Pemberton, J. A., Lloyd, C. E. M., Arthur, C. J., Johnes, P. J., Dickinson, M., Charlton, A. J., & Evershed, R. P. (2019). Untargeted characterisation of dissolved organic matter contributions to rivers from anthropogenic point sources using direct infusion- and high-performance liquid chromatography-Orbitrap mass spectrometry. *Rapid Communications in Mass Spectrometry*. https://doi.org/10.1002/rcm.8618

Untargeted characterisation of dissolved organic matter contributions to rivers from anthropogenic point sources using direct infusion- and high-performance liquid chromatography-Orbitrap mass spectrometry

Jonathan A. Pemberton[1], Charlotte E. M. Lloyd[1], Christopher J. Arthur[2], Penny J. Johnes[3], Michael Dickinson[4], Adrian J. Charlton[4], Richard P. Evershed[1,*]

[1]Organic Geochemistry Unit, School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK
[2]School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK
[3]School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK
[4]Fera Science Ltd, Sand Hutton, York, YO41 1LZ, UK

*Corresponding author: r.p.evershed@bristol.ac.uk

RATIONALE: Anthropogenic organic inputs to freshwaters can exert detrimental effects on aquatic ecosystems, raising growing concern for both environmental conservation and water security. Current regulation by the EU water framework directive (European Union, 2000/60/EC) relates to organic pollution by monitoring selected micropollutants, however, aquatic ecosystem responses requires a comprehensive understanding of dissolved organic matter (DOM) composition.[1] The introduction of high-resolution mass spectrometry (HRMS) is set to greatly increase our understanding of the composition of DOM of both natural and anthropogenic origin derived from diffuse and point sources.

METHODS: DOM was extracted from riverine and treated sewage effluent using solid phase extraction (SPE) and analysed using dissolved organic carbon (DOC) analysis, direct infusion-high resolution Orbitrap™ mass spectrometry (DI-HRMS) and high-performance liquid chromatography (HPLC/HRMS). The data obtained were analysed using univariate and multivariate statistics to demonstrate differences in background DOM, anthropogenic inputs and in-river mixing. Compound identifications were achieved based on MS^2 spectra searched against on-line databases.

RESULTS: DI-HRMS spectra showed the highly complex nature of all DOM SPE extracts. Classification and visualisation of extracts containing many thousands of individual compounds were achieved using PCA and hierarchical cluster analysis. Kruskal-Wallis analyses highlighted significant discriminating ions originating from the sewage treatment works for more in-depth investigation by HPLC/HRMS. The generation of MS^2 spectra in HPLC/HRMS provided the basis for identification of anthropogenic compounds including; pharmaceuticals, illicit drugs, metabolites and polymers, although many thousands of compounds remain unidentified.

CONCLUSIONS: This new approach enables comprehensive analysis of DOM in extracts without any preconceived ideas of the compounds which may be present. This approach has the potential to be used as a high throughput, qualitative, screening method to determine if the composition of point sources differs from that of the receiving water bodies, providing a new approach to the identification of hitherto unrecognised organic contribution to water bodies.

Keywords: Dissolved organic matter, sewage effluent, DI-HRMS, HPLC-HRMS, data visualisation, difference algorithm.

Introduction

Fresh surface water is a fundamental resource not only for drinking water and irrigation, but also for supporting terrestrial and aquatic ecosystems.[2] DOM is ubiquitous to all aquatic systems and is an extremely complex mixture of organic compounds although its composition has remained intractable due to the lack of suitable analytical methods.[3] DOM has been asserted to be a nutrient for autotrophs.[2,3] The range of compounds comprising DOM includes compounds generated naturally and through anthropogenic activities, and can include potentially toxic micropollutants which attract much attention in water quality legislation as these have been shown to have adverse impacts upon organisms within the aquatic ecosystems.[4–6] Despite individual anthropogenically derived compounds being at low concentration, the chronic exposure of stream biota to these compounds has been shown to have a wide range of acute ecotoxicological and chronic adverse effects on organisms.[4–7] These include the disruption of reproduction,[8,9] a reduction in biodiversity,[10] and, dysmorphia in the maturation of organisms.[11] Furthermore, different compounds which affect organisms in a similar way can work synergistically amplifying the impact.[7,12] Regulations only cover a very minor proportion of the commonly identified micropollutants, and many others almost certainly remain to be discovered.

Micropollutants have been identified in different discharges including sewage treatment works.[13,14] Sewage treatment works have been found to be a major gateway for the release of pharmaceuticals,[15] personal care products[16] and plasticisers[17] into the environment. The concentration and presence/absence of target compounds across different sewage treatment works has been found to vary between different sites and over time.[13,15,16,18,19] With over 9000 sewage treatment works in the UK and numerous other point sources the identification of potentially ecotoxicological compounds remains a challenge. Without identifying these micropollutants; the determination of ecotoxicity, effective mitigation solutions and environmental monitoring cannot be carried out.

The most common approach to the determination of organic compounds in both wastewater and the natural aquatic environment ecosystem is targeted analysis using MS approaches focusing on known or suspected compounds.[4,20] Optimised extraction methods are used to isolate and concentrate the target analytes with subsequent interrogation involving gas chromatography (GC)[21,22] or high-performance

liquid chromatography (HPLC)[13,23] linked MS. Targeted studies have largely focused on pharmaceuticals,[15,24] personal care products[16,25] and pesticides,[21,26] with their concentrations or load in the riverine environment being used to assess the effectiveness of sewage treatment and local sources.[13,15,27] The obvious limitation of targeted analysis is that it requires a predetermined list of known compounds. Targeted analysis will only determine the selected compounds and exclude other compounds originating from a point source or the environment. The use of electrospray ionisation (ESI) and HRMS has revolutionised the analysis of complex mixtures of water soluble compounds, such as DOM, allowing the exact mass of individual molecular ions to be determined.[28,29] The ionisation of intact molecules and their mass analysis using instruments with high resolving power and high mass accuracy means that each ion in a spectrum potentially corresponds to a unique compound (taking account of other adducts and isotopes). Application of this approach has revealed the extraordinary complexity and heterogeneity of DOM in the natural environment, as evidenced by the DI-HRMS spectra containing many thousands of resolved ions.[30,31]

One of the major challenges of utilising these HR mass spectra of DOM lies in the interrogation of the data. Attempts have been made to assign formulae to the observed ions in the spectra, using rule-based calculations.[32–34] All studies include carbon, oxygen, nitrogen and hydrogen, however, the inclusion of heteroatoms, e.g. P, Cl and S vary between studies.[30,35,36] Increasing the number of heteroatoms results in an exponential increase in the number of possible formulae for a single ion, resulting in a high level of uncertainty and false positives.[33] Isotopes and adducts, i.e. $[M+Na]^+$, $[M+K]^+$, $[M+Cl]^-$, will be present in all DI-HRMS spectra, but are rarely accounted for. Hence, despite the high mass resolution attainable using modern Fourier-transform ion cyclotron resonance (FTICR) or Orbitrap™ MS instruments, the exceptional complexity of the mass spectra obtained largely defies conventional approaches to handling these unusual data sets.

An alternative approach is to move toward data visualisation rather than more conventional peak identification approaches. One such approach is the use of van Krevelen diagrams. Such diagrams use the ratios of carbon:hydrogen and carbon:oxygen of the formulae assigned to ions as a basis for the comparison of DOM in water extracts. [37–41] These elemental ratios of formulae are used to classify ions to a compound class.[30,31,40,41] However, the interpretation of a van Krevelen diagram relies on the correct assignment of formulae, including appropriate numbers of heteroatoms. Incorrect assignments will lead to the inaccurate interpretations of differences in the composition of DOM extracts. Furthermore, a single ion in a DI-HRMS spectrum maybe the result of multiple isomers and therefore, the full complexity is not fully revealed. In addition, the correct classification using a van Krevelen diagram of a compound class for one isomer maybe incorrect for another isomer with the same formulae. Despite this van Krevelen diagrams have found utility in visualising differences in composition of DOM extracts

from different aquatic systems, addressing a range of questions relating DOM source and variability between ecosystems, e.g. differences between water bodies in different geographical locations.[38,42,43]

While van Krevelen diagrams have proved useful for visualising differences between DOMs extract chemistries, the approach is non-statistical and is rather restricted in truly exploiting the full complexity of the data, e.g. ion intensities and molecular species of unassigned formulae. An alternative, but still less widely applied approach, is multivariate statistics, in particular PCA of DI-HRMS spectra. The latter has been used to determine and visualise differences between the composition of DOM extracts from different SPE extraction methods[44] and different water bodies within the same pristine catchment.[42] PCA requires only the detected ions and their intensities in different DI-HRMS spectra to determine if extracts are different. However, this has not been applied to point sources in comparison to their receiving environment.

Herein, we address the challenge of how to deal with the question of the complexity of riverine DOM analysis by HRMS. We have taken a comprehensive approach in order to retain a broad view of DOM composition and developed a method for data reduction based on a difference algorithm to highlight complex anthropogenic DOM contributions against a natural or semi-natural DOM background. To achieve this, we first record DI-HRMS spectra of DOM recovered by SPE, then use PCA as a rapid qualitative screening method to determine if differences exist between DOM extracts of point sources and the receiving aquatic environment. Following this the difference algorithm, employing univariate statistics (Kruskal-Wallis analysis), was applied to allow the anthropogenic point source components to be identified in DI-HRMS spectra. Heatmaps and hierarchical cluster analysis are then used as data visualisation tools, which allow compositional differences to be recognised. The anthropogenic components highlighted through the untargeted difference analysis formed the basis for structural identification of specific molecular species by HPLC/HRMS/MS.

Experimental

Sampling

The sewage treatment works (at 51° 21' 21.8052" N, 2° 37' 2.262" W) is situated on the River Chew in Somerset, UK, which drains NW from its source at Chewton Mendip to the sewage works, and then NE to its confluence with the Bristol Avon at Keynsham (at 51° 25' 7.7196" N, 2° 29' 32.118" W). It is located downstream of Chew Valley Lake, a significant reservoir supplying water to the city of Bristol, UK. This site was selected to test and develop this method as it was close enough to the University of Bristol to allow rapid stabilisation of samples in cold storage following collection in the field (details of which are given below). The treatment methods used at the sewage treatment works include: (i) Preliminary solid removal of large particulates, (ii) Primary settling to further allow smaller particulates to flocculate and settle out of the water, (iii) Secondary treatment using trickle filter beds to biologically

break-down organic matter. The sewage treatment works has a tertiary treatment which includes phosphorous stripping. The final treated effluent is discharged into the river downstream of the reservoir.

Three comparative water samples (5.25 L) were collected in amber glass bottles. The first was taken ca. 60 m upstream of the sewage treatment outfall. The second was taken directly from the discharging sewage outfall and the third ca. 50 m downstream. Five procedural controls of HPLC grade water (1 L, Fischer Scientific, Loughborough, UK) were extracted with the water samples collected. The water was divided into 1 L aliquots, which were vacuum filtered using an all glass filter apparatus (47 mm, Merck Millipore, Feltham, UK) through glass fibre filters (0.5 µm, 47 mm, Advantec, Cole-Palmer, Hanwell, UK) within 24 hrs of collection. Both the filter and filtration apparatus were pre-combusted before use (450 ^{o}C, 4 h). An additional 20 ml of each water sample was filtered using the same apparatus and retained to determine the concentration of DOC. The filtered water samples (1 L) were acidified to pH 2 using hydrochloric acid (30%, TraceSelect, Sigma–Aldrich, Dorset, UK) and extracted using Oasis Hydrophilic-Lipophillic Balance (HLB) solid phase extraction cartridges (SPE, 400 mg bed mass, 60 µm particle size, Waters Ltd, Elstree, UK). The cartridges were conditioned using HPLC grade methanol (3 ml, Rathburn Chemicals Ltd. Walkerburn, UK) and HPLC grade water (3 mL, Fischer Scientific) before the acidified filtered water (1 L) was extracted. After extraction, the cartridges were rinsed with acidified HPLC grade water (3 mL, Fischer Scientific) and dried under vacuum for 30 min. The extracts were eluted from the SPE cartridges with HPLC grade methanol (6 x 1 mL, Rathburn Chemicals Ltd.) and dried under a steady stream of nitrogen. Dried extracts were dissolved in a mixture of HPLC grade methanol/water (1:1, *v/v*, 1 mL, Rathburn Chemicals Ltd**.,** Fischer Scientific).

An aliquot of each extract (100 µL) was mixed to create a pooled quality control (QC) and an aliquot of each extract (50 µL) was removed and dried under a steady stream of nitrogen for DOC analysis. The pooled QC and all extracts were then stored at –85 ^{o}C until required for analysis.

DOC analysis

The dried 50 µl aliquots of the extracts were dissolved in water (20 ml, MilliQ) before DOC analysis. Filtered water samples were analysed directly. All analyses were carried out using a Shimadzu TOC-L analyser using the non-purgeable organic carbon (NPOC) method recommended by Shimadzu for the analysis of environmental water samples. The mean of three to five injections of 150 µL, where the coefficient of variance for replicate injections was < 2%. The results are is presented in Table 1.

DI-HRMS analysis

DI-HRMS were recorded in positive ion mode using an Orbitrap™ Elite Hybrid Ion Trap-Orbitrap™ Mass Spectrometer (Thermo Scientific, Hemel Hempstead, UK) with a heated electrospray ionisation source (HESI).The instrument was calibrated using Thermo Scientific Pierce LTQ ESI Positive Ion

Calibration Solution. The instrument was calibrated and had a mass error of 3.2 ppm and resolution of m/Δm 197,389 at *m/z* 524.257 and upon tuning the S-lens radio frequency level was 61.81 %. Extracts were directly infused at a rate of 5 μL min^{-1} into the HESI. The source voltage was set to 3.0 kV, sheath gas (nitrogen) flow rate to 10 arbitrary units (arb), the auxiliary gas (nitrogen) flow rate to 5 arb and the sweep gas (nitrogen) flow rate to 5 arb and capillary temperature to 275 ^{o}C. The mass spectrometer was set to acquire in the mass range of *m/z* 150 to 2000 for 100 scans, and the ions detected were recorded in profile using the nominal resolving power "240,000". The maximum injection time was set at 200 ms and the automatic gain control (AGC) target was set to 1,000,000. The TIC was assessed for any losses in signal during analysis. Extracts were analysed in random order. The mixed QC and calibration solution was analysed after every 5 extracts the mass drift was 1.8 ppm over all analyses.

HPLC/HRMS and HPLC/HRMS/MS analysis

The SPE extracts (10 μL) were analysed using HPLC/HRMS using a Dionex Ulitmate HPLC system coupled to an Orbitrap™ Elite Hybrid Ion Trap-Orbitrap™ Mass Spectrometer (Thermo Scientific) with a HESI. Chromatographic separation used an ACE UltraCore Super C_{18} column (150 x 2.1 mm i.d., 25 Å particle size, Hichrom, Reading, UK). The column was kept at a constant temperature of 50 ^{o}C. A gradient program with HPLC grade water (Fischer Scientific) as mobile phase A and HPLC grade acetonitrile (Fischer Scientific) as mobile phase B both with a 0.1 % formic acid (Fischer Scientific) modifier was used. The flow rate was kept constant at 350 μL min^{-1}. The gradient program was as follows: 5% B for 1 min, 5% to 95% linear gradient for 30 min and 95% held for 5 min before returning to 5% in 1 minute and remaining at 5 % for 4 min. All spectra were recorded using the nominal resolving power at "120,000" in positive ion mode for the mass range *m/z* 150 to 2000 in centroid and the AGC target was set to 1,000,000. The source voltage was set to 3.5 kV, source temperature 80 ^{o}C, sheath gas (nitrogen) flow rate to 30 arbitrary units (arb), the auxiliary gas (nitrogen) flow rate to 10 arb and the sweep gas (nitrogen) flow rate to 10 arb and capillary temperature to 275 ^{o}C. Between each analysis a solvent blank of HPLC water (Fischer Scientific) was run to ensure that there was no carry over between samples.

The data dependant acquisition (DDA) method was used for the acquisition of MS^2 spectra for a target mass list of ions and their retention times. The HPLC method and source settings were consistent between the HPLC/HESI-HRMS and HPLC/HESI-HRMS/MS runs. Ions detected in the HPLC-MS within 10 ppm of the *m/z* of ions determined to be significant from the Kruskal-Wallis analysis were compiled into a target mass list of *m/z* and retention time. A stepwise method consisting of 7 scan events was used. A full scan event recorded using the nominal resolving power at "120,000" to identify the presence of a target mass ion. If a target mass ion was detected within the retention time range of 30 sec, then a series of 6 MS^2 scans were recorded in the Orbitrap™ using the nominal resolving power at

"7000" at different CID energies of 10, 20, 30, 40, 50 and 60 eV. The same target ion could be recorded twice before it was excluded for 30 sec.

Data processing

DI-HESI-HRMS files were converted from Thermo .raw to .mzML using MSConvert. All 100 scans were merged using an openMS spectramerger module in KNIME.[45,46] This was done as the XCMS for the peak picking of DI-HRMS expects a single mass spectrum. Ion picking and alignment was done using XCMS package (v 1.52.0) in R (v 3.4.0) to create a data matrix of ion intensities aligned by mass.[47] The changes in the mass accuracy across the analytical run were assessed using the accurate mass of standard ions and ions were aligned using a mass tolerance of 5 ppm. The ion had to be present in 3 out of 5 of the replicate DI-HRMS analyses.

Files from the HPLC/MS analysis were converted from Thermos .raw to .mzML using MSConvert.[48] Peak picking and alignment were performed using XCMS (v 1.52.0) package in R (v 3.4.0) to create a data matrix of sample intensities aligned by mass and retention time.[47,49,50] The method used for peak picking was the centWave algorithm which is recommended for peak picking and alignment of HPLC/HRMS data. Peaks were picked above a signal-to-noise ratio of 10, the mass tolerance allowed was 10 ppm and a retention time tolerance range of 15 to 60 s. The peaks were then aligned across samples if the mass was within 0.002 Da and retention times overlapped by 10 s.

HPLC/MS/MS files were converted from Thermos .raw to .mzML using MSConvert.[48] Peak picking was done using XCMS (v 1.52.0) package in R (v 3.4.0). [47,49,50] A data matrix of product ions and intensities was created corresponding to a specific precursor ion's mass, retention time and the fragmentation energy. Product ion spectra were compared to two databases mzCloud and MassBank.

Statistical analyses and visualisation methods

All calculation and visualisation of the statistical analyses of the DI-HRMS spectra were carried out using Mass Profiler Professional (Agilent Technologies Ltd, Abington, UK). The intensity of the ions was transformed using $\log_2$ scale. The position and clustering of the mixed QC in the PCA was used to determine if there were any changes caused by analytical variance or data processing. Once this was shown to be minimal the QC and blank data were removed and the PCA and hierarchical cluster analysis were calculated to determine the differences between the sample groups. Heat maps are generated automatically as part of the hierarchical cluster analysis and visualise the difference in ion intensity between extract mass spectra. Kruskal-Wallis analysis was then used to compare the sewage effluent and upstream DOM composition based on ion distributions and their intensities to determine statistically significant ions (based on p values) which vary between the mass spectra. Ions with a p value < 0.005 and were found to increase in intensity when comparing the upstream and sewage outfall

DI-HRMS spectra were compiled into a target list for further investigation using HPLC/HRMS (described above).

Results and Discussion

The analytical approach described above aimed to identify the complex array of anthropogenic compounds discharged in treated sewage DOM against a background of riverine DOM. One of our primary objectives was to retain a comprehensive overview of DOM composition in order that contributions that would be missed in targeted analyses can be routinely detected. This relates to our wider objective of developing a holistic understanding of the role of DOM in driving aquatic ecosystem ecology, rather than the more common goal of targeted analyses for the regulation of priority pollutants. The approach used proceeds in three phases: (i) DI-HRMS analysis of water samples to identify the ions derived from the sewage effluent DOM against the background of natural riverine DOM, (ii) application of statistical methods to allow significant compositional differences to be determined and visualised diagrammatically, and (iii) use of HPLC/HRMS to further explore the complexity of DOM to identify individual molecular species through MS^2 spectra.

DOC analyses of the DOM and SPE extracts

The DOC concentrations of the filtered water collected from each sampling site, the concentration of organic carbon recovered by SPE and hence, the extraction efficiency of the SPE are shown in Table 1. These data reveal little difference in the DOC concentrations and the SPE extracts of the water samples from the sewage outfall and the river. DOC concentrations were similar for all samples at ca. 3 mg C L^{-1}, sitting within the range of variation previously reported for UK rivers, including in this study, which ranged from 0.76 mg C L^{-1} in chalk catchments to >26 mg C L^{-1} in peat catchments.[51] These data emphasise the ineffectiveness of DOC concentrations in revealing differences in the composition of the DOM pool, where markedly different compound mixes can share similar DOC concentration. The DOC determinations do, however, provide a useful means of assessing the SPE recovery efficiencies of DOM from all three water samples, i.e. ca. 40%, which is typical of the recoveries recorded for the HLB phase in other studies.[44,52] It should be noted that this SPE phase was chosen as it has been widely used in targeted[13,15,53] and untargeted analyses[54,55] and passive sampling.[56,57] The similarities in DOC concentration and extraction efficiencies emphasise the need to explore alternative, i.e. molecular approaches, to gain an in depth understanding of the composition characteristics and potential ecological impacts in relation to DOM source.

DI-HRMS analysis of SPE extracts

The DI-HRMS spectra of the upstream, sewage outfall and downstream extracts are shown in Figure 1. The spectra of all three DOM SPE extracts show the remarkable complexity of the composition of both the riverine and sewage effluent extracts. The full mass range spectra show clear differences between

sources. The upstream spectrum (Figure 1(a.i)) shows a similar character to SPE extracts of DOM from other studies of riverine DOM.[30,58] The spectrum shows an extremely high density of ions in the range *m/z* 150 to 750, maximising at *m/z* 288.1956. In contrast, the sewage effluent and downstream extracts differ markedly in composition from the upstream extract. These two extracts are characterised by a prominent series of ions extending well-beyond *m/z* 1000 (Figure 1(b.i & c.i)). The differences in composition between these two extracts and the upstream DOM reveals a very significant contribution from the sewage works to the riverine DOM, suggesting overprinting of the river background DOM by the anthropogenic contribution. Preliminary assessment of this contribution reveals a prominent series of ions with a 58 Da mass defect with the intensities describing a slightly skewed normal distribution, suggestive of the presence of a polymer, or perhaps more correctly a mixture of oligomers.

Eighteen oligomeric series containing 156 ions were identified with a mass difference of 58.0419 $\pm$ 0.005 indicates a structural motif of [$CH_2CH_2CH_2O$], consistent with the presence of oligomers of the synthetic industrial polymer polypropylene glycol (PPG). Further investigation of the DI-HRMS spectrum points to the presence of a number of variants of PPG series, which will be discussed in detail below. Based on this preliminary assessment alone, the sewage outfall DOM has clearly profoundly affected the composition of the river DOM.

Figure 1(ii & iii) show examples of two selected mass range windows, i.e. *m/z* 250 to 300 (blue highlighted mass window in the Figure 1(i) spectra) and *m/z* 272.0 to 274.0 (purple highlighted mass window in Figure 1(i & ii) spectra), of the full DI-HRMS spectra. These spectra illustrate the exquisite compositional detail revealed through use of high mass resolution ($m/\Delta m$ = "240,000"), in particular, differences in composition between the upstream, sewage works discharge and downstream DOM. In Figure 1, the highlighted bars in the spectra for the three sampling locations show two narrower mass windows. Without any prior knowledge of the identities of the components giving rise to the various ions, simple visual comparisons between spectra offer insights into ions specific to the reservoir river outflow and sewage works DOM extracts. Figure 1(c.iii), clearly represents the effects of mixing of the two sources. Notable differences include the major ion at *m/z* 272.1642 present in the upstream DOM but absent from the sewage effluent. However, the downstream river DOM shows this ion at lower relative abundance due to the addition of compounds from the sewage outfall. In contrast, the dominant ion at *m/z* 274.2007 in the sewage works discharge spectrum remains the most abundant ion in the downstream extract despite dilution. All the other ions in the 4 amu mass window shown in Figure 1(iii) display similar behaviours relating to source specificity and dilution effects. However, it was quickly recognised that that continuing with manual comparisons of this sort across the full spectral range would be prohibitively time-consuming due to the many thousands of ions present in these mass spectra. Set out below is a new protocol for processing such a dataset to allow in depth interrogation of source contributions.

Statistical comparisons of DOM based on DI-HRMS spectra

The starting for the statistical analyses is to establish if differences exist between the compositions of extracts in relation to the ions present and their intensities. The latter proceeds with creation of a data matrix of the ions aligned by their accurate mass and intensities for each DI-HRMS spectrum. After this "peak picking" step the DI-HRMS spectra were aligned to reveal 3237 ions detectable above a *s/n* 5. PCA was then applied to the generated data matrix to initially assess whether differences existed in composition between the extracts; the results are shown in Figure 2. In both PCAs the extracts clearly cluster in their respective replicate extraction groups . Figure 2(a) shows the mixed QC (purple), clusters between the downstream and sewage effluent replicate extracts, showing it is compositionally more similar to the latter extracts than the upstream. The mixed QCs position on the PCA plot can be explained by the presence of the compounds contributed by the sewage outfall, but which are absent from the upstream extract. The DI-HRMS spectra of the mixed QC, recorded every 5 extracts analysed throughout the analytical run, plot close together in the PCA, confirming no major significant differences are attributable to analytical variance or data processing errors.

The PCA of the DOM extracts shown in Figure 2(b), highlights that there are distinct compositional differences between the upstream (green), sewage effluent (red), and downstream (blue) extracts; separation in principal component 1 (PC1) explains 46.9 % of the total variance. The sewage outfall and upstream extracts are end members, confirmed by PC1 showing they are least similar in composition. As expected, the downstream extract plots between these groups, which is consistent with it being a mixture of the point source and reservoir riverine DOM.

Hierarchical cluster analysis (Figure 3(a)) confirms that upstream, downstream and sewage outfall extracts cluster in their respective replicate groups. However, the dendrogram also shows that overall the downstream and sewage outfall are more similar in composition, as these separate further down the dendrogram than the downstream and upstream extracts. This further demonstrates the profound effect the sewage outfall point source had on the downstream riverine DOM composition.

The heatmap visualises the differences in intensity of all the detected ions not easily determined when comparing DI-HRMS spectra directly. The heatmap shows ions changing in intensity across the mass range of the DI-HRMS spectra. As discussed above when comparing the raw DI-HRMS spectra directly (Figure 1), ions were present in the downstream and sewage outfall spectra of higher mass (*m/z* >900), which were not seen in the upstream mass spectra; this can be clearly seen using the heatmap. Expanding the heatmap in this mass range shown in Figure 3(c), the ions in this area of the heatmap are represented consistently in red indicating a high intensity in the sewage outfall extract, blue indicating low intensity in the upstream extract, and yellow/orange in the downstream extract, showing that the intensity falls between the upstream and sewage outfall extracts. This demonstrates the expected behaviour of compounds originating from the point source, i.e. that these are highest concentration in

the sewage outfall, low concentration/absent upstream and diluted upon entering the river in proportion to the river flow.

The ions in the mass range used in Figure 1(iii) are shown in the expanded heatmap in the Figure 3(b). The contrasting changes in intensity for *m/z* 272.1642 and *m/z* 274.2007 (discussed above) can also be seen in the heat map, occurring consistently across all extraction replicates. In addition, using the heat map, more subtle changes can be seen, e.g. the ions *m/z* 273.1482 and 273.1670 exhibit the same high intensity in the sewage effluent and downstream extracts as shown by ion *m/z* 274.2007, which was not easily identifiable from directly comparing the DI-HRMS spectra. This visualisation tool creates a quick approach to compare changes in the intensity of particular ions between the DI-HRMS spectra and extraction replicates.

The upstream and sewage effluent DI-HRMS spectra were compared using Kruskal-Wallis analysis to highlight significant discriminating ions which differ between the mass spectra of the various extracts. A significance threshold p value <0.005 was chosen and only ions with a higher intensity in the sewage outfall when compared to the upstream were retained, as these compounds were deemed most likely to derive from the sewage outfall. It was found that of the 3237 ions detected, 510 ions were found to meet these criteria, hence, these ions were selected for further analysis by MS/MS. The complexity of the DI-HRMS spectra show there are multiple ions within a 1 Da mass range as illustrated by Figure 1(iii). Furthermore, each ion could be multiple structural isomers. Isolation of precursor ions for further MS^2 experiments from such a complex mixture would result in in chimeric product ion spectra, difficult to deconvolute and match to reference spectra. This made it unfeasible even with the Orbitrap™ MS to isolate a single ion from such a complex mixture,[59] therefore, tandem HPLC/HRMS/MS was used to identify specific components.

HPLC/HRMS and HPLC/HRMS/MS analyses

The total ion chromatograms (TICs) of each of the SPE extracts shown in Figure 4(d-f) suggest poor chromatographic separation, because there are no individually resolved chromatographic peaks in the TICs. However, plotting accurate mass extracted ion chromatograms (EICs) shows that individual compounds are separated chromatographically, and the apparently poor resolution actually arises from extensive co-elutions inevitable in these extremely complex mixtures. Thus, HPLC/HRMS offers the following possibilities: (i) further exploration of extract composition and attribution of components to source, and (ii) isolation of individual ions allowing MS^2 experiments to be carried out to identify compounds.

Even higher complexity is revealed through HPLC/HRMS than was apparent in the DI-HRMS. The "peak picking" algorithm detected 14,325 individual components across all extracts, which was recognised by aligning their unique masses and retention times (*m/z*@rt), producing a second data

matrix of peak areas. A components peak area was compared in ratio form across the three different extracts using a ternary plot (Figure 5). The components found in each of the three extracts show three main trends: (i) the green area of the ternary plot highlights components where < 5 % of the total peak area is attributable to the upstream extract, confirming these components derive from the sewage outfall and downstream extracts. As shown by the ternary plot most components have a higher contribution from the sewage outfall as these plot between 50-100 % on the axis of the sewage outfall. This reflects their absence/low abundance in the river background (upstream), high abundance in the sewage effluent, and reduced abundance downstream due to in-stream dilution. (ii) the blue area of the plot highlights components where < 5% of the total peak area is attributable to the sewage outfall. This shows that these components are predominantly found in the river (downstream and upstream extracts). (iii) the red area highlights components where >5 % of the peak areas is found in all three sources showing that these components are common to all SPE extracts. The ternary plot facilitates the overall comparison of the different components detected in the HPLC/HRMS analysis, which is simply not possible through manual direct comparison.

Turning to the second use of HPLC/HRMS we focussed on the 510 ions determined in the DI-HRMS as deriving uniquely from the sewage effluent. Using the accurate mass (± 5 ppm) 420 of these masses were detected as 681 components in the HPLC/HRMS. This showed that a substantial proportion of the individual ions in the DI-HRMS analysis comprise more than one structure and that 90 of the masses were undetectable in the HPLC/HRMS for a variety of reasons. The majority of 681 components detected in HPLC/HRMS analysis plot in the green area of the ternary plot, confirming that these components derive from the sewage outfall.

These 681 components were compiled into a target list and analysed by HPLC/HRMS/MS as described above. The chromatographic separation allows the isolation of individual compounds for which product ion spectra can be recorded over multiple collision energies. Ninety-six components were identified and the EICs of these are shown in Figure 4(a-c). The EICs of the identified compounds show that all 96 are only present in the downstream and sewage effluent extracts and none are detectable in the upstream extract, unequivocally confirming these compounds originate from the sewage treatment works. Interestingly, there is clearly a decrease in the peak area in the downstream extract compared to the sewage effluent extract, resulting from dilution of the point source by the river flow.

Of the 96 components identified, 72 related to the polymer PPG, eluded to above and discussed further below. The other 24 compounds were a mixture of pharmaceuticals, illicit drugs, flame retardants and metabolites, as summarised in Table 2. Twenty two of the compounds characterised have been previously identified in other sewage treatment effluents and/or surface water[13,15,60,61]. Two novel compounds were identified, namely the antiretroviral raltegravir and also piperine, which is a natural product derived from black pepper. The antiretroviral raltegravir was tentatively identified based on

multiple CID spectra recorded at a range of energies. Further evidence for the identification of raltegravir was obtained using higher energy collision dissociation at the same collision energies used to record the reference spectra recorded in mzCloud (10-100 eV).

As discussed above in relation to the DI-HRMS 18 series of ions were highlighted with a 58 Da mass defect in the downstream and sewage outfall extracts. The HPLC/HRMS TIC of these extracts showed no distinct series of chromatographically separated peaks with a normal distribution(s) which would be indicative of a synthetic polymer. However, using the accurate masses of the ions in each series (determined from DI-HRMS) the EIC shown in Figure 4(b-c), reveal 2 distinct normally distributed series of peaks, presumed to correspond to 2 series of oligomers. The first series elutes between 8 to 27 min with the most abundant oligomer eluting at 11.5 min and the second series eluting between 25 to 35 min with the abundant oligomer eluting at 31 min. Oligomeric ions from 5 of the series were found to coincide with the earlier eluting distribution and 3 were found to coincide with the later eluting series. This indicates that these co-eluting series are isotopes and adducts of the 2 different oligomer series.

The HPLC/HRMS/MS analyses of selected parent oligomeric ions from both series were found to produce series of product ions with a mass defect of 58.0419, consistent with the cleavage of the ether bond in PPG. Using the accurate mass of the parent and product ions it was possible to determine that each series had different end groups. For the earlier eluting series, the end groups were determined to be dihydroxy, while the later eluting series possessed hydroxyl and butyl end groups. It was not possible to identify the remaining 10 series of ions found to possess a 58 Da mass defect from the DI-HRMS spectra.

Conclusion

The results presented herein confirm the advantages of using an untargeted HRMS approach to the analysis of DOM contributed from point sources. The major findings of the research are:

(i) The DI-HRMS molecular ‘fingerprints’ of the DOM extracts of river water obtained using SPE reveal the exceptional compositional complexity and very wide range of DOM compounds in waters which are not currently quantified, identified or controlled under current water quality legislation. The DI-HRMS spectra of the DOM extracted from the upstream, downstream and sewage outfall water, show how a point source can dramatically alter the composition of the riverine DOM.

(ii) Manual assessments of the DOM composition, while revealing specific spectral features driving differences in DOM composition, emphasise the need to use chemometric statistical methods to interrogate data sets of this complexity.

(iii) PCA analysis of the DI-HRMS spectra was readily able to resolve the different DOM sources, including in-stream mixing. Hierarchical cluster analysis showed that the

composition of the downstream DI-HRMS spectra was more similar to the sewage outfall spectra than those of upstream extracts, confirming the importance of the point source contribution to the overall DOM.

(iv) Heatmapping facilitated visualisation of the changes in the intensity of ions between DI-HRMS spectra including the determination of ion intensity changes which were not readily identifiable directly comparing the DI-HRMS spectra.

(v) Comparison of the sewage outfall and upstream DI-HRMS spectra using Kruskal-Wallis analysis provided a critical statistical data reduction step to identify the most important molecular species driving the differences in composition between the DOM extracts.

(vi) HPLC/HRMS TIC shows extensive co-elution for the DOM extracts. However, EIC of individual ions showed that compounds were separated chromatographically, with peak picking revealing over 14,325 components. Ternary plotting provided a visual means of attributing components to sources.

(vii) A wide range of compounds were tentatively identified from the sewage outfall including pharmaceuticals, plasticisers, metabolites and illicit drugs. Many have been identified in previous studies as originating from sewage treatment works. Others remain to be investigated to determine their environmental behaviour and potential ecosystem impact in waters.

(viii) Industrially-produced oligomeric PPGs were identified using DI-HRMS and HPLC/HRMS in sewage effluent for the first-time.

Overall, the results demonstrate considerable value exists in combining DI-HESI-Orbitrap™-HRMS and HPLC/HESI-Orbitrap™-HRMS for the analysis of complex DOM extracts. Our approach also highlights the value of applying statistically approaches to the assessment of complex data sets to determine the components differing between sources. Such an approach would have value in assessing compositional differences of any point source in river systems or between temporal events driven biologically, seasonally and/or anthropogenically.

Acknowledgements

JAP was funded by a NERC Industrial Case Studentship between University of Bristol and Fera Science Ltd, UK. The research was performed within the context of the DOMAINE (Characterisation of the nature, origins and ecological significance of dissolved organic matter in freshwater ecosystems) project which is supported by a NERC Large Grant (NE/K010689/1). We thank the School of Chemistry Mass Spectrometry Facility for the access to the Orbitrap Elite mass spectrometer funded thorough an EPSRC capital award.

References

1. European Commission. EU Water Framework Directive 2000/60/EC. *Off J Eur Communities*. 2000;L327:1-72.
2. He W, Chen M, Schlautman MA, Hur J. Dynamic exchanges between DOM and POM pools in coastal and inland aquatic ecosystems: A review. *Sci Total Environ*. 2016;551-552:415-428. doi:10.1016/j.scitotenv.2016.02.031
3. Minor EC, Swenson MM, Mattson BM, Oyler AR. Structural characterization of dissolved organic matter: A review of current techniques for isolation and analysis. *Environ Sci Process Impacts*. 2014;16(9):2064-2079. doi:10.1039/c4em00062e
4. Petrie B, Barden R, Kasprzyk-Hordern B. A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring. *Water Res*. 2014;72:3-27. doi:10.1016/j.watres.2014.08.053
5. Pritchard JB. Aquatic toxicology: past, present, and prospects. *Environ Health Perspect*. 1993;100:249-257. doi:10.1289/ehp.93100249
6. Bunzel K, Kattwinkel M, Liess M. Effects of organic pollutants from wastewater treatment plants on aquatic invertebrate communities. *Water Res*. 2013;47(2):597-606. doi:10.1016/j.watres.2012.10.031
7. Zenker A, Cicero MR, Prestinaci F, Bottoni P, Carere M. Bioaccumulation and biomagnification potential of pharmaceuticals with a focus to the aquatic environment. *J Environ Manage*. 2014;133:378-387. doi:10.1016/j.jenvman.2013.12.017
8. Jobling S, Nolan M, Tyler CR, Brighty G, Sumpter JP. Widespread sexual disruption in wild fish. *Environ Sci Technol*. 1998;32(17):2498-2506. doi:10.1021/es9710870
9. Sonnenschein C, Soto AM. An updated review of environmental estrogen and androgen mimics and antagonists. *J Steroid Biochem Mol Biol*. 1998;65(1-6):143-150. doi:10.1016/S0960-0760(98)00027-2
10. Beketov MA, Kefford BJ, Schafer RB, Liess M. Pesticides reduce regional biodiversity of stream invertebrates. *Proc Natl Acad Sci*. 2013;110(27):11039-11043. doi:10.1073/pnas.1305618110
11. Capaldo A, Gay F, Lepretti M, et al. Effects of environmental cocaine concentrations on the skeletal muscle of the European eel (Anguilla anguilla). *Sci Total Environ*. 2018;640-641:862-873. doi:10.1016/j.scitotenv.2018.05.357
12. Relyea RA. A cocktail of contaminants: How mixtures of pesticides at low concentrations affect aquatic communities. *Oecologia*. 2009;159(2):363-376. doi:10.1007/s00442-008-1213-9
13. Loos R, Carvalho R, Comero S, et al. *EU Wide Monitoring Survey on Waste Water Treatment Plant Effluents*.; 2012. doi:10.2788/60663
14. Gavrilescu M, Demnerová K, Aamand J, Agathos S, Fava F. Emerging pollutants in the environment: Present and future challenges in biomonitoring, ecological risks and

bioremediation. *N Biotechnol*. 2015;32(1):147-156. doi:10.1016/j.nbt.2014.01.001

15. Santos LHMLM, Gros M, Rodriguez-Mozaz S, et al. Contribution of hospital effluents to the load of pharmaceuticals in urban wastewaters: Identification of ecologically relevant pharmaceuticals. *Sci Total Environ*. 2013;461-462:302-316. doi:10.1016/j.scitotenv.2013.04.077
16. Loraine GA, Pettigrove ME. Seasonal variations in concentrations of pharmaceuticals and personal care products in drinking water and reclaimed wastewater in Southern California. *Environ Sci Technol*. 2006;40(3):687-695. doi:10.1021/es051380x
17. Andresen JA, Grundmann A, Bester K. Organophosphorus flame retardants and plasticisers in surface waters. *Sci Total Environ*. 2004;332(1-3):155-166. doi:10.1016/j.scitotenv.2004.04.021
18. Golovko O, Kumar V, Fedorova G, Randak T, Grabic R. Seasonal changes in antibiotics, antidepressants/psychiatric drugs, antihistamines and lipid regulators in a wastewater treatment plant. *Chemosphere*. 2014;111:418-426. doi:10.1016/j.chemosphere.2014.03.132
19. Yates CA, Johnes PJ, Spencer RGM. Characterisation of treated effluent from four commonly employed wastewater treatment facilities: A UK case study. *J Environ Manage*. 2019;232:919-927. doi:10.1016/j.jenvman.2018.12.006
20. Schymanski EL, Singer HP, Slobodnik J, et al. Non-target screening with high-resolution mass spectrometry: Critical review using a collaborative trial on water analysis. *Anal Bioanal Chem*. 2015;407(21):6237-6255. doi:10.1007/s00216-015-8681-7
21. Tanabe A, Mitobe H, Kawata K, Yasuhara A, Shibamoto T. Seasonal and spatial studies on pesticide residues in surface waters of the Shinano River in Japan. *J Agric Food Chem*. 2001;49(8):3847-3852. doi:10.1021/jf010025x
22. Peček G, Pavlović DM, Babić S. Development and validation of a SPE-GC-MS method for the determination of pesticides in surface water. *Int J Environ Anal Chem*. 2013;93(12):1311-1328. doi:10.1080/03067319.2012.736976
23. Ruff M, Mueller MS, Loos M, Singer HP. Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass-spectrometry - Identification of unknown sources and compounds. *Water Res*. 2015;87:145-154. doi:10.1016/j.watres.2015.09.017
24. Gros M, Petrović M, Barceló D. Tracing pharmaceutical residues of different therapeutic classes in environmental waters by using liquid chromatography/quadrupole-linear ion trap mass spectrometry and automated library searching. *Anal Chem*. 2009;81(3):898-912. doi:10.1021/ac801358e
25. Mills GA, Vrana B, Allan I, Alvarez DA, Huckins JN, Greenwood R. Trends in monitoring pharmaceuticals and personal-care products in the aquatic environment by use of passive sampling devices. *Anal Bioanal Chem*. 2007;387(4):1153-1157. doi:10.1007/s00216-006-0773-

y

26. Boyd-Boland AA, Magdic S, Pawliszyn JB. Simultaneous determination of 60 pesticides in water using solid-phase microextraction and gas chromatography–mass spectrometry. *Analyst*. 1996;121(7):929-937. doi:10.1039/AN9962100929
27. Jiang J-Q, Zhou Z, Sharma VK. Occurrence, transportation, monitoring and treatment of emerging micro-pollutants in waste water — A review from global views. *Microchem J*. 2013;110:292-300. doi:10.1016/j.microc.2013.04.014
28. Hawkes JA, Dittmar T, Patriarca C, Tranvik L, Bergquist J. Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. *Anal Chem*. 2016;88(15):7698-7704. doi:10.1021/acs.analchem.6b01624
29. Koch BP, Ludwichowski KU, Kattner G, Dittmar T, Witt M. Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Mar Chem*. 2008;111(3-4):233-241. doi:10.1016/j.marchem.2008.05.008
30. Osborne DM, Podgorski DC, Bronk DA, et al. Molecular-level characterization of reactive and refractory dissolved natural organic nitrogen compounds by atmospheric pressure photoionization coupled to Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom*. 2013;27(8):851-858. doi:10.1002/rcm.6521
31. Li H, Minor EC. Dissolved organic matter in Lake Superior: insights into the effects of extraction methods on chemical composition. *Environ Sci Process Impacts*. 2015;17(10):1829-1840. doi:10.1039/C5EM00199D
32. Koch BP, Witt M, Engbrodt R, Dittmar T, Kattner G. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochim Cosmochim Acta*. 2005;69(13):3299-3308. doi:10.1016/j.gca.2005.02.027
33. Koch BP, Dittmar T, Witt M, Kattner G. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Anal Chem*. 2007;79(4):1758-1763. doi:10.1021/ac061949s
34. Herzsprung P, Hertkorn N, von Tümpling W, Harir M, Friese K, Schmitt-Kopplin P. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Anal Bioanal Chem*. 2014;406(30):7977-7987. doi:10.1007/s00216-014-8249-y
35. Phungsai P, Kurisu F, Kasuga I, Furumai H. Molecular characterization of low molecular weight dissolved organic matter in water reclamation processes using Orbitrap mass spectrometry. *Water Res*. 2016;100:526-536. doi:10.1016/j.watres.2016.05.047
36. Gonsior M, Schmitt-Kopplin P, Stavklint H, Richardson SD, Hertkorn N, Bastviken D. Changes in dissolved organic matter during the treatment processes of a drinking water plant in sweden

and formation of previously unknown disinfection byproducts. *Environ Sci Technol*. 2014;48(21):12714-12722. doi:10.1021/es504349p

37. D'Andrilli J, Dittmar T, Koch BP, Purcell JM, Marshall AG, Cooper WT. Comprehensive characterization of marine dissolved organic matter by fourier transform ion cyclotron resonance mass spectrometry with electrospray and atmospheric pressure photoionization. *Rapid Commun Mass Spectrom*. 2010;24(5):643-650. doi:10.1002/rcm.4421
38. Mosher JJ, Kaplan LA, Podgorski DC, McKenna AM, Marshall AG. Longitudinal shifts in dissolved organic matter chemogeography and chemodiversity within headwater streams: a river continuum reprise. *Biogeochemistry*. 2015;124(1-3):371-385. doi:10.1007/s10533-015-0103-6
39. Herzsprung P, von Tümpling W, Wendt-Potthoff K, et al. High field FT-ICR mass spectrometry data sets enlighten qualitative DOM alteration in lake sediment porewater profiles. *Org Geochem*. 2017;108:51-60. doi:10.1016/j.orggeochem.2017.03.010
40. Hockaday WC, Purcell JM, Marshall AG, Baldock JA, Hatcher PG. Electrospray and photoionization mass spectrometry for the characterization of organic matter in natural waters: a qualitative assessment. *Limnol Oceanogr Methods*. 2009;7(1):81-95. doi:10.4319/lom.2009.7.81
41. Kim S, Kramer RW, Hatcher PG. Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Anal Chem*. 2003;75(20):5336-5344. doi:10.1021/ac034415p
42. Dubinenkov I, Flerus R, Schmitt-Kopplin P, Kattner G, Koch BP. Origin-specific molecular signatures of dissolved organic matter in the Lena Delta. *Biogeochemistry*. 2015;123(1-2):1-14. doi:10.1007/s10533-014-0049-0
43. Lu Y, Li X, Mesfioui R, et al. Use of ESI-FTICR-ms to characterize dissolved organic matter in headwater streams draining forest-dominated and pasture-dominated watersheds. *PLoS One*. 2015;10(12):e0145639. doi:10.1371/journal.pone.0145639
44. Li Y, Harir M, Uhl J, et al. How representative are dissolved organic matter (DOM) extracts? A comprehensive study of sorbent selectivity for DOM isolation. *Water Res*. 2017;116:316-323. doi:10.1016/j.watres.2017.03.038
45. Röst HL, Sachsenberg T, Aiche S, et al. OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nat Methods*. 2016;13(9):741-748. doi:10.1038/nmeth.3959
46. Berthold MR, Cebron N, Dill F, et al. KNIME - the Konstanz information miner. *ACM SIGKDD Explor Newsl*. 2009;11(1):26. doi:10.1145/1656274.1656280
47. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. 2006;78(3):779-787. doi:10.1021/ac051437y
48. Chambers MC, MacLean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and

proteomics. *Nat Biotechnol*. 2012;30(10):918-920. doi:10.1038/nbt.2377

49. Tautenhahn R, Bottcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*. 2008;9:16. doi:10.1186/1471-2105-9-504
50. Benton HP, Want EJ, Ebbels TMD. Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. *Bioinformatics*. 2010;26(19):2488-2489. doi:10.1093/bioinformatics/btq441
51. Yates CA, Johnes PJ, Owen AT, et al. Variation in dissolved organic matter (DOM) stoichiometry in U.K. freshwaters: Assessing the influence of land cover and soil C:N ratio on DOM composition. *Limnol Oceanogr*. 2019:1-13. doi:10.1002/lno.11186
52. Waska H, Koschinsky A, Ruiz Chancho MJ, Dittmar T. Investigating the potential of solid-phase extraction and Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) for the isolation and identification of dissolved metal-organic complexes from natural waters. *Mar Chem*. 2015;173:78-92. doi:10.1016/j.marchem.2014.10.001
53. Zhou JL, Maskaoui K, Lufadeju A. Optimization of antibiotic analysis in water by solid-phase extraction and high performance liquid chromatography-mass spectrometry/mass spectrometry. *Anal Chim Acta*. 2012;731:32-39. doi:10.1016/j.aca.2012.04.021
54. D'Archivio AA, Fanelli M, Mazzeo P, Ruggieri F. Comparison of different sorbents for multiresidue solid-phase extraction of 16 pesticides from groundwater coupled with high-performance liquid chromatography. *Talanta*. 2007;71(1):25-30. doi:10.1016/j.talanta.2006.03.016
55. Raeke J, Lechtenfeld OJ, Wagner M, Herzsprung P, Reemtsma T. Selectivity of solid phase extraction of freshwater dissolved organic matter and its effect on ultrahigh resolution mass spectra. *Environ Sci Process Impacts*. 2016;18(7):918-927. doi:10.1039/C6EM00200E
56. Li H, Helm PA, Paterson G, Metcalfe CD. The effects of dissolved organic matter and pH on sampling rates for polar organic chemical integrative samplers (POCIS). *Chemosphere*. 2011;83(3):271-280. doi:10.1016/j.chemosphere.2010.12.071
57. Jeong Y, Schäffer A, Smith K. Equilibrium partitioning of organic compounds to OASIS HLB®as a function of compound concentration, pH, temperature and salinity. *Chemosphere*. 2017;174:297-305. doi:10.1016/j.chemosphere.2017.01.116
58. Kim S, Simpson AJ, Kujawinski EB, Freitas MA, Hatcher PG. High resolution electrospray ionization mass spectrometry and 2D solution NMR for the analysis of DOM extracted by C18solid phase disk. *Org Geochem*. 2003;34(9):1325-1335. doi:10.1016/S0146-6380(03)00101-3
59. Viant MR, Sommer U. Mass spectrometry based environmental metabolomics: A primer and review. *Metabolomics*. 2013;9(SUPPL.1):144-158. doi:10.1007/s11306-012-0412-x
60. Karolak S, Nefau T, Bailly E, Solgadi A, Levi Y. Estimation of illicit drugs consumption by

wastewater analysis in Paris area (France). *Forensic Sci Int*. 2010;200(1-3):153-160. doi:10.1016/j.forsciint.2010.04.007

61. Sheng L-H, Chen H-R, Huo Y-B, et al. Simultaneous Determination of 24 Antidepressant Drugs and Their Metabolites in Wastewater by Ultra-High Performance Liquid Chromatography–Tandem Mass Spectrometry. *Molecules*. 2014;19(1):1212-1222. doi:10.3390/molecules19011212

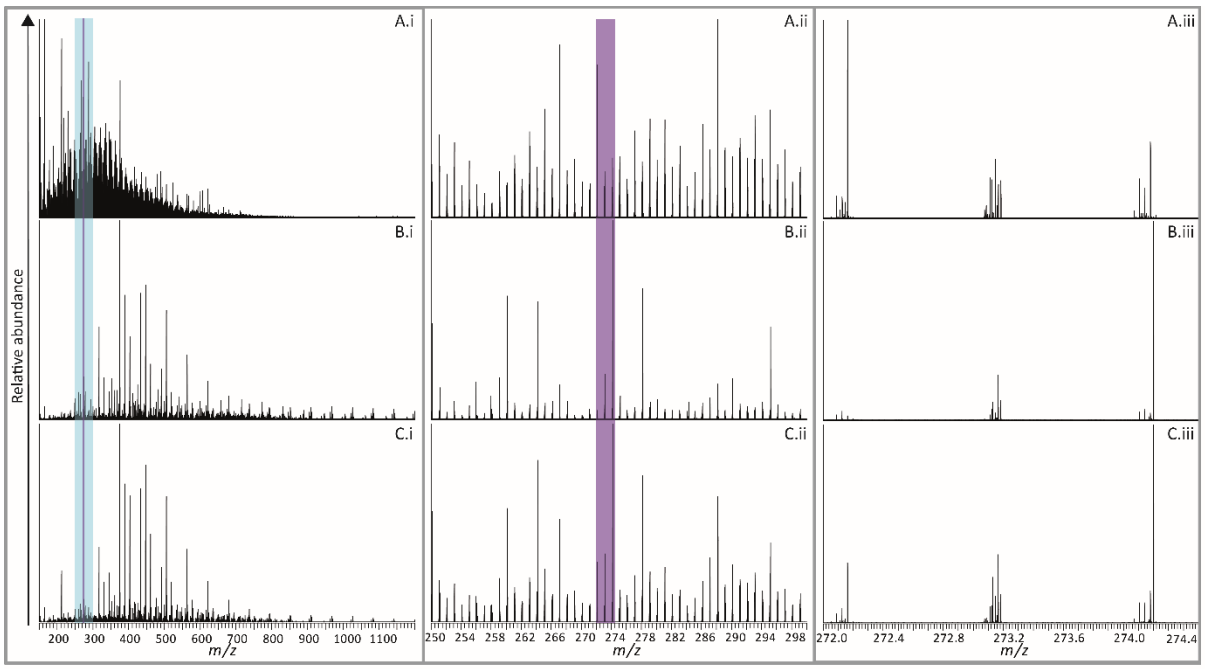

Figure 1. DI-HRMS spectra of the (A) upstream, (B) sewage outfall, and (C) downstream, SPE extracts displaying the mass ranges *m/z* 150 to 1200 (A.i, B.i, and C.i) *m/z* 250 to 300 (A.ii, B.ii, and C.ii) and *m/z* 272.0 to 274.4 (A.iii, B.iii, and C.iii).

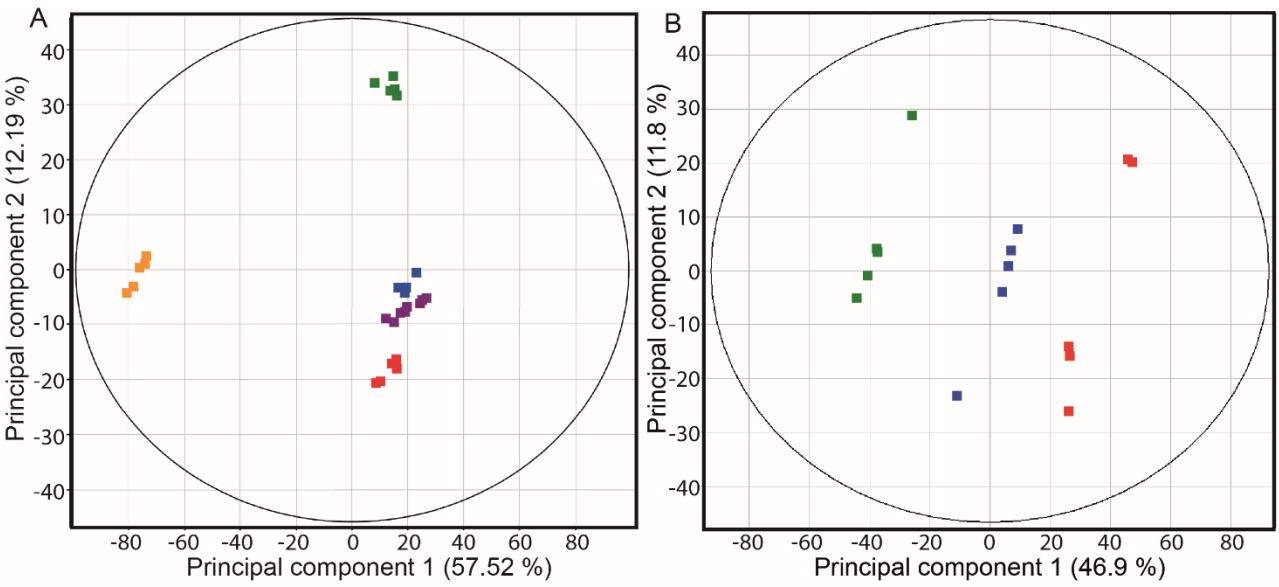

Figure 2. (A) PCA of the DI-HRMS spectra of the upstream (■), sewage outfall (■), downstream (■), blank (■) and mixed quality control (■) DOM extracts. (B) PCA of the DI-HRMS spectra of the upstream (■), sewage outfall (■), and downstream (■) DOM extracts.

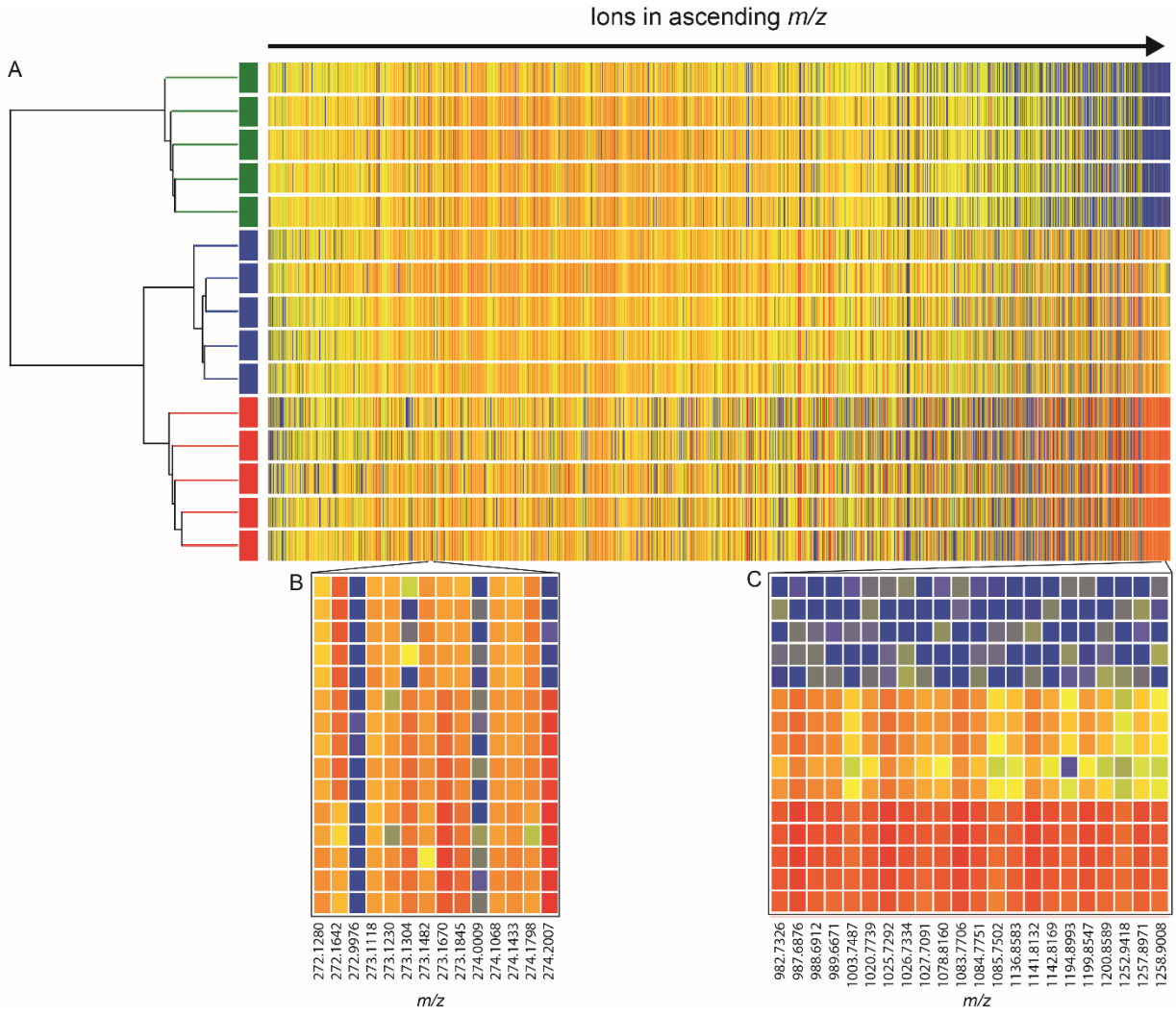

Figure 3. (A) Hierarchical cluster analysis of the upstream (■), sewage outfall (■), and downstream (■) DOM extracts and heatmap of the ions detected in the DI-HRMS. Comparison of the $\log_2$ of the intensity of the ions represented by colour with higher intensity hotter (red) and lower intensity colder (blue). Two narrower mass ranges (B) *m/z* 272.0 to 274.5 and (C) *m/z* 982.0 to 1259.0 from the heatmap.

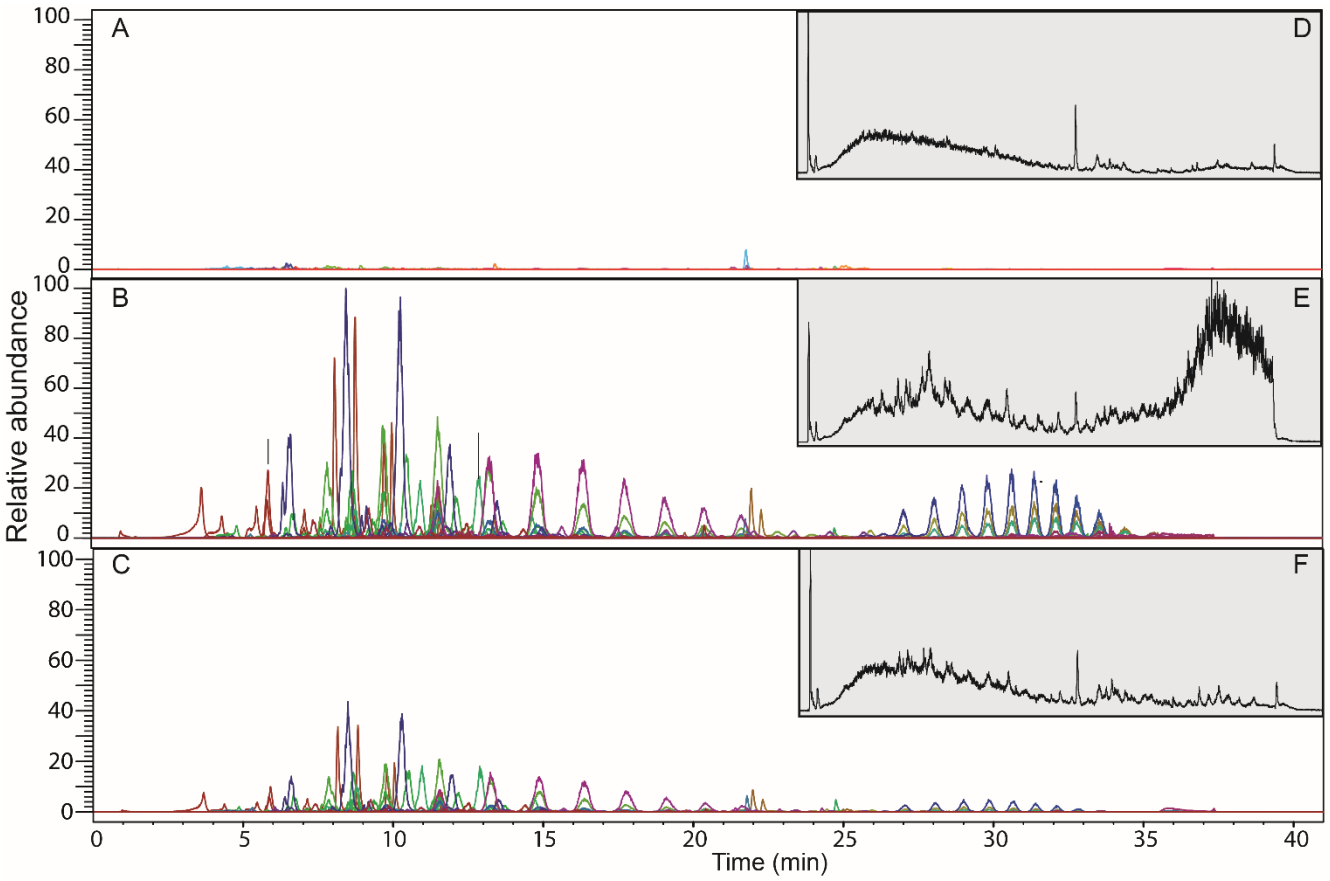

Figure 4. (A) EIC of mass of the precursor ions identified in the upstream SPE extract, (B) EIC of the precursor ions identified in the sewage effluent SPE extract, (C) EIC of the precursor ions identified in the downstream SPE extract, (D) TIC of upstream SPE extracts, (E) TIC of the sewage outfall SPE extracts, and (F) TIC of the downstream SPE extracts.

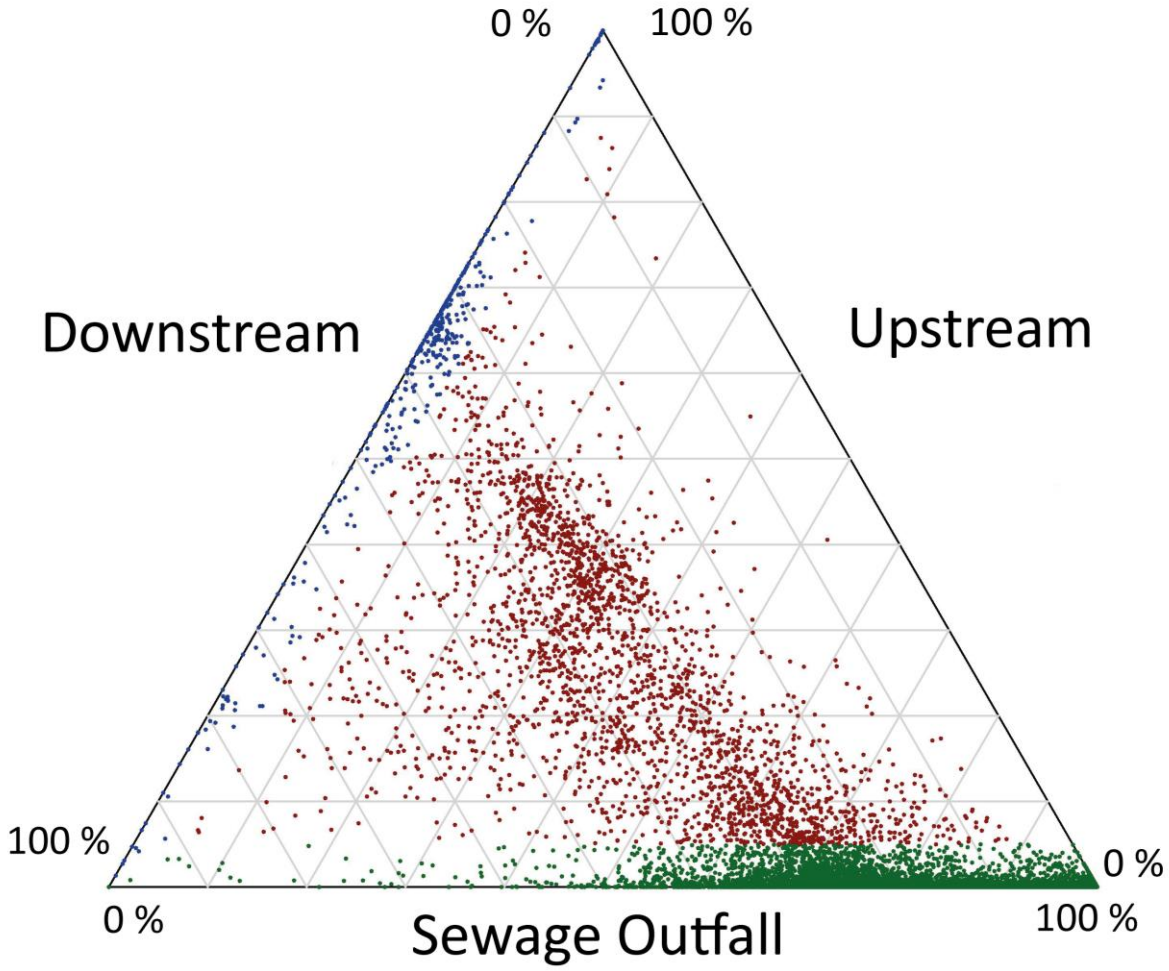

Figure 5. Ternary plot of ratios the peak areas of individual components detected in the SPE extracts. Green points highlight peaks where < 5% of the component derives from the upstream contribution. Blue area highlights peaks where < 5% of the component derives from sewage outfall. Red highlights peaks where > 5% of the component can be attributed to all three sources.

Table 1. DOC of the filtered water, concentration of organic carbon extracted using SPE and the extraction efficiency of the extraction procedure

	DOC filtered water (mg C L^{-1})	DOC concentration of extracts (mg C L^{-1})	Extract efficiencies (%)
Upstream	3.39	1.50 ± 0.04	42.11 ± 0.5
Sewage Outfall	3.19	1.29 ± 0.05	40.49 ± 1.7
Downstream	3.45	1.40 ± 0.06	40.34 ± 0.8
Blank	N/A	0.18 ± 0.02	N/A

Table 2. Summary of the 24 identified compounds from the sewage effluent extract.

Precursor (*m/z*)	Retention Time (min)	Fragmentation energy (eV)	p value	Formulae	Compound	Product ions (*m/z*)
300.1592	3.67	40	0.0013	$C_{18}H_{21}NO_3$	Codeine	282.1497, 267.1260, 253.1231, 243.1023, 225.0917, 215.1074, 199.0760, 193.0648, 187.0754, 183.0811, 175.0760, 165.0701, 161.0603
268.1544	3.93	30	0.0016	$C_{14}H_{21}NO_4$	Atenolol acid	250.1441, 233.1176, 226.1079, 208.0971,191.0706, 165.0547, 145.0471, 116.1067, 98.0960
325.1915	5.47	30	0.0013	$C_{20}H_{24}N_2O_2$	Quinine	307.1782, 279.1521, 278.1570, 264.1315, 253.1296, 226.1199, 210.0940, 202.0851, 198.0880, 186.0918, 184.0739, 174.0926, 172.0744, 166.1228, 160.0798, 134.0914, 110.0951
290.1392	5.89	30	0.0013	$C_{16}H_{19}NO_4$	Benzoylecgonine	272.1288, 168.1023, 150.0917, 124.1123, 122.0964, 119.0493, 91.0545
256.0152	5.9	40	0.0013	$C_9H_7Cl_2N_5$	Lamotrigine	229.0052, 221.0468, 220.0390, 213.9925, 210.9831, 193.0408, 186.9827, 185.9878, 183.9712, 179.0245, 173.9880, 171.9716, 166.0299, 165.0214, 158.9768, 151.0190
266.1657	6.07	40	0.0013	$C_{17}H_{19}N_3$	Mirtazapine	235.1230, 223.1230, 209.1073, 195.0917
304.1549	7.04	40	0.0013	$C_{17}H_{21}NO_4$	Cocaine	272.1281, 182.1176, 150.0913, 108.0807
253.0978	7.76	30	0.0013	$C_{15}H_{12}N_2O_2$	Carbamazepine 10,11-epoxide	254.0817, 236.071, 210.09187,180.0809
278.2113	8.14	30	0.0013	$C_{17}H_{27}NO_2$	Venlafaxine	261.206, 215.1435, 121.0641
373.1586	8.75	30	0.0013	$C_{20}H_{24}N_2O_3S$	Desacetyl diltiazem	373.1580, 328.1002, 223.0900, 178.0321, 150.04
260.1647	8.82	30	0.0013	$C_{16}H_{21}NO_2$	Propranolol	242.1540, 218.1171, 183.0804, 157.0647, 132.1020, 116.1067, 98.0961, 86.0960
325.1711	9.68	30	0.0024	$C_{20}H_{21}FN_2O$	Citalopram	325.1721, 307.1614, 280.1139, 262.1033, 234.0721, 166.0656, 156.0813, 116.0496, 109.0449
415.1456	10.06	30	0.0013	$C_{17}H_{20}F_6N_2O_3$	Flecainide	415.1454, 398.1189, 386.12,370.0870, 332.1345, 330.05569,318.0558, 315.1075 301.0297
264.1752	11.15	30	0.0013	$C_{19}H_{21}N$	Nortriptyline	264.0840, 233.1331, 191.0860, 155.0861, 117.0700, 105.0700, 91.0543
278.1909	11.29	30	0.0013	$C_{20}H_{23}N$	Amitriptyline	278.1918, 233.1332, 191.0861, 179.0859, 155.0861, 117.0701, 105.0700, 91.0543
502.2957	11.53	30	0.0013	$C_{32}H_{39}NO_4$	Fexofenadine	484.2830, 466.2726,262.1591, 250.5923, 246.1489, 233.1174,171.1168
237.1028	11.61	30	0.0013	$C_{15}H_{12}N_2O$	Carbamazepine	237.0708, 220.0758, 194.0966, 192.0810
192.1388	13.54	30	0.0041	$C_{12}H_{17}NO$	*N*,*N*-Diethyl-3-methylbenzamide, (DEET)	192.13829, 119.05, 100.07569, 91.0542
445.1636	13.21	30	0.0013	$C_{20}H_{21}FN_6O_5$	Raltegravir	361.1326, 318.1261, 278.0944, 253.0943, 236.0678, 193.07800 168.0780, 140.0824, 109.0451
286.1443	16.63	30	0.0013	$C_{17}H_{19}NO_3$	Piperine	287.1490, 215.1071 ,201.0551, 173.0599, 150.0919, 135.0443, 112.0757
399.2512	21.99	40	0.0013	$C_{18}H_{39}O_7P$	Tri(butoxyethyl) phosphate	299.1627, 243.1001, 225.0894, 199.0736, 143.0108, 124.0100, 101.0963, 98.9841
273.1855	22.32	30	0.0013	$C_{18}H_{24}O_2$	Galaxolidone	255.1743, 227.1794, 203.107, 175.1117

2