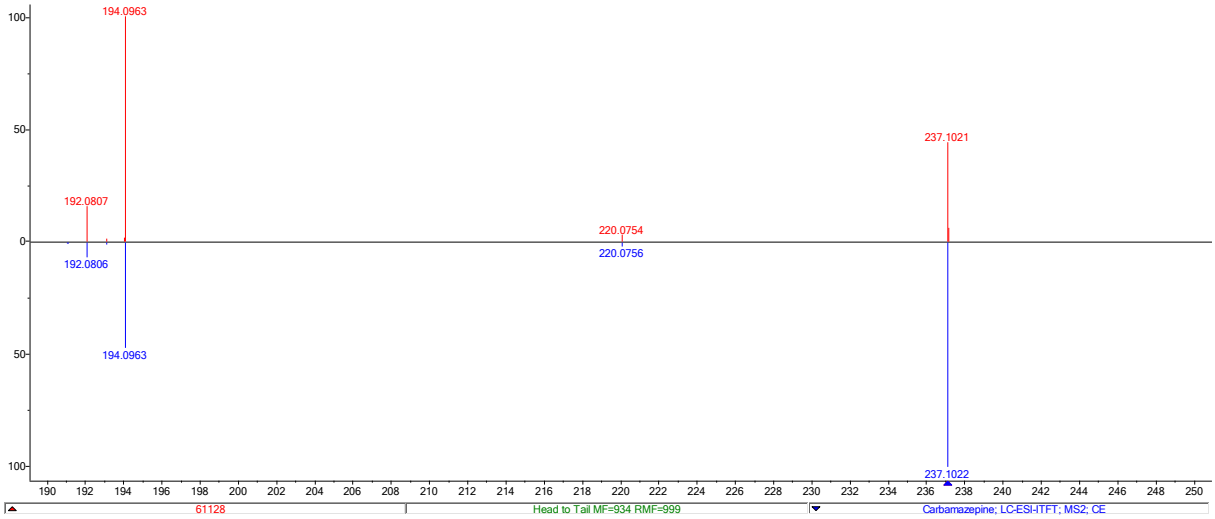


Name: Metformin
Formula: C4H11N5
PubChemID: 4091
Compound class: pharmaceutical

WWTP stream: PR-Comm
Fate: average elimination 100% in PR, all in BIO

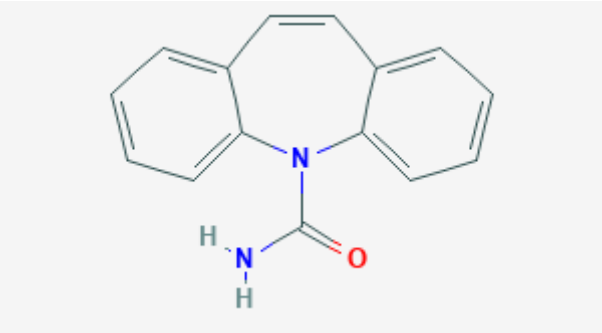


Profile 61128



Name: Carbamazepine
Formula: C15H12N2O
PubChemID: 2554
Compound class: pharmaceutical

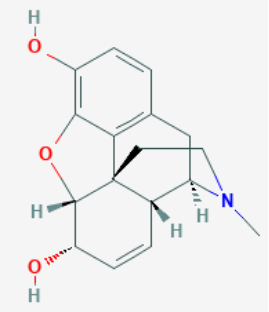
WWTP stream: PR-Chem
Fate: 91% average elimination in PR, mostly in PAC



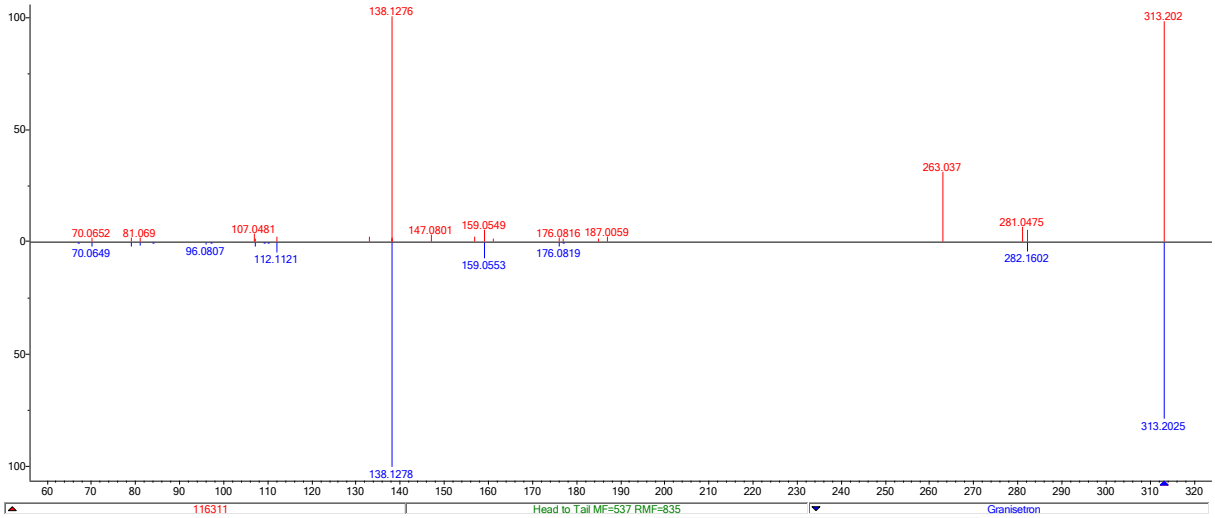
Profile 96527

Name: Morphine
Formula: C17H19NO3
PubChemID: 5288826
Compound class: opioid

WWTP stream: PR-Comm
Fate: Average elimination 100% in PR, mostly in BIO, fully removed in OZO

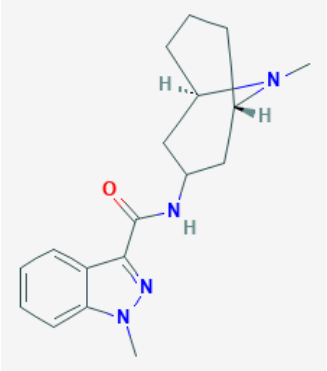


Profile 116311

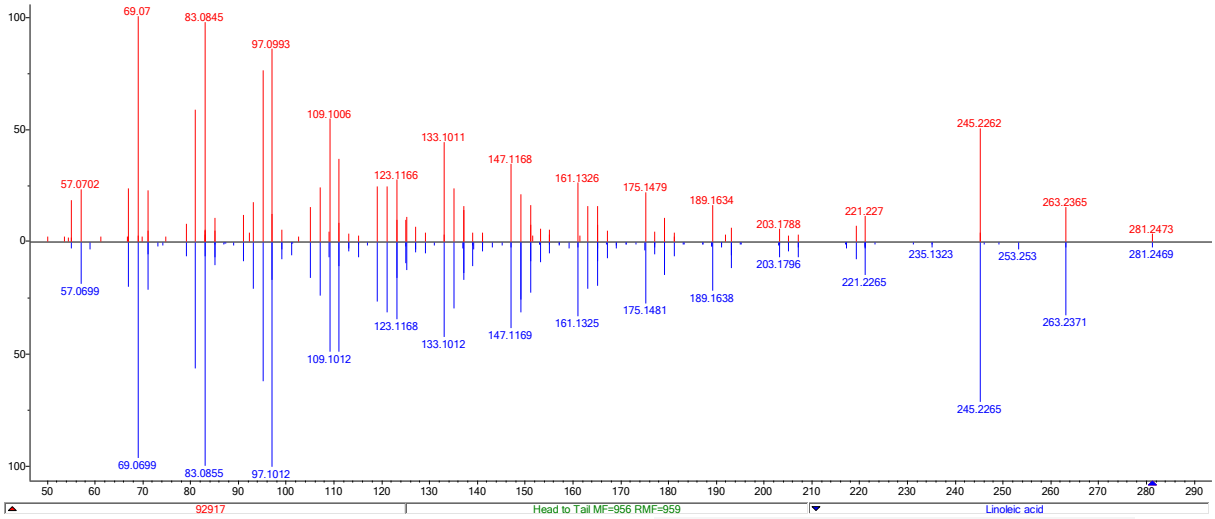


Name: Granisetron
Formula: C18H24N4O
PubChemID: 5284566
Compound class: pharmaceutical

WWTP stream: PR-Chem
Fate: 98% average elimination in PR, mostly in BIO, some in OZO

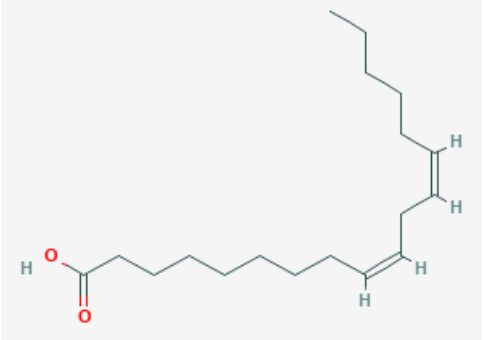


Profile 92917

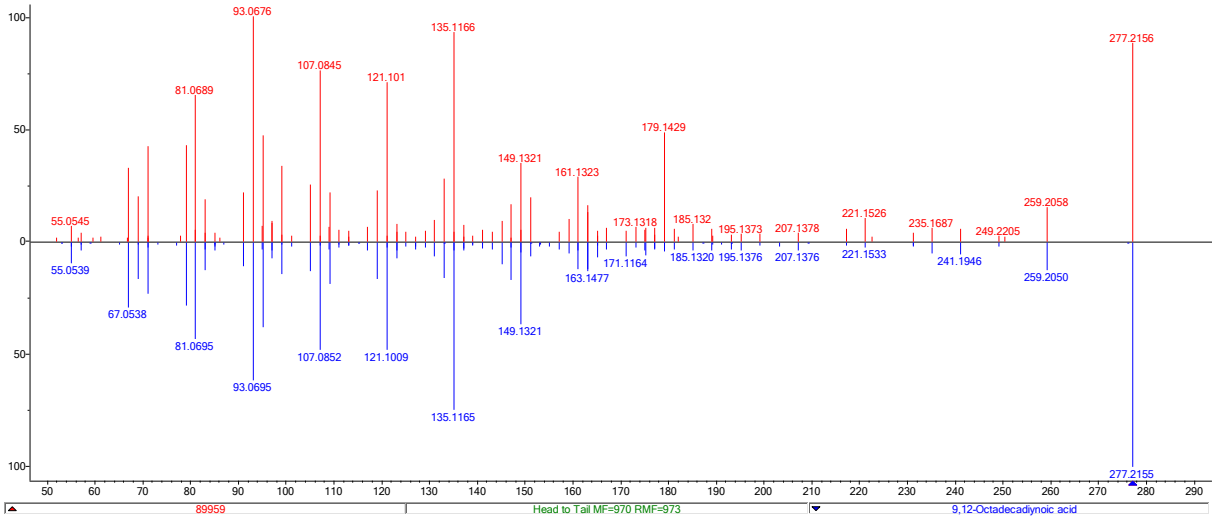


Name: Linoleic acid
Formula: C18H32O2
PubChemID: 5280450
Compound class: Fatty acid / possible industrial compound

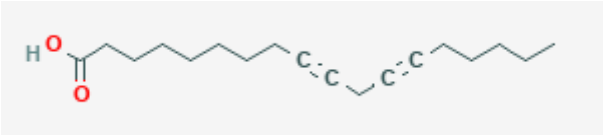
WWTP stream: PR-Chem
Fate: Average elimination 77% in seelcted PR samples, mainly in BIO



Profile 89959

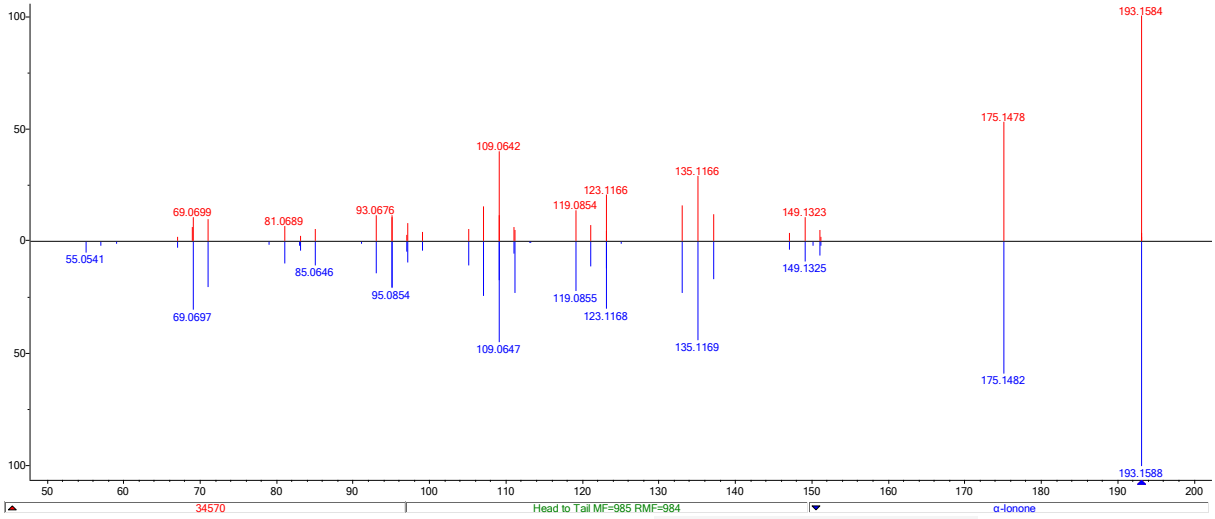


Name: 9,12-Octadecadiynoic acid
Formula: C18H28O2
PubChemID: 1931
Compound class: fatty acid

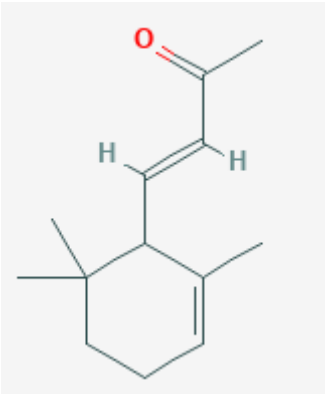


WWTP stream: PR-Comm
Fate: 100% average elimination. Some formation in BIO, removal in OZO

Profile 34570

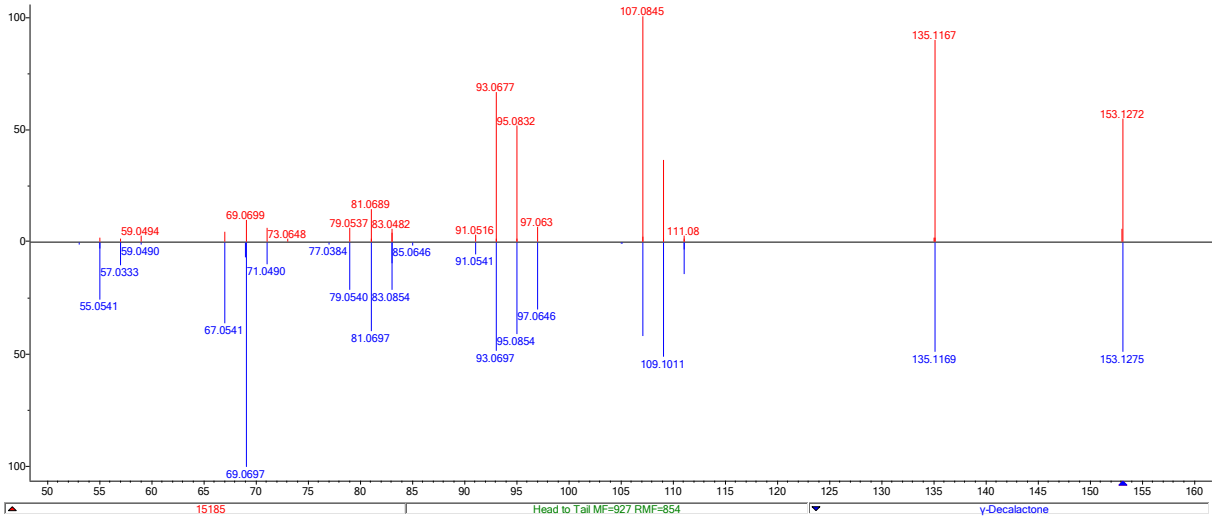


Name: alpha-Ionone
Formula: C13H20O
PubChemID: 5282108
Compound class: Food additive / fragrance

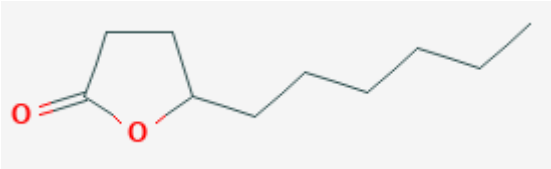


WWTP stream: PR-Chem
Fate: Mixed results in PR, sometimes removed in BIO, sometimes persistant

Profile 15185

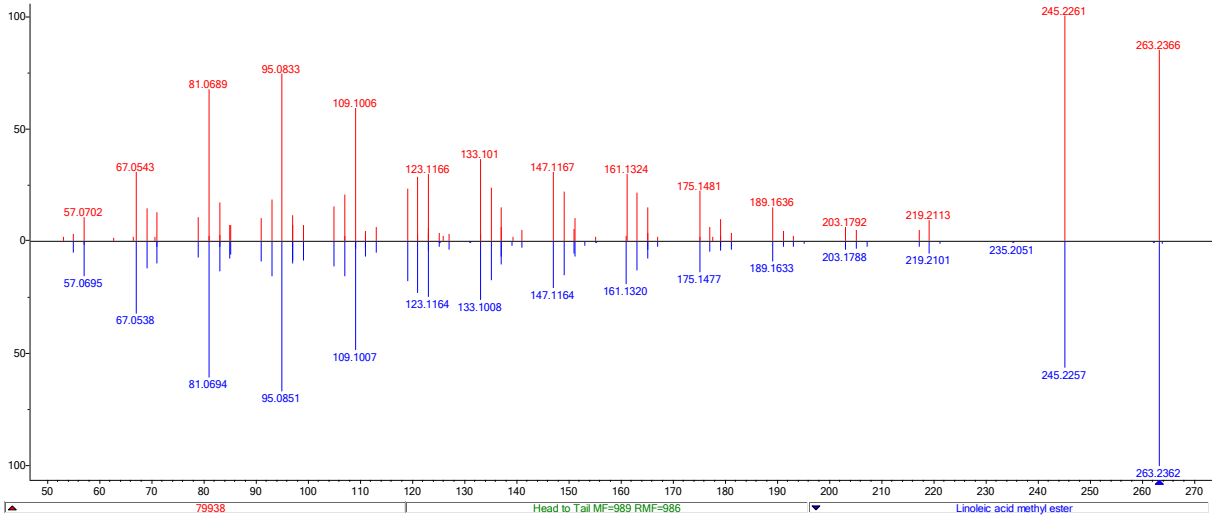


Name: gamma-Decalactone (multiple good candidates)
Formula: C10H18O2
PubChemID: 12813
Compound class: Natural food additive



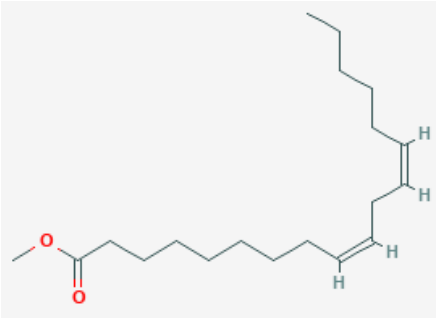
WWTP stream: PR-Chem
Fate: Average elimination 90% in PR, in BIO

Profile 79938

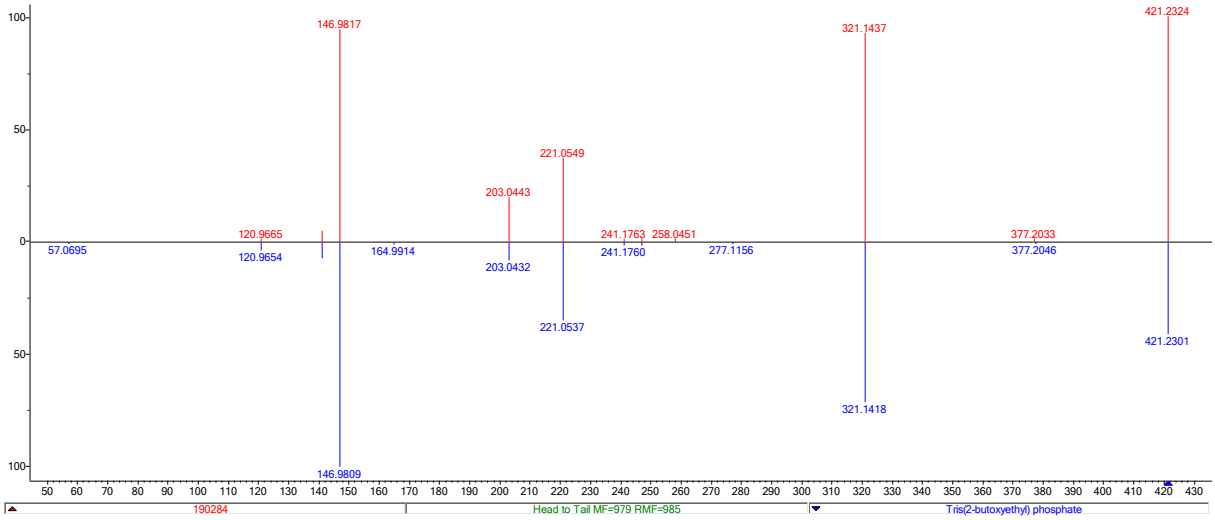


Name: Linoleic acid methyl ester [M+H-CH4O]+
Formula: C19H34O2
PubChemID: 5284421
Compound class: cosmetic, flavor and food additive

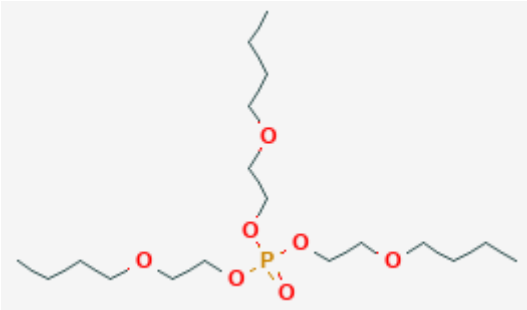
WWTP stream: PR-Chem
Fate: 90% average elimination in highest PR samples, mostly in BIO.



Profile 190284

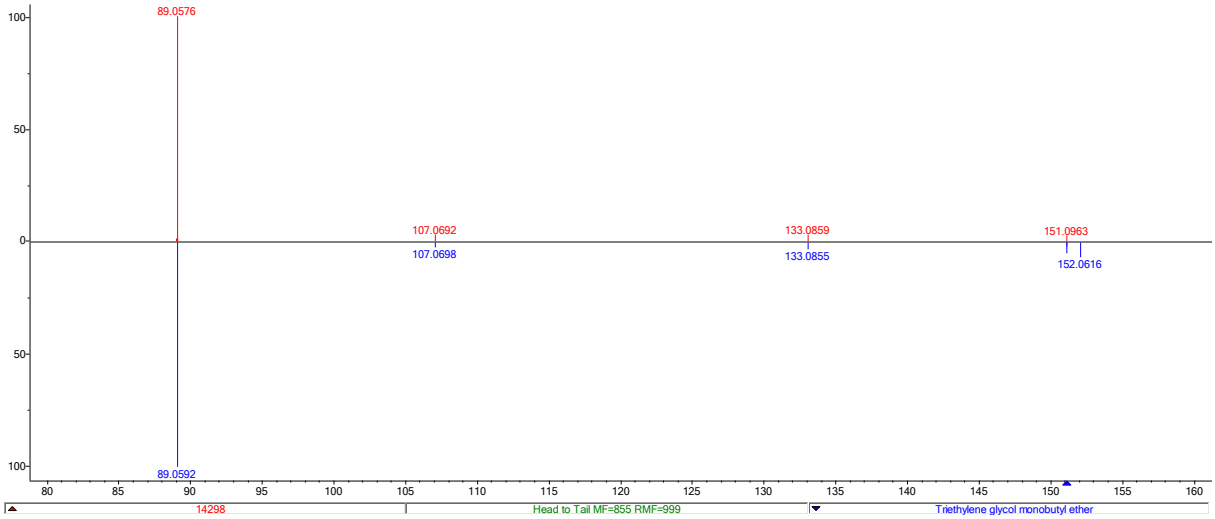


Name: Tris(2-butoxyethyl) phosphate
Formula: C18H39O7P
PubChemID: 6540
Compound class: flame retardant

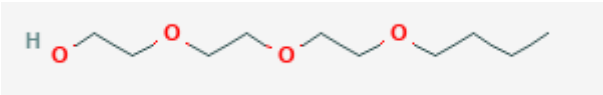


WWTP stream: PR-Chem
Fate: Average elimination 74% in PR, mostly in BIO

Profile 14298

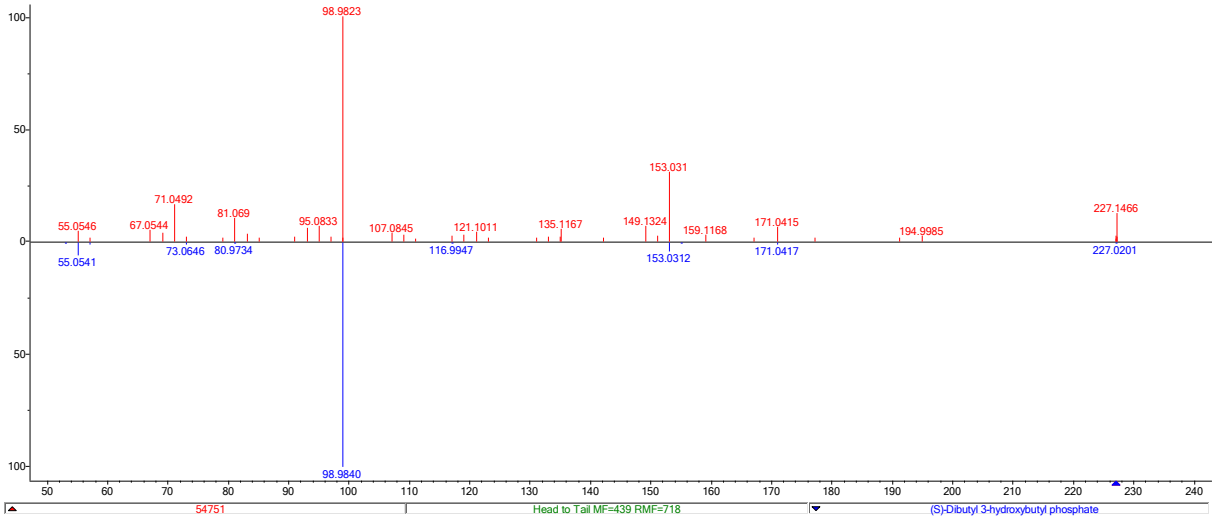


Name: Triethylene glycol monobutyl ether
Formula: C10H22O4
PubChemID: 8923
Compound class: Solvent

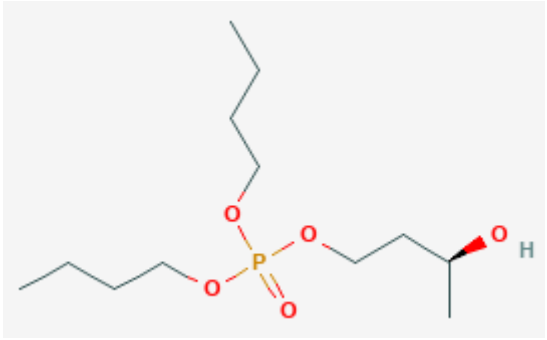


WWTP stream: PR-Chem
Fate: Average elimination in PR 100%, all in BIO, slight reformation in OZO

Profile 54751

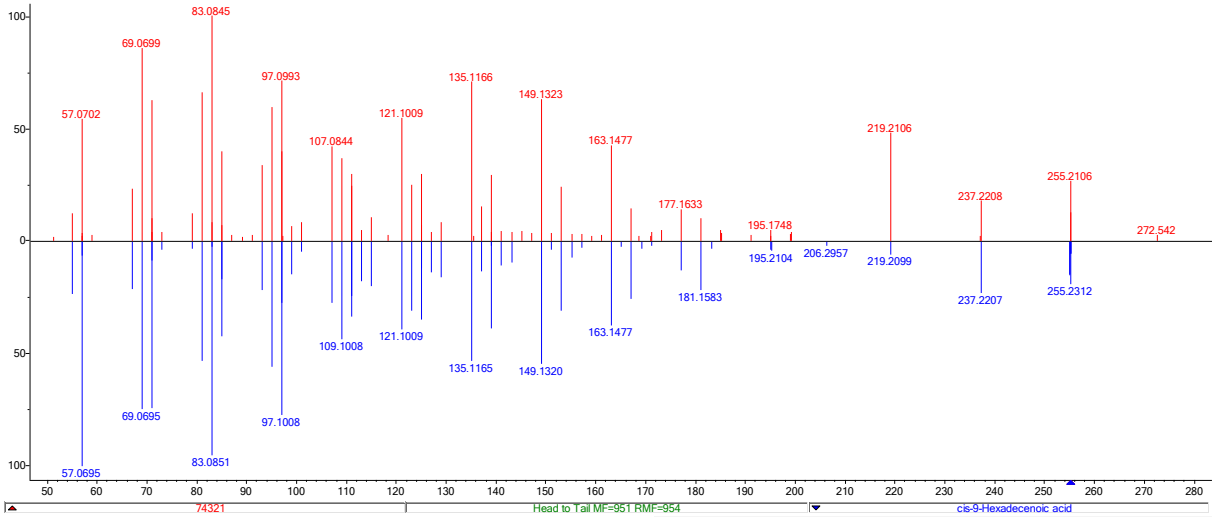


Name: (S)-Dibutyl-3-hydroxybutyl phosphate
Formula: C12H27O5P
PubChemID: 124306374
Compound class:

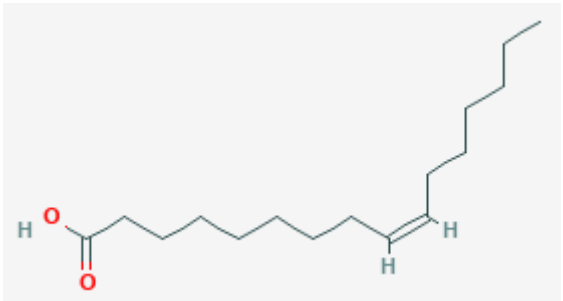


WWTP stream: PR-Chem
Fate: 67% average elimination in PR, mostly in BIO

Profile 74321

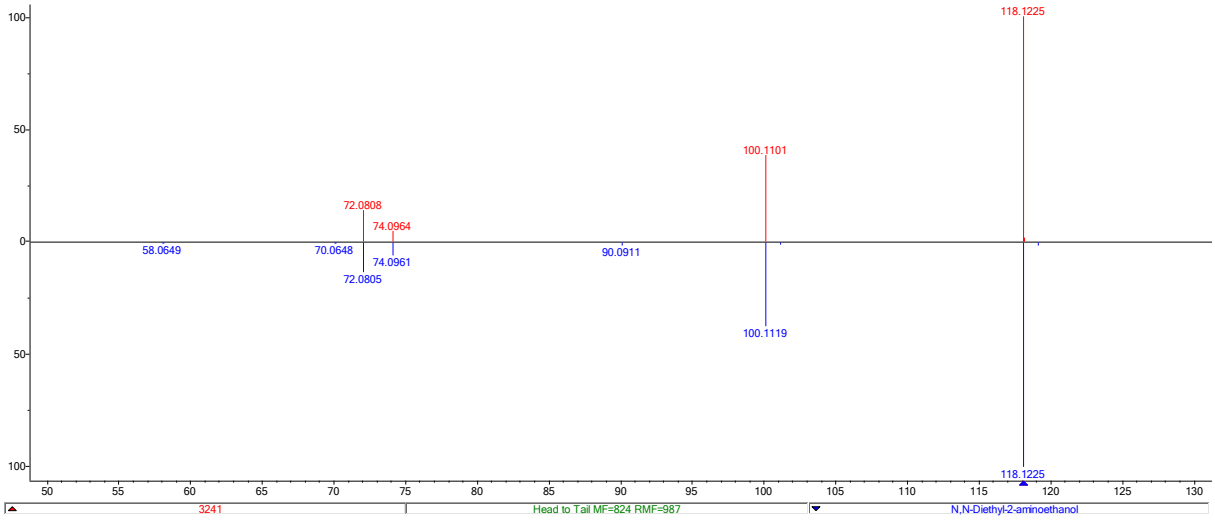


Name: cis-9-hexadecenoic acid
Formula: C16H30O2
PubChemID: 445638
Compound class: pharmaceutical intermediate

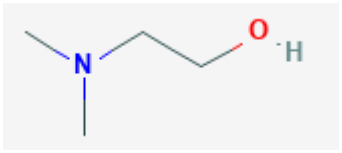


WWTP stream: PR-Comm
Fate: Mixed results in PR, sometimes removed in BIO, sometimes persistant

Profile 3241

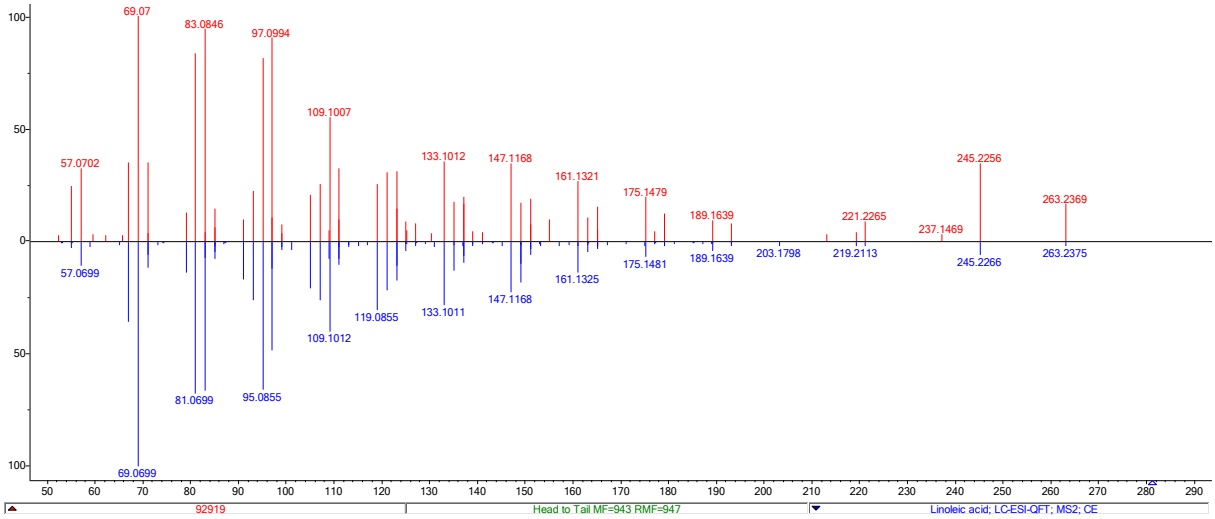


Name: N,N-Dimethyl-2-aminoethanol
Formula: C6H15NO
PubChemID: 7902
Compound class: multiple uses



WWTP stream: PR-Chem
Fate: Average elimination 100% in PR

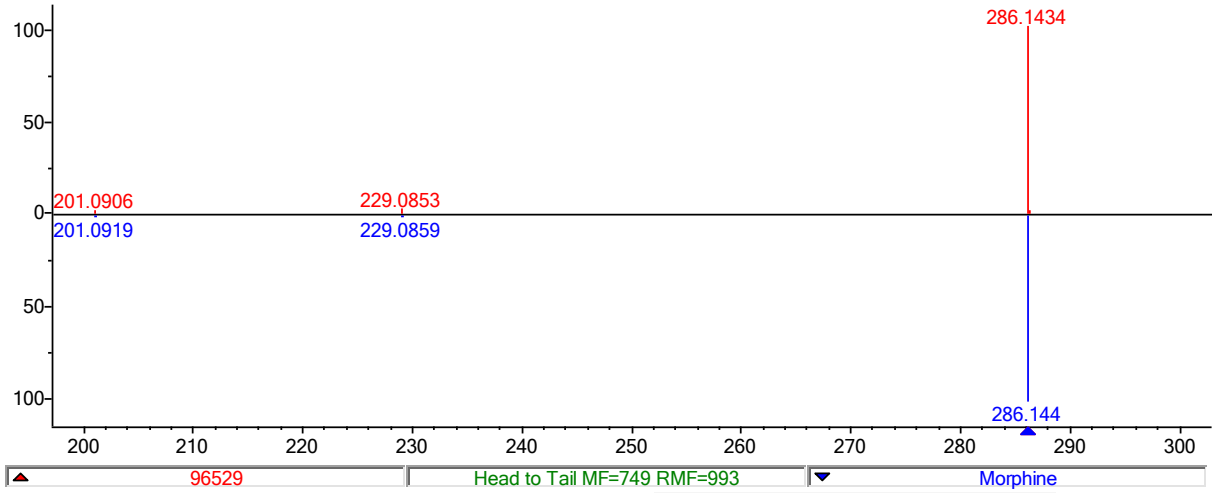
Profile 92919



Name: Linoleic acid
Formula: C18H32O2
PubChemID:
Compound class: fatty acid

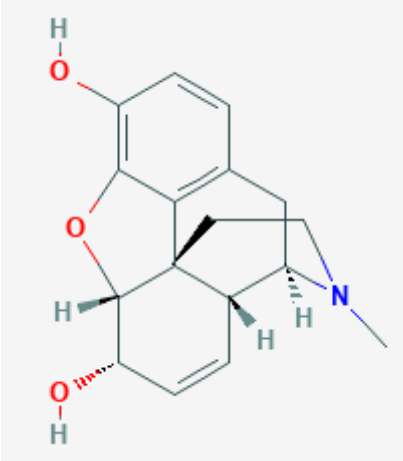
WWTP stream: 92919
Fate: 98% average elimination in PR, mostly in BIO, some in OZO

Profile 96529

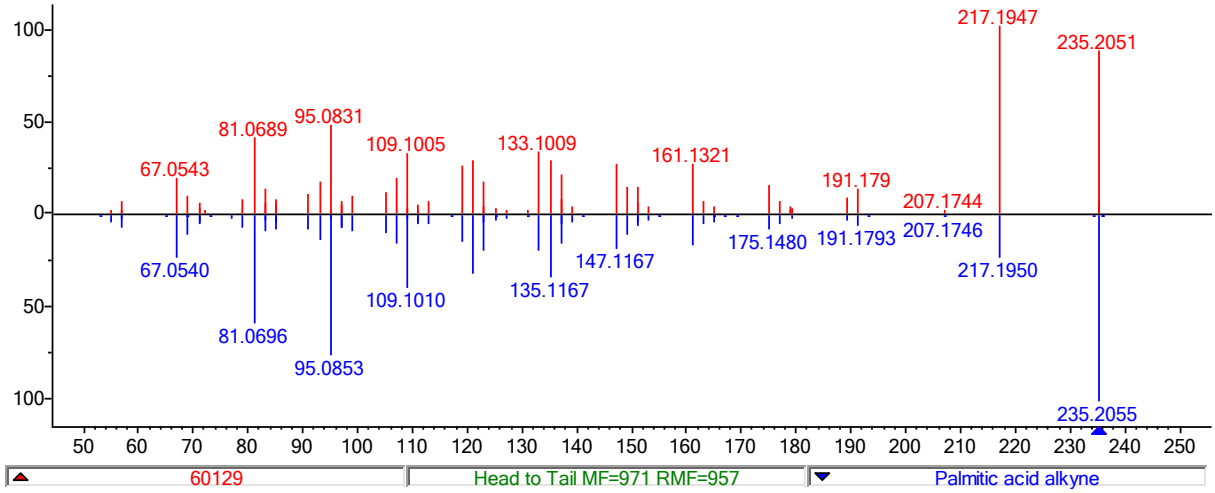


Name: Morphine
Formula: C17H19NO3
PubChemID: 5288826

WWTP stream: PR-Comm
Fate: 100% average elimination, all in BIO



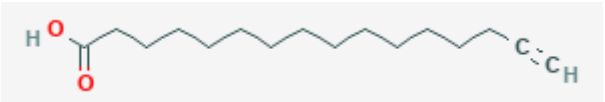
Profile 60129



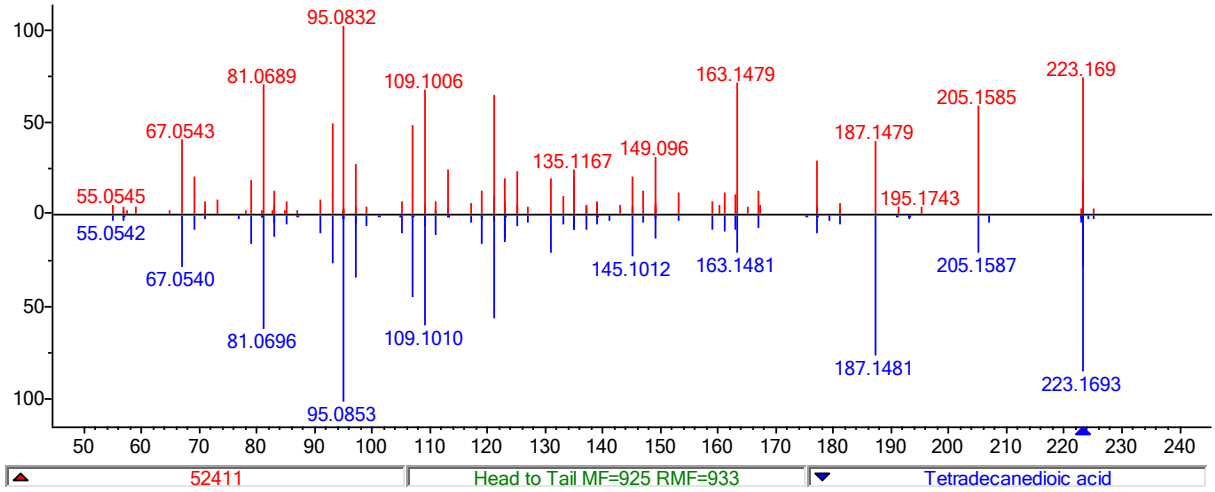
Name: Palmitic acid alkyne
Formula: C16H28O2
PubChemID: 127256
Compound class: fatty acid

WWTP stream: PR-Comm

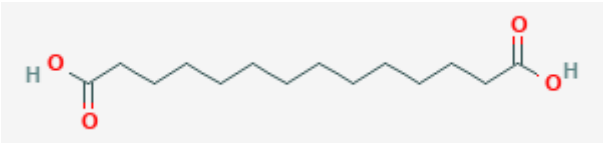
Fate: 85% average elimination in selected samples. Removed in BIO in all but 1 sample, where formation was observed.



Profile 52411

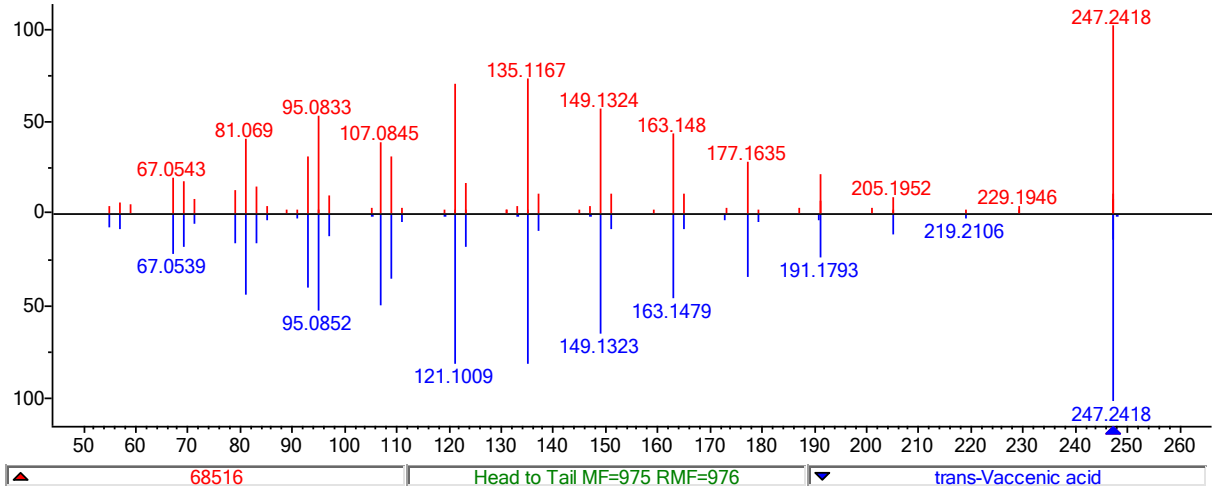


Name: Tetradecanedoic acid
Formula: C14H26O4
PubChemID: 13185
Compound class: metabolite / industrial chemical

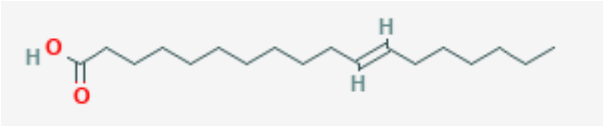


WWTP stream: PR-Chem
Fate: In the majority of samples, elimination >70%, formation observed in 3 samples

Profile 68516

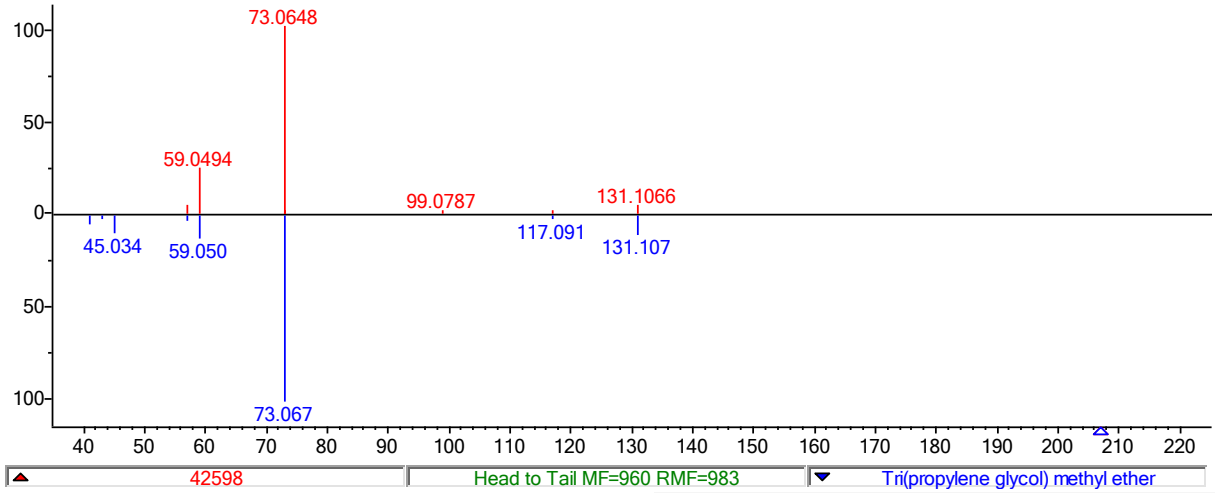


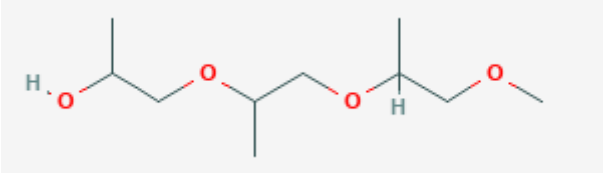
Name: trans-Vaccenic acid
Formula: C18H34O2
PubChemID: 5281127
Compound class: Metabolite of rumen, found in dairy and meat products



WWTP stream: PR-Chem
Fate: 99% average elimination in PR, mostly in BIO

Profile 42598

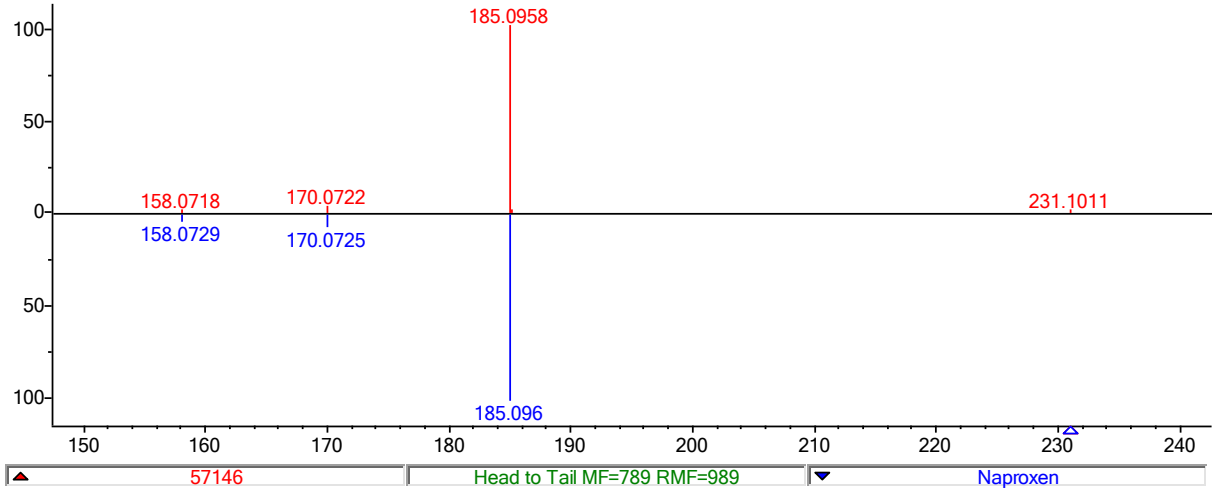


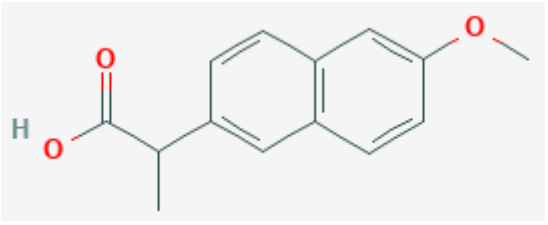


Name: Tri(propylene glycol) methyl ether
Formula: C10H22O4
PubChemID: 30111
Compound class:propylene glycol ether

WWTP stream: PR-Chem
Fate: 91% average elimination in PR, mostly in BIO

Profile 57146

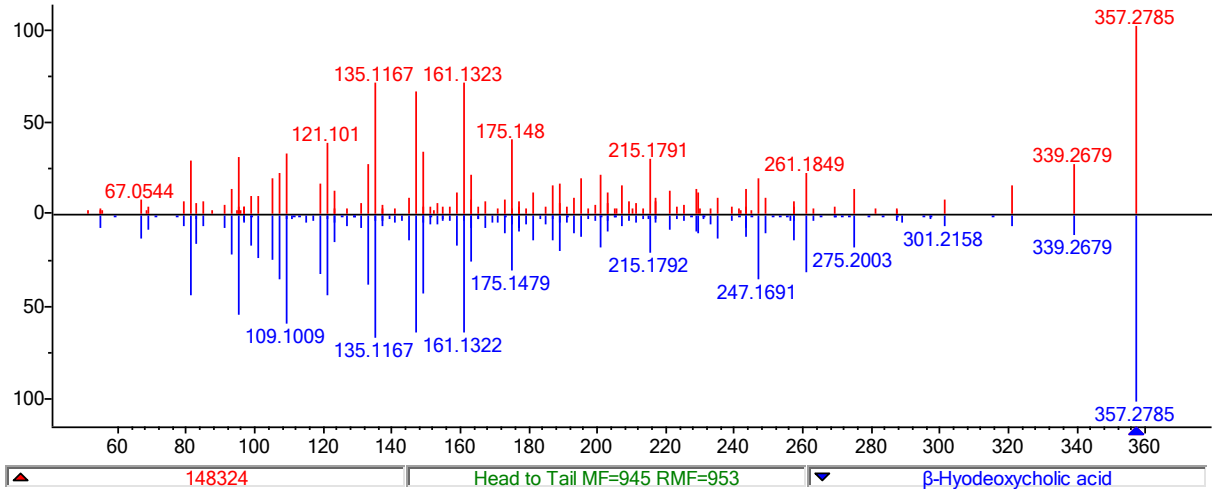




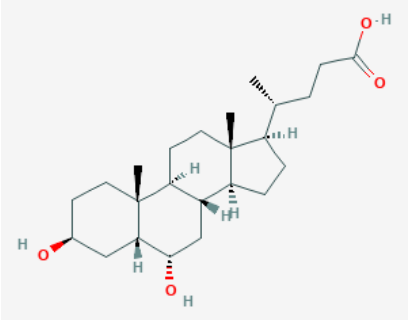
Name: Naproxen
Formula: C14H14O3
PubChemID: 1302
Compound class: pharmaceutical

WWTP stream: PR-Chem
Fate: 100% average elimination in PR, mostly in BIO

Profile 148324

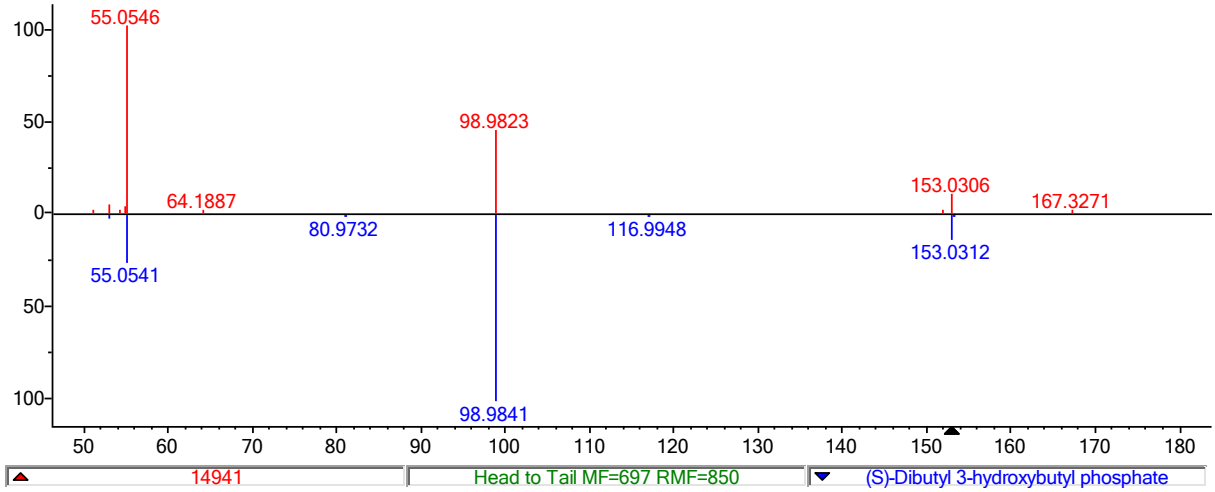


Name: beta-hyodexycholic acid
Formula: C24H40O4
PubChemID: 5283822
Compound class: pharmaceutical

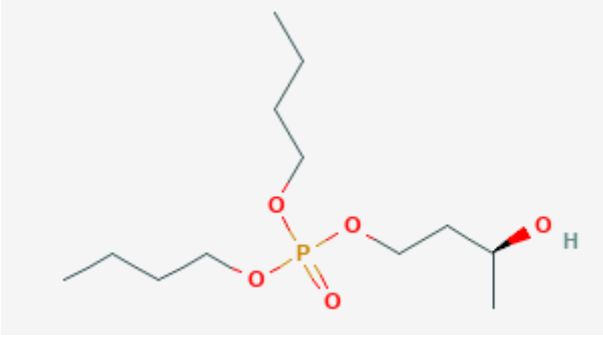


WWTP stream: PR-Comm
Fate: In detected sample, average elimination 100% in PR

Profile 14941

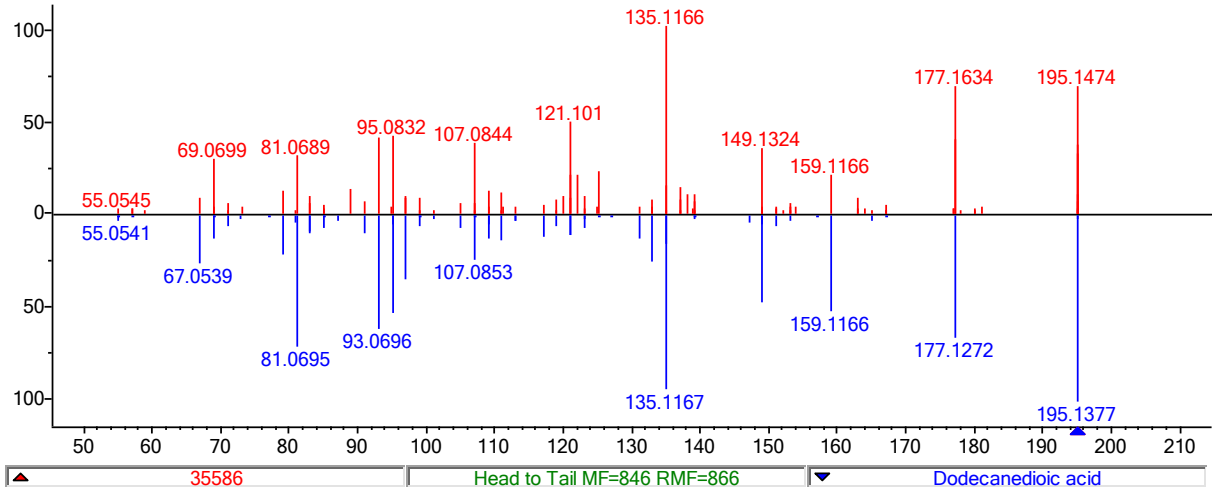


Name: (S)-Dibutyl 3-hydroxybutyl phosphate
Formula: C12H27O5P
PubChemID: 124306374
Compound class: Organophosphate ester

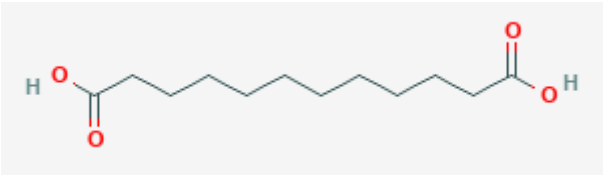


WWTP stream: PR-Chem
Fate: 80% average elimination in detected PR

Profile 35586

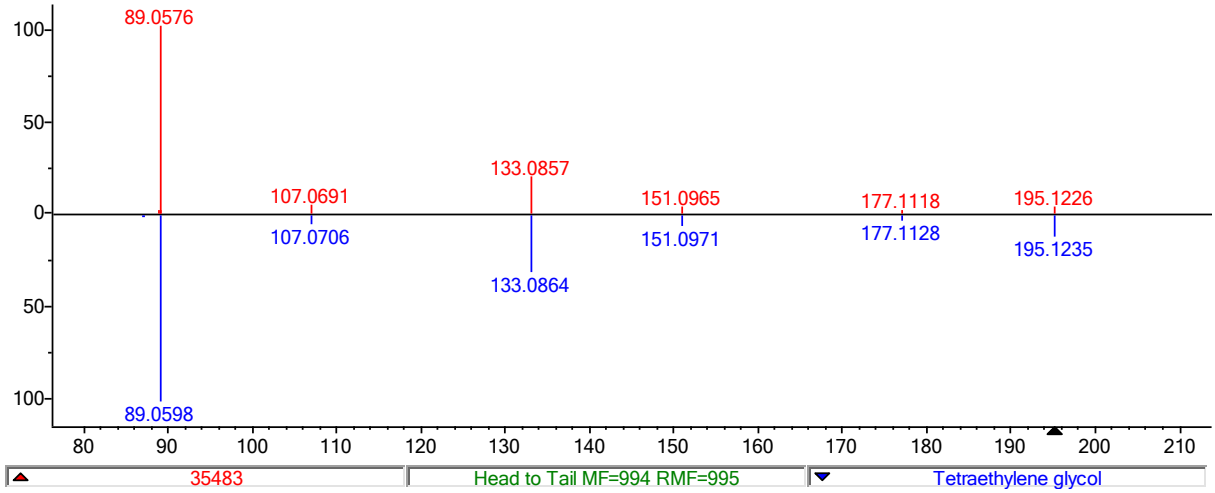


Name: Dodecanedioic acid
Formula: C12H22O4
PubChemID:12736
Compound class: fatty acid, multiple uses

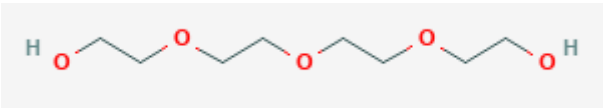


WWTP stream: PR-Chem
Fate: Mixed results in PR, sometimes removed in BIO, sometimes formation

Profile 35483

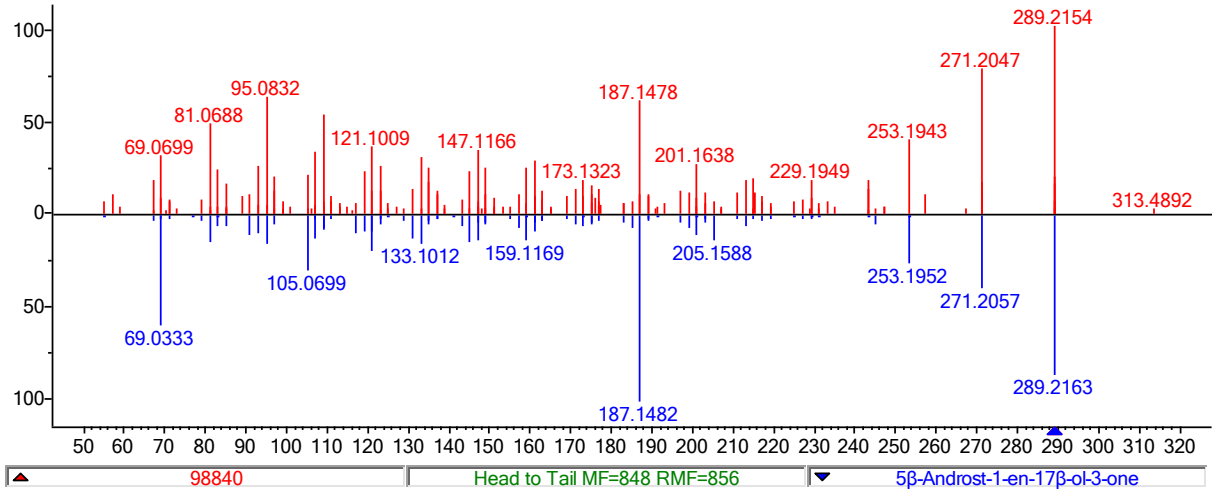


Name: Tetraethylene glycol
Formula: C8H18O5
PubChemID: 8200
Compound class: Industrial solvent



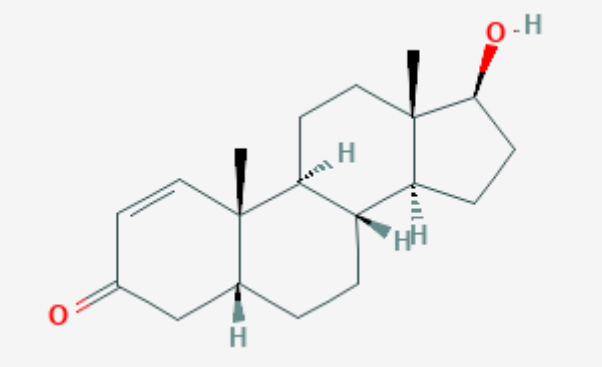
WWTP stream: PR-Chem
Fate: Average 100% elimination in PR samples, all in BIO

Profile 98840

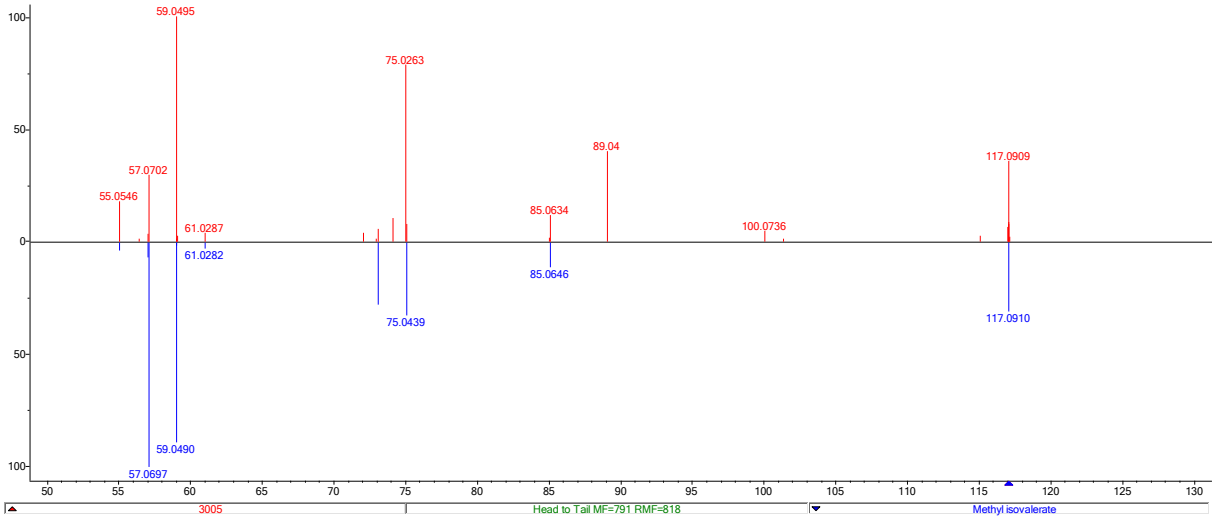


Name: 5β-Androst-1-en-17β-ol-3-one
Formula: C19H28O
PubChemID: 12133279
Compound class: Bovin metabolite

WWTP stream: PR-Chem
Fate: 98% average elimination in PR, mostly in BIO

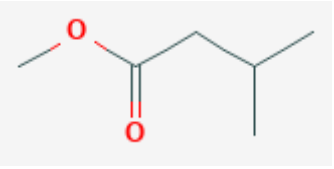


Profile 3005

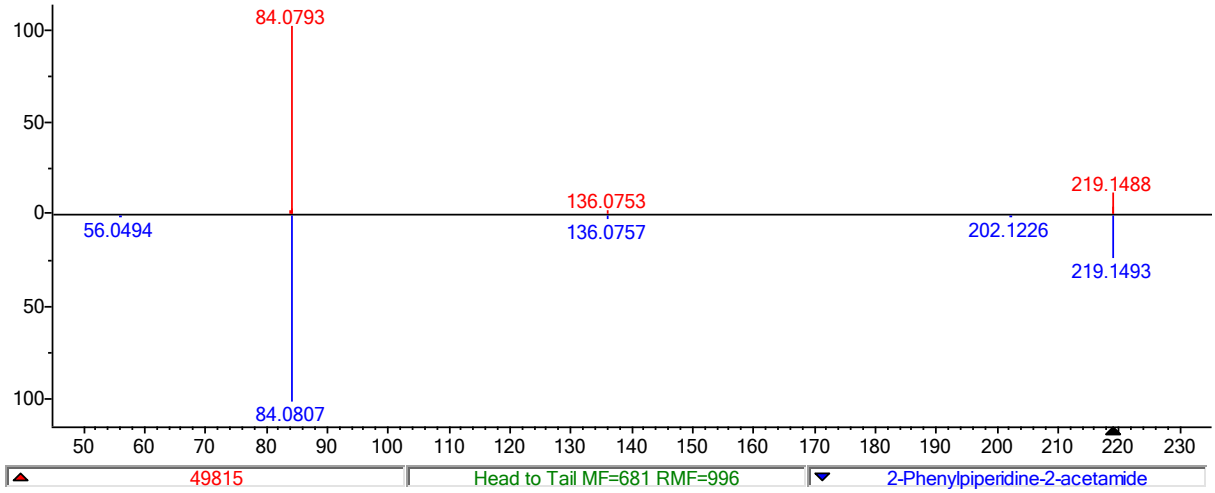


Name: Methyl isovalerate
Formula: C6H12O2
PubChemID: 11160
Compound class: Flavoring

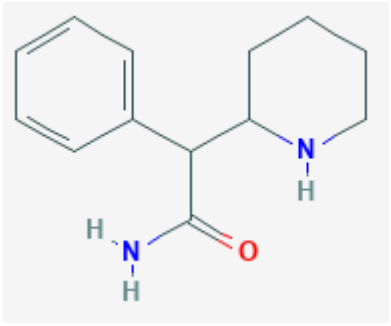
WWTP stream: PR-Chem
Fate: 99% average elimination in PR, all in BIO



Profile 49815

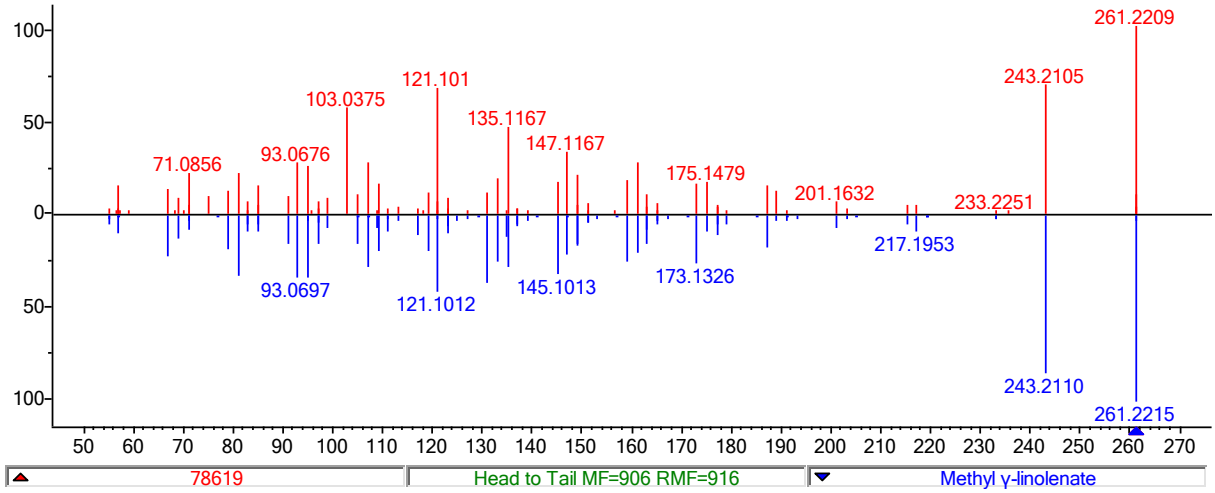


Name: 2-phenylpiperidine-2-acetamide
Formula: C13H18N2O
PubChemID: 86862
Compound class:

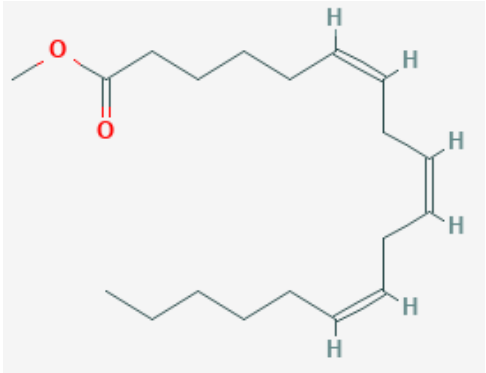


WWTP stream: PR-Comm
Fate: Average elimination 99% in PR, partly in BIO, partly in OZO

Profile 78619

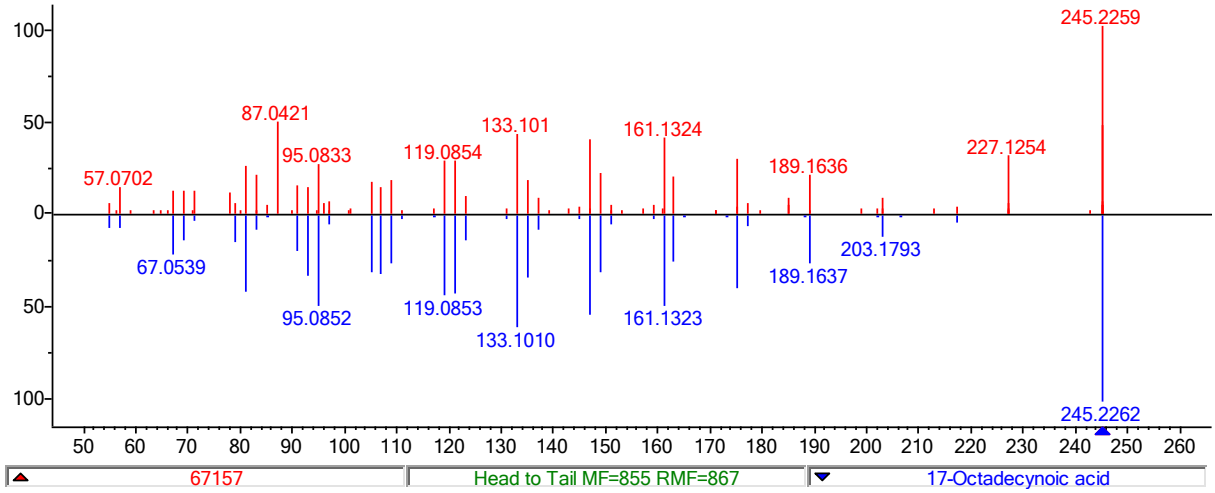


Name: Methyl gamma-linolenate
Formula: C19H32O2
PubChemID: 6439889
Compound class: antineoplastic agent

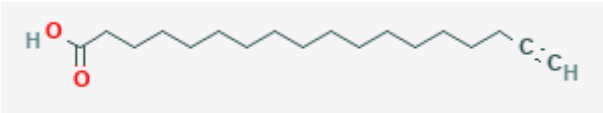


WWTP: PR-Chem
Fate: 78% average elimination in PR, mostly in BIO

Profile 67157

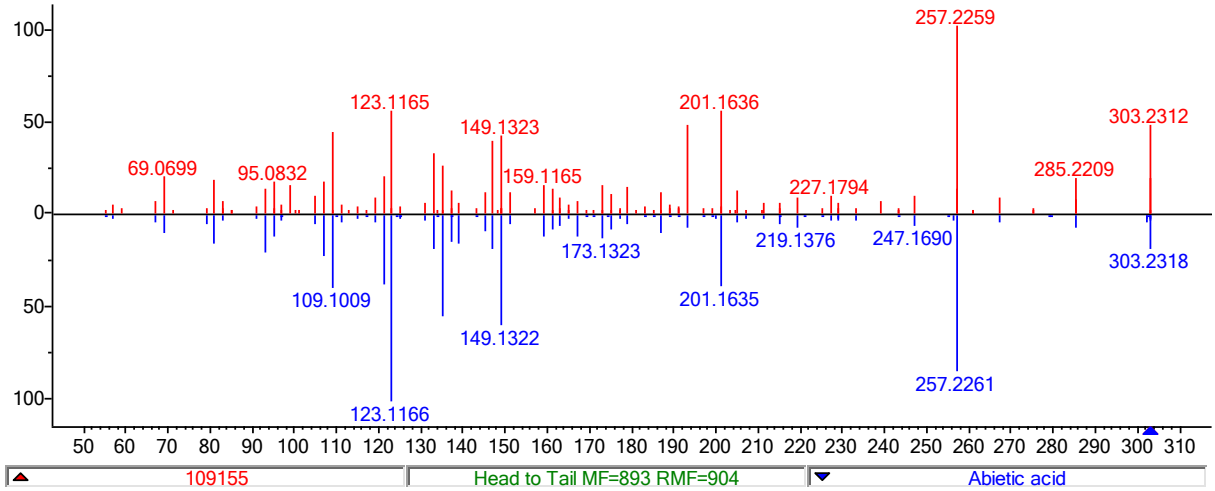


Name: 17-Octadecynoic acid (multiple candidates)
Formula: C18H32O2
PubChemID: 1449
Compound class: pharmaceutical?



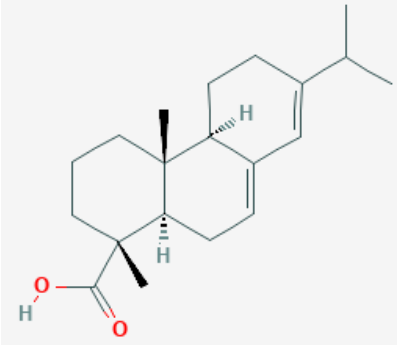
WWTP stream: PR-Chem
Fate: In detected samples, average elimination 82% in PR, mostly in BIO

Profile 109155

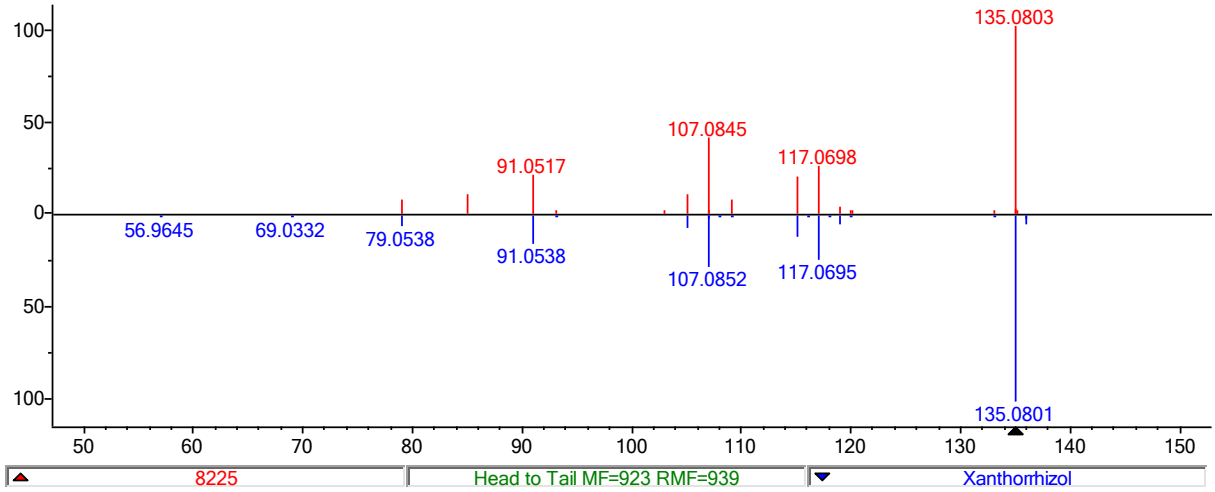


Name: Abietic acid
Formula: C20H30O2
PubChemID: 10569
Compound class: Industrial compound

WWTP stream: PR-Chem
Fate: 88% average elimination in detected PR samples, all in BIO

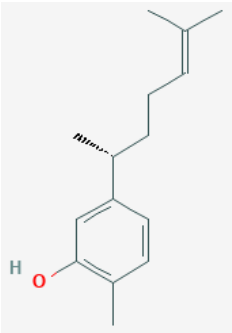


Profile 8225

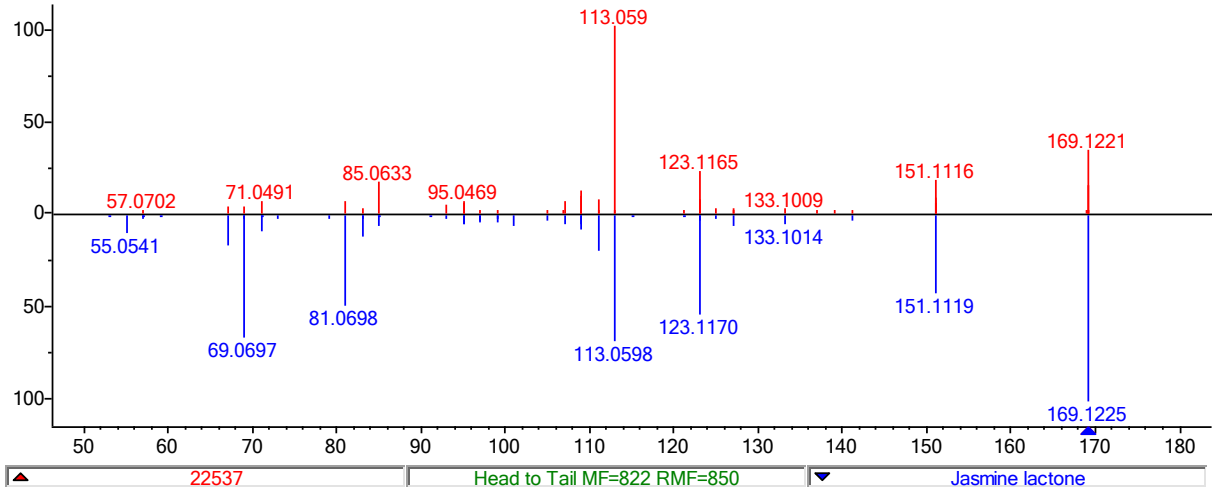


Name: Xanthorrhizol
Formula: C15H22O
PubChemID: 93135
Compound class: nature medicine

WWTP stream: PR-Chem
Fate: 100% average elimination in PR, all in BIO

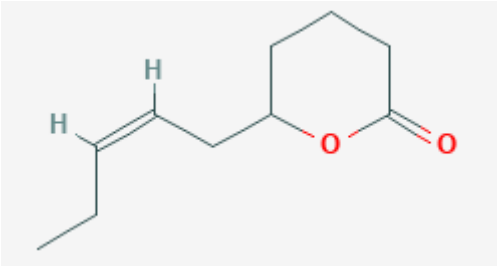


Profile 22537

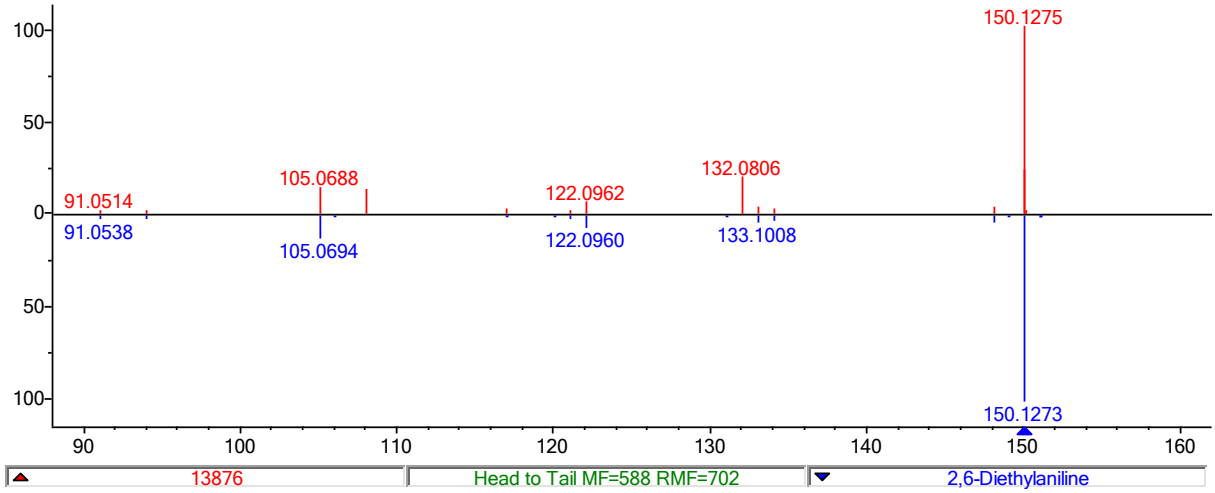


Name: Jasmine lactone
Formula: C10H16O2
PubChemID: 5352626
Compound class: food additive and flavoring

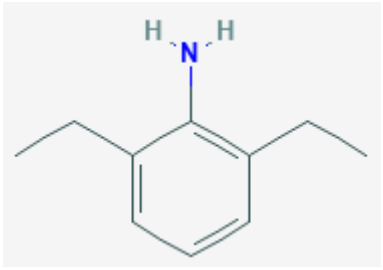
WWTP stream: PR-Chem
Fate: 75% average elimination in PR, mostly in BIO



Profile 13876

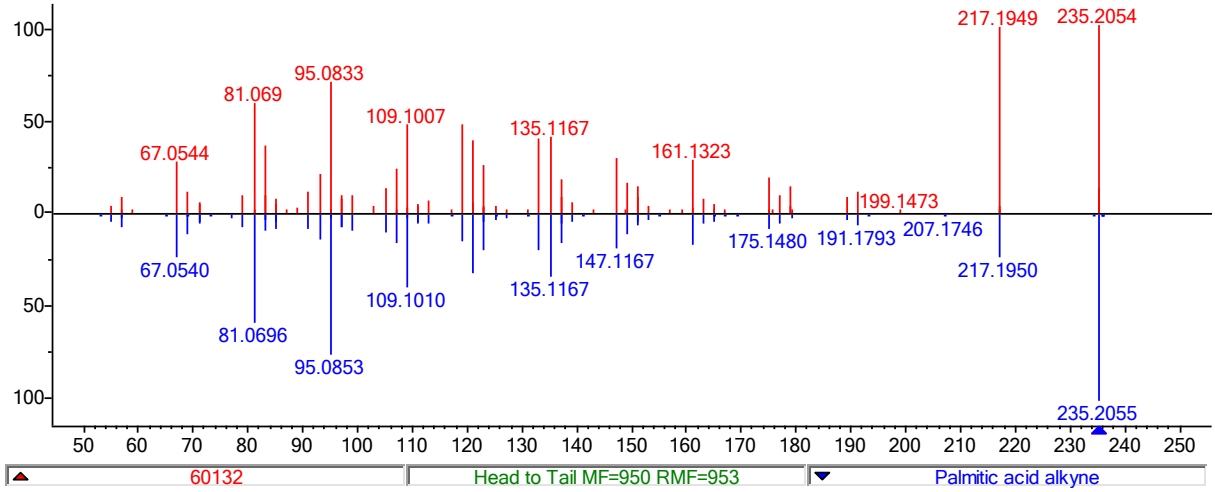


Name: 2,6-Diethylaniline
Formula: C10H15N
PubChemID: 11369
Compound class: multiple



WWTP stream: PR-Chem
Fate: 100% average elimination in PR, mostly in BIO

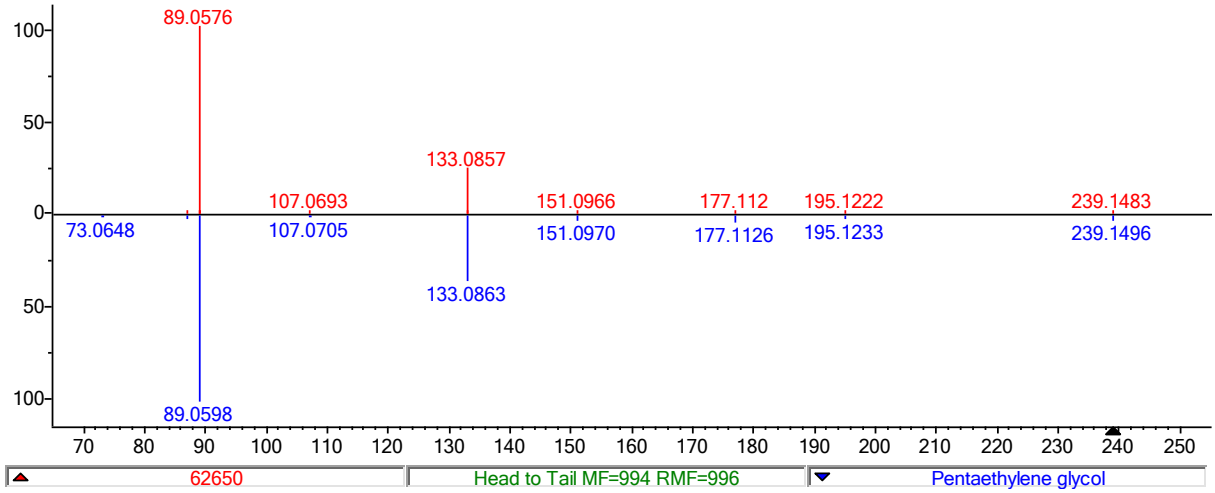
Profile 60132



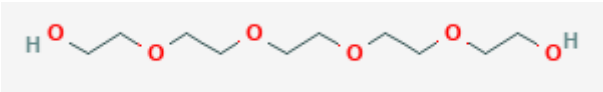
Name: Palmitic acid alkyne
Formula: C16H28O2
PubChemID:
Compound class: fatty acid

WWTP stream: PR-Comm
Fate: In detected samples, average elimination 100% in PR, all in BIO

Profile 62650

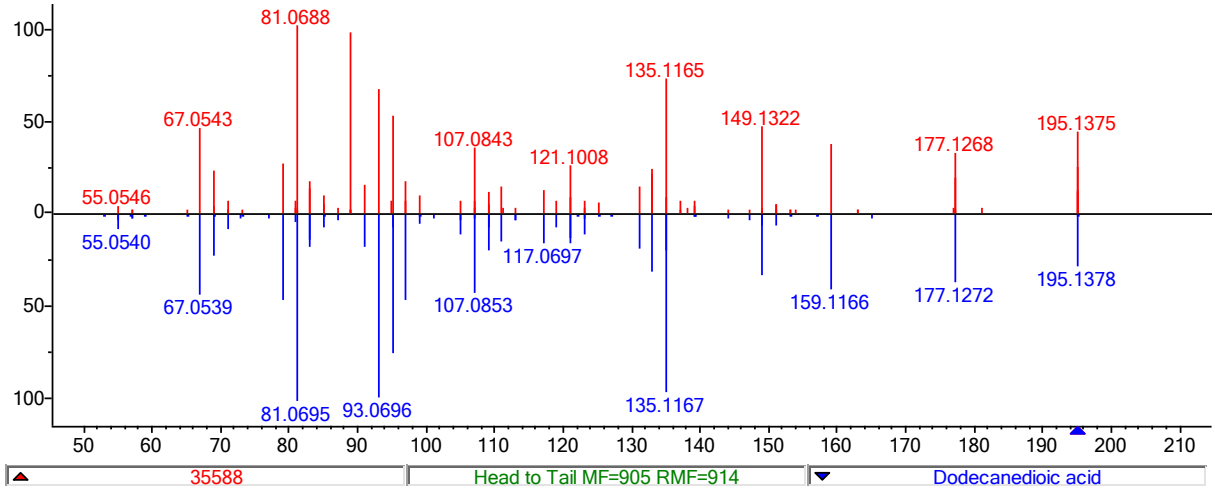


Name: Pentaethylene glycol
Formula: C10H22O6
PubChemID: 62551
Compound class: Industrial solvent

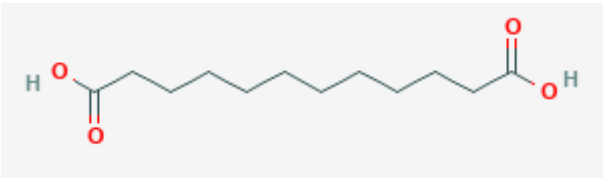


WWTP stream: PR-Chem
Fate: average elimination 100% in PR, all in BIO

Profile 35588

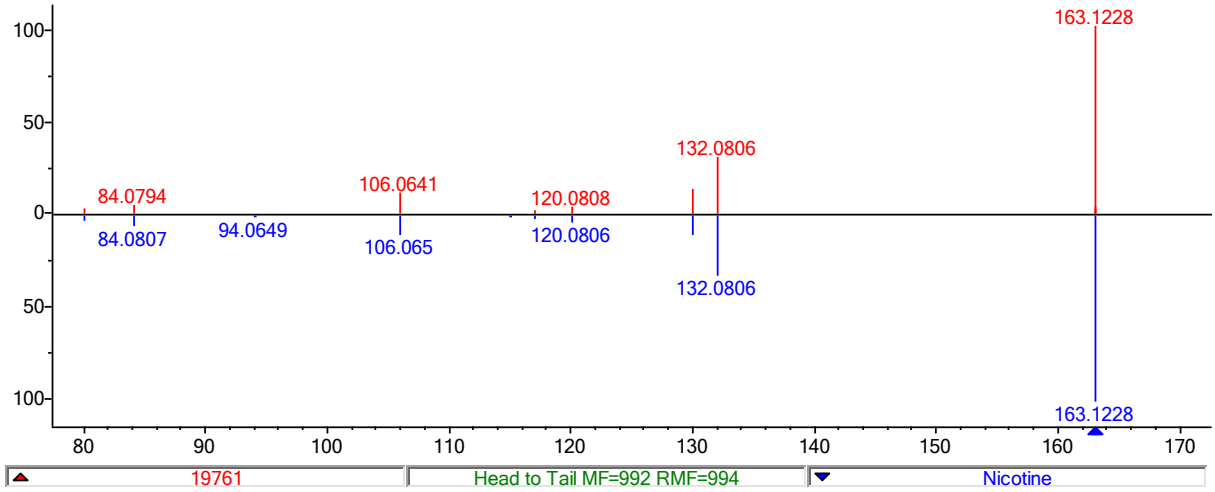


Name: Dodecanedioic acid
Formula: C12H22O4
PubChemID: 12736
Compound class: fatty acid

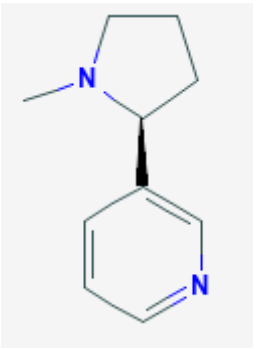


WWTP stream: PR-Chem
Fate: In detected samples, 100% elimination in PR, all in BIO

Profile 19761



Name: Nicotine
Formula: C10H14N2
PubChemID: 89594
Compound class: tobacco ingredient



WWTP stream: PR-Chem
Fate: In detected samples, average elimination 93% in PR, mostly in BIO

Section S5. Sensitivity analysis of automated trend assignment

To perform the automated trend assignment, cutoffs were selected to define the high, middle, and low domains. A sensitivity analysis was performed to estimate how important these cutoffs were to the final result. Four different cutoffs were considered; cutoff_01 was selected and the results of this scenario are presented in the main manuscript.

Table S 12. Summary of cutoffs considered in sensitivity analysis of pattern recognition algorithm

Name	Lower cutoff	Upper cutoff
Cutoff_01	20%	60%
Cutoff_02	30%	70%
Cutoff_03	20%	80%
Cutoff_04	10%	90%

Comparison of major trend results from different cutoff scenarios

As shown in Figure S 19 below, using the different cutoff values changed the number of features within a major trend by less than 2% (average 0.45±0.48% absolute change). The largest changes were observed in Trend 1 (*i.e.*, influent features removed in BIO) and in Trend 10 (*i.e.*, persistent features) and Trend 9 (*i.e.*, influent features removed in post treatment), particularly for cutoff scenarios 2 and 4. Here the changes in the percent of features in Trend 1 or Trend 10 was inversed between cutoff_02 and cutoff_04, which can be explained by increasing the lower cutoff value to 30% or decreasing it to 10%, respectively, compared to the reference of 20% in cutoff_01. Decreasing the lower cutoff from 20 to 10% resulted in 1.7% more persistant features in Trend 1 (removal in Bio), since 124 non-target features no longer met threshold of 90% removal to be assigned to Trend 1. It can be confirmed that the lower cutoff value, and not the upper cutoff value, is most likely the reason for this difference because with cutoff_02, where the upper cutoff was varied but the lower cutoff value was kept constant, no difference was observed to the reference cutoff_01.

Additionally, those trends most consistantly affected by the change in cutoff values were Trend 11 (*i.e.*, likely biological TPs, formed in both biological treatment and in biological post-treatment) and Trend 12 (*i.e.*, “unclassified“ trend), with both seeing decreases in the number of non-target features with this trend in cutoff_02, cutoff_03, and cutoff_04, in relative to the reference scenario, cutoff_01. However, these decreases were very small (0.17±0.02% and 0.35±0.10%, respectively).

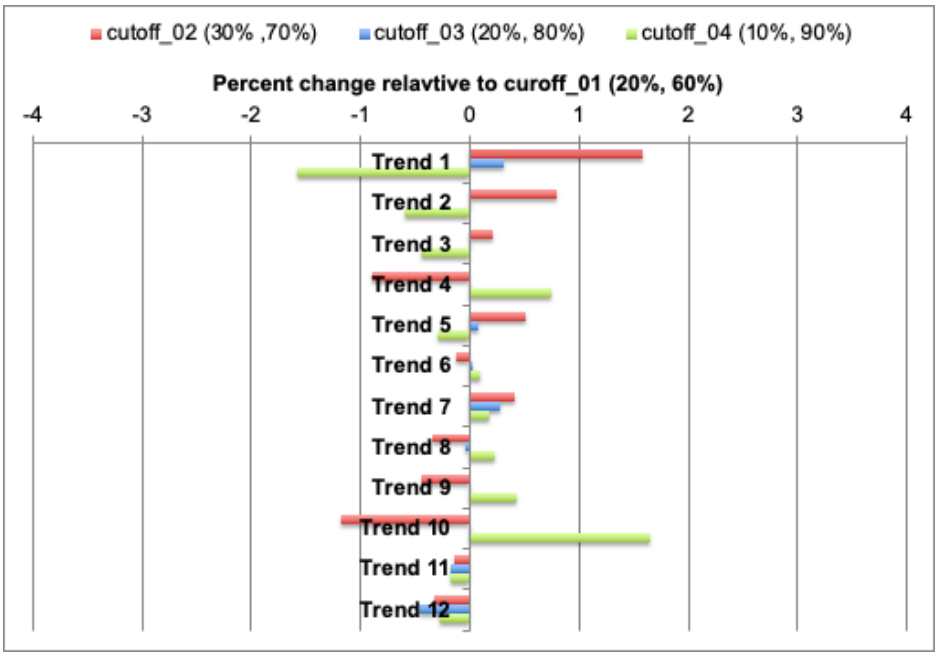


Figure S 19. Tornado plot, visualizing the difference in number of nontarget features in each major trend with different cutoff values in the automated trend assignment algorithm. On the x-axis is percent change relative to cutoff_01 (20%, 60%), while on the y-axis the 12 major trends are listed.

In addition to the percent increase or decrease observed in the number of non-target features per major trend, the number of non-target features per trend was compared across the 4 cutoff scenarios (Figure S 20). No large differences are observable in the distribution of features across the major trends among the different cutoff scenarios. The major trend most affected by the different cutoffs are the number of persistent feautres (*i.e.*, Trend 10), the reasons for which are explained above. In general, the conclusions drawn from cutoff_01 scenario appear to be quite robust and supports its implementation.

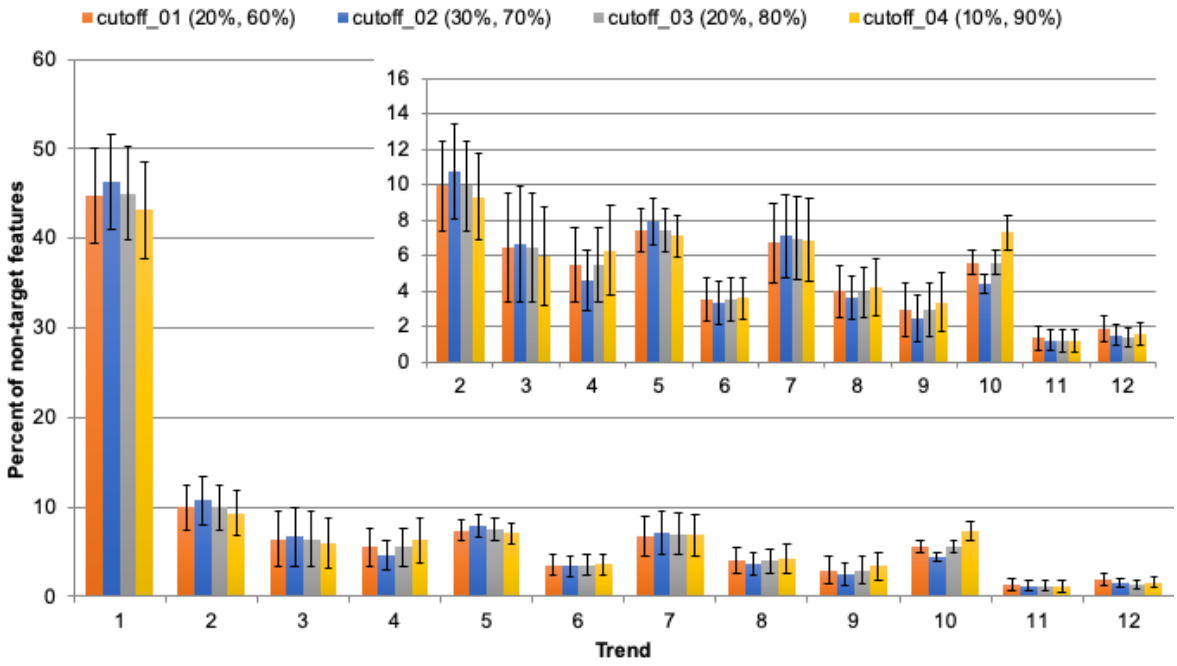


Figure S 20. Barchart of non-target features in each major trend with different cutoff values in automated trend assignment algorithm. Standard deviation are calculated across all 24 sampling dates. On x-axis are the 12 major trends, on the y-axis is the percent of non-target features. In the inset is a close up of Trends 2-12.

Section S6. Results of Trend Analysis

For each non-target profile detected on a sampling date, a minor trend was assigned automatically as described in Section S3. These minor trends were then classified as 1 of 12 major trends, as shown in Table S 5. The number of features with each major trend on each sampling day is shown in Figure S 21 and in

Table S 13.

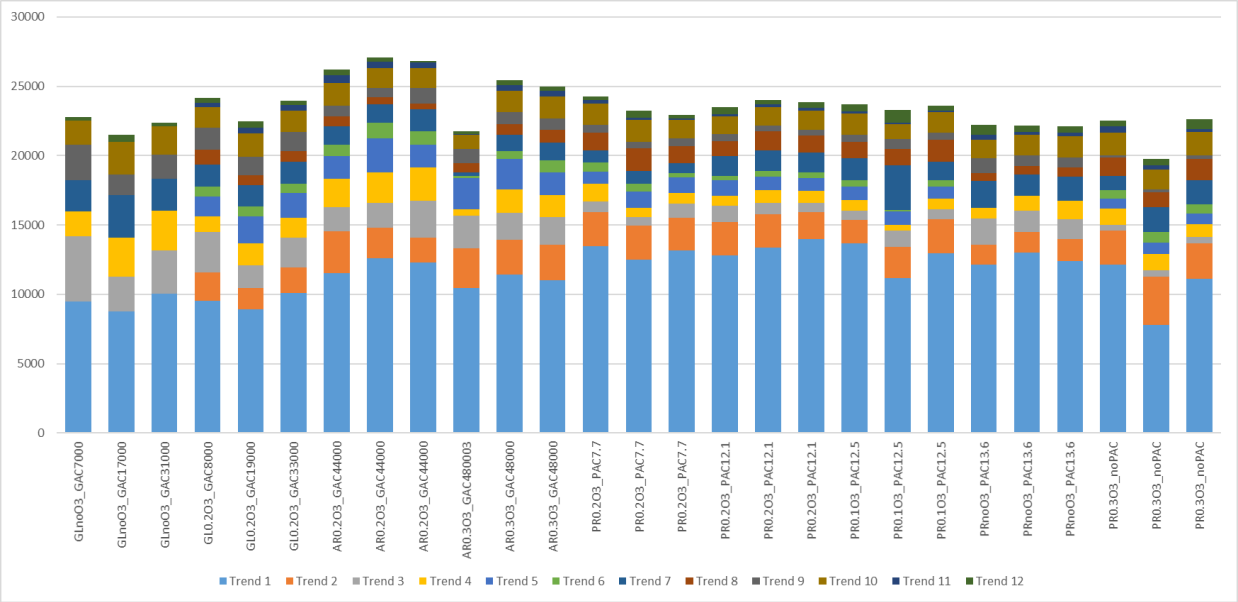


Figure S 21. Distribution of non-target features in trends across 27 sampling dates. On the x-axis is the sample, indicated with the two letter abbreviation for the WWTP, applied ozone dose and GAC or PAC dose. On the y-axis is the number of non-target features and the assigned major trend is colored as indicated in the legend. Due to the influence of rain in the first sampling of $AR_{0.3O3\_GAC44000}$, which is also clearly visible due to the lower number of features detected, this sample was removed from the data set.

Table S 13. Distribution of non-target features in major trends across 27 sampling dates (available in separate Excel).

Table S 14. Summary of non-target features in each step of the WWTP. Reported are number of features, cumulative feature intensity, mean m/z of non-target features and mean retention time (rt) in minutes of non-target features (available in separate Excel).

During the initial data processing, it was observed that a number (7942) of nontarget features were assigned a Trend 5 or Trend 6, which are both related to formation during ozonation, in the samples in PR where no ozonation was applied. These features were therefore assumed to be false positives and were removed from the entire dataset.

Although the assignment of minor trends into major trends simplified the interpretation, it did not illuminate exactly what was happening at each step of WWTP, making it difficult to evaluate the impact of each treatment step. Therefore an approach was applied to consolidate trend further, such that trends of non-target features observed after each treatment step were binned according to their trend in the respective step and is explained here in detail. All non-target features detected in the influent, regardless of assigned trend, were binned to the category “From influent”. For any non-target feature detected after biological treatment, features with a Trend 1 were called “From influent, removed >80%” due to a large intensity decrease during biological treatment. Features with Trend 8, 9, 10 remained in the category “From influent”, since no large intensity change was observed during biological treatment and all other non-target features detected at this sampling point were called “From Biological treatment”, due to their apparent formation during this treatment step. After ozonation, non-target features with Trend 1 and 8 were called “From influent, removed >80%”, since the features with a Trend 8 were removed >80% during ozonation. Features with Trend 9 and 10 were still listed as “From influent”. Features with Trend 3 and 4 were still called “From Biological treatment”, while features with Trend 2 and 11 are now called

“From Biological treatment, removed >80%”, due to their intensity decrease during ozonation. Features in all other Trends were called “From ozonation”. Finally, in the effluent, features with Trend 9 were moved to the “From influent, removed >80%” category and features with Trend 3 were moved to “From Biological treatment, removed >80%”, both due to their elimination during post-treatment. A distinction could then also be made between Trend 5 and Trend 6 features, namely the former were now “From ozonation, removed >80%”, while the latter remained “From ozonation”. Features with Trends 7 and 11 were called “From post-treatment” due to their intensity increase during this treatment step. The assignment of features for each treatment step is summarized in Table S 15.

Table S 15. Binning of major trends in each treatment step. The columns are the four WWTP sampling points (INF – in the influent; BIO – after biological treatment; OZO – after ozonation; EFF – after post-treatment, in the effluent). Rows are the seven defined classes used for visualization (see Figure S22).

	INF	BIO	OZO	EFF
From influent	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	8, 9, 10	9, 10	10
From influent, removed >80%		1	1, 8	1, 8, 9
From biological treatment		2, 3, 4, 5, 6, 7, 11	3, 4	4
From biological treatment, removed >80%			2, 11	2, 3
From ozonation			5, 6, 7	6
From ozonation, removed >80%				5
From post-treatment				7, 11

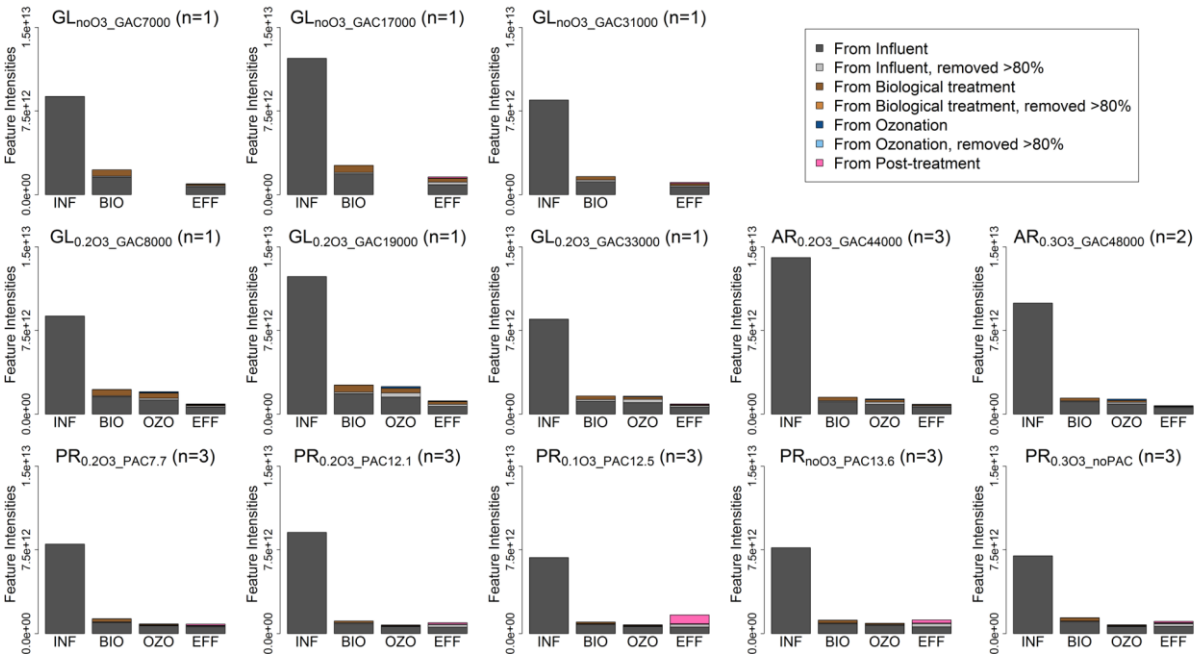


Figure S 22. Cumulative intensity of non-target features at each step of the WWTP (INF: influent; BIO: after biological treatment; OZO: after ozonation; EFF: after post-treatment, effluent). The average for each sample setting is shown; the sample settings are shown in the title of each graph. GL, granular activated carbon (GAC) filtration without pre-ozonation (top row). GL and AR, pre-ozonation followed by GAC filtration (middle row, ordered by increasing GAC bed volumes). PR, pre-ozonation followed by powdered activated carbon (PAC) dosed onto sand filter (bottom row). Colors indicate the fate of features in each treatment step as shown in the legend. Number of samples at each setting shown in the respective title.

Ozonation Transformation Products

Non-target features with either a Trend 5 or Trend 6 were considered possible ozonation transformation products, due to their formation during ozonation. Additionally a suspect screening of known 999 OTPs (SI, Table S8) was used to focus on a wide set of know OTPs. The fate of both the suspect OTPs and non-target OTPs were then assessed in the different post-treatments by comparing the number of well-removed OTPs (*i.e.*, Trend 5) with the number of stable OTPs (*i.e.*, Trend 6) detected.

Table S 16. Number of suspect and non-target OTPs detected on each of the 20 sampling dates were ozonation was applied. Sampling campaign is listed, along with the applied ozone does, the granular activated carbon (GAC) bed volumes (BVs), and the powdered activated carbon (PAC) dose in mg/L. The absolute number of detections and the percent of features that were removed >80% (Trend 5) in the respective post-treatment are given.

Sampling Campaign	Ozone dose ($gO_3/gDOC$)	GAC BV	PAC (mg/L)	Number detected		Percent removal	
				suspect OTPs	non-target OTPs	suspect OTPs	non-target OTPs
$GL_{0.2O3\_GAC8000}$	0.22	8415		43	2134	65	66
$GL_{0.2O3\_GAC19000}$	0.2	19121		55	2674	47	72
$GL_{0.2O3\_GAC33000}$	0.18	32853		53	2447	58	74
$AR_{0.2O3\_GAC44000}$	0.13	44483		49	2455	57	68
$AR_{0.2O3\_GAC44000}$	0.13	44555		43	3579	51	69
$AR_{0.2O3\_GAC44000}$	0.18	44627		51	2627	45	64
$AR_{0.3O3\_GAC48000}$	0.35	48082		54	2745	59	79
$AR_{0.3O3\_GAC48000}$	0.29	48154		53	2493	43	66
$PR_{0.2O3\_PAC7.7}$	0.19		5.8	56	1571	50	58
$PR_{0.2O3\_PAC7.7}$	0.27		8.8	61	1736	54	67
$PR_{0.2O3\_PAC7.7}$	0.24		8.6	63	1443	51	77
$PR_{0.2O3\_PAC12.1}$	0.2		13.6	55	1468	82	78
$PR_{0.2O3\_PAC12.1}$	0.2		9.9	56	1402	77	68
$PR_{0.2O3\_PAC12.1}$	0.21		12.9	50	1297	76	68
$PR_{0.1O3\_PAC12.5}$	0.09		9.3	49	1431	76	70
$PR_{0.1O3\_PAC12.5}$	0.09		11.8	56	1108	89	90
$PR_{0.1O3\_PAC12.5}$	0.08		16.3	55	1349	67	66
$PR_{0.3O3\_noPAC}$	0.29		0	52	1324	40	55
$PR_{0.3O3\_noPAC}$	0.27		0	51	1553	45	51
$PR_{0.3O3\_noPAC}$	0.21		0	57	1385	44	55
Average ± Standard Deviation				**53±5**	**1911±676**	**59±15**	**68±9**

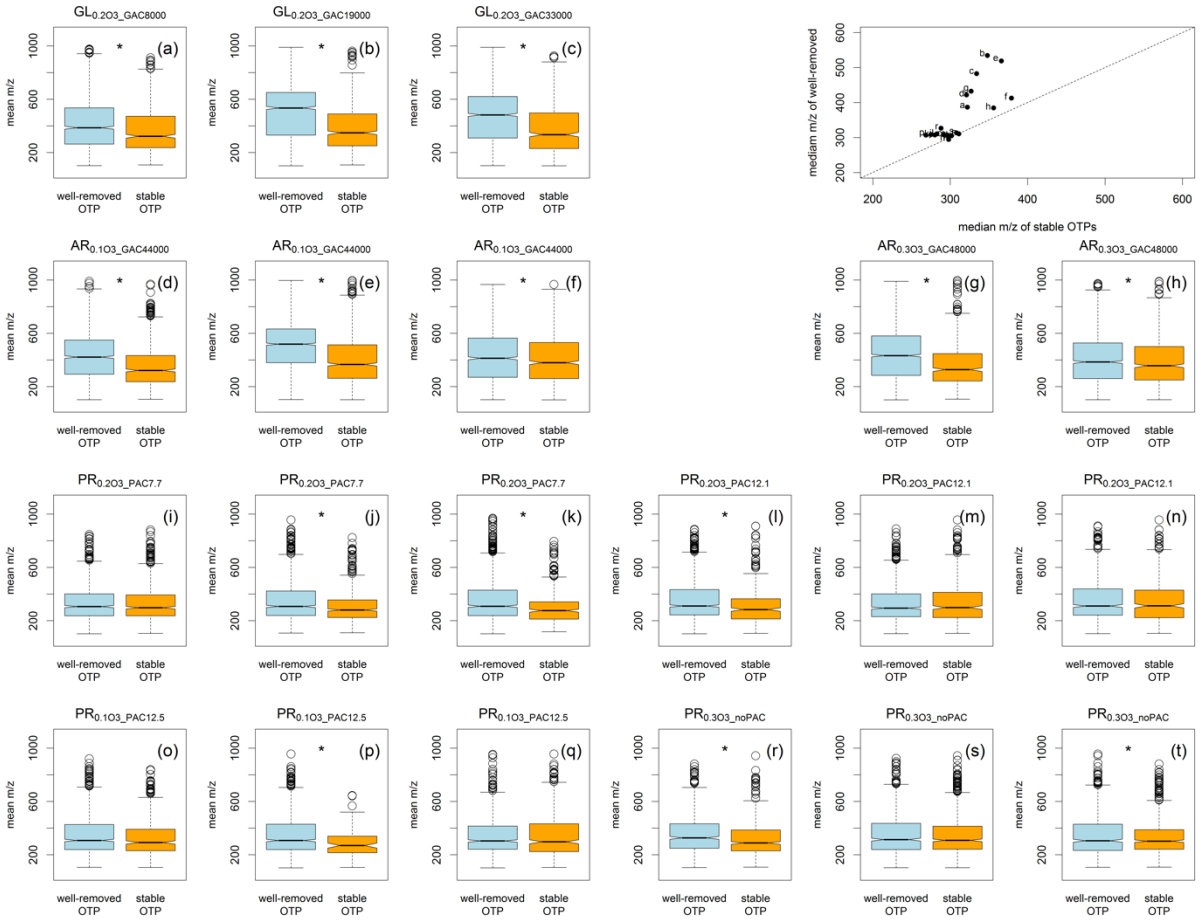


Figure S 23. Comparison of *m/z* values for well-removed OTPs (in blue) and for stable OTPs (in orange) for each sampling date. In the upper right panel is a scatter plot of the median *m/z* of stable OTPs vs. the median *m/z* of well-removed OTPs for each sampling date, as indicated by the letter and the corresponding plot. Signficant differences are indicated by a star (t-test, p<0.05).

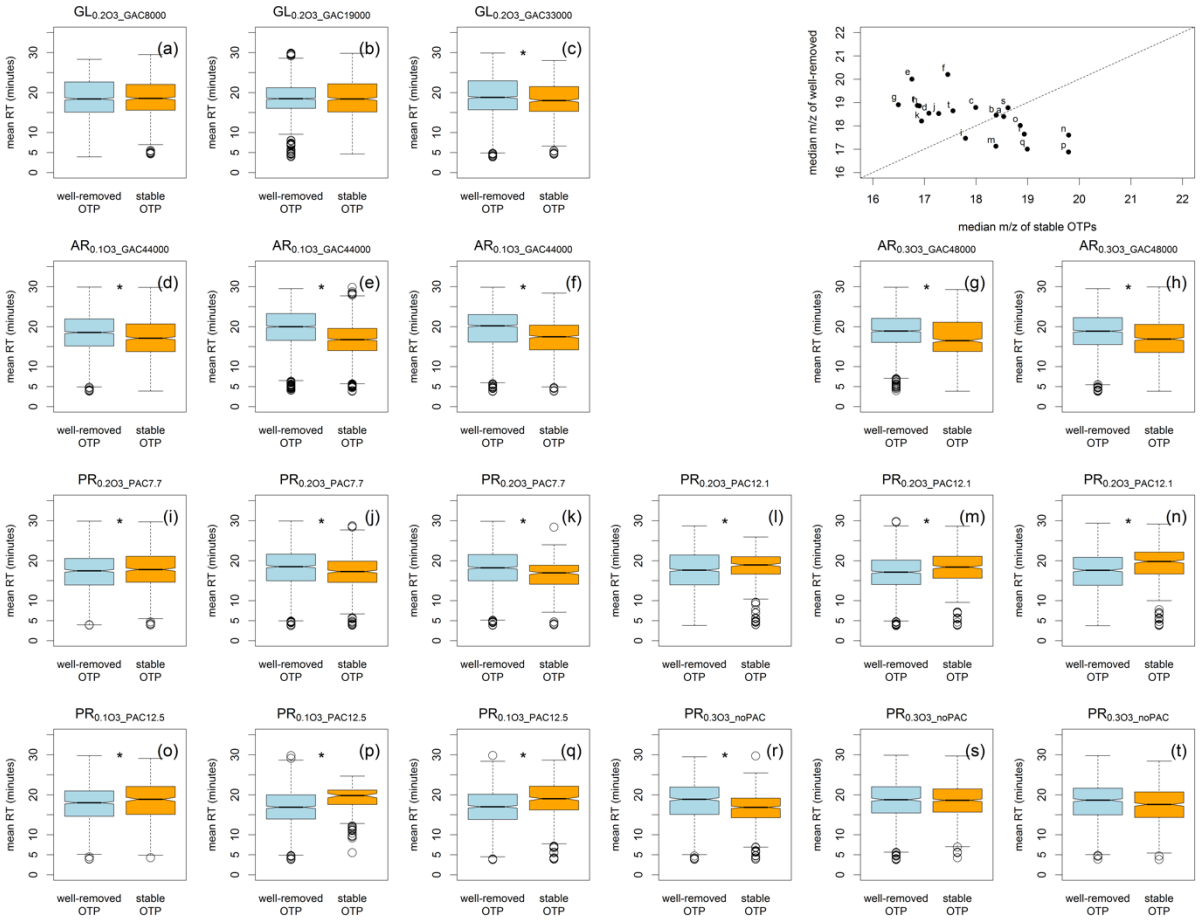


Figure S 24. Comparison of retention times (RT) of well-removed OTPs (in blue) and for stable OTPs (in orange) for each sampling date. In the upper right panel is a scatter plot of the median retention time of stable OTPs vs. the median retention time of well-removed OTPs for each sampling date, as indicated by the letter and the corresponding plot. Signficant differences are indicated by a star (t-test, p<0.05).

Section S7: Linkage Analysis

For a further characterization of the removal of different OTPs, linkage analysis was applied to link potential OTPs to their respective potential parent compounds through a set of 45 known ozonation reactions. Reactions were selected based on transformations known to occur during ozonation (Schollée et al. (2018); Schollée et al. (in prep)). Detections of different reaction types followed similar trends in all sampling dates and the most commonly detected reaction types were hydration, dehydrogenation, hydrogenation, addition of one oxygen, and demethylation (Figure S 25). Hydrogenation, addition of one oxygen, and demethylation was also among the most detected reaction types in a previous ozonation linkage analysis (Schollée et al. 2018). Reactions with an additional of one or two oxygens or carboxylation have been seen to also potentially occur in the transformation of DOM during ozonation (Remucal et al. 2020).

Differences in OTP removal among the post-treatments for the different reaction types were visualized as a scatter plot (Figure S 26a and b). If reactions fall below the 1:1 line, then these OTPs are removed better in the post-treatment plotted on the x-axis and if they are above the 1:1 line, then OTPs of this reaction type are removed better in the post-treatment plotted on the y-axis.Figure S 26a compares OTP removal in the fresh GAC ($GL_{0.2O3\_GAC8000}$, x axis) versus GAC at higher BVs (y axis, see legend). In Figure S 26b, correspondingly, PAC doses are compared to the lowest PAC dose ($PR_{0.2O3\_PAC7.7}$, x axis). Hierarchical cluster analysis (HCA) was also applied, to determine which reaction types have similar patterns with regard to OTP removal in the different post-treatments (Figure 4c). Reaction types were assigned to one of five clusters, as indicated by colors on the left side and number on the right side of the heatmap.

Figure S 26a shows that among the GAC post-treatments, the majority of OTP linkages were best removed with the fresh GAC, as most reaction types from all sampling dates are below the 1:1 line. Some exceptions can be seen, however. OTP types from cluster 1, which included (+O-CHN, $-C_2H_6$, $+O_5$), appeared to be better removed at higher GAC BVs. Similarly, reactions from cluster 5, including +O3-H, $+O_4$-H, and $+O_5$-H, were better removed at higher GAC BVs, especially in the two AR samplings (BV >40,000). Given that these reactions appear to be better removed at higher GAC BVs, we can conclude that for these types of OTPs, biological degradation is the major removal mechanism and/or that they are poorly removed through sorption. This is in line with expectations, since many of the reactions in clusters 1 and 5 include highly oxygenated OTPs. For PR, OTP removal was clearly lower when no PAC was dosed onto the sand filter and clearly higher for almost all reactions at the two settings with higher PAC doses (Figure S 26b). There appeared to be no difference among the different reaction types, which can be explained in that there is no change in the removal mechanisms with higher doses, only more removal capacity.

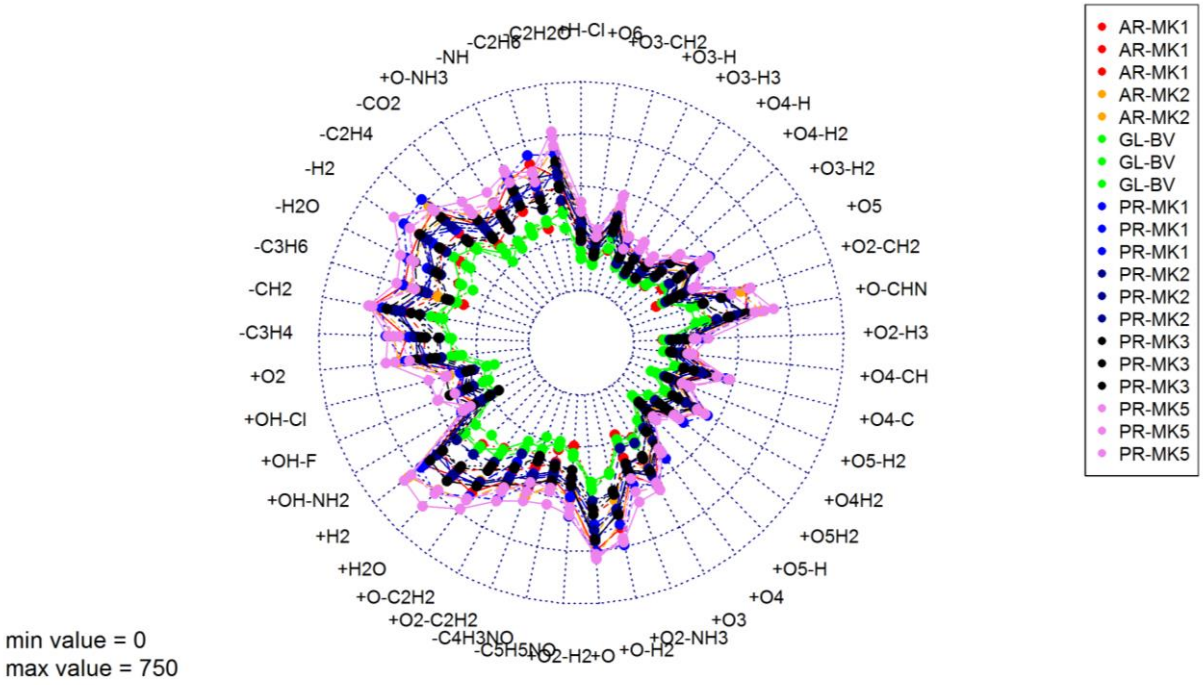


Figure S 25. Distribution of detected reaction types with the linkage analysis. Reaction types are shown along the perimeter of the circle. Detected reactions for each sampling date are shown separately and indicated by color, as shown in the legend.

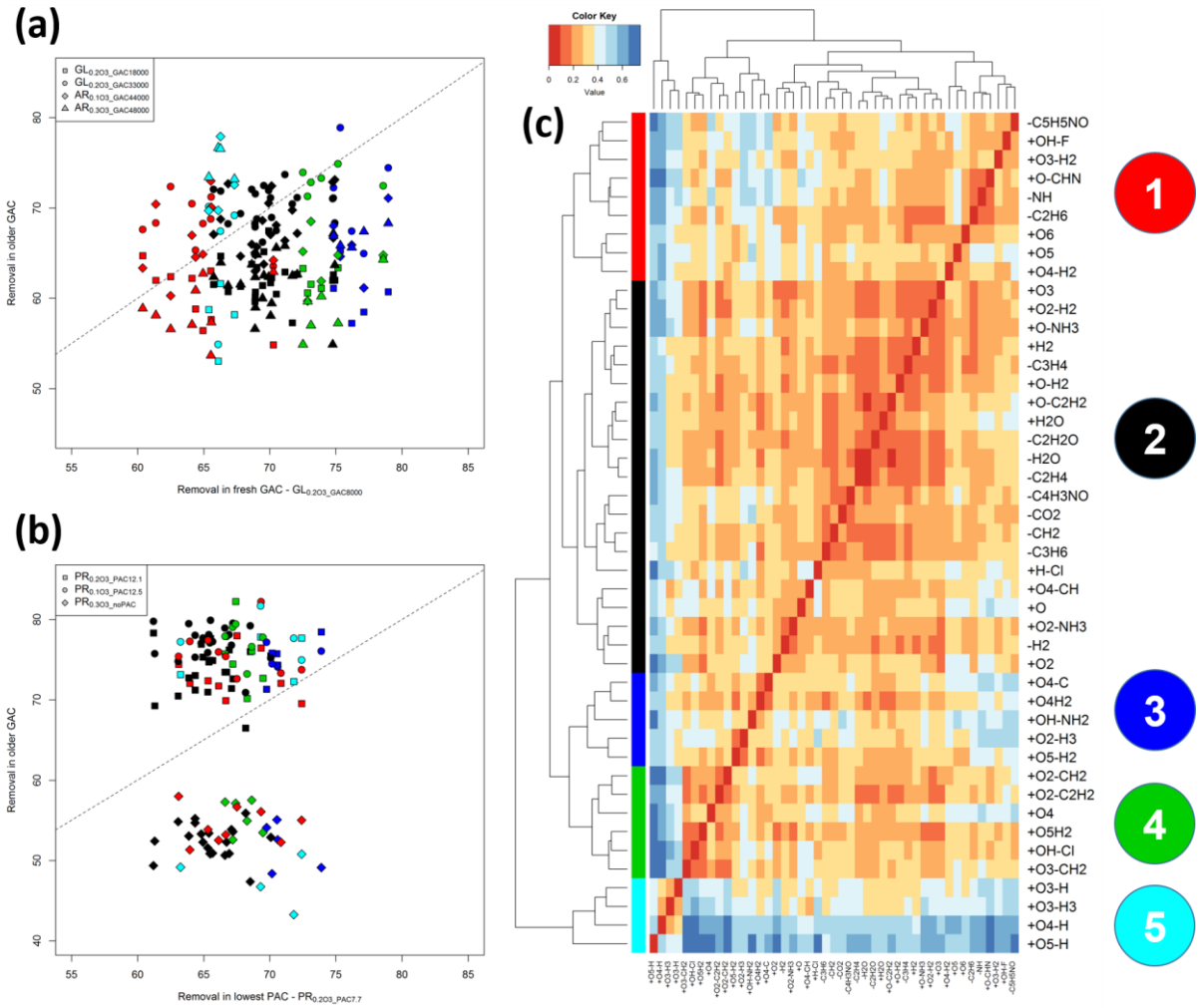


Figure S 26. Removal of potential OTPs of various reactions types in the different post-treatments. In (a, on the left/top), OTP removal in the post-treatment with the freshest GAC ($GL_{0.2O3\_GAC8000}$) is always plotted on the x-axis. On the y-axis is the OTP removal during other GAC post-treatments ($GL_{0.2O3\_GAC18000}$, $GL_{0.2O3\_GAC33000}$, $AR_{0.1O3\_GAC44000}$, $AR_{0.3O3\_GAC48000}$). In (b, on the left/bottom), OTP removal in the PR-MK1 is plotted ($PR_{0.2O3\_PAC7.7}$), versus OTP removal in the three other PAC doses on the y-axis ($PR_{0.2O3\_PAC13.3}$, $PR_{0.2O3\_PAC13.4}$, $PR_{0.3O3\_noPAC}$). In c, a heatmap of the linkage reaction types based on log2 of the ratio between well-removed and stable OTPs at each sampling date. Five clusters were defined and are indicated by color along the left side of the plot and by number and color and number on the right side of the plot. Fromation of clusters appears to be driven by the removal in fresh GAC, as can be seen in the correlation the removal in the fresh GAC (x-axis in (a)) with the defined clusters.

Section S8: References

(R-SIG-DB), R.S.I.G.o.D., Wickham, H. and Müller, K. (2018) DBI: R Database Interface. R package version 1.0.0, https://CRAN.R-project.org/package=DBI.

Abellán, M.N., Gebhardt, W. and Schröder, H.F. (2008) Detection and identification of degradation products of sulfamethoxazole by means of LC/MS and −MSⁿ after ozone treatment. Water Science and Technology 58(9), 1803-1812.

Acero, J.L., Stemmler, K. and von Gunten, U. (2000) Degradation Kinetics of Atrazine and Its Degradation Products with Ozone and OH Radicals: A Predictive Tool for Drinking Water Treatment. Environmental Science & Technology 34(4), 591-597.

Adler, D., Gläser, C., Nenadic, O., Oehlschlägel, J. and Zucchini, W. (2018) ff: Memory-Efficient Storage of Large Data on Disk and Fast Access Functions. R package version 2.2-14, https://CRAN.R-project.org/package=ff.

Albergamo, V., Schollée, J.E., Schymanski, E.L., Helmus, R., Timmer, H., Hollender, J. and de Voogt, P. (2019) Nontarget Screening Reveals Time Trends of Polar Micropollutants in a Riverbank Filtration System. Environmental Science & Technology 53(13), 7584-7594.

Analytics, R. and Weston, S. (2018) iterators: Provides Iterator Construct for R. R package version 1.0.10, https://CRAN.R-project.org/package=iterators.

Andreozzi, R., Caprio, V., Marotta, R. and Vogna, D. (2003) Paracetamol oxidation from aqueous solutions by means of ozonation and H2O2/UV system. Water Research 37(5), 993-1004.

Bache, S.M. and Wickham, H. (2014) magrittr: A Forward-Pipe Operator for R. R package version 1.5, https://CRAN.R-project.org/package=magrittr.

Badawy, A., Mostafa, N., El-Aziz, B., El-Aleem, A. and Nesrine, L. (2011) Stability indicating spectrophotometric method for determination of rosuvastatin in the presence of its acid degradation products by derivative spectrophotometric techniques. J Adv Pharm Res 2, 44-55.

Bader, T., Schulz, W., Kümmerer, K. and Winzenbacher, R. (2016) General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. Analytica Chimica Acta 935, 173-186.

Barron, E., Deborde, M., Rabouan, S., Mazellier, P. and Legube, B. (2006) Kinetic and mechanistic investigations of progesterone reaction with ozone. Water Research 40(11), 2181-2189.

Bates, D. and Maechler, M. (2019) Matrix: Sprace and Dense Matrix Classes and Methods. R package verion 1.2-17., https://CRAN.R-project.org/package=Matrix.

Bengtsson, H. (2003) The R.oo package - Object-Oriented Programming with References Using Standard R Code. Hornik, K., Leisch, F. and Zeileis, A. (eds), pp. ISSN 1609-1395X.

Bengtsson, H. (2019) R.utils: Various Programming Utilities. R package version 2.8.0, https://CRAN.R-project.org/package=R.utils.

Benitez, F.J., Acero, J.L., Real, F.J., Roldán, G. and Rodríguez, E. (2015) Ozonation of benzotriazole and methylindole: Kinetic modeling, identification of intermediates and reaction mechanisms. Journal of Hazardous materials 282, 224-232.

Benner, J. and Ternes, T.A. (2009) Ozonation of Metoprolol: Elucidation of Oxidation Pathways and Major Oxidation Products. Environmental Science & Technology 43(14), 5472-5480.

Bianchini, R.M., Castellano, P.M. and Kaufman, T.S. (2011) Characterization of two new potential impurities of Valsartan obtained under photodegradation stress condition. Journal of Pharmaceutical and Biomedical Analysis 56(1), 16-22.

Bocker, S., Letzel, M., Liptak, Z. and Pervukhin, A. (2006) Decomposing metabolomic isotope patterns. Proc. of Workshop on Algorithms in Bioinformatics (WABI 2006), volume 4175 of Lect Notes Comput Sci, 12 - 23.

Bocker, S., Letzel, M., Liptak, Z. and Pervukhin, A. (2009) SIRIUS: Decomposing isotope patterns for metabolite identification. Bioinformatics 25(2), 218 - 224.

Bocker, S. and Liptak, Z. (2007) A fast and simple algorithm for the Money Changing Problem. Algorithmica 48(4), 413 - 432.

Bocker, S., Liptak, Z., Martin, M., Pervukhin, A. and Sudek, H. (2008) DECOMP--from interpreting mass spectrometry peaks to solving the money changing problem. Bioinformatics 24(4), 591 - 593.

Bogler, S. (2019) Abatement of organic micropollutants during ozonation and activated carbon filtration in WWTP Altenrhein, Swiss Federal Institute of Technology (ETH Zürich), Duebendorf, Switzerland.

Bollmann, A.F., Seitz, W., Prasse, C., Lucke, T., Schulz, W. and Ternes, T. (2016) Occurrence and fate of amisulpride, sulpiride, and lamotrigine in municipal wastewater treatment plants with biological treatment and ozonation. Journal of Hazardous materials 320, 204-215.

Bolstad, B. (2018a) preprocessCore: A Collection of pre-processing functions. R package verion 1.44.0, https://github.com/bmbolstad/preprocessCore.

Bolstad, B.M. (2018b) affyio: Tools for parsing Arrymetrix data files. R package version 1.52.0, https://github.com/bmbolstad/affyio.

Borowska, E., Bourgin, M., Hollender, J., Kienle, C., McArdell, C.S. and von Gunten, U. (2016) Oxidation of cetirizine, fexofenadine and hydrochlorothiazide during ozonation: Kinetics and transformation products. Water Research 94, 350-362.

Boule, P., Meunier, L., Bonnemoy, F., Boulkamh, A., Zertal, A. and Lavedrine, B. (2002) Direct phototransformation of aromatic pesticides in aqueous solution. International Journal of Photoenergy 4, 690475.

Bourgin, M., Beck, B., Boehler, M., Borowska, E., Fleiner, J., Salhi, E., Teichler, R., von Gunten, U., Siegrist, H. and McArdell, C.S. (2018) Evaluation of a full-scale wastewater treatment plant upgraded with ozonation and biological post-treatments: Abatement of micropollutants, formation of transformation products and oxidation by-products. Water Research 129, 486-498.

Broeckling, C.D., Afsar, F.A., Neumann, S., Ben-Hur, A. and Prenni, J.E. (2014) RAMClust: A Novel Feature Clustering Method Enables Spectral-Matching-Based Annotation for Metabolomics Data. Analytical Chemistry 86(14), 6812-6817.

Bryan, J. (2016) cellranger: Translate Spreadsheet Cell Ranges to Rows and Columns. R package version 1.1.0, https://CRAN.R-project.org/package=cellranger.

Bryan, J., Hester, J., Robinson, D. and Wickham, H. (2019) reprex: Prepare Reproducible Example Code via the Clipboard. R package version 0.3.0, https://CRAN.R-project.org/package=reprex.

Calza, P., Medana, C., Raso, E., Giancotti, V. and Minero, C. (2011) Characterization of phenazone transformation products on light-activated TiO2 surface by high-resolution mass spectrometry. Rapid Communications in Mass Spectrometry 25(19), 2923-2932.

Calza, P., Sakkas, V.A., Medana, C., Vlachou, A.D., Dal Bello, F. and Albanis, T.A. (2013) Chemometric assessment and investigation of mechanism involved in photo-Fenton and TiO2 photocatalytic degradation of the artificial sweetener sucralose in aqueous media. Applied Catalysis B: Environmental 129, 71-79.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L. and Mallick, P. (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotech 30(10), 918-920.

Chang, W. (2019a) fastmap: Fast Implementation of a Key-Value Store. R package version 1.0.0, https://r-lib.github.io/fastmap/.

Chang, W. (2019b) R6: Encapsulated Classes with Reference Semantics. R package version 2.4.0, https://CRAN.R-project.org/package=R6.

Chang, W., Cheng, J., Allaire, J., Xie, Y. and McPherson, J. (2019) shiny: Web Application Framework for R. R package version 1.3.0, https://CRAN.R-project.org/package=shiny.

Chen, X., Richard, J., Liu, Y., Dopp, E., Tuerk, J. and Bester, K. (2012) Ozonation products of triclosan in advanced wastewater treatment. Water Research 46(7), 2247-2256.

Cheng, J. (2018) promises: Abstractions for Promise-Based Asynchronous Programming. R package version 1.0.1, https://CRAN.R-project.org/package=promises.

Cheng, J., Bravo, H.C., Ooms, J. and Chang, W. (2019) httpuv: HTTP and WebSocket Server Library. R package verion 1.5.1, https://CRAN.R-project.org/package=httpuv.

Cheng, J. and Chang, W. (2019) later: Utilities for Delaying Function Execution. R package version 0.8.0, https://CRAN.R-project.org/package=later.

Chiaia-Hernández, A.C., Günthardt, B.F., Frey, M.P. and Hollender, J. (2017) Unravelling Contaminants in the Anthropocene Using Statistical Analysis of Liquid Chromatography–High-Resolution Mass Spectrometry Nontarget Screening Data Recorded in Lake Sediments. Environmental Science & Technology 51(21), 12547-12556.

Christophoridis, C., Nika, M.-C., Aalizadeh, R. and Thomaidis, N.S. (2016) Ozonation of ranitidine: Effect of experimental parameters and identification of transformation products. Science of the Total Environment 557–558, 170-182.

Coelho, A.D., Sans, C., Agüera, A., Gómez, M.J., Esplugas, S. and Dezotti, M. (2009) Effects of ozone pre-treatment on diclofenac: Intermediates, biodegradability and toxicity assessment. Science of the Total Environment 407(11), 3572-3578.

Corporation, M. and Weston, S. (2018) doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.14, https://CRAN.R-project.org/package=doParallel.

Couture-Beil, A. (2018) rjson: JSON for R. R package version 0.2.20, https://CRAN.R-project.org/package=rjson.

Csardi, G. (2018) prettyunits: Pretty, Human Readable Formatting of Quantities. R package version 1.0.2, https://CRAN.R-project.org/package=prettyunits.

Csárdi, G. (2017) crayon: Colored Terminal Output. R package 1.3.4, https://CRAN.R-project.org/package=crayon.

Csárdi, G. (2018) pkgconfig: Private Configuration for 'R' Packages. R package version 2.0.2, https://CRAN.R-project.org/package=pkgconfig.

Csárdi, G. (2019) cli: Helpers for Developing Command Lind Interfaces. R package version 1.1.0, https://CRAN.R-package.org/package=cli.

Csardi, G. and Chang, W. (2019) processx: Execute and Control System Processes. R package version 3.3.0, https://CRAN.R-project.org/package=processx.

Csárdi, G. and Chang, W. (2019) callr: Call R from R. R package version 3.2.0, https://CRAN.R-package.org/package=callr.

Csárdi, G., Müller, K. and Hester, J. (2018) desc: Manipulate DESCRIPTION Files. R package version 1.2.0, https://CRAN.R-project.org/package=desc.

Csardi, G., Team, R.D.C., Wickham, H., Chang, W., Flight, R.M., Müller, K. and Hester, J. (2018) sessioninfo: R Session Information. R package version 1.1.1, https://CRAN.R-project.org/package=sessioninfo.

Csardi, G., Wickham, H., Chang, W., Hester, J., Morgan, M. and Tenenbaum, D. (2019) remotes: R Package Installation from Remote Repositories, Including 'GitHub'. R package version 2.0.3, https://CRAN.R-project.org/package=remotes.

Dahl, D.B., Scott, D., Roosen, C., Magnusson, A. and Swinton, J. (2018) xtable: Export Tables to LaTaX or HTML. R package version 1.8-3, https://CRAN.R-project.org/package=xtable.

Dantas, R.F., Canterino, M., Marotta, R., Sans, C., Esplugas, S. and Andreozzi, R. (2007) Bezafibrate removal by means of ozonation: Primary intermediates, kinetics, and toxicity assessment. Water Research 41(12), 2525-2532.

Dantas, R.F., Sans, C. and Esplugas, S. (2011) Ozonation of Propranolol: Transformation, Biodegradability, and Toxicity Assessment. Journal of Environmental Engineering 137(8), 754-759.

DeWitte, B., Dewulf, J., Demeestere, K., Van De Vyvere, V., De Wispelaere, P. and Van Langenhove, H. (2008) Ozonation of Ciprofloxacin in Water: HRMS Identification of Reaction Products and Pathways. Environmental Science & Technology 42(13), 4889-4895.

Diehle, M., Gebhardt, W., Pinnekamp, J., Schaffer, A. and Linnemann, V. (2019) Ozonation of valsartan: Structural elucidation and environmental properties of transformation products. Chemosphere 216, 437-448.

Dodd, M.C., Rentsch, D., Singer, H.P., Kohler, H.-P.E. and Gunten, U.v. (2010) Transformation of β-Lactam Antibacterial Agents during Aqueous Ozonation: Reaction Pathways and Quantitative Bioassay of Biologically-Active Oxidation Products. Environmental Science & Technology 44(15), 5940-5948.

Dutky, S. (2013) bitops: Bitwise Operations. R package version 1.0-6, S original by Steve Dutky initial R port, extensions by Martin Maechler; revised and modified by Steve Dutk, https://CRAN.R-project.org/package=bitops.

Eddelbuettel, D. (2013) Seamless R and C++ Integration with Rcpp, Springer, New York.

Eddelbuettel, D. and Balamuta, J.J. (2017) Extending R with C++: A Brief Introduction to Rcpp. PeerJ Preprints 5, e3188v3181.

Eddelbuettel, D. and Francois, R. (2011) Rcpp: Seamless R and C++ Integration. 2011 40(8), 18.

Eddelbuettel, D., Lucas, A., Tuszynski, J., Bengtsson, H., Urbanek, S., Frasca, M., Lewis, B., Stokely, M., Muehleisen, H., Murdoch, D., Hester, J., Wu, W., Kou, Q., Onkelinx, T., Lang, M., Simko, V., Hornik, K., Neal, R. and Bell, K. (2019) digest: Create Compact Hash Digests of R Objects. R package version 0.6.19, https://CRAN.R-project.org/package=digest.

Favier, M., Dewil, R., Van Eyck, K., Van Schepdael, A. and Cabooter, D. (2015) High-resolution MS and MSn investigation of ozone oxidation products from phenazone-type pharmaceuticals and metabolites. Chemosphere 136, 32-41.

Feng, J., Zheng, Z., Luan, J., Zhang, J. and Wang, L. (2008) Degradation of diuron in aqueous solution by ozonation. Journal of Environmental Science and Health, Part B 43(7), 576-587.

Gagolewski, M. (2019) R package stringi: Character string processing facilities, http://www.gagolewski.com/software/stringi.

Gaslam, B. (2018) fansi: ANSI Control Sequence Aware String Functions. R package version 0.4.0, https://CRAN.R-project.org/package=fansi.

Gatto, L. (2018) ProtGenerics: S4 generic functions for Bioconductor proteomics infrastructure. R package version 1.14.0, https://github.com/lgatto/ProtGenerics.

Gatto, L. and Lilley, K.S. (2011) MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics 28(2), 288-289.

Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A. (2004) affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20(3), 307-315.

Gibb, S. (2015) readMzXmlData: Reads Mass Spectrometry Data in mzXML Format. R package version 2.8.1, https://CRAN.R-project.org/package=readMzXmlData.

Gibb, S. and Strimmer, K. (2012) MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinformatics 28(17), 2270-2271.

GmbH, M.S. (2018a) XLConnect: Excel Connector for R. R package version 0.2-15, https://CRAN.R-project.org/package=XLConnect.

GmbH, M.S. (2018b) XLConnectJars: JAR Dependencies for the XLConnect Package. R package version 0.2-15, https://CRAN.R-project.org/package=XLConnectJars.

Gómez-Ramos, M.d.M., Mezcua, M., Agüera, A., Fernández-Alba, A.R., Gonzalo, S., Rodríguez, A. and Rosal, R. (2011) Chemical and toxicological evolution of the antibiotic sulfamethoxazole under ozone treatment in water solution. Journal of Hazardous materials 192(1), 18-25.

Götz, C., Otto, J. and Singer, H. (2015) Überprüfung des Reinigungseffekts. Auswahl geeigneter organischer Spurenstoffe. Aqua & Gas 95(2), 34-40.

Grolemund, G. and Wickham, H. (2011) Dates and Times Made Easy with lubridate. 2011 40(3), 25.

Guha, R. (2007) Chemical Informatics Functionality in R. 2007 18(5), 16.

Guha, R. (2017) rcdklibs: The CDK Libraries Packaged for R. R package version 2.0, https://CRAN.R-project.org/package=rcdklibs.

Guha, R. (2018) fingerprint: Functions to Operate on Binary Fingerprint Data. R package version 3.5.7, https://CRAN.R-project.org/package=fingerprint.

Gulde, R., Rutsch, M., Clerc, B., Schollée, J.E., von Gunten, U. and McArdell, C.S. (submitted) How to use lab experiments to efficiently identify ozonation transformation products in wastewater treatment.

Hastie, T., Tibshirani, R., Narasimhan, B. and Chu, G. (2018) impute: Imputation for microarray data. R package version 1.56.0.

Henry, L. (2020) lifecycle: Manage the Life Cycle of your Package Functions. R package version 0.2.0, https://CRAN.R-project.org/package=lifecycle.

Henry, L. and Wickham, H. (2018) tidyselect: Select from a Set of Strings. R package version 0.2.5, https://CRAN.R-project.org/package=tidyselect.

Henry, L. and Wickham, H. (2019) purrr: Functional Programming Tools. R package version 0.3.2, https://CRAN.R-project.org/package=purrr.

Henry, L. and Wickham, H. (2020) rlang: Functions for Base Types and Core R and 'Tidyverse' Features. R package version 0.4.5, https://CRAN.R-project.org/package=rlang.

Hester, J. (2019) glue: Interpreted String Literals. R package version 1.3.1, https://CRAN.R-project.org/package=glue.

Hester, J., Müller, K., Ushey, K., Wickham, H. and Chang, W. (2018) withr: Run Code 'With' Temporarily Modified Global State. R package version 2.1.2, https://CRAN.R-project.org/package=withr.

Hester, J. and Wickham, H. (2019) fs: Cross-Platform File System Operations Based on 'libuv'. R package version 1.2.7, https://CRAN.R-project.org/package=fs.

Hörsing, M., Kosjek, T., Andersen, H.R., Heath, E. and Ledin, A. (2012) Fate of citalopram during water treatment with O3, ClO2, UV and fenton oxidation. Chemosphere 89(2), 129-135.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K.D., Irizarry, R.A., Lawrence, M., Love, M.I., MacDonald, J., Obenchain, V., Oleś, A.K., Pagès, H., Reyes, A., Shannon, P., Smyth, G.K., Tenenbaum, D., Waldron, L. and Morgan, M. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods 12(2), 115-121.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. and Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18(suppl_1), S96-S104.

Hübner, U., Seiwert, B., Reemtsma, T. and Jekel, M. (2014) Ozonation products of carbamazepine and their removal from secondary effluents by soil aquifer treatment – Indications from column experiments. Water Research 49, 34-43.

Inc., R.a. (2017) htmltools: Tools for HTML. R package version 0.3.6, https://CRAN.R-project.org/package=htmltools.

Keen, O.S., Ferrer, I., Michael Thurman, E. and Linden, K.G. (2014) Degradation pathways of lamotrigine under advanced treatment by direct UV photolysis, hydroxyl radicals, and ozone. Chemosphere 117(Supplement C), 316-323.

Krahnstöver, T., Wintgens, T. and Deiniger, P. (2018) Spurenstoffentfernung durch die Kombination von Ozonung und Pulveraktickohleadsorption mit anschliessender Raumfiltration ("Aktifilt Plus"), Fachhochschule Nordwestschweiz Hochschule für Life Sciences, Basel, Switzerland.

Kuang, J., Huang, J., Wang, B., Cao, Q., Deng, S. and Yu, G. (2013) Ozonation of trimethoprim in aqueous solution: Identification of reaction products and their toxicity. Water Research 47(8), 2863-2872.

Kuhn, M., Wickham, H. and Vaughan, D. (2018) generics: Common S3 Generics not Provided by Base R Methods Related to Model Fitting. R package version 0.0.2, https://CRAN.R-project.org/package=generics.

Lajeunesse, A., Blais, M., Barbeau, B., Sauvé, S. and Gagnon, C. (2013) Ozone oxidation of antidepressants in wastewater –Treatment evaluation and characterization of new by-products by LC-QToFMS. Chemistry Central Journal 7(1), 1-11.

Lang, D.T. and team, C. (2019a) RCurl: General Network (HTTP/FTP/...) Client Interface for R. R package version 1.95-4.12, https://CRAN.R-project.org/package=RCurl.

Lang, D.T. and team, C. (2019b) XML: Tools for Parsing and Generating XML Within R and S-Plus. R package version 3.98-1.20, https://CRAN.R-project.org/package=XML.

Lang, M. and Team, R.D.C. (2019c) backports: Reimplementations of Function Introduced Since R-3.0.0. R package version 1.1.4, https://CRAN.R-project.org/package=backports.

Lange, F., Cornelissen, S., Kubac, D., Sein, M.M., von Sonntag, J., Hannich, C.B., Golloch, A., Heipieper, H.J., Möder, M. and von Sonntag, C. (2006) Degradation of macrolide antibiotics by ozone: A mechanistic case study with clarithromycin. Chemosphere 65(1), 17-23.

Langfelder, P., Zhang, B. and Horvath, S. (2016) dynamicTreeCut: Methods for Detection of Clusters in Hierarchical Clustering Dendrograms. R package version 1.63-1, https://CRAN.R-project.org/package=dynamicTreeCut.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. (2013) Software for Computing and Annotating Genomic Ranges. PLOS Computational Biology 9(8), e1003118.

Lester, Y., Mamane, H., Zucker, I. and Avisar, D. (2013) Treating wastewater from a pharmaceutical formulation facility by biological process and ozone. Water Research 47(13), 4349-4356.

Lisec, J. (2018) InterpretMSSpectrum: Interpreting High Resolution Mass Spectra. R package version 1.2, https://CRAN.R-project.org/package=InterpretMSSpectrum.

Loden, J., Daeschler, D., Rodola', G. and Csardi, G. (2018) ps: List, Query, Manipulate System Processes. R package version 1.3.0, https://CRAN.R-project.org/package=ps.

Loos, M. (2015) nontarget: Detecting Isotope, Adduct and Homologue Relations in LC-MS Data. R package version 1.9, http://CRAN.R-project.org/package=nontarget.

Loos, M. (2016) enviPick: Peak picking for high resolution mass spectrometry data. R package version 1.5., http://CRAN.R-project.org/package=enviPick.

Loos, M. and Corona, F. (2014) nontargetData: Quantized simulation data of isotope pattern centroids. R package version 1.1, https://CRAN.R-project.org/package=nontargetData.

Loos, M., Gerber, C., Corona, F., Hollender, J. and Singer, H. (2015) Accelerated Isotope Fine Structure Calculation Using Pruned Transition Trees. Analytical Chemistry 87(11), 5738-5744.

Madhavan, J., Grieser, F. and Ashokkumar, M. (2010) Combined advanced oxidation processes for the synergistic degradation of ibuprofen in aqueous environments. Journal of Hazardous materials 178(1), 202-208.

Marotta, R., Spasiano, D., Di Somma, I. and Andreozzi, R. (2013) Photodegradation of naproxen and its photoproducts in aqueous solution at 254 nm: A kinetic investigation. Water Research 47(1), 373-383.

Mawhinney, D.B., Vanderford, B.J. and Snyder, S.A. (2012) Transformation of 1H-Benzotriazole by Ozone in Aqueous Solution. Environmental Science & Technology 46(13), 7102-7111.

McArdell, C.S., Böhler, M., Hernandez, A., Oltramare, C., Büeler, A. and Siegrist, H. (2020) Pilotversuche zur erweiterten Abwasserbehandlung mit granulierter Aktivkohle (GAK) und kombiniert mit Teilozonung (O3/GAK) auf der ARA Glarnerland (AVG), Ergänzende Untersuchungen zur PAK-Dosierung in die biologische Stufe mit S::Select®-Verfahren in Kombination mit nachfolgender GAK, Eawag, Dübendorf, Switzerland.

McDowell, D.C., Huber, M.M., Wagner, M., von Gunten, U. and Ternes, T.A. (2005) Ozonation of Carbamazepine in Drinking Water: Identification and Kinetic Study of Major Oxidation Products. Environmental Science & Technology 39(20), 8014-8022.

Mehta, S., Shah, R.P. and Singh, S. (2010) Strategy for identification and characterization of small quantities of drug degradation products using LC and LC-MS: Application to valsartan, a model drug. Drug Testing and Analysis 2(2), 82-90.

Mestankova, H., Escher, B., Schirmer, K., von Gunten, U. and Canonica, S. (2011) Evolution of algal toxicity during (photo)oxidative degradation of diuron. Aquatic Toxicology 101(2), 466-473.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. and Leisch, F. (2019) e1071: Misc Functions of the Department of Statistics, Probability Theory Group (formerly E1071), TU Wien. R package 1.7-1, https://CRAN.R-project.org/package=e1071.

Miao, H.-F., Cao, M., Xu, D.-Y., Ren, H.-Y., Zhao, M.-X., Huang, Z.-X. and Ruan, W.-Q. (2015) Degradation of phenazone in aqueous solution with ozone: Influencing factors and degradation pathways. Chemosphere 119, 326-333.

Microsoft and Weston, S. (2017) foreach: Provides Foreach Looping Construct for R. R package version 1.4.4, https://CRAN.R-project.org/package=foreach.

Morgan, M. (2018) zlibbioc: An R packaged zlib-1.2.5. R package version 1.28.0, http://bioconductor.org/packages/release/bioc/html/Zlibbioc.html.

Morgan, M. (2019) BiocManager: Access the Bioconductor Project Package Repository. R package version 1.30.10, https://CRAN.R-project.org/package=BiocManager.

Morgan, M., Obenchain, V., Lang, M., Thompson, R. and Turaga, N. (2019) BiocParallel: Bioconductor facilities for parallel evaluation. R package version 1.16.6, https://github.com/Bioconductor/BiocParallel.

Müller, A., Weiss, S.C., Beißwenger, J., Leukhardt, H.G., Schulz, W., Seitz, W., Ruck, W.K.L. and Weber, W.H. (2012) Identification of ozonation by-products of 4- and 5-methyl-1H-benzotriazole during the treatment of surface water to drinking water. Water Research 46(3), 679-690.

Müller, K. (2018a) hms: Pretty Time of Day. R package version 0.4.2, https://CRAN.R-project.org/package=hms.

Müller, K. (2018b) rprojroot: Finding Files in Project Subdirectories. R package version 1.3-2, https://CRAN.R-project.org/package=rprojroot.

Müller, K. and Wickham, H. (2018) pillar: Coloured Formatting for Columns. R package version 1.3.1, https://CRAN.R-project.org/package=pillar.

Müller, K. and Wickham, H. (2019) tibble: Simple Data Frames. R package version 2.1.1, https://CRAN.R-project.org/package=tibble.

Müllner, D. (2013) fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. Journal of Statistical Software 53(9), 1-18.

Nijs, V., Fang, F., LLC, T.T. and Allen, J. (2019) shinyAce: Ace Editor Bindings for Shiny. R package version 0.3.3, https://CRAN.R-project.org/package=shinyAce.

Oehlschlägel, J. (2018) bit: A Class for Vectors of 1-Bit Booleans. R package version 1.1-14, https://CRAN.R-project.org/package=bit.

Oltramare, C., Hernandez, A., Mangold, S., Boehler, M. and McArdell, C.S. (in prep) Abatement of micropollutants in wastewater treatment upgraded with pilot-scale GAC and ozonation/GAC treatment.

Ooms, J. (2014) The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. arXiv:1403.2805 [stat.CO], https://arxiv.org/abs/1403.2805.

Pagès, H., Lawrence, M. and Aboyoun, P. (2018) S4Vectors: S4 implementation of vecotr-like and list-like objects. R package version 0.20.1.

Pedersen, T.L., Petyuk, V.A., Gatto, L. and Gibb, S. (2019) mzID: An mzIdentML parser for R. R package version 1.20.1.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and Team, R.D.C. (2018) _nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-137., https://CRAN.R-project.org/package=nlme.

Radjenović, J., Petrović, M. and Barceló, D. (2009) Complementary mass spectrometry and bioassays for evaluating pharmaceutical-transformation products in treatment of drinking water and wastewater. TRAC Trends in Analytical Chemistry 28(5), 562-580.

Razavi, B., Song, W., Cooper, W.J., Greaves, J. and Jeong, J. (2009) Free-Radical-Induced Oxidative and Reductive Degradation of Fibrate Pharmaceuticals: Kinetic Studies and Degradation Mechanisms. The Journal of Physical Chemistry A 113(7), 1287-1294.

Remucal, C.K., Salhi, E., Walpen, N. and von Gunten, U. (2020) Molecular-Level Transformation of Dissolved Organic Matter during Oxidation by Ozone and Hydroxyl Radical. Environmental Science & Technology.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47-e47.

Robinson, D. and Hayes, A. (2019) broom: Convert Statistical Analysis Objects into Tidy Tibbles. R package version 0.5.2, https://CRAN.R-project.org/package=broom.

Rodayan, A., Roy, R. and Yargeau, V. (2010) Oxidation products of sulfamethoxazole in ozonated secondary effluent. Journal of Hazardous materials 177(1–3), 237-243.

Salgado, R., Pereira, V.J., Carvalho, G., Soeiro, R., Gaffney, V., Almeida, C., Cardoso, V.V., Ferreira, E., Benoliel, M.J., Ternes, T.A., Oehmen, A., Reis, M.A.M. and Noronha, J.P. (2013) Photodegradation kinetics and transformation products of ketoprofen, diclofenac and atenolol in pure water and treated wastewater. Journal of Hazardous materials 244-245, 516-527.

Santoke, H., Song, W., Cooper, W.J., Greaves, J. and Miller, G.E. (2009) Free-Radical-Induced Oxidative and Reductive Degradation of Fluoroquinolone Pharmaceuticals: Kinetic Studies and Degradation Mechanism. The Journal of Physical Chemistry A 113(27), 7846-7851.

Sarkar, D. (2008) Lattice: Multivariate Data Visualization with R, Springer, New York.

Schollée, J.E., Bourgin, M., von Gunten, U., McArdell, C.S. and Hollender, J. (2018) Non-target screening to trace ozonation transformation products in a wastewater treatment train including different post-treatments. Water Research 142, 267-278.

Schollée, J.E., Gulde, R., von Gunten, U. and McArdell, C.S. (in prep) High-throughput suspect screening of predicted ozonation transformation products in wastewater through *in silico* fragmentation.

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P. and Hollender, J. (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environmental Science & Technology 48(4), 2097-2098.

Sideris, T. (2019) Evaluation of organic micro-pollutant abatement in advanced wastewater treatment. Evaluation of organic micro-pollutant abatement in advanced wastewater treatment, Swiss Federal Institute of Technology (ETH Zürich), Duebendorf, Switzerland.

Šojić, D., Despotović, V., Orčić, D., Szabó, E., Arany, E., Armaković, S., Illés, E., Gajda-Schrantz, K., Dombi, A., Alapi, T., Sajben-Nagy, E., Palágyi, A., Vágvölgyi, C., Manczinger, L., Bjelica, L. and Abramović, B. (2012) Degradation of thiamethoxam and metoprolol by UV, O3 and UV/O3 hybrid processes: Kinetics, degradation intermediates and toxicity. Journal of Hydrology 472-473, 314-327.

Song, W., Chen, W., Cooper, W.J., Greaves, J. and Miller, G.E. (2008) Free-Radical Destruction of β-Lactam Antibiotics in Aqueous Solution. The Journal of Physical Chemistry A 112(32), 7411-7417.

Stacklies, W., Redestig, H., Scholz, M., Walther, D. and Selbig, J. (2007) pcaMethods—a bioconductor package providing PCA methods for incomplete data. Bioinformatics 23(9), 1164-1167.

Stephens, J., Simonov, K., Xie, Y., Dong, Z., Wickham, H., Horner, J., reikcoh, Beasley, W., O'Connor, B. and Warnes, G.R. (2018) yaml: Methods to Convert R Data to YAML and Back. R package version 2.2.0, https://CRAN.R-project.org/package=yaml.

Stravs, M.A. (2015) RMassScreening: Suspect screening and time series in LC-HRMS data. R package version 0.2.

Stravs, M.A., Schymanski, E.L., Singer, H.P. and Hollender, J. (2013) Automatic recalibration and processing of tandem mass spectra using formula annotation. Journal of Mass Spectrometry 48(1), 89-99.

Szabó, R.K., Megyeri, C., Illés, E., Gajda-Schrantz, K., Mazellier, P. and Dombi, A. (2011) Phototransformation of ibuprofen and ketoprofen in aqueous solutions. Chemosphere 84(11), 1658-1663.

Tay, K.S., Rahman, N.A. and Abas, M.R.B. (2011) Characterization of atenolol transformation products in ozonation by using rapid resolution high-performance liquid chromatography/quadrupole-time-of-flight mass spectrometry. Microchemical Journal 99(2), 312-326.

Tay, K.S., Rahman, N.A. and Abas, M.R.B. (2012) Ozonation of metoprolol in aqueous solution: ozonation by-products and mechanisms of degradation. Environmental Science and Pollution Research 20(5), 3115-3121.

Tekle-Röttering, A., Reisz, E., Jewell, K.S., Lutze, H.V., Ternes, T.A., Schmidt, W. and Schmidt, T.C. (2016) Ozonation of pyridine and other N-heterocyclic aromatic compounds: Kinetics, stoichiometry, identification of products and elucidation of pathways. Water Research 102, 582-593.

Tierney, L. (2018) codetools: Code Analysis Tools for R. R package version 0.2-16, https://CRAN.R-project.org/package=codetools.

Topalov, A., Molnár-Gábor, D., Kosanić, M. and Abramović, B. (2000) Photomineralization of the herbicide mecoprop dissolved in water sensitized by TiO2. Water Research 34(5), 1473-1478.

Tuszynski, J. (2019) caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc.. R package version 1.17.1.2, https://CRAN.R-project.org/package=caTools.

Urbanek, S. (2013) png: Read and write PNG images. R package version 0.1-7, https://CRAN.R-project.org/package=png.

Urbanek, S. (2015) base64enc: Tools for base64 encoding. R package version 0.1-3, https://CRAN.R-project.org/package=base64enc.

Urbanek, S. (2019) rJava: Low-Level R to Java Interface. R package version 0.9-11, https://CRAN.R-project.org/package=rJava.

Ushey, K., Allaire, J., Wickham, H. and Ritchie, G. (2019) rstudioapi: Safely Access the RStudio API. R package version 0.10, https://CRAN.R-project.org/package=rstudioapi.

Venables, W.N. and Ripley, B.D. (2002) Modern Applied Statistics with S., Springer, New York.

Vogna, D., Marotta, R., Napolitano, A., Andreozzi, R. and d'Ischia, M. (2004) Advanced oxidation of the pharmaceutical drug diclofenac with UV/H2O2 and ozone. Water Research 38(2), 414-422.

von Sonntag, C. and von Gunten, U. (2012) Chemistry of Ozone in Water and Wastewater Treatment: From Basic Principles to Applications, IWA Pub.

Wand, M. (2015) KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995). R package 2.23-15, https://CRAN.R-project.org/package=KernSmooth.

Warnes, G.R., Bolker, B., Bonebakker, B., Gentleman, R., Liaw, W.H.A., lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M. and Venables, B. (2019) gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.1, https://CRAN.R-project/org/package=gplots.

Warnes, G.R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., MacQueen, D., Magnusson, A., Rogers, J. and others (2017) gdata: Various R Programming Tools for Data Manipulation. R package version 2.18.0, https://CRAN.R-project.org/package=gdata.

Warnes, G.R., Bolker, B. and Lumley, T. (2018) gtools: Various R Programming Tools. R package version 3.8.1, https://CRAN.R-project.org/package=gtools.

Weston, S. and Wickham, H. (2014) itertools: Iterator Tools. R package version 0.1-3, https://CRAN.R-project.org/package=itertools.

Wickham, C. (2018a) munsell: Utilities for Using Munsell Colours. R package version 0.5.0, https://CRAN.R-project.org/package=munsell.

Wickham, H. (2007) Reshaping Data with the reshape Package. 2007 21(12), 20.

Wickham, H. (2011a) The Split-Apply-Combine Strategy for Data Analysis. 2011 40(1), 29.

Wickham, H. (2011b) testthat: Get Started with Testing. The R Journal 3, 5-10.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York.

Wickham, H. (2017) tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1, https://CRAN.R-project.org/package=tidyverse.

Wickham, H. (2018b) httr: Tools for Working with URLs and HTTP. R package version 1.4.0, https://CRAN.R-project.org/package=httr.

Wickham, H. (2018c) scales: Scale Functions for Visualization. R package version 1.0.0, https://CRAN.R-project.org/package=scales.

Wickham, H. (2018d) vctrs: Vector Helpers. R package version 0.1.0, https://CRAN.R-project.org/package=vctrs.

Wickham, H. (2019a) assertthat: Easy Pre and Post Assertions. R package version 0.2.1, https://CRAN.R-project.org/package=assertthat.

Wickham, H. (2019b) ellipsis: Tools for Working with .... R package version 0.1.0, https://CRAN.R-project.org/package=ellipsis.

Wickham, H. (2019c) forcats: Tools for Working with Categorical Variables (Factors). R package version 0.4.0, https://CRAN.R-project.org/package=forcats.

Wickham, H. (2019d) lazyeval: Lazy (Non-Standard) Evaluation. R package version 0.2.2, https://CRAN.R-project.org/package=lazyeval.

Wickham, H. (2019e) modelr: Modelling Functions that Work with the Pipe. R package version 0.1.5, https://CRAN.R-project.org/package=modelr.

Wickham, H. (2019f) rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.3, https://CRAN.R-project.org/package=rvest.

Wickham, H. (2019g) stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0, https://CRAN.R-project.org/package=stringr.

Wickham, H. and Bryan, J. (2019a) readxl: Read Excel Files. R package version 1.3.1, https://CRAN.R-project.org/package=readxl.

Wickham, H. and Bryan, J. (2019b) usethis: Automate Package and Project Setup. R package version 1.5.0, https://CRAN.R-project.org/package=usethis.

Wickham, H., Francois, R., Henry, L. and Müller, K. (2019a) dplyr: A Grammer of Data Manipulation. R package version 0.8.0.1, https://CRAN.R-project.org/package=dplyr.

Wickham, H. and Henry, L. (2019) tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. R package version 0.8.3, https://CRAN.R-project.org/package=tidyr.

Wickham, H. and Hester, J. (2019) pkgbuild: Find Tools Needed to Build R Packages. R package version 1.0.3, https://CRAN.R-project.org/package=pkgbuild.

Wickham, H., Hester, J. and Chang, W. (2018a) pkgload: Simulate Package Installation and Attach. R package version 1.0.2, https://CRAN.R-project.org/package=pkgload.

Wickham, H., Hester, J. and Chang, W. (2019b) devtools: Tools to Make Developing R Packages Easier. R package version 2.0.2, https://CRAN.R-project.org/package=devtools.

Wickham, H., Hester, J. and Francois, R. (2018b) readr: Read Rectangular Text Data. R package version 1.3.1, https://CRAN.R-project.org/package=readr.

Wickham, H., Hester, J., Müller, K. and Cook, D. (2017) memoise: Memoisation of Functions. R package version 1.1.0, https://CRAN.R-project.org/package=memoise.

Wickham, H., Hester, J. and Ooms, J. (2018c) xml2: Parse XML. R package version 1.2.0, https://CRAN.R-project.org/package=xml2.

Wickham, H. and Miller, E. (2019) haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.1.0, https://CRAN.R-project.org/package=haven.

Wickham, H. and Pedersen, T.L. (2019) gtable: Arrange 'Grobs' in Tables. R package version 0.3.0, https://CRAN.R-project.org/package=gtable.

Wickham, H. and Ruiz, E. (2019) dbplyr: A 'dplyr' Back End for Databases. R package version 1.4.2, https://CRAN.R-project.org/package=dbplyr.

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73(1), 3-36.

Xie, Y. (2018) mime: Map Filenames to MIME Types. R package version 0.6, https://CRAN.R-project.org/package=mime.

Zeileis, A., Fischer, J.C., Hornik, K. and Ihaka, R. (2019) colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes. Archive, a.o.E.-P. (ed), http://arxiv.org/abs/1903.06490.

Zimmermann, S.G., Schmukat, A., Schulz, M., Benner, J., Gunten, U.v. and Ternes, T.A. (2012) Kinetic and Mechanistic Investigations of the Oxidation of Tramadol by Ferrate and Ozone. Environmental Science & Technology 46(2), 876-884.