



# A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment

Yaroslav Verkh<sup>a</sup>, Marko Rozman<sup>a, b</sup>, Mira Petrovic<sup>a, c, \*</sup>

<sup>a</sup> Catalan Institute for Water Research (ICRA), Carrer Emili Grahit 101, 17003 Girona, Spain

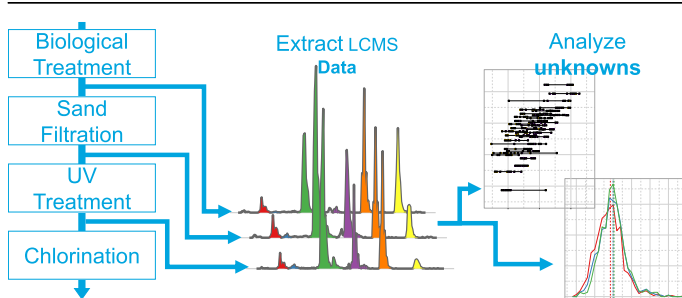
<sup>b</sup> Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia

<sup>c</sup> Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

## HIGHLIGHTS

- Secondary treatment removed 67% influent features while 24% new effluent appeared.
- The biodegradable organic matter differed chemically from the recalcitrant.
- Kendrick plot uncovered the removal of CH<sub>2</sub> and C<sub>2</sub>H<sub>4</sub>O homologs.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 7 August 2017

Received in revised form

12 February 2018

Accepted 16 February 2018

Available online 17 February 2018

Handling Editor: Keith Maruya

### Keywords:

LC-HR mass spectrometry

Non-targeted analysis

Wastewater treatment

Dissolved organic matter

Statistical analysis

Direct injection

## ABSTRACT

The dissolved organic matter (DOM) in wastewater is typically described by a limited number of concentration measurements of select DOM fractions or micro-contaminants, which determine the removal efficiency in a wastewater treatment. Current methods do not necessarily reflect the true performance of the treatment with regard to environmental and public health risk. Herein we describe the development and application of a non-targeted liquid chromatography-high resolution mass spectrometry (LC-HRMS) data analysis for the evaluation of wastewater treatment processes. Our data analysis approach was applied to a real wastewater system with secondary biological treatment and tertiary treatment consisting of sand filtration, UV-treatment, and chlorination. We identified significant changes in DOM during wastewater treatment. The secondary treatment removed 1617 of 2409 (67%) detected molecular features (grouped isotopologues belonging to the same molecule) from the influent while 255 of 1047 (24%) new molecular features appeared in the secondary effluent. A reduction in the number of large molecules (>450 Da) and an increase in unsaturated molecular features of the effluent organic matter was observed. Van Krevelen plots revealed the distribution of unsaturation and heteroatoms and Kendrick mass defect plots uncovered  $-CH_2-$  homologous series implying a removal of heavy constituents in that fraction. The demonstrated approach is a step towards a more comprehensive monitoring of DOM in wastewater and contributes to the understanding of current treatment technologies.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Wastewater DOM represents a complex, heterogenic mixture of polysaccharides, proteins, lipids, nucleic acids, soluble microbial

\* Corresponding author. Catalan Institute for Water Research (ICRA), Carrer Emili Grahit 101, 17003 Girona, Spain.

E-mail address: [mpetrovic@icra.cat](mailto:mpetrovic@icra.cat) (M. Petrovic).

products and anthropogenic organic chemicals. Anthropogenic compounds in wastewater include, among others, surfactants, personal care products, pharmaceuticals, biocides, pesticides, and industrial chemicals. Additionally, there is a wide range of biologically active transformation products (TPs), intermediates, metabolites (Michael-Kordatou et al., 2015) and disinfection by-products in the case of tertiary treatment (Richardson and Postigo, 2016). Some of these compounds can be hazardous even at a low concentration and may raise concerns regarding their release into the environment (Daughton, 2004). The composition of wastewater DOM is dependent upon the type of wastewater (municipal, industrial, hospital's effluent, runoff from fields, etc.) and the nature of the treatment process used in wastewater treatment plants (WWTP) (Deblonde et al., 2011). Currently, the efficiency of DOM removal at a WWTP is evaluated through measurements of the chemical oxygen demand (COD), biological oxygen demand (BOD) and the total organic carbon (TOC). Additional specialized technologies for the prioritized fractions of DOM include: measuring the dissolved organic halide to estimate the halogenated organics, the assessment of aromaticity using specific UV absorbance ( $SUVA_{254}$ ), size exclusion chromatography (SEC) to identify mass/size distributions of C- or N-containing constituents or excitation-emission-matrix fluorescence used to identify substance classes in natural OM (Chen et al., 2003; Clesceri et al., 1998; Huber et al., 2011). However, while these techniques reveal the chemical characteristics to a certain extent and the abundance of organics in WWTP influent and effluent, they do not provide information on the presence of unique organic substances and need to be combined into one data stream. Thus there is a need for new strategies to assess the quality of wastewater treatment (Prasse et al., 2015).

Different MS methods were developed for the analysis of wastewater treatment constituents. Initially, those focused on the detection of a small number of contaminants and occasionally on their TPs (Richardson and Ternes, 2014). By definition, these approaches omit thousands of DOM constituents which are present in the influent or emerge during the treatment process. Since some are potentially hazardous, overlooking these compounds limits the understanding of the impact of the effluent organic matter on the environment. Moreover, monitoring of the entire molecular complement or even a sub-complement of wastewater offers a possibility for a more comprehensive evaluation of the organic content in wastewater and a deeper understanding of the treatment processes and DOM transformation (Hollender et al., 2017). This new understanding will allow us to learn about the shortcomings of the treatment processes themselves and propose evidence-based improvement strategies.

The ability of high-resolution mass spectrometry (HRMS) to identify small amounts of organic chemicals from increasingly complex mixtures can provide information on wastewater DOM. For example, HRMS suspect screenings attempted to identify dominant signals using chemical databases and *in silico* prediction to find the structure of unique chemicals (Causanilles et al., 2017). Due to the complexity of wastewater, an HRMS analysis yields  $10^3$ – $10^5$  signals. This makes a manual structural identification of so many unique substances nearly impossible. Therefore this HRMS methodology generally uncovers only a small fraction of compounds and omits the unknown majority of wastewater DOM (Wode et al., 2015). Even without a tentative structural identification of particular substances, the large number of signals with assigned elemental compositions can be used to uncover physicochemical changes in wastewater treatment (Maizel and Remucal, 2017; Nürenberg et al., 2015).

Non-targeted analysis reduces the challenge of manual data treatment as for example present in a suspect screening, however without omitting it completely. The described HRMS data

treatment emerged from the fields of petroleomics and characterization of natural organic matter (NOM) (Hughey et al., 2001; Remucal et al., 2012; Sleighter and Hatcher, 2008) and was applied for the organic matter in processed water (Fang et al., 2017) and wastewater (Maizel and Remucal, 2017; Phungsai et al., 2016). Yet, the challenges of a non-targeted method remain the incomplete exclusion of noise signals from datasets and the exclusion of true DOM signals. These arise from the necessarily complex HRMS data acquisition, extraction, and clean-up methodology. Often DOM analyses do not include a chromatographic separation that might enhance the resolution of the spectral data. A direct infusion, without a separation on a chromatographic column, can simplify the procedure. Yet, the retention time is an additional variable to distinguish molecular features and an advantage of LC over an injection without a separation. The separation on the column is also beneficial for the reduction of the matrix effect compared to a direct infusion. Separation simplifies the mixture thus increases the chance of detecting low-intensity signals (Iparraguirre et al., 2014; Taylor, 2005). Meanwhile, measurement without sample pre-concentration does not lead to a loss of DOM, which is inevitable during a pre-concentration step (Li et al., 2016). Also, the use of LC-ESI-MS additionally reduces the discovered DOM, since it mostly detects medium-polar compounds (Aral et al., 2017).

HRMS non-targeted analysis includes a series of methods to sieve through large amounts of data. For example, in the van Krevelen diagram the atomic ratio X/C, where X is an element of interest, is plotted against H/C. In petroleomics and NOM chemistry the correlation between areas in this plot and functional classes of compounds led to the elucidation of the chemical composition of organic matter (Kim et al., 2003; Lu et al., 2010; Minor et al., 2012; Zhang et al., 2012a). A comparison of multiple samples revealed the transformation of matter (as oxidation of DOM) (Cortés-Francisco and Caixach, 2013; Herzsprung et al., 2012). Van Krevelen plot applied to wastewater treatment revealed a possible transformation of DOM (Maizel and Remucal, 2017). On the other hand, heterogeneous DOM (found in wastewater or eutrophic river) leads to a less structured distribution of points obscuring the graphic nature of the van Krevelen plot. For instance, the difficulty of exploring the graphical nature of the van Krevelen plot of the heterogeneous organic matter was experienced outside the field of water chemistry (Marshall et al., 2018).

The monitoring of double bond equivalents (DBE) in wastewater treatment can be applied to estimate the quality of the process and to recognize hydrophobicity-altering reactions as hydrolysis or oxidation (Cortés-Francisco et al., 2014). DBE reflect the level of unsaturation by double bonds in an organic molecule using only the counts of H, C, O, N, and halogen atoms in a molecule. DBE do not always apply to predict aromaticity since they can include double bonds with heteroatoms. Other models to predict unsaturation of a molecule as DBE divided by the number of C atoms, DBE minus oxygen atoms (DBE-O) and the aromaticity index were proposed (Cortés-Francisco and Caixach, 2013; Koch and Dittmar, 2006). DBE-O correlate especially well with the saturation of oxygen-rich organic compounds, but they represent a more abstract measure of unsaturation than DBE.

Mass defects are widely applied in HRMS analysis. Identification of homolog series using a non-targeted mass defect analysis was used to map surfactants in wastewater (Loos and Singer, 2017). Homologs occur in NOM (Hughey et al., 2001) but more importantly in classes of anthropogenic substances like surfactants, polyfluorinated compounds or chlorine substitute series (Jobst et al., 2013). The Kendrick Mass Defect (KMD), which sets the exact mass of a chosen molecular fragment to a nominal value (like 14.015 Da–14.000 Da for  $-CH_2-$ ) is able to identify homologous series for various structural moieties as  $-C_2H_4O-$  or  $-H/ + Cl$ . The

pattern recognition in Kendrick plots previously elucidated reactions and heteroatom distributions (Hughey et al., 2001; Jobst et al., 2013; Zhang et al., 2012b).

In this work, we present an LC-HRMS data treatment of wastewater DOM that enables us to monitor and to gain a deeper understanding of wastewater treatment processes. It consists of HRMS data processing using MZmine 2.26 (Pluskal et al., 2010) for data extraction and R (R Core Team, 2017). We obtained a list of molecular features with assigned molecular formulae where prediction rules, as isotopic pattern score, pre-defined atomic ratios, and heuristic rules allowed it. The obtained results were assessed with analytical tools such as van Krevelen and KMD diagrams, and the observation of DBE-O, mass or intensity shifts in subsets of features using our own R scripts. Here the data treatment is adapted to fingerprint the heterogenic mixture of DOM before and after wastewater treatment. We discuss the introduced HRMS methodology, compare it to other DOM analytics, and test its value on a real wastewater treatment system (secondary biological treatment followed by tertiary treatment train with sand filtration, UV-treatment, and chlorination).

## 2. Materials and methods

### 2.1. Reagents and sampling

HPLC grade solvents methanol, water, and acetonitrile were purchased from Fisher (Germany) and the buffer solution was prepared using HPLC grade 98–100% formic acid (Merck, Germany). A 24 h composite secondary influent and an effluent sample with a corresponding hydrological retention time correction were taken from a WWTP in Castell d'Aro, Spain. Additionally, six grab samples of two different time series with 24 h between them were taken in the tertiary treatment which included sand filtration, UV-treatment and chlorination steps. A grab sample of the secondary effluent was taken which corresponded in time to a set of the tertiary treatment samples. The short residence time of the tertiary treatment rendered composite sampling difficult.

The description of the WWTP can be found in the [supplementary information \(SI\)](#). The samples were filtered under vacuum using 1.0  $\mu\text{m}$  and 0.45  $\mu\text{m}$  Hydrophilic Polyvinylidene Fluoride Durapore<sup>®</sup> membrane filters (Merck Millipore Ltd). A mixture of 32 detected internal standards (IS, isotopically labelled pharmaceuticals and antibiotics) was used to evaluate the ion suppression caused by the matrix, and to estimate variations in the instrumental response from injection to injection. A correction of the matrix effect using an intensity normalization of spectra was not attempted. A normalization leads to worse results in replicates compared to the unaltered spectra.

### 2.2. LC-HRMS analysis

LC-HRMS analysis was performed on an LTQ-Orbitrap Velos<sup>™</sup> coupled with the Aria TLX-1 HPLC system (Thermo Fisher Scientific, USA). The system was controlled via Aria software, version 1.6, under the Xcalibur 2.2 software (Thermo Fisher Scientific, USA). The chromatographic separation was achieved on Acquity UPLC<sup>®</sup> BEH C<sub>18</sub> (2.1 mm  $\times$  50 mm, 1.7  $\mu\text{m}$  particle size, Waters UK) chromatographic column in both the positive ionization (PI) and the negative ionization (NI) modes. 50  $\mu\text{L}$  of the filtered sample was injected into the LCMS. A solvent gradient with formic acid 0.1% in acetonitrile and an aqueous solution of formic acid 0.1% were used in both PI and NI modes. MS parameters and LC solvent gradients are provided in [SI](#).

### 2.3. Data extraction evaluation

The mass tolerance for the processing of LC-MS spectra was set to 5 ppm in MZmine. A lower  $m/z$  tolerance was avoided to avoid the negative impact of too stringent tolerance margins on the instrument's performance (Erve et al., 2009; Makarov et al., 2006). The elution time of extracted ion chromatograms was between 0.05 and 8.00 min. Margins of retention time <0.5 min and  $m/z < 5$  ppm were used to exclude duplicate molecular features detected in both modes, thanks to the same chromatographic method in PI and NI modes. Signals with signal-to-baseline < 3 were ignored to exclude instrumental noise.  $m/z$  mass error of <5 ppm for data extraction was experimentally confirmed with IS. Only monoisotopic signals with at least one consecutive isotopologue were analyzed. This parameter excludes random noise signals and allows the consecutive molecular formula prediction only for signals with an isotopic pattern. Only protonated or deprotonated ions with a charge  $\pm 1$  or  $\pm 2$  were analyzed. Molecular features that corresponded to their adducts  $[\text{M}+\text{NH}_4]^+$ ,  $[\text{M}+\text{Na}]^+$ ,  $[\text{M}+2\text{Na}]^{2+}$ ,  $[\text{M}+\text{K}]^+$ ,  $[\text{M}+2\text{K}]^{2+}$ ,  $[\text{M}+\text{CH}_3\text{OH}]^+$  in PI mode and  $[\text{M}-\text{H}_2\text{O}-\text{H}]^-$ ,  $[\text{M}-2(\text{H}_2\text{O})-\text{H}]^-$ ,  $[\text{M}-2\text{H}+\text{K}]^-$ ,  $[\text{M}-2\text{H}+\text{Na}]^-$ ,  $[\text{M}-\text{Cl}]^-$ ,  $[\text{M}-\text{H}+\text{HAc}]^-$ ,  $[\text{M}+\text{Br}]^-$ ,  $[\text{M}-\text{H}+\text{FA}]^-$  in NI mode were removed. The intensity of the adduct features was not added to the intensity of the primary molecular features because it does not offer a significant improvement of the data (Nürenberg et al., 2015). The IS showed prevalent ionization  $[\text{M}+\text{H}]^+$  in PI and  $[\text{M}-\text{H}]^-$  in NI modes.

The formulae were predicted using the atomic ranges: C<sub>1-80</sub>, H<sub>1-100</sub>, O<sub>0-20</sub>, N<sub>0-15</sub>, S<sub>0-4</sub>, Cl<sub>0-4</sub>, Br<sub>0-4</sub>, isotopic pattern score >60%, and mass error <5 ppm. Atomic ranges for the prediction of molecular formula of micro-contaminants and DOM were derived from a database of 14631 micro-contaminants (Aalizadeh et al., 2017). Formulae were predicted and, where discovered, assigned to molecules with a neutral mass under 1000 Da to prevent a high rate of false predictions for heavy molecules. Additionally, to the heuristic rules applied by the software, the atomic ratios of generated formulae were H/C < 3.2, O/C < 1.2, N/C < 1.3, and S/C < 0.8 corresponding to 99.7% of registered small molecules (Kind and Fiehn, 2007). Using R we selected a molecular formula candidate with the smallest absolute mass deviation from 10% of highest isotopic pattern scores to represent the elemental composition of a molecular feature. MZmine data extraction and data cleaning in R are described in detail in (Yaroslav Verkh et al. (in preparation)).

79 pharmaceuticals were measured in PI and NI modes using the same spectrometric conditions as the samples in this study to check the applicability of the formula prediction. The pharmaceuticals' samples underwent the same signal detection,  $m/z$  correction, isotope grouping, and formula prediction. A subsequent targeted screening for  $[\text{M}+\text{H}]^+$  in PI and  $[\text{M}-\text{H}]^-$  in NI modes with an  $m/z$  error <5 ppm was used to identify the molecular features of the pharmaceuticals within the non-targeted data. 69 out of 79 (87%) molecular formulae were predicted correctly.

The baseline was corrected by subtracting the intensity of a signal in the solvent blank from the matching signal in the sample, but only for signals with an intensity  $\text{Ratio}_{\text{Sample:Blank}} > 3$ , while signals with a lower ratio were ignored. This approach accounts for intensity deviations caused by the matrix suppression (Phungsai et al., 2016). The coefficient of variation (CV) of feature intensity among triplicates of a sample filtered out random noise with a conservative value of 30%. The CV of IS in the blank, influent and effluent of the secondary treatment are below the 30% threshold independent of the sample matrix for 97% of records. The 2 outliers out of 32 IS were ronidazole-*d*<sub>3</sub> and simvastatin-*d*<sub>6</sub>. The presented results are for the combined data of PI and NI modes unless stated otherwise.

## 2.4. Analytical tools

The estimation of the statistically significant differences between the properties of molecular features of any two samples was performed under the assumption of independent data unless stated otherwise. The statistical significance was tested with a two-sided Welch two-sample *t*-test for variables with a distribution resembling a normal one or skewed distributions with sample sizes >50 for a confidence interval 95% and  $\alpha$  0.05. Skewed distributions were log-transformed where necessary using  $\log_2$  to obtain a normal distribution. Differences in location shifts of non-normally distributed variables were tested with a Wilcoxon rank-sum test with continuity correction.

The KMD results presented herein were programmed in R and tested with the data of surfactants ( $-CH_2-$  and  $-C_2H_4O-$  moieties) and halogenated substances ( $-CF_2-$  and  $-H/ + Cl$  moieties) using the NORMAN database (Schymanski et al., 2014; Trier Xenia et al., 2015) and PCB suspects (Grabowska, 2010). The homologs of real substances were retained having KMD error <5 ppm and retention time <2 min among the homologs of a series. In real wastewater data, the homolog series of  $-CH_2-$  were constricted to at least seven homologs per series in the influent and three homologs per series in effluent of secondary treatment. Additionally, the effluent only contained series with KMD detected in influent to prioritize the comparison of KMD homologs among the samples.

## 3. Results and discussion

### 3.1. Validity of tertiary grab samples

Two grab samples of each tertiary treatment step were used to check the consistency of the sampling, before comparing the results of secondary and tertiary treatment. The sample pairs were compared using four variables. Mass, retention time (RT), logarithmic intensity (Int) served as raw data outputs and the count of N atoms as a measure of elemental properties. The sample pairs were considered significantly similar upon exceeding a *p*-value of 0.05 in hypothesis testing. The grab samples of the sand filtration ( $p_{\text{mass}} = 0.71$ ,  $p_{\text{RT}} = 0.54$ ,  $p_{\text{Int}} = 0.30$ ,  $p_{\text{N}} = 0.75$ ) and the chlorination ( $p_{\text{mass}} = 0.87$ ,  $p_{\text{RT}} = 0.62$ ,  $p_{\text{Int}} = 0.95$ ,  $p_{\text{N}} = 0.54$ ) were significantly similar. The UV-treatment samples had an intensity difference of 1.2 times, but otherwise were significantly similar ( $p_{\text{mass}} = 0.86$ ,  $p_{\text{RT}} = 0.71$ ,  $p_{\text{Int}} = 0.03$ ,  $p_{\text{N}} = 0.44$ ). Grab samples were largely stable over the 24 h period due to the hypothesis testing.

Composite secondary effluent and grab secondary effluent differed in mass and retention time distribution ( $p_{\text{mass}} = 0.03$ ,  $p_{\text{RT}} = 0.72 \cdot 10^{-3}$ ,  $p_{\text{Int}} = 0.06$ ,  $p_{\text{N}} = 0.71$ ). The difference, given the same data extraction, shows that care has to be exercised when comparing molecular features between composite samples and grab samples in the study.

Differences in the treatment series of grab secondary effluent and tertiary treatment corresponding to the same time series were explored using one way ANOVA tests. The stages did not differ significantly in their properties by exceeding the *p*-value of 0.05, except considering the retention time. ( $\text{anova}_{\text{mass}}$ :  $F(3, 4763) = 1.09$ ,  $p = 0.35$ ;  $\text{anova}_{\text{RT}}$ :  $F(3, 4763) = 2.83$ ,  $p = 0.04$ ;  $\text{anova}_{\text{Int}}$ :  $F(3, 4763) = 0.80$ ,  $p = 0.49$ ;  $\text{anova}_{\text{N}}$ :  $F(3, 2099) = 0.31$ ,  $p = 0.82$ ; homogeneity of variance was confirmed.) This shows that the detected complement of features is largely stable during the tertiary treatment. DOM and small molecules have been shown to react in advanced treatment technologies (Papageorgiou et al., 2017; Richardson and Postigo, 2016). However, the tertiary treatments of sand filtration, UV, and chlorination are designed to remove the suspended particles of biomass from the secondary treatment and to remove biological constituents and can

underperform with respect to small molecules (Paredes et al., 2016; Zhang et al., 2017).

### 3.2. Distribution of substances

In the secondary treatment, 1617 of 2409 (67%) detected molecular features were removed and 255 of 1047 (24%) new features appeared in the secondary effluent (Table 1). Biological treatment involves both metabolic and co-metabolic reactions that form TPs. This process explains a large number of removed and appeared molecular features. The tertiary treatment removed a minute portion of detected molecular features but added new 379 of 1106 features (34% of tertiary effluent). Compared to the conventional COD, BOD or TOC measurements, which give an average estimate of the DOM content, the presented method tracks changes at the level of subsets of DOM constituents. This approach can benefit specialized treatments for which the appearance or removal of DOM fractions is more important than the bulk removal. For example, the tracking of fractions of soluble microbial products might improve the understanding of membrane fouling in reverse osmosis and nanofiltration (Jarusutthirak and Amy, 2006).

### 3.3. Mass changes

After recognizing interesting subsets of features we explored mass changes within the subsets since mass is the most intrinsic property of molecular features in HRMS. The secondary effluent shows a mean mass drop of 129 Da ( $p < 2.2 \cdot 10^{-16}$ ) compared to the influent. There is no significant mass drop between the secondary and tertiary treatments ( $p = 0.27$ ).

However, the mean mass of appeared features drops by 163 Da compared to the disappeared ones ( $p = 7.26 \cdot 10^{-13}$ ) (Fig. 1). The drop is considerably larger for this transformed DOM compared to the bulk DOM in secondary treatment, showing that major changes can be masked by bulk properties. Meanwhile, features detected both in influent and effluent of tertiary treatment do not show a mean mass change which hints at their recalcitrant nature. Often, SEC combined with a carbon detector is used to estimate the weight distributions of DOM and the bulk rejection estimation with experimental ease (Jarusutthirak and Amy, 2006). However, the SEC spectra lack the high resolution of LC-HRMS, the possibility to extract subsets of signals of interest and the possibility to correlate masses of individual substances in these fractions to atomic properties such as unsaturation or atomic ratios of DOM.

### 3.4. Intensity changes

The possibility to connect feature subsets to the spectrometric intensity of individual molecular features is also a powerful incentive for non-targeted analysis. The mean absolute intensity decreases by 3.38 times in the secondary effluent ( $p < 2.20 \cdot 10^{-16}$ ). Biological treatment is known to achieve high TOC removal efficiencies for the entire DOM (Evgenidou et al., 2015), yet many medium polar substances of environmental concern are either recalcitrant or are transformed without mineralization (Prasse et al., 2015). A high number of molecular features can be detected in HRMS of effluent wastewater even while controlling DOC and BOD (Tseng et al., 2013). Therefore, it is interesting that the mean of the absolute intensity differences of 792 features, detected before and after secondary treatment, drops by 2.5 times ( $p < 2.20 \cdot 10^{-16}$  for paired data) in effluent compared to the influent. This value is lower than for the overall intensity decrease hinting at the recalcitrant nature of these constituents. Together with a high removal of molecular features in the secondary treatment, the data agree with previous work (Park et al., 2009). A fraction of recalcitrant

**Table 1**

The number of the total detected molecular features, the percentage of all features in a fraction compared to a number of influent features, and only the features with an assigned molecular formula in fractions of the samples. The fractions of both secondary and tertiary effluent relate to the influent.

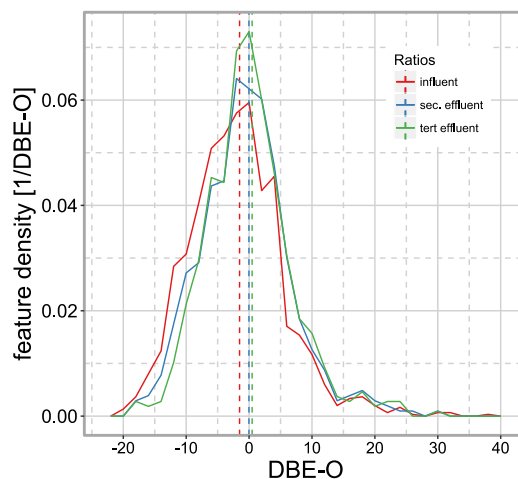
Fraction	Total features	% of influent	Features with assigned formula
Influent <sub>Total</sub>	2409	100	1495
Influent <sub>Disappeared</sub>	1617	67	1069
Secondary effluent <sub>Total</sub>	1047	44	515
Secondary effluent <sub>Appeared</sub>	255	11	89
Secondary effluent <sub>Recalcitrant</sub>	792	33	426
Tertiary effluent <sub>Total</sub>	1106	46	541
Tertiary effluent <sub>Appeared</sub>	379	16	161
Tertiary effluent <sub>Recalcitrant</sub>	727	30	380

features shows an intensity increase, which can be connected to a lower matrix effect in the effluent. There is no significant difference in feature intensity between the secondary and tertiary effluents of the measured samples ( $p = 0.66$ ), again hinting at their recalcitrant nature.

### 3.5. Chemical profile

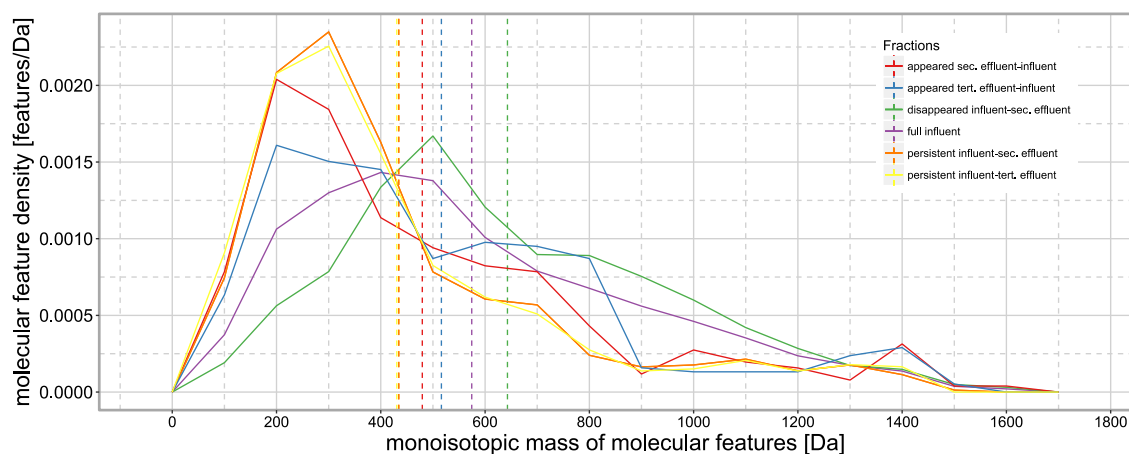
The feature subset analysis was supplemented with DBE-O exploration. The average DBE-O increased by 1.5 ( $p = 6.92 \cdot 10^{-5}$ ) in secondary effluent (Fig. 2). Moreover, the appearing molecular features show a significant increase of DBE-O of 4 compared to the disappearing ( $p = 1.00 \cdot 10^{-5}$ ). These fractions are interesting for the estimation of metabolic activity and the observations are supported by the documented accumulation of aromaticity in biological treatment (Park et al., 2009). The established estimation of aromaticity in wastewater DOM using SUVA<sub>254</sub> allows recognizing the presence of aromatic compounds. The presented method, however, estimates the unsaturation of DOM and expands the analysis to not UV-active molecules in subsets of molecular features.

Additionally to the explored unsaturation by DBE-O, the chemical properties of DOM were investigated using the van Krevelen plot. It estimated types of wastewater DOM in secondary treatment depending on the characteristic regions in the diagram (Fig. 3), yet no significant difference was observed between secondary and tertiary treatments. The diagram deviates from a distribution in a previous study of a biological treatment (Maizel and Remucal, 2017), where the majority of molecular features falls

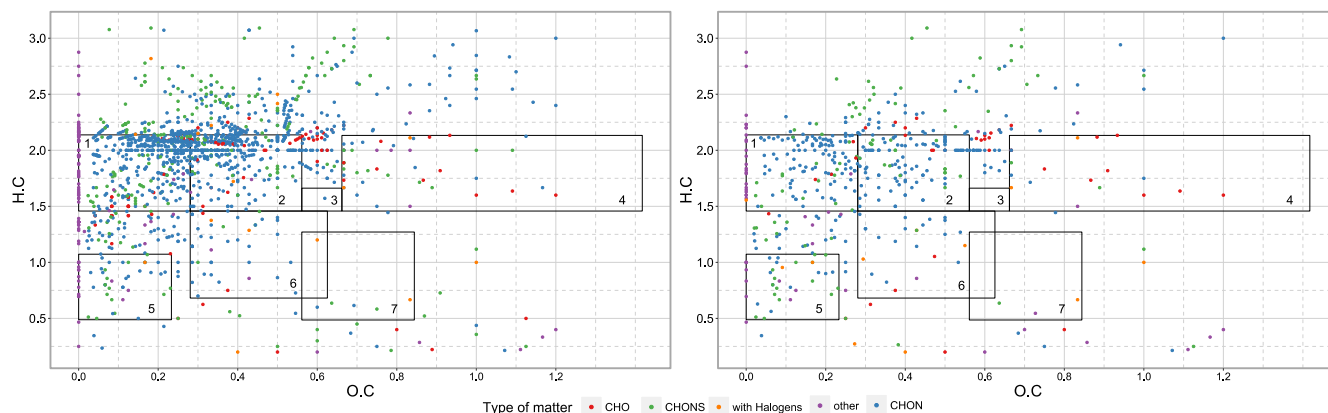


**Fig. 2.** Density distribution of DBE-O of molecular features with assigned molecular formulae in influent and effluent of secondary treatment, and effluent of tertiary treatment. The dashed lines show means of distributions.

within the seven regions. Our van Krevelen plot shows many features with  $H/C > 2$ , while the literature's features do not exceed  $H/C \leq 2$ , which is caused by different HRMS acquisition and approach to the prediction of elemental composition. Different to the literature, our plot shows features with  $H/C < 0.5$  and  $O/C$  up to 1.2, which would correspond to highly oxygenated, unsaturated matter. Here, the different MS acquisition and data treatment might have led to highlighting different subsets of DOM. In our analysis, DOM is largely aliphatic and the oxygen content drops with increasing unsaturation. The overall feature density decreases in the secondary treatment. But, none of the seven regions displayed a disproportional change of feature density in course of treatment compared to other regions. This indicates that DOM composition stayed largely constant throughout the treatment.  $N/C$  in the van Krevelen plot correlates with  $H/C$ , providing a link between the unsaturation and the content of nitrogen (Fig S1 in SI).  $N/C$  of DOM did not display regional changes in the secondary treatment, showing that the detected molecular features were not subject to a transformation detectable by in the patterns of the plot. In Fig. 3 The CHONS group shows a spread to either higher aliphatic or lower aromatic  $H/C$  values compared to CHO and CHON which suggests that there is a level of clustering by the elemental



**Fig. 1.** Density distribution of molecular feature masses in the subsets of features in secondary influent, secondary and tertiary effluents compared to the full complement of the influent. The dashed lines show means of distributions.



**Fig. 3.** Van Krevelen plot of molecular features in secondary influent and effluent. Elemental composition shown as color, where “other” signifies the not explicitly mentioned combinations of elements C, H, O, N, S. The rectangles, reproduced from (Maizel and Remucal, 2017), depict NOM regions of 1 - lipids, 2 - proteins, 3 - amino sugars, 4 - carbohydrates, 5 - condensed hydrocarbons, 6 - lignin, and 7 - tannin. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

composition which has to be explored in the future. Our LCMS acquisition, as well as data treatment, prioritized compounds of anthropogenic origin. Many molecular features in our analysis do not belong to the regions of DOM applied by Maizel et al., and they do not follow the literature's observations for NOM and effluent DOM. We suggest that due to the shifted focus of the analysis we see a complement of features which might be of anthropogenic origin and therefore does not look like NOM in the diagram. This focus on the anthropogenic fraction in the van Krevelen analysis adds to the extensively researched field of NOM and effluent DOM.

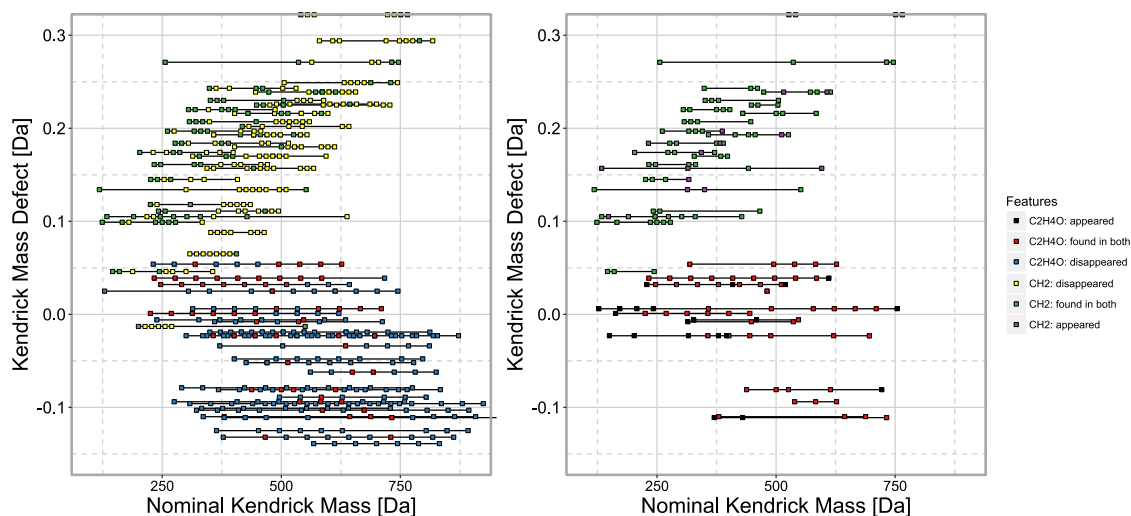
### 3.6. Kendrick mass defect

The exploration of transformation pairs above helped us to select the most interesting moieties for the KMD analysis. We uncovered 31 homologous  $-CH_2-$  series in the secondary treatment (Fig. 4), which often correspond to surfactants in a municipal wastewater treatment. The plot showed a removal of heavy constituents. This removal of heavy homologs, however, was not observed for 38 series of  $-C_2H_4O-$  homologs. Most  $-C_2H_4O-$  mass series were completely removed, while a few were almost completely recalcitrant in the secondary effluent. There are more

$-CH_2-$  and  $-C_2H_4O-$  homologous series in raw data than shown in the plots, but we prioritized pronounced series by involving a strict set of rules. We ranked the exclusion of fake homologs over losing true homologs. Additionally, we set the minimum length of a homolog series for  $-CH_2-$  and  $-C_2H_4O-$   $<7$  in secondary influent which is a high number of homologs in one series. In case of abundant surfactant series, this prioritization improves the interpretation using the graphic nature of the Kendrick plot. We take into account that the mechanism of TP formation for DOM is more complicated than a sequential cleavage of  $-CH_2-$  (Federle and Itrich, 2006; Vega Morales et al., 2009). Also, the detection of homolog series is not straightforward and depends both on the quality of acquired data and the restrictions of the picking algorithm (Loos and Singer, 2017). In this study, Kendrick plots helped to uncover prominent homologous series of prescreened structural moieties of interest which will be used for further prioritized research.

### 3.7. Benefits and challenges

The focus of the study was not DOM enrichment but data analysis. While being aware of the possible shortcomings of the



**Fig. 4.** 31  $-CH_2-$  and 38  $-C_2H_4O-$  homolog series in secondary influent (A) with more than six molecular features in a series and their development in secondary effluent (B) for PI mode data.

applied pre-treatment and LC-HRMS approach, the analysis detects significant changes of various DOM attributes during WWT. The fractions of high interest, such as anthropogenic organics, are retained by the procedure, as signified by the occurrence of IS throughout the chromatographic range. The focus of data treatment is to extract many features while reducing inclusion of noise. Although noise signals cannot be avoided in big datasets (Nürenberg et al., 2015), properly selected settings lower their amount. The application of retention time to distinguish molecular features facilitates the task. Similarly, the prediction of elemental composition for DOM is challenging, especially for high-molecular-weight compounds. Predictions were improved by using heuristic rules and isotopic information. The prediction of P- and F-containing formulas remains difficult even having a low mass error (Kind and Fiehn, 2006).

#### 4. Conclusion

We tested and applied non-targeted LC-MS data extraction and analysis of direct samples to fingerprint a multi-stage wastewater treatment system. Significant changes of DOM were identified. Secondary treatment removed 1617 of 2409 (67%) detected molecular features while 255 of 1047 (24%) features appeared after secondary treatment. Subsets of molecular features within the treatment stages exhibited significant changes. The unsaturation increased in secondary treatment by 1.5 DBE-O in the bulk DOM and by 4.0 DBE-O comparing the subsets of disappeared and appeared molecular features. The van Krevelen plots indicated highly oxygenated, unsaturated matter and highly aliphatic, nitrogen-rich matter. The distribution of features did not follow the established regions for NOM and was attributed to anthropogenic compounds found in wastewater. KMD plots uncovered 31 prioritized, abundant homologous series for the  $-CH_2-$  moiety. The fraction of  $-CH_2-$  series exhibited a preferential removal of homologs at high masses. On the other hand, 38  $-C_2H_2O-$  series were largely removed, suggesting complex behavior of heterogeneous wastewater DOM. Altogether, the presented not-targeted analysis combined various existing techniques to analyze wastewater. It explored the bulk properties of wastewater LC-HRMS data and the subsets of features within the samples. The approach has the potential to describe a chemical fingerprint of DOM transformation in wastewater. However, the analysis is still limited by the scope of LC-MS detection, signal suppression in the samples and the precision of data extraction.

#### Conflicts of interest

The authors declare no competing financial interest.

#### Acknowledgement

This work has been supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642904 - TreatRec ITN-EID project. The sampling by Lluís Corominas (ICRA, Spain) is greatly acknowledged. We thank Antoni Ginebreda (IDAEA, Spain) for the advice on the application of statistical tools and Peter Daldorff (Atkins, UK) for proofreading. The authors are part of the Consolidated Research Groups of the Generalitat de Catalunya (2014 SGR 291—ICRA).

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.chemosphere.2018.02.095>.

#### References

- Aalizadeh, R., Schymanski, E., Thomaidis, N., Williams, A., Glowacka, N., Cirka, L., 2017. Merged Norman Suspect List "SusDat" [WWW Document]. Norman Netw. [http://www.norman-network.com/sites/default/files/files/suspectListExchange/NORMAN\\_SusDat\\_MergedSuspects24052017.xlsx](http://www.norman-network.com/sites/default/files/files/suspectListExchange/NORMAN_SusDat_MergedSuspects24052017.xlsx) (accessed 7.20.17).
- Aral, H., Çelik, K.S., Altındağ, R., Aral, T., 2017. Synthesis, characterization, and application of a novel multifunctional stationary phase for hydrophilic interaction/reversed phase mixed-mode chromatography. *Talanta* 174, 703–714. <https://doi.org/10.1016/j.talanta.2017.07.014>.
- Causanilles, A., Kinyua, J., Ruttkies, C., van Nuijs, A.L.N., Emke, E., Covaci, A., de Voogt, P., 2017. Qualitative screening for new psychoactive substances in wastewater collected during a city festival using liquid chromatography coupled to high-resolution mass spectrometry. *Chemosphere* 184, 1186–1193. <https://doi.org/10.1016/j.chemosphere.2017.06.101>.
- Chen, W., Westerhoff, P., Leenheer, J.A., Booksh, K., 2003. Fluorescence excitation–emission matrix regional integration to quantify spectra for dissolved organic matter. *Environ. Sci. Technol.* 37, 5701–5710. <https://doi.org/10.1021/es034354c>.
- Clesceri, L.S., Eaton, A.D., Greenberg, A.E., 1998. *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association.
- Cortés-Francisco, N., Caixach, J., 2013. Molecular characterization of dissolved organic matter through a desalination process by high resolution mass spectrometry. *Environ. Sci. Technol.* 47, 9619–9627. <https://doi.org/10.1021/es4000388>.
- Cortés-Francisco, N., Harir, M., Lucio, M., Ribera, G., Martínez-Lladó, X., Rovira, M., Schmitt-Kopplin, P., Hertkorn, N., Caixach, J., 2014. High-field FT-ICR mass spectrometry and NMR spectroscopy to characterize DOM removal through a nanofiltration pilot plant. *Water Res.* 67, 154–165. <https://doi.org/10.1016/j.watres.2014.08.046>.
- Daughton, C.G., 2004. Non-regulated water contaminants: emerging research. *Environ. Impact Assess. Rev.* 24, 711–732. <https://doi.org/10.1016/j.eiar.2004.06.003>.
- Deblonde, T., Cossu-Leguille, C., Hartemann, P., 2011. Emerging pollutants in wastewater: a review of the literature. *Int. J. Hyg Environ. Health* 214, 442–448. <https://doi.org/10.1016/j.ijheh.2011.08.002>.
- Erve, J.C.L., Gu, M., Wang, Y., DeMaio, W., Talaat, R.E., 2009. Spectral accuracy of molecular ions in an LTQ/Orbitrap mass spectrometer and implications for elemental composition determination. *J. Am. Soc. Mass Spectrom.* 20, 2058–2069. <https://doi.org/10.1016/j.jasms.2009.07.014>.
- Evgenidou, E.N., Konstantinou, I.K., Lambropoulou, D.A., 2015. Occurrence and removal of transformation products of PPCPs and illicit drugs in wastewaters: a review. *Sci. Total Environ.* 505, 905–926. <https://doi.org/10.1016/j.scitotenv.2014.10.021>.
- Fang, Z., He, C., Li, Y., Chung, K.H., Xu, C., Shi, Q., 2017. Fractionation and characterization of dissolved organic matter (DOM) in refinery wastewater by revised phase retention and ion-exchange adsorption solid phase extraction followed by ESI FT-ICR MS. *Talanta* 162, 466–473. <https://doi.org/10.1016/j.talanta.2016.10.064>.
- Federle, T.W., Itrich, N.R., 2006. Fate of free and linear alcohol-ethoxylate-derived fatty alcohols in activated sludge. *Ecotoxicol. Environ. Saf.* 64, 30–41. <https://doi.org/10.1016/j.ecoenv.2005.05.008>.
- Grabowska, I., 2010. Polychlorinated biphenyls (pcbs) in Poland: occurrence, determination and degradation. *Pol. J. Environ. Stud.* 19.
- Herzsprung, P., von Tümpling, W., Hertkorn, N., Harir, M., Büttner, O., Bravidor, J., Friese, K., Schmitt-Kopplin, P., 2012. Variations of dom quality in inflows of a drinking water reservoir: linking of van krevelen diagrams with eemf spectra by rank correlation. *Environ. Sci. Technol.* 46, 5511–5518. <https://doi.org/10.1021/es300345c>.
- Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? *Environ. Sci. Technol.* 51, 11505–11512. <https://doi.org/10.1021/acs.est.7b02184>.
- Huber, S.A., Balz, A., Abert, M., Pronk, W., 2011. Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography – organic carbon detection – organic nitrogen detection (LC-OCD-OND). *Water Res.* 45, 879–885. <https://doi.org/10.1016/j.watres.2010.09.023>.
- Hughey, C.A., Hendrickson, C.L., Rodgers, R.P., Marshall, A.G., Qian, K., 2001. Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* 73, 4676–4681. <https://doi.org/10.1021/ac010560w>.
- Iparraguirre, A., Navarro, P., Rodil, R., Prieto, A., Olivares, M., Etxebarria, N., Zuloaga, O., 2014. Matrix effect during the membrane-assisted solvent extraction coupled to liquid chromatography tandem mass spectrometry for the determination of a variety of endocrine disrupting compounds in wastewater. *J. Chromatogr. A* 1356, 163–170. <https://doi.org/10.1016/j.chroma.2014.06.051>.
- Jarusutthirak, C., Amy, G., 2006. Role of soluble microbial products (SMP) in membrane fouling and flux decline. *Environ. Sci. Technol.* 40, 969–974.
- Jobst, K.J., Shen, L., Reiner, E.J., Taguchi, V.Y., Helm, P.A., McCrindle, R., Backus, S., 2013. The use of mass defect plots for the identification of (novel) halogenated contaminants in the environment. *Anal. Bioanal. Chem.* 405, 3289–3297. <https://doi.org/10.1007/s00216-013-6735-2>.
- Kim, S., Kramer, R.W., Hatcher, P.G., 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* 75, 5336–5344. <https://doi.org/10.1021/>

- ac034415p.
- Kind, T., Fiehn, O., 2006. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinf.* 7, 234. <https://doi.org/10.1186/1471-2105-7-234>.
- Kind, T., Fiehn, O., 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinf.* 8, 105. <https://doi.org/10.1186/1471-2105-8-105>.
- Koch, B.P., Dittmar, T., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* 20, 926–932. <https://doi.org/10.1002/rcm.2386>.
- Li, Y., Harir, M., Lucio, M., Kanawati, B., Smirnov, K., Flerus, R., Koch, B.P., Schmitt-Kopplin, P., Hertkorn, N., 2016. Proposed guidelines for solid phase extraction of suwannee river dissolved organic matter. *Anal. Chem.* 88, 6680–6688. <https://doi.org/10.1021/acs.analchem.5b04501>.
- Loos, M., Singer, H., 2017. Nontargeted homologue series extraction from hyphenated high resolution mass spectrometry data. *J. Cheminf.* 9 <https://doi.org/10.1186/s13321-017-0197-z>.
- Lu, M., Zhang, Z., Qiao, W., Wei, X., Guan, Yueming, Ma, Q., Guan, Yingchun, 2010. Remediation of petroleum-contaminated soil after composting by sequential treatment with Fenton-like oxidation and biodegradation. *Bioresour. Technol.* 101, 2106–2113. <https://doi.org/10.1016/j.biortech.2009.11.002>.
- Maizel, A.C., Remucal, C.K., 2017. The effect of advanced secondary municipal wastewater treatment on the molecular composition of dissolved organic matter. *Water Res.* 122, 42–52. <https://doi.org/10.1016/j.watres.2017.05.055>.
- Makarov, A., Denisov, E., Lange, O., Horning, S., 2006. Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom.* 17, 977–982. <https://doi.org/10.1016/j.jasms.2006.03.006>.
- Marshall, J.W., Schmitt-Kopplin, P., Schuetz, N., Moritz, F., Roullier-Gall, C., Uhl, J., Colyer, A., Jones, L.L., Rychlik, M., Taylor, A.J., 2018. Monitoring chemical changes during food sterilisation using ultrahigh resolution mass spectrometry. *Food Chem.* 242, 316–322. <https://doi.org/10.1016/j.foodchem.2017.09.074>.
- Michael-Kordatou, I., Michael, C., Duan, X., He, X., Dionysiou, D.D., Mills, M.A., Fatta-Kassinos, D., 2015. Dissolved effluent organic matter: characteristics and potential implications in wastewater treatment and reuse applications. *Water Res.* 77, 213–248. <https://doi.org/10.1016/j.watres.2015.03.011>.
- Minor, E.C., Steinbring, C.J., Longnecker, K., Kujawinski, E.B., 2012. Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry. *Org. Geochem.* 43, 1–11. <https://doi.org/10.1016/j.orggeochem.2011.11.007>.
- Nürnberg, G., Schulz, M., Kunkel, U., Ternes, T.A., 2015. Development and validation of a generic nontarget method based on liquid chromatography – high resolution mass spectrometry analysis for the evaluation of different wastewater treatment options. *J. Chromatogr. A* 1426, 77–90. <https://doi.org/10.1016/j.chroma.2015.11.014>.
- Papageorgiou, A., Stylianou, S.K., Kaffes, P., Zouboulis, A.I., Voutsas, D., 2017. Effects of ozonation pretreatment on natural organic matter and wastewater derived organic matter – possible implications on the formation of ozonation by-products. *Chemosphere* 170, 33–40. <https://doi.org/10.1016/j.chemosphere.2016.12.005>.
- Paredes, L., Fernandez-Fontaina, E., Lema, J.M., Omil, F., Carballa, M., 2016. Understanding the fate of organic micropollutants in sand and granular activated carbon biofiltration systems. *Sci. Total Environ.* 551–552, 640–648. <https://doi.org/10.1016/j.scitotenv.2016.02.008>.
- Park, M.-H., Lee, T.-H., Lee, B.-M., Hur, J., Park, D.-H., 2009. Spectroscopic and chromatographic characterization of wastewater organic matter from a biological treatment plant. *Sensors* 10, 254–265. <https://doi.org/10.3390/s100100254>.
- Phungsai, P., Kurisu, F., Kasuga, I., Furumai, H., 2016. Molecular characterization of low molecular weight dissolved organic matter in water reclamation processes using Orbitrap mass spectrometry. *Water Res.* 100, 526–536. <https://doi.org/10.1016/j.watres.2016.05.047>.
- Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M., 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinf.* 11, 395.
- Prasse, C., Stalter, D., Schulte-Oehlmann, U., Oehlmann, J., Ternes, T.A., 2015. Spoilt for choice: a critical review on the chemical and biological assessment of current wastewater treatment technologies. *Water Res.* 87, 237–270. <https://doi.org/10.1016/j.watres.2015.09.023>.
- R Core Team, 2017. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Remucal, C.K., Cory, R.M., Sander, M., McNeill, K., 2012. Low molecular weight components in an aquatic humic substance as characterized by membrane dialysis and orbitrap mass spectrometry. *Environ. Sci. Technol.* 46, 9350–9359. <https://doi.org/10.1021/es302468g>.
- Richardson, S.D., Postigo, C., 2016. Discovery of new emerging Dbps by high-resolution mass spectrometry. In: *Comprehensive Analytical Chemistry*. Elsevier, pp. 335–356. <https://doi.org/10.1016/bs.coac.2016.01.008>.
- Richardson, S.D., Ternes, T.A., 2014. Water analysis: emerging contaminants and current issues. *Anal. Chem.* 86, 2813–2848. <https://doi.org/10.1021/ac500508t>.
- Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J., 2014. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. *Environ. Sci. Technol.* 48, 1811–1818. <https://doi.org/10.1021/es4044374>.
- Sleighter, R.L., Hatcher, P.G., 2008. Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Mar. Chem.* 110, 140–152. <https://doi.org/10.1016/j.marchem.2008.04.008>.
- Taylor, P.J., 2005. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry. *Clin. Biochem.* 38, 328–334. <https://doi.org/10.1016/j.clinbiochem.2004.11.007>.
- Tseng, L.Y., Gonsior, M., Schmitt-Kopplin, P., Cooper, W.J., Pitt, P., Rosso, D., 2013. Molecular characteristics and differences of effluent organic matter from parallel activated sludge and integrated fixed-film activated sludge (IFAS) processes. *Environ. Sci. Technol.* <https://doi.org/10.1021/es4002482>, 130827102639005.
- Vega Morales, T., Torres Padrón, M.E., Sosa Ferrera, Z., Santana Rodríguez, J.J., 2009. Determination of alkylphenol ethoxylates and their degradation products in liquid and solid samples. *TrAC Trends Anal. Chem.* 28, 1186–1200. <https://doi.org/10.1016/j.trac.2009.07.011>.
- Yaroslav Verkh, Marko Rozman, Mira Petrovic, n.d. Extraction and cleansing of data for a non-targeted analysis of high-resolution mass spectrometry data of wastewater. MethodX (in preparation).
- Wode, F., van Baar, P., Dünnebier, U., Hecht, F., Taute, T., Jekel, M., Reemtsma, T., 2015. Search for over 2000 current and legacy micropollutants on a wastewater infiltration site with a UPLC-high resolution MS target screening method. *Water Res.* 69, 274–283. <https://doi.org/10.1016/j.watres.2014.11.034>.
- Xenia, Trier, Lunderberg, David, Peaslee, Graham, Wang, Zhanyun, 2015. PFAS Suspect List (Fluorinated Substances) [WWW Document]. Norman Netw. <http://www.norman-network.com/sites/default/files/files/suspectListExchange/PFAS%20csv%20database%2011.26.15.csv> (accessed 3.2.17).
- Zhang, H., Zhang, Y., Shi, Q., Hu, J., Chu, M., Yu, J., Yang, M., 2012a. Study on transformation of natural organic matter in source water during chlorination and its chlorinated products using ultrahigh resolution mass spectrometry. *Environ. Sci. Technol.* 46, 4396–4402. <https://doi.org/10.1021/es203587q>.
- Zhang, H., Zhang, Y., Shi, Q., Ren, S., Yu, J., Ji, F., Luo, W., Yang, M., 2012b. Characterization of low molecular weight dissolved natural organic matter along the treatment trait of a waterworks using Fourier transform ion cyclotron resonance mass spectrometry. *Water Res.* 46, 5197–5204. <https://doi.org/10.1016/j.watres.2012.07.004>.
- Zhang, Y., Zhang, N., Zhao, P., Niu, Z., 2017. Characteristics of molecular weight distribution of dissolved organic matter in bromide-containing water and disinfection by-product formation properties during treatment processes. *J. Environ. Sci.* <https://doi.org/10.1016/j.jes.2017.03.013>.