Profiling of wastewater treatment systems by statistical non-targeted high resolution mass spectrometry analysis of dissolved organic matter

Yaroslav Verkh[†], Marko Rozman[†,‡] and Mira Petrovic*[,†,§]

[†]Catalan Institute for Water Research (ICRA), Carrer Emili Grahit 101, 17003 Girona, Spain.

[‡]Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb.

[§]Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

ABSTRACT: The quality monitoring of the dissolved organic matter (DOM) in wastewater was, hitherto, largely confined to a limited number of concentration measurements of select DOM fractions or micro-contaminants to determine the removal efficiency of a wastewater treatment. Current methods do not necessarily reflect the true performance with regard to environmental and public health risk, especially for complex treatment technologies, which generate transformation- and by-products. Herein we describe the development and application of a non-targeted analysis of wastewater using liquid chromatography-high resolution mass spectrometry (LC-HRMS) data. We apply the workflow to a nanofiltration (NF) and a biological (BT) treatments to show the merits of our approach on an example of two real wastewater treatments with diametrically different physicochemical properties. The method determined removal preferences in NF for large and aromatic substances and indicated an increase of aromaticity in the effluent organic matter after BT. Additionally, the method showed > 1300 gains and/or losses for relevant metabolic moieties, e.g. CH_2, H_2O, etc. We sorted wastewater DOM into 5 aromaticity and heteroatom depending classes using the van Krevelen plot and obtained 23 CH_2-homolog series utilizing the Kendrick mass defect. The demonstrated approach is a step towards the monitoring of the entire DOM in wastewater and contributes to understanding of the current treatment technologies. Furthermore, it allows to shift the hitherto somewhat reactive targeted monitoring approach to a proactive non-targeted one, where the entire bulk of the wastewater DOM and not select, few priority substances are considered.

Introduction. Wastewater DOM represents a complex, heterogenic mixture of polysaccharides, proteins, lipids, nucleic acids, soluble microbial products and synthetic organic chemicals. The fraction of the synthetic organic chemicals includes, among others, surfactants, personal care products, pharmaceuticals, biocides, pesticides, industrial chemicals, but also a wide range of biologically active transformation products (TPs), intermediates and metabolites[1]. Many of these compounds can be hazardous even at a low concentration and may raise concerns regarding their release into the environment[2]. The composition of wastewater DOM is related to the type of wastewater (municipal, industrial, hospital's effluent, runoff from fields, etc.) and the nature of the treatment process applied at wastewater treatment plants (WWTP)[3]. Currently, the efficiency of DOM treatment at WWTP is evaluated through measurements of the chemical oxygen demand (COD), biological oxygen demand (BOD) and the total organic carbon (TOC) with additional specialized tools for the prioritized fractions of DOM, e.g. the dissolved organic halide to estimate the halogenated organics, the assessment of the aromaticity using the specific UV absorbance ($SUVA_{254}$) for the UV active fractions of DOM, size exclusion chromatography (SEC) to identify mass/size distributions of C-or N-containing constituents or excitation-emission-matrix fluorescence that was used to identify substance classes in natural OM[4–6]. However, while they reveal the chemical characteristics to a certain extend and the abundance of organics in WWTP influent and effluent, they do not provide information on the presence of unique organic substances and need a lot of expert knowledge to be united into one data stream. Therefore new strategies are needed to assess the overall quality of wastewater treatment[7].

So far, the research community is mostly focused on a detection of a small number of hand-picked contaminants and occasionally on their TPs[8]. By definition these approaches largely neglect hundreds to thousands of DOM constituents, present in the influent or appearing during the treatment process. Overlooking them limits our understanding of the impact of the effluent organic matter on the environment, since some of them are potentially hazardous. 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 under-

standing of the treatment processes and DOM transformation. This in turn 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 contribute to a "full description" of wastewater DOM. However, HRMS analysis results in a high amount of data that is difficult to process and analyse. The present HRMS approaches (e.g. suspect screening) attempt to identify selected dominant signals using chemical databases and thus elucidate the structure of unique chemicals[9]. Due to the complexity of wastewater an HRMS analysis yields $10^3 - 10^5$ signals, which makes a manual structural identification of that many unique substances nearly impossible. Therefore the methodology generally uncovers only a small fraction of compounds and omits the unknown majority of wastewater DOM[10]. Even without a tentative structural identification of particular substances the large number of signals with assigned elemental compositions can be used to uncover physiochemical changes in the treatment[11]. Yet, while omitting the problem of the manual treatment of data in a suspect screening, the challenge of a non-targeted method is the occurrence of false positives and negatives that arises through the necessarily complex HRMS data extraction and clean-up methodology.

In this work, we present a statistical fingerprinting: LC-HRMS data treatment of wastewater DOM that enables to monitor and to gain a deeper understanding of wastewater treatment processes. The statistical HRMS data treatment emerged in the fields of petroleomics and characterization of natural organic matter (NOM)[12,13], however generally without the application of LC, which may enhance the resolution of the spectral data. Here it is adapted to fingerprint the heterogenic mixture of DOM before and after wastewater treatments. We discuss the introduced methodology, compare it to conventional DOM analytics and test it on two real wastewater treatment systems, NF and a secondary BT.

METHODS

Reagents, sampling and pre-concentration. HPLC grade solvents methanol, water and acetonitrile were purchased from Fisher (Germany) and buffer solutions were prepared using HPLC grade ammonium acetate (Fisher, Germany), ammonium formate (Acros Organics, Spain), 30 % $NH_{3(aq)}$ and 37 % $HCl_{(aq)}$ (Panreac, Spain) and 98-100 % formic acid (Merck, Germany). Feed (influent) and permeate (effluent) of NF treatment were taken from a pilot WWTP in Quart, Spain. NF was applied as a tertiary treatment and its feed corresponds to the effluent of a secondary BT. The influent and effluent of a secondary BT used in this study were taken from a municipal WWTP in Celrà, Spain. Subsequently, the samples were pre-concentrated following a modified procedure from Dittmar et al.[14]. The details of the treatment works and the pre-concentration procedure are given in the Supporting Information (SI, Sections 1 & 2). The properties of pharmaceutical internal

standards (IS) added to BT samples are listed in SI, Table S-3.

LC HRMS analysis. HPLC-HRMS analysis was performed on a LTQ-Orbitrap VelosTM coupled with the Aria TLX-1 HPLC system (Thermo Fisher Scientific, USA). The sample injection, separation, and MS acquisition were carried out automatically. The chromatographic separation was achieved on Acquity UPLC® HSS T_3 (2.1 mm × 50 mm, 1.8 µm particle size, Waters UK) in the positive ion mode (PI) and on Acquity UPLC® BEH C_{18} (2.1 mm × 50 mm, 1.7 µm particle size, Waters UK) chromatographic column in the negative ion mode (NI). Detailed parameters and experimental solvent gradients can be found in SI, Section 3.

Data extraction with *MzMine 2.17* and *R*. The mass tolerance for the processing of LC-MS spectra was set to 0.001 Da or 5 ppm. The duration boundaries of the extracted ion chromatograms were 0.030-0.800 min. Two consecutive formula prediction runs were performed with the following atomic ranges: First run: C_{1-60}, H_{3-80}, O_{0-20}, N_{0-10}, P_{0-1}, S_{0-2}, Cl_{0-4}, Br_{0-2}, F_0; Second run: C_{1-80}, H_{3-100}, O_{0-20}, N_{0-15}, P_{0-2}, S_{0-2}, Cl_{0-4}, Br_{0-2}, F_{0-2}. Applied *Mzmine 2* modules, the cleaning-up in *R* and the testing of data are described in detail in SI, Section 4.

Data extraction evaluation. The signal threshold for the most software modules was set slightly higher than an established baseline in the raw spectra to exclude the instrumental noise. The applied precision of < 5 ppm for m/z signals during the data extraction was experimentally confirmed with IS (SI, Table S-4). Only features - multiple isotopologues corresponding to the same substance and represented by the monoisotopic signal with the lowest m/z value, were retained in the data sets. This way we made sure that we excluded random noise signals and allowed only substances with an isotopic pattern to be used for the molecular formula prediction. We removed features that corresponded to common adducts under the given ionisation conditions and retained only the protonated or deprotonated singly-charged ions. According to the recommendation of Nürenberg et al.[11] the intensity of the adduct features was not added to the intensity of the primary features, because it doesn't offer a significant improvement of the data. IS showed in all cases a prevalent $[M±H]^±$ ionisation in NF and BT samples.

To find the optimal parameters for the prediction of the molecular formula for DOM and in particular for organic micro-contaminants we modelled the prediction with a theoretical set of micro-contaminants. A database of pharmaceuticals, pesticides and miscellaneous micro-contaminants, which are relevant for the wastewater treatment in the EU, was composed from the lists of high-priority suspects (SI, Figure S-1)[10,15,16]. We confirmed the sensibility of the found atomic ranges using experimental data of twelve pharmaceuticals in PI and NI mode, which were measured at the same conditions as the samples in this study (SI, Table S-5).

We performed a correction of the baseline in a latter step by subtracting the intensity of a signal in the solvent blank from the matching signal in the sample. This ensured that

a correction was applied only to the relevant chromatographic regions. We filtered the triplicate samples by the coefficient of variation (CV) of intensities of features. A conservative value of 10 % offers a good balance between keeping the data precise and an exclusion of true positives.

The estimation of the statistical significance between the variables of samples was performed under the assumption of independent data sets. We executed parametric significance tests for variables with a distribution resembling a normal one for a confidence interval of 95 %. Skewed distributions of variables were log-transformed where necessary using the common logarithm to obtain a normal distribution. More details on the data extraction can be found in SI Section 5.

Analytical tools. To gain an insight into the chemical composition of wastewater NOM we used the van Krevelen plot. It is a widely used tool in the HRMS analysis of organic matter, in which ratio X/C, where X is an element of interest, is plotted against H/C. In petroleomics and NOM chemistry the correlation of particular areas in the plot and functional classes of compounds led to the elucidation of the chemical composition of the organic matter[17,18–20] and a comparison of multiple samples revealed big scale reactions (e.g. oxidation of OM)[21,22]. To identify homologous series for multiple molecular moieties, e.g. $-CH_2-$ or $-C_2H_4O-$ we used the Kendrick Mass Defect (KMD), which is a difference between the Kendrick mass and the measured IUPAC mass. The Kendrick mass is the measured IUPAC mass multiplied by the factor of the nominal and exact IUPAC masses of a molecular moiety. Pattern recognition in Kendrick plots previously elucidated reactions and heteroatom content in the system[12,23,24]. We trained our Kendrick statistics tool for the search of true homologs using the lists of priority substances from the NORMAN database and obtained positive results, since the true homologs were largely retained and false positives were removed.

We successfully applied the previously published correlation between the percentage of hydrogen in a formula (%H) and Relative Mass Defect (RMD) to calculate %H for features without an assigned elemental composition[25]. We calculated a linear fit for the plot restricted to values of RMD > 0 and %H > 3 (SI, Figure S-2 & S-3). Considering only features with RMD > 0 to not strain the model for the NF, we estimated %H for 45 features without a predicted elemental formula in the influent and for 8 in the effluent for the NF data set. Also, this tool presents a quality test of the elemental composition assignment by comparing whether the RMD from the raw m/z data and the estimated %H correlate well, which is the case for the majority of data in our non-target analysis for both NF and BT. The majority of the features in the discarded cluster of %H < 3 and RMD < 50 are unsaturated with H/C < 0.6. The positive RMD is mainly contributed by hydrogen in organic matter and in an aromatic, H-poor matter with a moderate to high level of oxygen a neutral to negative RMD is expected. Features with -320 < RMD < 50 do not follow the linear regression in the %H vs RMD plot and can belong to unsaturated matter or to non-natural substances with a high content of heteroatoms e.g. halogenated chemicals. A screening for heavy heteroatoms in this cluster can uncover substances with wrongfully assigned elemental compositions and focus the research on halogenated DOM.

RESULTS AND DISCUSSION

Distribution of substances. NF is a wastewater treatment process, which relies mainly on the physical retention of DOM constituents present in a feed stream, while no or few chemical or biological transformations are expected to occur. The process is well documented in the scientific literature and it is a good system to test a new, complex analytical procedure such as the presented non-target analysis. In contrast, BT is diametrically different treatment technology in which we expect the occurrence of many TPs due to the bio-chemical DOM transformation by bacteria. The test of the non-targeted analysis on the system allows to estimate its validity for chemically complex systems. The inspection of both systems evaluates the applicability spectrum of this statistical non-target analysis.

Table 1. The number of the total detected features and only the features with an assigned elemental composition in various fractions of NF and BT DOM.

Fraction	NF$_{All}$	NF$_{Formula}$	BT$_{All}$	BT$_{Formula}$
Influent$_{Total}$	2763	2619	2860	2752
Influent$_{Persistent}$	432	423	1303	1264
Influent$_{Disappear}$	2331	2196	1557	1488
Effluent$_{Total}$	741	719	2328	2259
Effluent$_{Persistent}$	432	423	1303	1264
Effluent$_{Appeared}$	309	296	1025	995

In NF 2331 of 2763 (84 %) of detected features were removed by the treatment process, while in the BT 1557 of 2860 (54 %) features were removed (Table 1). Although the number of influent features in NF and BT is similar, the number of effluent BT features is three times higher in the BT. A small number of new substances identified in NF effluent is in a good agreement with fundamental characteristic of the NF treatment since the NF is largely a physicochemical treatment technology and does not induce a major transformation of DOM. In contrast, BT involves both metabolic and co-metabolic processes and leads to the formation of a large variety of TPs[7]. The high number of the appearing TPs in the effluent DOM (36% compared to the influent DOM) illustrates well this fundamental trait of BT. Compared to the conventional COD, BOD or TOC measurements, which give an average estimate of the DOM content, our methodology offers a better insight into a complex DOM mixture, because we track changes at the level of the fractions and individual constituents of DOM.
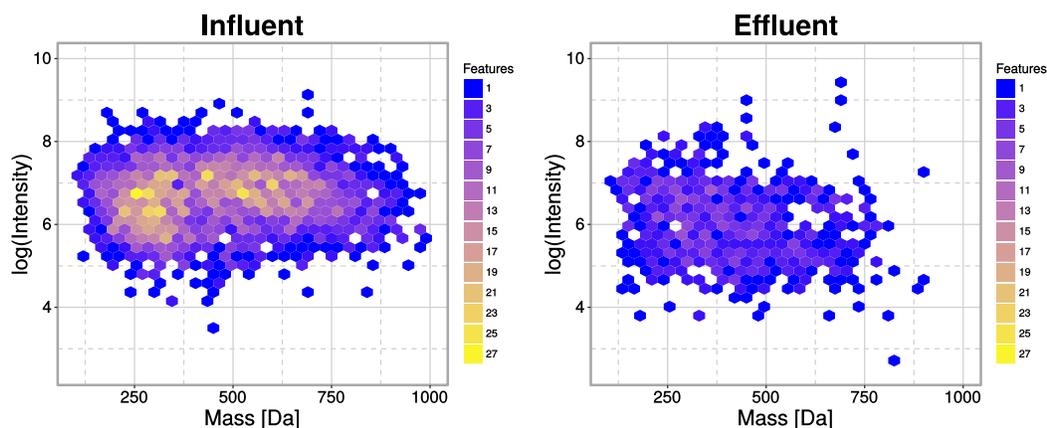
Figure 1. Distribution of masses and logarithmic absolute intensities of the 2763 influent and 732 effluent NF DOM features.
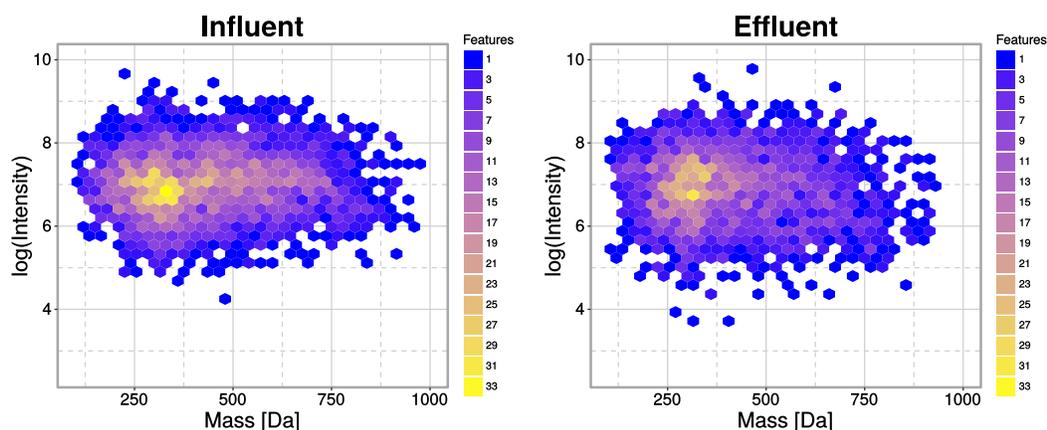

Figure 2. Distribution of masses and logarithmic absolute intensities of the 2860 influent and 2328 effluent BT DOM features.

Mass changes. In NF we observed a significant decrease of 45 Da in the mean mass after the treatment. A calculation of the proportions of the total number of features for mass fractions of respectively 100 Da each, suggests a more efficient rejection of substances with masses above 500 Da, since the feature fractions drop after the treatment for the mass fractions > 500 Da (SI, Figure S-4). Figure 1 illustrates the preferential removal of heavy substances with an almost complete removal of feature density in the region of high masses of the effluent. As an approximation, it can be stated that the mass of organic compounds is proportional to their size, hence NF works well for the large molecules in DOM. The previously researched main causes of the preferential size exclusion in NF are steric hindrance and physical sieving by pores, which corroborates our finds[26]. Also, our analysis was able to locate the threshold of ≈ 500 Da at which the size restriction effects of the NF membrane start to limit the removal, as observed with the decrease of feature proportions for mass ranges.

BT offers a different set of observations, with a mean mass drop of 14 Da in the effluent of BT compared to the influent (Figure 2). However, with our methodology we are able to digitally extract the relevant fractions: the features that were actually transformed. We calculated a mean drop of 30 Da for the appeared features compared to the disappearing ones, which is twice as big for the transformed

DOM compared to the bulk BT DOM. Often SEC combined with a carbon detector is used to estimate the weight distributions of DOM and the bulk rejection estimation with a relative experimental ease[27]. However the SEC spectra do not have the high resolution of the present non-targeted analysis and the possibility to extract fractions of substances digitally, which is a complicated task in hands-on experiments and the possibility to connect masses of individual substances in these fractions to atomic properties as the unsaturation or atomic ratios of DOM.

Intensity changes. Our approach pinpoints the intensity drop of remaining compounds in NF, i.e. features detected in both influent and effluent. We observed a significant drop of 11.8 times for the mean absolute intensity of the remaining 432 features in the effluent (SI, Figure S-5). The mass range of substances that appear in the effluent of NF is similar to the mass range of the features removed from the influent suggesting only a limited transformation of DOM, which may be attributed to metabolism of a possibly existing biofilm on the fouling NF membranes[27]. In the scope of the present research, the existence of such a biofilm could though not be experimentally confirmed but seems plausible as no biocide was employed in the treatment train. In this light, the analysis of disappearing and appearing compounds, which depicts the chemical reactivity of the DOM in wastewater treatment, may allow the monitoring of the activity of a biofilm in advanced water

treatment plants for water recycling or also in seawater desalination. This is of high application interest as biofilm control and the development of early warning tools remains one of the challenges in these processes.

No such clear distinctions in the intensity profile were found in BT under current experimental conditions. The bulk intensities of the influent and the effluent show little variation. BT is known to achieve high TOC removal efficiencies for the entire DOM[28], yet multiple reports show that many medium polar substances of environmental concern are either recalcitrant or are transformed without

mineralization[7]. Since our DOM extraction and analytical method are concentrated on the fraction of similarly small, polar substances we observe an overall low drop of intensity and the number of compounds in the effluent OM (Figure 2). Interestingly, the persistent features show a median increase of effluent-to-influent intensity ratio of separate features of 1.4 times. This effect can be caused by the increase of the effluent intensity due to a weaker ion suppression in the effluent, but also an accumulation of substances that were already abundant in the influent is possible.
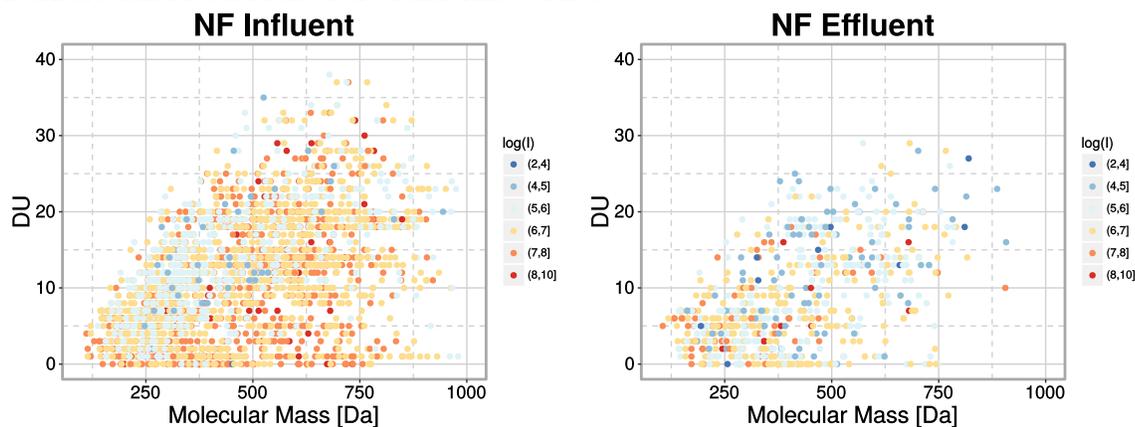

Figure 3. Mass, degree of unsaturation and ranges of logarithmic absolute intensity for 2619 and 719 features with the assigned elemental composition in the influent and effluent NF DOM, respectively.
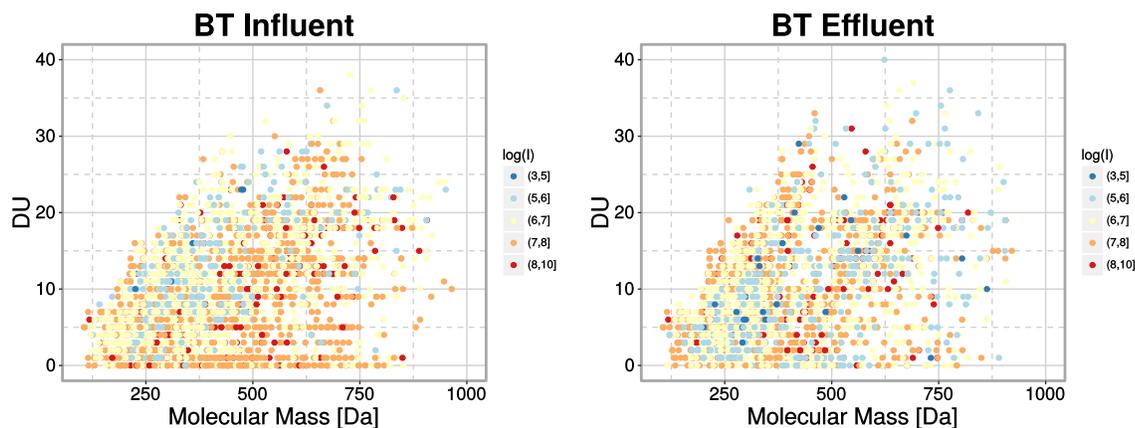

Figure 4. Mass, degree of unsaturation and ranges of logarithmic absolute intensity for 2752 and 2259 features with the assigned elemental composition in the influent and effluent BT DOM, respectively.

Chemical profile. We calculated the degree of unsaturation (DU), which estimates the amount of double bonds and rings in organic compounds from their elemental composition and correlates with the saturation of organic substances. A mean DU drop of 2.5 double bond equivalents shows that unsaturated compounds are preferentially removed compared to the saturated ones (Figure 3). This correlates well with previously researched dependency of substance rejection on the polarity in NF, specifically on the adsorption of hydrophobic substances onto the surface of the membrane[26]. A comparison of the total DOM in BT influent and effluent does not show a readable pattern and only a slight DU increase of 0.6 (Figure 4), yet our calcula-

tion for appearing vs. disappearing features shows a significant increase by 1.4 double bond equivalents, which is in line with the documented accumulation of aromaticity in BT[29]. The monitoring of DU in wastewater treatment can be applied to estimate the quality of the process, to recognize hydrophobicity altering reactions as hydrolysis or oxidation and conceptually to even support process control. The established estimation of aromaticity in wastewater DOM using $SUVA_{254}$ performs well. Our methodology, however, gives insight into the unsaturation of DOM while also expanding the analysis to the not UV active DOM constituents and allows to connect the aromaticity of individ-

ual compounds to properties such as the content of heteroatoms (N, O) or the molecular weight in a an experimentally straight forward way.

We explored the chemical reactivity in BT by monitoring gains or losses of simple structural addition/elimination and de/conjugation moieties between the total influent features and those appearing in the effluent using previously explored metabolic logic[30,31] (Table 2). Herein, we applied the principle to DOM fractions, which are the most relevant for the monitoring of TPs. BT showed a high level of chemical reactivity with 1303 potential transformation pairs for the monitored moieties. Presumably losses of CH_2, C_2H_4 and C_2H_4O are connected with the degradation of oligomeric surfactants, as was shown previously for CH_2[30]. The monitoring of reaction pairs can help us to estimate the chemical activity of a treatment system. E.g., the elimination/addition on a double bond can assess the mechanisms of the shifts of unsaturation.

Table 2. Number of features in the influent corresponding to gains/losses of molecular moieties in the appeared features in the effluent of BT.

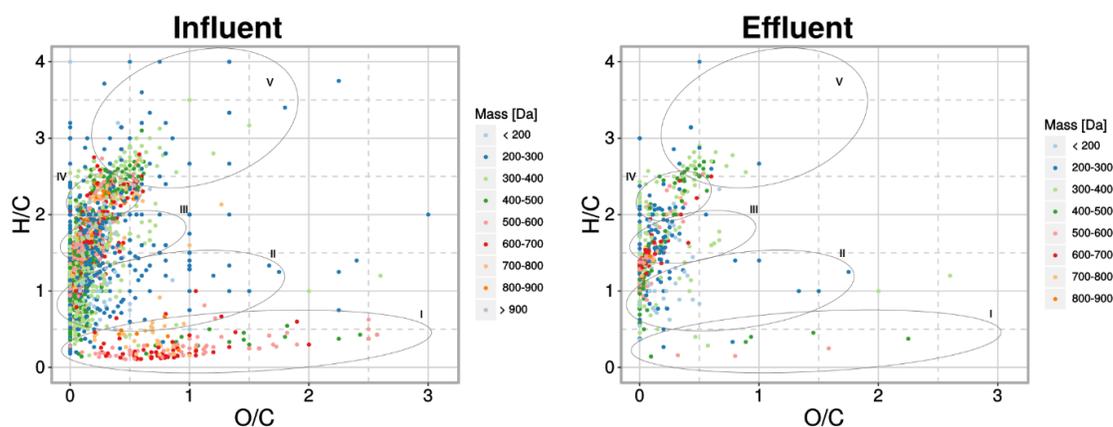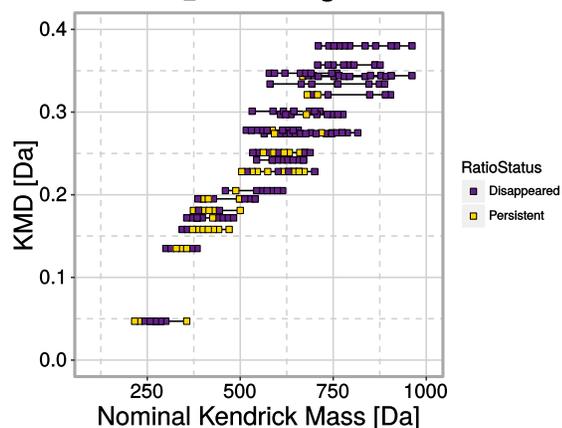| Reaction	Moiety	Reaction Pairs
Demethylation	$- CH_2$	192
Deethylation	$- C_2H_4$	161
Deethoxylation	$- C_2H_4O$	127
Hydroxylation	$+ O$	146
Alcohol Oxidation	$+ O/- 2H$	131
Hydrogenation	$+ H_2$	130
Dehydrogenation	$- H_2$	125
Dehydration	$- H_2O$	108
Acetylation	$+ C_2H_2O$	98
Deacetylation	$- C_2H_2O$	80
Glucuronidation	$+ C_6H_8O_6$	3
Deglucuronidation	$- C_6H_8O_6$	2

Figure 5. A van Krevelen plot of the influent and effluent NF DOM features with assigned elemental formulas. The colour represents mass ranges.

We explored the characteristic regions of heterogeneous wastewater DOM in the van Krevelen plot of NF treatment (Figure 5). The aromatic region **I** was confirmed by a low H/C values and contains many heavy substances. The region **II** of mixed oxygen-poor aromatics and aliphatics contains the majority of the detected DOM. The oxygen-rich aliphatics **III** are substances with a small mass and a moderate amount of N compared to other groups (SI, Figure S-6). The aliphatic region **IV** with a very high H/C and a moderate oxygen content contains mainly substances with a high mass and a low N content. The highly aliphatic region **V** with a variable O content contains rather small masses and a high amount of N. The van Krevelen plot of BT exhibits a similar pattern as in NF, even the distribution of masses in the plot (SI, Figure S-7). The similarity despite the different origin of the wastewater suggests that areas in the plot correspond to general matter types of DOM. Additionally, NF influent underwent a secondary BT, accounting for similarities between BT effluent and NF.

However, even despite the overall resemblance in BT and NF a comparison of the region **I** of the van Krevelen plot showed that the amount of constituents in the region drops by 5 % in NF and rises by 3 % of total features in BT after the treatment, which provides us with a fingerprint of the treatment for this particular region and correlates well with the DU findings above (SI, Figure S-8). There are established analytical methods to extract elemental information, e.g. elemental analysis or the estimation of phenols or surfactants in DOM. However, our methodology offers the extraction of an extended elemental information for hundreds of individual DOM constituents in an experimentally straight forward way. More importantly, even without a precise identification of clusters the plot helps to estimate the chemical changes in the system e.g. oxidation of OM or transformation of polar constituents into non-polar ones, which can be observed as the clusters shift or change the number of constituents throughout the treatment.

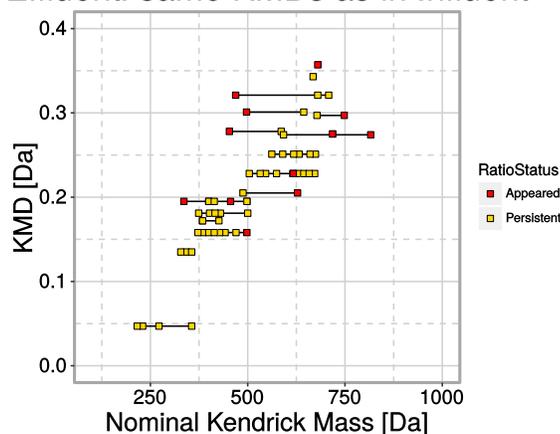Figure 6. 23 -*CH₂*- homolog series in the Influent and their development in the Effluent of NF.

Kendrick mass defect. We used the Kendrick Mass Defect (KMD) to explore homolog series, which occur in NOM[12] but more importantly in classes of synthetic substances like surfactants, polyfluorinated compounds or chlorine substitute series[24]. We found numerous CH_2-(Figure 6) and C_2H_4O-series in NF and BT treatments, which often correspond to surfactants in municipal wastewater treatment. The depicted homologs for the CH_2-moiety show different developments in NF and BT. While NF largely removes the substances, prioritising heavy constituents, in the BT we see a less complete removal of features, but new homologs appear at lower masses (SI, Figure S-9), which corresponds well to the formation of TPs in BT. This way our methodology can pinpoint not only whether substances are removed, but also suggest a mechanism of transformation – in case of BT presumably a sequential cleavage of CH_2-units as also suggested by the numerous CH_2 moiety losses observed in the effluent (Table 2).

Benefits and challenges. We attempted to record spectra of the untreated samples, yet due to a low sensitivity caused by a severe matrix suppression we decided to pre-concentrate the samples. Our pre-concentration procedure so far restricts the extractable DOM to 50–60 % of the total content. LC-MS additionally reduces that estimate, since it mostly detects moderately polar compounds. Yet, the fractions of the highest interest, e.g. the synthetic organics, are retained by the procedure and the high number of identified constituents ensures a good statistical representation of the sample. The focus of the method is to extract possibly many features without introducing false positives. Although false positives certainly cannot be avoided in big data sets[11], we can reduce their occurrence considering good settings, e.g. for the molecular formula prediction (SI, Table S-5). Our balanced set of parameters neglects minor signals, but prevents the introduction of noise. The ion suppression becomes increasingly more problematic in complex LC-MS samples, which leads to non-linear intensity deviations among samples. We can estimate the removal of DOM with LC-MS data, even taking into account the high (70 %) ion suppression effect that arises in the influent of NF. In BT where both the influent

and effluent samples show a high (80 %) suppression compared to the blank we can still compare influent to the effluent with the mean influent-to-effluent ratio of IS being 0.84, however showing broad variation visualizing the nonlinear nature of ion suppression. The previously researched data extraction procedure for a non-targeted analysis of wastewater treatments and the corresponding challenges shows that data sets with a high amount of true positives can be extracted under the prerequisite of good settings, however the analysis can be challenged by the matrix suppression in complex samples[11]. We expand the toolbox of a non-targeted analysis by concentrating on the development of a set of analytical tools to characterise wastewater DOM. The issue will be addressed in the future by standardization using a wastewater DOM specific IS mixture.

Once a feature is detected the prediction of its elemental composition is challenging, especially for the high molecular weight compounds. This was improved by using heuristic rules and isotopic information. We tested the prediction using experimental HRMS data and produced a valid set of parameters. The prediction of P- and F-containing formulas remains difficult. Despite the complexity of its prediction the elemental information is necessary for a non-targeted LM-CS analysis, because it uncovers the chemical properties of DOM, e.g. substance classes.

The transition of the van Krevelen plot into wastewater research is challenging. Wastewater DOM is a heterogeneous mixture compared to the DOM of natural waters, which consists of numerous homologs and structurally similar matter[18,22]. The substances in wastewater come from many sources and often contain synthetic compounds, which are inconsistent with the structure of natural organics. Similarly, Kendrick statistics can be impacted by the conservative settings, which remove low-intensity homologs. KMD statistics require the knowledge of the addition/elimination to prioritize the search, but are difficult to apply for unknown unknowns. Yet, it is a good bridge between the non-targeted and suspect screenings and can be used to rapidly screen wastewater for priority sub-

stances. Also, using a homolog series we can assign elemental formulas to high masses by extrapolation from low masses in the series, thus avoiding the statistical error of the formula prediction for masses above 400 Da.

CONCLUSION We successfully tested and applied non-targeted LC-MS data extraction and analysis to fingerprint two wastewater treatment systems, namely a NF and a secondary BT. Significant changes of DOM were identified after both treatments. In NF a removal of 84 % of detected substances and a preferential rejection of large and unsaturated molecules were discovered. The analysis of the chemically more complex BT gave insight into the chemical transformation of DOM during the treatment. We uncovered mass shifts in the fractions of BT DOM and the addition or elimination of molecular moieties (e.g. $+O$, $-H_2$) in course of the treatment. The methodology test showed that this non-targeted analysis can be well applied to find critical changes in the system and connect these to the individual detected compounds. Van Krevelen plots grouped the constituents into clusters depending on their assigned elemental composition, the quality of which was tested using the linear dependency of the RMD on the percentage of hydrogen in a molecule. Kendrick-type mass defects uncovered homolog series for addition/elimination moieties and traced the differences in the removal between NF and BT. The presented not-targeted analysis holds the potential to rapidly create a chemical picture of DOM in wastewater – a fingerprint of a treatment technology. However, the analysis is still limited by the extraction procedure, scope of LC-MS detection, the signal suppression in the samples and the precision of settings in the data extraction – challenges to be addressed in the future.

Supporting Information. The Supporting Information material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Tel.: (+34) 972183380. Fax: (+34) 972183248. E-mail: mpetrovic@icra.cat.
Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENT

REFERENCES

(1) Michael-Kordatou, I.; Michael, C.; Duan, X.; He, X.; Dionysiou, D. D.; Mills, M. A.; Fatta-Kassinos, D. *Water Res.* **2015**, *77*, 213–248.

(2) Daughton, C. G. *Environ. Impact Assess. Rev.* **2004**, *24* (7–8), 711–732.

(3) Deblonde, T.; Cossu-Leguille, C.; Hartemann, P. *Int. J. Hyg. Environ. Health* **2011**, *214* (6), 442–448.

(4) Clesceri, L. S.; Eaton, A. D.; Greenberg, A. E. *Standard Methods for the Examination of Water and Wastewater*; American Public Health Association, 1998.

(5) Huber, S. A.; Balz, A.; Abert, M.; Pronk, W. *Water Res.* **2011**, *45* (2), 879–885.

(6) Chen, W.; Westerhoff, P.; Leenheer, J. A.; Booksh, K. *Environ. Sci. Technol.* **2003**, *37* (24), 5701–5710.

(7) Prasse, C.; Stalter, D.; Schulte-Oehlmann, U.; Oehlmann, J.; Ternes, T. A. *Water Res.* **2015**, *87*, 237–270.

(8) Richardson, S. D.; Ternes, T. A. *Anal. Chem.* **2014**, *86* (6), 2813–2848.

(9) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098.

(10) Wode, F.; van Baar, P.; Dünnbier, U.; Hecht, F.; Taute, T.; Jekel, M.; Reemtsma, T. *Water Res.* **2015**, *69*, 274–283.

(11) Nürenberg, G.; Schulz, M.; Kunkel, U.; Ternes, T. A. *J. Chromatogr. A* **2015**, *1426*, 77–90.

(12) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. *Anal. Chem.* **2001**, *73* (19), 4676–4681.

(13) Sleighter, R. L.; Hatcher, P. G. *Mar. Chem.* **2008**, *110* (3–4), 140–152.

(14) Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. *Limnol. Oceanogr. Methods* **2008**, *6* (6), 230–235.

(15) Stravs, M. A.; Schymanski, E. L.; Singer, H. P.; Hollender, J. *J. Mass Spectrom.* **2013**, *48* (1), 89–99.

(16) Schymanski, E. L.; Singer, H. P.; Longrée, P.; Loos, M.; Ruff, M.; Stravs, M. A.; Ripollés Vidal, C.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48* (3), 1811–1818.

(17) Kim, S.; Kramer, R. W.; Hatcher, P. G. *Anal. Chem.* **2003**, *75* (20), 5336–5344.

(18) Minor, E. C.; Steinbring, C. J.; Longnecker, K.; Kujawinski, E. B. *Org. Geochem.* **2012**, *43*, 1–11.

(19) Lu, M.; Zhang, Z.; Qiao, W.; Wei, X.; Guan, Y.; Ma, Q.; Guan, Y. *Bioresour. Technol.* **2010**, *101* (7), 2106–2113.

(20) Zhang, H.; Zhang, Y.; Shi, Q.; Hu, J.; Chu, M.; Yu, J.; Yang, M. *Environ. Sci. Technol.* **2012**, *46* (8), 4396–4402.

(21) Cortés-Francisco, N.; Caixach, J. *Environ. Sci. Technol.* **2013**, *47* (17), 9619–9627.

(22) Herzsprung, P.; von Tümpling, W.; Hertkorn, N.; Harir, M.; Büttner, O.; Bravidor, J.; Friese, K.; Schmitt-Kopplin, P. *Environ. Sci. Technol.* **2012**, *46* (10), 5511–5518.

(23) Zhang, H.; Zhang, Y.; Shi, Q.; Ren, S.; Yu, J.; Ji, F.; Luo, W.; Yang, M. *Water Res.* **2012**, *46* (16), 5197–5204.

(24) Jobst, K. J.; Shen, L.; Reiner, E. J.; Taguchi, V. Y.; Helm, P. A.; McCrindle, R.; Backus, S. *Anal. Bioanal. Chem.* **2013**, *405* (10), 3289–3297.

(25) Sleno, L. *J. Mass Spectrom.* **2012**, *47* (2), 226–236.

(26) Bellona, C.; Drewes, J. E.; Xu, P.; Amy, G. *Water Res.* **2004**, *38* (12), 2795–2809.
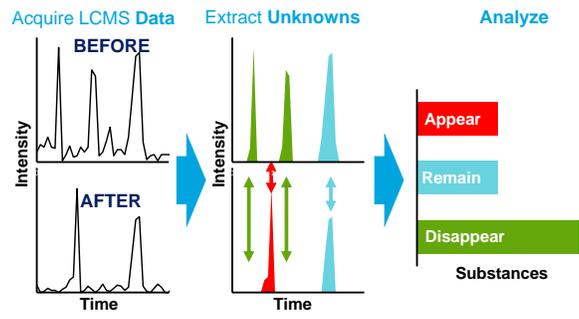
(27) Jarusutthirak, C.; Amy, G. *Environ. Sci. Technol.* **2006**, *40* (3), 969–974.

(28) Evgenidou, E. N.; Konstantinou, I. K.; Lambropoulou, D. A. *Sci. Total Environ.* **2015**, *505*, 905–926.

(29) Park, M.-H.; Lee, T.-H.; Lee, B.-M.; Hur, J.; Park, D.-H. *Sensors* **2009**, *10* (1), 254–265.

(30) Schollée, J. E.; Schymanski, E. L.; Avak, S. E.; Loos, M.; Hollender, J. *Anal. Chem.* **2015**, *87* (24), 12121–12129.

(31) Gallert, C.; Winter, J. In *Environmental Biotechnology*; Wiley-VCH Verlag GmbH & Co. KGaA, 2004; pp 1–48.

Table of Contents artwork