This is the accepted manuscript version of the contribution published as:

Herzsprung, P., Kamjunke, N., Wilske, C., Friese, K., Boehrer, B., Rinke, K., Lechtenfeld, O.J., von Tümpling, W. (2023):

Data evaluation strategy for identification of key molecular formulas in dissolved organic matter as proxies for biogeochemical reactivity based on abundance differences from ultrahigh resolution mass spectrometry *Water Res.* **232**, art. 119672

The publisher's version is available at:

http://dx.doi.org/10.1016/j.watres.2023.119672

1	Data evaluation strategy for identification of key molecular formulas in dissolved organic
2	matter as proxies for biogeochemical reactivity based on abundance differences from
3	ultrahigh resolution mass spectrometry
4	
5	Peter Herzsprung,* ^{,†} Norbert Kamjunke, [‡] Christin Wilske, [‡] Kurt Friese, [†] Bertram Boehrer, [†]
6	Karsten Rinke. [†] Oliver J. Lechtenfeld. ^{§,∥} Wolf von Tümpling [‡]
7	
8	† UFZ – Helmholtz Centre for Environmental Research, Department Lake Research,
9	Brückstraße 3a, D-39114 Magdeburg, Germany
10	* UEZ Ualmhaltz Cantra for Environmental Passarch Department Diver Ecology
10	¹ UTZ – Hemmonz Centre for Environmental Research, Department River Ecology,
11	Bruckstraße 3a, D-39114 Magdeburg, Germany
12	§ UFZ – Helmholtz Centre for Environmental Research, Department Analytical Chemistry,
13	Permoserstr. 15, D-04318 Leipzig, Germany
14	UFE Helmholtz Centre for Environmental Research, ProVIS – Centre for Chemical Microscopy,
15	Permoserstr. 15, D-04318 Leipzig, Germany
16	
17	*Corresponding author:
18	Peter Herzsprung (UFZ – Helmholtz Centre for Environmental Research, Department Lake
19	Research, Brückstraße 3a, D-39114 Magdeburg, Germany, e-mail: peter.herzsprung@ufz.de
20	
-	

- 22 Abstract
- 23

The molecular composition of dissolved organic matter (DOM) is of relevance for global 24 carbon cycling and important for drinking water processing also. The detection of variation of 25 DOM composition as function of time and space from a methodological viewpoint is essential 26 to observe DOM processing and was addressed so far. High resolution concerning DOM quality 27 was achieved with Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS). 28 However almost none of the existing FTICR-MS data sets were evaluated addressing the fate 29 30 of single mass features / molecular formulas (MFs) abundance during experiments. In contrast to former studies we analyze the function of MF abundance of time and space for such MFs 31 which are present in all samples and which were formerly claimed as recalcitrant in not all but 32 a great number of studies. For the first time the reactivity of MFs was directly compared by 33 their abundance differences using a simple equation, the relative intensity difference (δRI). 34 Search strategies to find out the maximum δRI values are introduced. The corresponding MFs 35 36 will be regarded as key MFs (KEY-MFs). In order to test this new approach data from a recent photo degradation experiment were combined with monitoring surveys conducted in two 37 38 drinking water reservoirs. The SRI values varied over one order of magnitude (more than fivefold). MFs like C₉H₁₂O₆ and C₁₀H₁₄O₆ revealed high biogeochemical reactivity as photo 39 products. Some of the KEY-MFs were identical with MFs identified as disinfection byproducts 40 precursors in recent studies. Other KEY-MFs were oxygen-rich and relatively unsaturated 41 (poly-phenol-like) and hence relevant to flocculation procedures. 42

43

44 Key words

45

- 46 Dissolved organic matter (DOM), key molecular formulas, high-resolution mass spectrometry
- 47 (HRMS), drinking water reservoir, relative intensity difference

49 **1.** Introduction

50

The molecular composition and properties of dissolved organic matter (DOM) play a critical 51 role in a number of biogeochemical processes such as metal redox cycling (Dadi et al., 2017; 52 Riedel et al., 2013), microbial transformations (Lechtenfeld et al., 2015; Ohno et al., 2014; 53 Riedel et al., 2016) and photochemical processes(Bittar et al., 2015; Gonsior et al., 2013; Holt 54 et al., 2021; Maizel et al., 2017; Riedel et al., 2016; Wilske et al., 2020). In natural systems, 55 these processes are difficult to disentangle (Kellerman et al., 2015). From the analytical point 56 57 of view, processes can be exclusively observed by concentration variations of components (for example inorganic ions, organic molecules). The variation as function of time is one possibility 58 59 for investigation of processes. A negative or positive temporal trend of a specific component abundance is defined as degradation and product formation, respectively. Also spatial 60 61 abundance variations can be interpreted as processes. Depending on the reference system, a component can be regarded as degraded / produced, if its concentration is decreased / increased 62 63 in one location of a system compared to another location of the same system. Such a system can be a reservoir and one location could be the epilimnion and the other one the hypolimnion 64 65 if the reservoir is thermally stratified. However, without any additional information, it is difficult to interpret concentration differences between epi- and hypolimnion. If a component 66 is depleted / accumulated in the epilimnion, the question is, was it consumed / produced or did 67 it keep unchanged and was as alternative interpretation produced / consumed in the 68 hypolimnion or even released / resorbed from the sediment pore-water by diffusion into the 69 hypolimnion / diffusion to the sediment. 70

Another aspect is the detection of DOM concentration differences themselves. While the 71 quantification of single components like nutrients (nitrate, ammonium), metal ions, anions or 72 trace organic compounds like pharmaceuticals or pesticides in analytical standard methods are 73 74 well established, the analysis of DOM comprises considerable difficulties due to the complexity 75 of its composition. DOM probably consists of thousands or even millions of different organic compounds and the identification of the isomeric structure of each molecule is still far from any 76 instrumental analytical realization (Hawkes et al., 2018; Herzsprung et al., 2012). The up to 77 date highest analytical resolution of DOM can be achieved by Fourier transform-ion cyclotron 78 resonance mass spectrometry (FTICR-MS) (Hertkorn et al., 2006; Reemtsma, 2009). This 79 method is principally appropriate to differentiate abundances of mass features, to which 80 molecular formulas (MFs) can be assigned. 81

Various limitations exist to determine the enormous complexity of organic molecules with 82 83 FTICR-MS. In combination with sampling and sample preparation it seems challenging to quantify each single compound in the totality of the DOM in order to investigate transformation 84 processes. By far not all of the DOM components can be detected by FTICR-MS. Because 85 desalting is required, a solid phase extraction (SPE) or dialysis or reverse osmosis must be 86 installed before analysis in the mass spectrometer. Molecules with high water solubility / 87 polarity are not retained on the cartridge whereas highly hydrophobic molecules might be 88 retained but no longer eluted with an organic solvent (Raeke et al., 2016). Regarding this 89 90 procedure, the FTICR-MS analyzed DOM is designated as SPEDOM (only the extractable DOM) (Dittmar et al., 2008; Raeke et al., 2016). Even the SPEDOM composition cannot be 91 92 analyzed completely. The mass peak intensity is a function of ionization efficiency. This efficiency will not be a function of the assigned elemental composition, but of each possible 93 94 isomer. Any data evaluation which bases on comparison of component abundances derived from FTICR-MS spectra from different samples has to suggest the same isomeric composition 95 96 of MF in each considered sample. The lower the dissolved organic carbon (DOC) concentration and DOM quality differences between compared samples the better the MF abundances will be 97 98 evaluable.

99 Being aware of these limitations, we present a new attempt on a semi-quantitative level to show100 how DOM transformation processes can be followed.

So far, DOM processes were often described by the use of bulk, e.g. optical parameters 101 102 (Bracchini et al., 2010; Howard et al., 2021; Painter et al., 2018; Zheng et al., 2020) providing not sufficient resolution on a molecular basis. More chemical resolution can be achieved by 103 NMR, which was used for investigation of DOM transformation. (Kida et al., 2021) NMR can 104 be regarded as bulk parameter as well, because many molecules having the same structural 105 subunits will contribute to a signal. Also studies using FTICR-MS results often used bulk 106 107 parameters like the intensity weighted average number of components, H/C or O/C ratios, NOSC (nominal oxidation state of carbon), and others or statistical methods like PCA (Berg et 108 109 al., 2022; Casas-Ruiz et al., 2020; Fudyma et al., 2021; Kellerman et al., 2014; Martinez-Perez et al., 2017; Osterholz et al., 2018; Wagner et al., 2015). From those studies no details about 110 the fate of single MF are available but trends which parts of the DOM (as specified by its 111 112 oxygenation, aromaticity, molecular weight...) are reactive.

Some studies, including our recent approaches, used formula by formula rank correlations with external parameters like the fluorescence intensity, bacterial activity or chlorophyll or even the radiation dose.(Herzsprung et al., 2012; Herzsprung et al., 2020; Kamjunke et al., 2019; Lavonen et al., 2015; Stubbins et al., 2014) The resulting correlation coefficients and p-values
can indicate biogeochemical processes. The dimensions (orders of magnitude) of MF
abundance differences are not explored yet.

A more successful strategy addressing the fate of single MF are experiments, where the start 119 solution is compared with the reacted solution. The MFs are categorized into a) "removed" 120 (present in the start solution but not in the end solution), b) "produced" (present in the end 121 solution but not in the start solution) and c) "resistant" (present in both solutions). (Yuan et al., 122 123 2017) (Tang et al., 2022) This experiment design cannot distinguish between large abundance 124 MFs which were no more found in the second sample and low abundance MFs (just above the 125 signal to noise ratio in the one sample and just below the signal to noise ratio in the other 126 sample). It may happen that a large start MF abundance corresponds to a just above the S/N ratio in the second sample, but regarded finally as resistant. In addition, many of the process 127 128 oriented FTICR-MS studies do not report about the fate of single MF although the data are principally available, but not for the readership. In many studies there is a lack of providing raw 129 130 data, MFs and their intensities.

This is a research gap in combination with the incomplete evaluated "resistant group": We 131 hypothesize therefore that the resistant MF group (present both in a start and end sample or 132 more generally in all samples of an experiment as function of time or space) can be resolved 133 for products and degraded MFs with pregnant abundance differences by introduction of new 134 and appropriate evaluation techniques. Within this common presence pool we expect that a 135 fraction remains with reliable resistance. It is clear that FTICR-MS studies were performed 136 where this common presence pool was included and DOM transformations were evaluated by 137 statistical means as mentioned above. The abundance differences, which will be specified in an 138 equation in the methods section as δRI , will be regarded as proxy for the reactivity of a MF. 139 We hypothesize that MFs with extreme δRI exist, which we name key molecular formulas 140 (KEY-MFs). The KEY-MFs may be of interest for the DOM biogeochemistry community while 141 the microbiology community already uses key players for investigation which are the bacteria 142 143 mainly responsible for certain metabolism. (Tanuwidjaja et al., 2021)

Our hypothesis will be addressed by design of search strategies in FTICR-MS example data sets. The used examples originate from photochemical experiment (time variation) and drinking water reservoir monitoring (space and time variation). (Herzsprung et al., 2020; Wilske et al., 2021; Wilske et al., 2020) Drinking water processing is an important environmental and health issue and DOM components are suggested to influence flocculation costs and to act as disinfection byproduct precursors. (Phungsai et al., 2018; Lavonen et al., 2013) Such issues

- justify the analytical expense (FTICR-MS) and advanced evaluation effort (search for KEY-MFs). While it is not intended to provide a complete list with all elemental compositions and their quantified influence on drinking water processing in the present study, we want to demonstrate the possibility to make DOM transformations more comprehensible. Our intension is to provide new methodological evaluation tools for understanding DOM turnover processes from available FTICR-MS elemental formula data sets. With the new generated knowledge already existing elemental formula data sets might be reanalyzed.
- 157 Our new judge- and assessment methodology is structured as follows:
- 158 I) Calculation of relative intensity differences (δ RIs) (2.1) from experiments where 159 abundance differences are expected between start and end sample (function of time) or 160 surface and bottom (function of space).
- 161 II) Simple search of KEY-MFs use of data pairs (start / end) or (surface / bottom) (2.2)
- 162 III) Advanced search of KEY-MFs use of data with more time / seasonal / space resolution
 163 (2.3)
- IV) Use of δRI ranking for prove of multivariate similarities between the samples by
 hierarchical cluster analysis (HCA) (2.4)
- V) Deeper statistical prove. A statistically control was performed with description in SI 3 and
 execution in "1-s2.0-Sxxx-mmc3.xlsx" using the Cox Stuart (Cox and Stuart, 1955) and
 the Wilcoxon (Wilcoxon, 1945) test.
- 169 All steps are explained using following data sets:
- 170 Photo experiment (PHOT) published in (Herzsprung et al., 2020)
- 171 Reservoir Muldenberg monitoring (MULD) (Wilske et al., 2021)
- 172 Reservoir Rappbode monitoring (RAPP) (Herzsprung et al., 2020)

174 **2.** Methods

175 2.1 Calculation of relative (percentage) intensity differences δRIs (I)

Samples have to be selected, where a process relevant abundance difference is expected. This 176 can be a kinetical experiment, where the end DOM quality is expected to be different from the 177 start (time series). Another possibility are monitoring experiments, where different MF 178 abundances can be expected at different locations (variation with space). A lake or reservoir 179 can be considered, where during stratification different DOM quality between surface and 180 bottom can be suggested. Also different depths in sediment pore water might be compared, 181 where different DOM quality might be induced by desorption / adsorption or microbial 182 processes. 183

184 Thus, for calculation of δ RIs each two samples (a sample pair) is required, which can be named 185 sample X and sample Y.

186 The general Equation would be:

187 $\delta RI = \{ [intensity (sample X) - intensity (sample Y)] / [intensity (sample Y)] \} \cdot 100\%$

188 The used RIs base on the normalization on the total intensity. (Flerus et al., 2012)

In our example, the δR is as key values are calculated by sample pairs of end versus start 189 190 intensity in a photo experiment (PHOT) and surface versus bottom intensity in the reservoir 191 monitoring. The calculation principle is visualized in Fig. 1 using examples of sample pairs from PHOT (end/start) and monitoring in reservoirs, sampling month = July, (surface/bottom) 192 from MULD and RAPP. Background information about the three selected example data sets is 193 available in SI 1.1. These three data sets were merged to a joined data set (finally named as total 194 data set) where only the shared components (total common presence) in all samples were 195 considered. 196

As limitation one has to suggest that each component has the same isomeric composition in all samples. This need of assumption is not only valid for our study but for all studies which used direct infusion FTICR-MS (without chromatographic separation and without mass fragmentation). A further limitation of the equation (calculating relative differences) is the fact that the information is lost whether these were major abundance MFs (high importance for bulk DOM) or low abundance MFs (little importance for bulk DOM).



Fig. 1 Visualization of the calculation of δRI. A, B: RI as function of time (PHOT, start,
end); RI as function of space in MULD (C, D) and in RAPP (E, F), lake surface and bottom in
July

208

209 2.2 Simple search of KEY-MFs (II)

210 The simple search bases on the idea to calculate the ranking of the δRI values for each

- experiment and combining them to a rank sum for each MF. By suggestion that the different
- experiments have to be combined for the KEY-MF search, the KEY-MFs are simply the
- components with extreme positive or extreme negative rank sum values.
- In our case the data subset was used as described in SI 1.2 (sample X means end sample of
- 215 PHOT or surface sample from MULD or RAPP; Y means start sample from PHOT or bottom
- from MULD or RAPP, see Fig. 1). For each data set (PHOT, MULD, RAPP) the ranks of the
- δ RI are calculated for all 1111 MFs. Finally the rank sum for each MF is calculated as shown
- in "1-s2.0-Sxxx-mmc2.xlsx", sheet "simple_search", column M. As shown in the

- 219 "simple_search" sheet we selected the first ten with the lowest rank sum (Screenshot 6, SI
- 220 2.3.1) and the last ten with the highest rank sum (using autofilter in excel represents one
- simple possibility). The lowest rank sum KEY-MF are defined to have positive reactivity
- 222 (products), the highest rank sum KEY-MF are defined to have negative reactivity (degraded).
- 223 It is clear that the number ten (of KEY-MFs) was selected randomly (just to show example
- results) and one has to reflect about the weighting of the experiments which contribute to the
- rank sum.
- This simple search method can be extended by including more specific steps depending on the special data sets and the scientific question.
- 228

229 2.3 Advanced search of KEY-MFs (III)

230 If more data as function of time and / or space are available, the search of KEY-MFs can be

elaborated. While the simple search strategy addresses experiments which generated data

pairs (start / end, surface / bottom for example) our advanced search is targeted to more valid

233 interpretation of temporal and spatial biogeochemical processing.

234

235 **2.3.1 Experiments with higher time resolution**

If more than two data points in a kinetical experiment are available, the question is which data
can be used as data pair X and Y. Here we introduce one solution (of potential more
possibilities) using median values.

- As an example we use data from PHOT (SI 1.2). The reaction course is divided into three
- 240 phases: Starting (four measurements), middle (five measurements), end (four measurements).
- Sample X is defined as the median value of the end phase and sample Y as the median value
- of the start phase (SI 2.1, equation 1, Fig. S1). It is clear that median values are more robust
- 243 from a statistical and experimental point of view.

244

245 2.3.2 Monitoring with space and seasonal resolution

In the reservoir samples which we use here as example several sample pairs X / Y (surface /

bottom) are available, particularly at different seasons. The δRI values are calculated for all

selected seasons in separate using equation 2 (SI 2.1), both for MULD and RAPP. For the

advanced KEY-MF search, only months with lake stratification were used, because only at

- such times differences between X and Y can be expected. The temperature data, which are the
- basis for suggestion of stratification, are reported in the recent studies together with the data
- sets MULD and RAPP. For MULD three months (June, July, August) and for RAPP four
- 253 months (June, July, August, October) were combined by calculation of the median δRI values
- using equation 3 (SI 2.1).
- 255

256 2.3.3 Combination of 2.3.1 and 2.3.2 for ranking calculation

257 The ranking of the median δRI values as a normalization offers the possibility to combine

both data sets. In our case, it is calculated in "1-s2.0-Sxxx-mmc2.xlsx", sheet

259 "advanced_search", column O, P, Q, with the rank sum in column M. Before the smallest and

260 highest rank sum values are searched for KEY-MFs additional criteria are recommended to

improve the robustness with regard to biogeochemical process interpretation.

262

263 2.3.4 Advanced criteria, search of MFs with identical sign for δRI

For different sampling dates during the total stratification period in the reservoir monitoring one can assign biogeochemical reactivity only for components with identical sign for δ RI on each of the considered sampling dates (for example in MULD for all three months, June, July, August with lake stratification). Regarding PHOT, we expect for photo products that all experiment end time points (T₉ – T₁₂) show higher intensity compared to all start time points (T₀ – T₃) and for photo degraded components the opposite.

The principle of the following search equations is the question (programmed in excel) if a groupof values is (for reservoirs pairwise) larger or smaller compared to the opposite group. The

272 details are shown in SI 2.2 (eqs. 4a, b; 5a, b; 6a, b).

273 The result of such "equation question" as excel equation (reservoir as an example)

274 $(=IF(AND(RI_{surface}Jun>RI_{bottom}Jun;RI_{surface}Jun>RI_{bottom}Jun;RI_{surface}Aug>RI_{bottom}Aug);1;0))$

is in principle "true" or "false" (1 or 0). Using autofilter, the three data sets (MULD, RAPP,

276 PHOT) can be combined by the same sign of reactivity, if for all three the same result (true

277 resp. false) was found. The implementation of all six equations is documented in the

²⁷⁸ "advanced_criteria" sheet of "1-s2.0-Sxxx-mmc2.xlsx" with column addresses shown in SI 2.2.

If eqs. 4a, 5a, 6a have the result "1", the component has positive reactivity (product). For eqs.
4b, 5b, 6b the result "1" means negative reactivity (degradation). The result "0" for any of these
eqs. means no reactivity.

The possible combination of results of such equation questions can be translated in DOM reactivity of the MFs using reactivity codes as follows: 101010 = photo product, 6 characters, equation sequence 4a|4b|5a|5b|6a|6b| as explained in SI 2.2 (Screenshot 1 – 5). 010101 = photo degraded; 000000 = non reactive; 001010 = microbial product (no photo reactivity but positive reactivity in both reservoirs); 000101 = microbial degraded (no photo reactivity but negative reactivity in both reservoirs); all remaining codes like 000100 mean that no clear reactivity can be assigned to the corresponding MFs.

The assigned biogeochemical reactivities to all considered 1111 MFs are balanced (number of MFs) in Table 1 and shown individually for all MFs in the "advanced_criteria" column M (reactivity codes shown in column N).

292

293 2.3.5 Combination of median δ RI values (2.3.3) with advanced criteria (2.3.4)

By this process it can be verified that the via rank sum searched KEY-MF (which address certain biogeochemical processing) have the correct reactivity code. The reactivity code is appropriate to select all MFs with the same defined biogeochemical reactivity.

Practically it means that the reactivity codes and corresponding biogeochemical reactivities
were copied from the "advanced_criteria" sheet to the "advanced_search" sheet in "1-s2.0Sxxx-mmc2.xlsx". The advanced KEY-MF search continued with following steps:

300 Search of photo product KEY-MF: reactivity code is set to 101010 (column J, using

autofilter). Then the ten lowest values in column M (rank sum) are searched and the cells in

column L labelled with "photo product" (Screenshot 7, SI 2.3.2). The found KEY-MFs are

shown in Screenshot 8.

Search of photo degraded KEY-MF: reactivity code is set to 010101. Then the ten highest
rank sum values are searched and the cells labelled with "photo degraded".

Search of microbial products: reactivity code is set to 001010. An additional rank sum value

307 column is available (column N). Only the ranks of both reservoirs were added MF-wise, but

the PHOT ranks were not considered. Then the ten lowest rank sum values are searched and

309 the cells labelled with "microbial product".

These MF- data can be extracted and their PHOT median δ RI values can be plotted versus the reservoir median δ RI values. In such plot the KEY-MFs can be highlighted and the order of magnitude of δ RI differences between KEY-MFs and residual MFs can be visualized as

shown in Fig. 3.

314

315 2.4 Prove of multivariate similarities using HCA (IV)

As shown for the simple KEY-MF search (2.2) and the advanced search (2.3) the δ RI values can be used for calculation of their ranking for each of the 1111 MFs. A HCA based on the ranked dataset distance is an option to find out similarities and dissimilarities between the sample pairs. Here the PHOT (median end / start) sample pair is compared with reservoir sample pairs (surface / bottom) for months with and without stratification, for all 1111 MFs. The ranking is calculated in the 1-s2.0-Sxxx-mmc2.xlsx, sheet "ranking_for_HCA", column M (PHOT), columns N – T (MULD), columns U – AA (RAPP).

The ranks were used for calculation of a HCA using Ward's minimum variance method with squared Euclidean distance for classification of all single sampling dates (each seven for the reservoirs and one for PHOT) (Ward, 1963). The statistical assessments were executed in StatisticaTM Version 13.3.

The received dendrogram (Fig. 2) clearly separates sampling dates during stratification in the reservoirs. This implies that the subdivision of the waterbody by temperature gradient as the main driver for stratification is reflected in the DOM quality analysis. PHOT data clustered together with the reservoir sampling campaigns during stratification. This is one more indicator for the influence of the thermocline.



334

Fig. 2 Dendrogram HCA showing the distributions of δ RI. for photo experiment (PHOT) and the two reservoirs Rappbode (RAPP) and Muldenberg (MULD) in different months; red: stratification; blue: no stratification

338

339340 3. Results and Discussion

The focus is on the methodological outcome of the presented data evaluation. The selected data sets / experiments / studies are used as examples. The δ RIs are used for discussion how MFs illustrate biogeochemical processes (3.1). Maximum δ RIs represent KEY-MFs. The corresponding RI versus time or space plots (3.2) are appropriate examples to confirm suggested processes. In 3.3 RI versus time or space plots are presented as examples to challenge the limitations of process interpretation. Drinking water reservoirs are one example (3.4) where the KEY-MF search can be of environmental benefit.

348

349

350 **3.1 Evaluation of the calculated δRIs**

PHOT clustered (dendrogram in Fig. 2) together with the reservoir sampling campaigns during stratification. Within this cluster, the MULD (Jun – Aug) samples clustered separately from the RAPP (Jun – Oct) samples. This is not surprising since biogeochemical cycling might be different in reservoirs and depends on climatic conditions. The October sample from RAPP clustered together with the other sampling dates during stratification. At this time point RAPP was stratified in contrast to MULD in October. The MULD October sample clustered together with time points from both reservoirs during a period without stratification. Without
stratification the samples are clustered relatively mixed (MULD Feb; RAPP Dec; MULD Dec;
MULD Oct) and (MULD Mar; RAPP Jan; RAPP Mar). This suggests that the evaluated δRIs
behave randomly indicating that there is no potential to effectively evaluate biogeochemical
processes (they are not visible).

362

The next step is to find out which of the 1111 components were mainly responsible for the observed differences in ranking between late spring / summer months (stratification) and late autumn / winter months (no stratification). One possibility to search for significant behavior of components was the search for components with positive or negative δ RIs in all months with stratification (2.3.4; SI 2.2, eqs. 5a,b; 6a,b). A balance of this analysis is shown in Table 1. The

- 368 reactivities are defined and explained in 2.3.4.
- 369
- 370

Table 1 Balances of reactivities of all 1111 components under consideration

Identification of	Reactivity	PHOT	MULD	RAPP	Number of
reactivity	code				MFs
Non reactive	000000	0	0	0	102
not assigned	000001	0	0	negative	54
not assigned	000010	0	0	positive	29
not assigned	000100	0	negative	0	70
Microbially degraded	000101	0	negative	negative	78
not assigned	000110	0	negative	positive	9
not assigned	001000	0	positive	0	52
not assigned	001001	0	positive	negative	14
Microbial product	001010	0	positive	positive	71
not assigned	010000	negative	0	0	85
not assigned	010001	negative	0	negative	84
not assigned	010010	negative	0	positive	1
not assigned	010100	negative	negative	0	47
Photo degraded	010101	negative	negative	negative	155
not assigned	011000	negative	positive	0	63
not assigned	011001	negative	positive	negative	11
not assigned	011010	negative	positive	positive	12
not assigned	100000	positive	0	0	9
not assigned	100010	positive	0	positive	38
not assigned	100100	positive	negative	0	1
not assigned	100110	positive	negative	positive	2
not assigned	101000	positive	positive	0	8
Photo product	101010	positive	positive	positive	116

373 Reactivity codes are shown in 1-s2.0-Sxxx-mmc2.xlsx, sheet "advanced_criteria", column N

As largest group 155 MFs were found with negative reactivity in all summer months of both reservoirs and in PHOT. The second largest group contains 116 MFs identified as photo products (positive reactivity in all three data sets). The non reactive group contains 102 MFs.

377 71 components were identified as microbial products and 78 as microbially degraded. The 378 reactivity of all other components remains unclear and will not be further considered in this article. The found reactivities can be compared with the results from our recent study (columns 379 H - L in the "advanced_criteria" sheet of 1-s2.0-Sxxx-mmc2.xlsx) which used rank correlation 380 with radiation dose (PHOT) and Chla (RAPP Reservoir monitoring). A balance is shown in 381 382 Table S4. 91 % (106 of 116) of the photo products were confirmed (new results in column M) by the former study, 61 % of the photo degraded components, 52 % of the microbial products, 383 384 47 % of the microbially degraded components and 41 % of the non-reactive components. This demonstrates that principally reasonable reactivities can be found and assigned but many 385 386 compounds remain where no clear reactivity can be derived. For some components different reactivities were found. As an example, 6 components were found to be photo products in the 387 388 present study but in the cited study (Herzsprung et al., 2020) they were calculated as microbial products (Chla⁺Rad⁰). This is not surprising because we expect that the reactivity assignment 389 is not completely independent from the applied search or calculation algorithms. 390

391

392 **3.2 Search of KEY-MFs**

It is not intended to allocate reactivities to all components considered in the present study. 393 394 Instead we want to show how to find the components showing the largest reactivity (highest δ RI values.) in all selected (three) experiments. A simple way (simple search) is the use of the 395 calculated ranking as shown in 1-s2.0-Sxxx-mmc2.xlsx, sheet "simple_search". Here the 396 evidently most reactive MFs are highlighted after combination of three experiments but without 397 any further suggestions. In the advanced search (2.3) additional selection criteria were 398 introduced, the expectation that KEY-MFs show the same reactivity in all three compared data 399 400 sets (except for the suggested microbial transformations, where no photo reactivity was found). 401 The sum of ranks for KEY_MFs for both the simple (2.2) and the advanced (2.3) search shown 402 in Table S2. The simple search did not show exactly the same KEY-MFs like the advanced search. In the following we rely on the KEY-MFs found in the advanced search. Ten 403 404 components with the lowest sum value of ranks (products, C₉H₁₂O₆; C₉H₁₂O₅; C₁₀H₁₆O₆; $C_{10}H_{16}O_7$; $C_{11}H_{14}O_7$; $C_{11}H_{18}O_7$; $C_{10}H_{14}O_6$; $C_{10}H_{14}O_5$; $C_{11}H_{16}O_5$; $C_9H_{12}O_4$) and ten with the 405 406 highest sum value (degraded, C₂₀H₁₆O₁₄; C₁₉H₁₄O₁₃; C₁₇H₁₄O₁₃; C₁₈H₁₄O₁₃; C₁₇H₁₂O₁₂;

- 407 $C_{18}H_{12}O_{12}$; $C_{20}H_{16}O_{13}$; $C_{18}H_{16}O_{14}$; $C_{19}H_{16}O_{14}$; $C_{21}H_{16}O_{13}$) were listed as searched in 408 "advanced_search" sheet, using autofilter for column J (reactivity code) and M (rank sum).
- 409 The ten components with the lowest rank sum, all photo products, are relative aliphatic (H/C >
- 410 1.3) and their number of carbon atoms is low (9 11). In contrast, the found photo degraded 411 components with the highest rank sum are relative aromatic (H/C < 1) and oxygen-rich (O/C >
- 412 0.6) with higher number of carbon atoms (17-23).
- 413 While the ranking of δR is shown in Table S2 is a robust search which component is more or less reactive, it does not provide information about the quantitative relations, e.g. if a component 414 has 20% or 100% more reactivity. In order to give a first impression of the dimensions of the 415 relations of reactivities between components, PHOT δ RIs (equation 1, SI 2.1) are plotted versus 416 the reservoir median δRIs (equation 3, SI 2.1) and shown in Fig. 3. The photo product KEY-417 418 MFs found in Table S2 with potential high reactivity are shown by their formulas. Key photo products are highlighted in Fig. 3, key photo degraded MFs in Fig. S2. Evidently photo products 419 exist with more than five-fold δRI compared to others from this group (Fig. 3, for example 420 $C_9H_{12}O_6$). All results depend on the definition of the equations (1 - 3) and selection of samples. 421 The intention is not to present an optimized calculation but moreover a reasonable idea (of 422 423 many possible ones) to compare components reactivities in a simple (semi-quantitative) 424 manner. The search strategy using δRI is in the focus.
- The absolute δ RI values of degraded components are lower compared to those values of the photo products (Fig. S2 versus Fig. 3), partly caused by the calculation algorithm. The 20 listed KEY-MFs (Table S2) show in all data sets comparably high absolute δ RI values within the respective reactivity groups.
- Accordingly microbial products can be searched by the lowest rank sum, if the value for PHOT
 is excluded from the rank sum calculation (Fig. S3 and Table S3) and the reactivity code is set
 as 001010 in the "advanced_search" sheet.
- 432 KEY-MFs can be further checked if the RIs can be combined to meaningful profiles using all 433 available data from the joint data set. RIs are plotted versus radiation dose (PHOT) and versus 434 season and depth (MULD, RAPP). For this purpose a procedure (programmed in Excel, 1-s2.0-435 Sxxx-mmc2.xlsx) for displaying DOM component profiles is introduced in SI 12. In Fig. S9 436 the plots of the photo product having lowest rank sum (Table S2) are shown. The RI of the 437 photo product $C_9H_{12}O_6$ increased nearly monotonously (in PHOT, Fig. S9 A) and the final 438 intensity nearly doubled compared to the initial intensity.



445

Fig. 3 δRI_{PHOT} plotted versus median δRI_{Reservoir}; green: Reservoir Rappbode, blue: Reservoir
Muldenberg; red: KEY-MFs Rappbode, pink: key comp. Muldenberg. Only the coordinates of
photo products and photo degraded components are shown (see Table 1); as KEY-MF only the
photo products are highlighted

The data from RAPP (Fig. S9 E,G) showed a significantly elevated $C_{9}H_{12}O_{6}$ intensity at the 446 surface water (epilimnion) during stratification months for this component. The same effect 447 was observed in MULD samples (Fig. S9 C). Here the stratification ended already in October. 448 According to the available data, MULD showed elevated epilminetic C₉H₁₂O₆ intensity at the 449 beginning of the following stratification period in 2016. As photo degraded component, 450 $C_{20}H_{16}O_{14}$ showed nearly monotonous decreasing intensity in PHOT (Fig. S19 A). Both in 451 RAPP and MULD samples, depleted intensity was measured during stratification (Fig. S19 452 C,E,G). The profiles of all 30 found KEY-MFs are shown systematically arranged in SI 9-11 453 454 (Figs. S14-S18, photo products; Figs. S19-S23, photo degraded components; Figs. S24-S28, microbial products). As an example for a microbial product the profiles from $C_{10}H_{10}O_7$ (Fig. S7, 455 456 A,C,E,G) show elevated epilimnion intensities in both reservoirs but rather decreasing intensity 457 in PHOT.

458

Regarding the $C_9H_{12}O_6$ profile in Fig. S9 A, it can be shown which properties we expect from a photo product. First, the intensity must be significantly higher at the end of the irradiation experiment (Fig. S9 A). Second, we expect that in natural waters or managed waters (lakes or reservoirs) the intensity is increased in a water body (e.g. epilimnion) receiving more solar radiation dose compared to a water body with attenuated radiation (hypolimnion). This is the case for $C_9H_{12}O_6$ in Fig. S9 C (MULD) and E, G (RAPP). As shown in the methods section 465 (Fig. 1), this expectation can be written as equations and used as search algorithm. Table 1466 shows that reasonable reactivity groups can be found.

As discussed in our recent study,(Herzsprung et al., 2020) it is hardly possible to discriminate between exclusively photochemically produced (or degraded) compounds and those where microbial activity contributed. In addition, one has to consider that photo products might be microbially available (possible priming effect). (Bianchi, 2012; Fudyma et al., 2021; Kuzyakov,

- 471 2010; Olefeldt et al., 2013) This discussion is addressed in SI 5.2 (Fig. S4).
- 472
- 473

474 **3.3 Limitations of assignment of reactivity types**

475 Threshold criteria have to be set for search routines. According to the defined equations 4-6 (SI 2.2), which can be regarded as thresholds, MFs could be assigned to reactivity groups (Table 1). 476 477 The MF C₁₀H₁₀O₇ shows clearly no trend of intensity increase over the total period of PHOT and the final intensity was smaller compared to the start intensity (Fig. S7 A). Photo production 478 479 of this MF can be excluded from that plot. Since epilimnetic intensity increase was observed in both reservoirs for this MF, microbial production can be suggested. A similar MF, C₁₀H₁₂O₇ 480 481 showed significant intensity increase during PHOT and epilimnetic intensity increase in both 482 reservoirs (Fig. S7 B). As conclusion, both C10H10O7 and C10H12O7 must have different biogeochemical reactivity as it was calculated by the defined thresholds. C₈H₁₀O₆ and C₈H₁₀O₅ 483 also have similar compositions. $C_8H_{10}O_6$ has been calculated as microbial product and $C_8H_{10}O_5$ 484 as photo product. However the photo degradation profiles of these two MFs (Fig. S8 A, B) are 485 not such different as those profiles of the compared pair $C_{10}H_{10}O_7$ and $C_{10}H_{12}O_7$ (Fig. S7 A, B). 486 Only the first five time points show different reactivity (increasing intensity for C₈H₁₀O₅ and 487 more or less constant intensity for $C_8H_{10}O_6$ (Fig. S8 A). The final 8 time points show nearly no 488 RI variation for both MFs. These observations show, as an example, that $C_8H_{10}O_6$ is not rather 489 490 appropriate to be selected as KEY-MF (microbial product, see Fig. S3 and Table S3) concerning clarity of reactivity assignment. As a counterexample, for $C_{10}H_{10}O_7$ it is much clearer that it 491 492 must have been microbially produced rather than photochemically (Fig. S7 A, C, E).

The selected threshold criteria are just an idea amongst many other possibilities. It is clear that assignment of reactivity types to MFs will depend on threshold criteria. As shown above and in the SI many profiles were found which confirmed that selected threshold criteria were to some extend reasonable, but not in all cases.

497 For further elucidation six MFs were selected as an example to compare them for assigning498 clear biogeochemical reactivity by discussion of the RI versus radiation dose or versus depth

- plots and consulting rank correlation data from our previous study(Herzsprung et al., 2020).
 For this purpose Figs. S9-S11 are discussed subsequently to Fig. S11 in SI 6.
- It can be finally stated that pure statistical analysis is useful (Herlemann et al., 2014; Osterholz et al., 2018) to describe biogeochemical turnover of DOM but evidently not sufficient to confirm clear reactivity for all assigned MFs. Plotting of the RIs versus time and / or space can achieve more clarity of DOM reactivity.
- 505 Our intension was just to implement a methodology to find out MFs with maximum δ RI. The 506 three selected data sets are appropriate to derive clear statements about processes from RI versus 507 time or space plots. It is clear that processes can be easier derived from batch experiments (here 508 photo degradation) or a stratified lake or reservoir with the assumption of low water mass 509 exchange. The process oriented interpretation of δ RI in rivers or marine ecosystems might be 510 more complicated. However search of maximum δ RI will contribute to understanding of DOM 511 transformations.
- 512

513 **3.4 Practicality of the KEY-MF search with respect to drinking water production**

MFs from reservoir monitoring were already identified to be potential precursors of disinfection 514 515 byproducts as discussed recently (Herzsprung et al., 2020; Phungsai et al., 2016; 2018). With 516 our KEY-MF search it can be verified which precursors are most important and which show high reactivity, as indicated by δRI . For example, $C_{10}H_{14}O_6$ has been identified as a key photo 517 product (Fig. 4 A) and precursor (Phungsai et al., 2016). For MULD, its intensity remained at 518 519 relative low level (red profile, Fig. 4 C) at the water subtraction depth and the bottom layers between spring and late summer (August 2015). In September, after mixing homogenized its 520 vertical distribution, it subsequently dropped continuously over the wintertime. In contrast to 521 this, the polyphenol-like MF C₁₉H₁₄O₁₃ decreased from August to September (Fig. 4 D) and 522 afterwards increased. This knowledge might be used by the water works authority, e.g. by 523 adapting the water subtraction depth in order to avoid undesired DOM MFs, e.g. disinfection 524 byproduct precursors or polyphenol-likes. The latter ones would cause more flocculation costs 525 526 (because of precipitation with iron or aluminum salts) (Andersson et al., 2020; Lavonen et al., 2015; Zhang et al., 2012). If higher spatial resolution is available like in RAPP (8 – 9 different 527 depths, Fig. 4 E,F,G,H), a better adaptation of the subtraction depth is possible. It seems also 528 practical to simply use lake temperature data, which reflect the stratification situation as a proxy 529 530 if the seasonal patterns remain relatively stable within the year. Our results suggest that photo products, which can act as disinfection byproduct precursors (Phungsai et al., 2018), generally 531 532 accumulate in eplimnetic waters and vice versa polyphenol-likes are depleted there because of

- 533 photo degradation. δ RIs of KEY-MFs evidently can be used as proxies for the properties of
- 534 DOM (precursor, flocculation relevance).
- 535 The relevance of the KEY-MFs addressing their abundance in the mass spectrum compared to
- all other assigned MFs has been tested. The found KEY-MFs have mainly abundances much
- 537 larger than the signal to noise ratio (details in SI7 / Fig. S12).



Fig. 4 intensity versus time or depths of the MFs $C_{10}H_{14}O_6$ (photo product) and $C_{19}H_{14}O_{13}$ (photo degraded) – the intensities at the water subtraction depths in Reservoir Muldenberg are shown (red line, C, D)

541 **4.** Conclusions

This study presents an evaluation method for FTICR-MS data sets in order to find out KEYMFs which reveal high biogeochemical reactivity within the very complex DOM pool. The
main outcomes are:

- The hypothesis that the pool with apparently resistant MFs (present in all samples of an
 experiment) can be resolved to products, degraded MFs and reliable resistant MFs could
 be convincingly confirmed
- The search strategies confirmed in part biogeochemical reactivities of MFs which were
 statistically evaluated (by rank correlation) in recent studies
- RI versus time or space plots for single MF are appropriate to proof the suggested biogeochemical reactivities
- The δRI values from different data sets can be plotted versus each other (photo
 experiment versus reservoir stratification) and let space for biogeochemical
 interpretations
- Photo products with extreme δRI values can be found as KEY-MFs
- Some KEY-MFs could be identified as potential disinfection byproduct precursors or
 important for flocculation processes
- 558

The reactivity of MFs can be discriminated in a (at least) semi-quantitative way, just stating component A is more reactive compared to component B under preconditions C. The resulting question is, whether it is reasonable in any case to evaluate all mass peak data which are available in its completeness or would it be more effective to select KEY-MFs as proxies for DOM transformation processes. It is promising to further differentiate only KEY-MFs by chromatographic fractionation or MS / MS fragmentation to get more insights into the isomeric composition and finally into their variations.

566 Supporting information

- 567 Equations for KEY-MF search, description of statistics, RI versus time or depth plots in 1-
- 568 s2.0-Sxxx-mmc1.docx. All elemental formula data, calculations, production of plots in 1-s2.0-
- 569 Sxxx-mmc2.xlsx. Statistical calculations in 1-s2.0-Sxxx-mmc3.xlsx

570 Acknowledgements

- 571 This research did not receive any specific grant from funding agencies in the public,
- 572 commercial, or not-for-profit sectors.

- 573 5. References
- 574
- Andersson, A., Lavonen, E., Harir, M., Gonsior, M., Hertkorn, N., Schmitt-Kopplin, P., Kylin, H. and
 Bastviken, D. 2020. Selective removal of natural organic matter during drinking water
 production changes the composition of disinfection by-products. Environmental ScienceWater Research & Technology 6(3), 779-794.
- Berg, S.M., Peterson, B.D., McMahon, K.D. and Remucal, C.K. 2022. Spatial and Temporal Variability
 of Dissolved Organic Matter Molecular Composition in a Stratified Eutrophic Lake. Journal of
 Geophysical Research-Biogeosciences 127(1).
- Bianchi, T.S. 2012. The role of terrestrially derived organic carbon in the coastal ocean: A changing
 paradigm and the priming effect (vol 108, pg 19473, 2011). Proceedings of the National
 Academy of Sciences of the United States of America 109(13).
- Bittar, T.B., Stubbins, A., Vieira, A.A.H. and Mopper, K. 2015. Characterization and photodegradation
 of dissolved organic matter (DOM) from a tropical lake and its dominant primary producer,
 the cyanobacteria Microcystis aeruginosa. Mar. Chem. 177, 205-217.
- Bracchini, L., Dattilo, A.M., Hull, V., Loiselle, S.A., Nannicini, L., Picchi, M.P., Ricci, M., Santinelli, C.,
 Seritti, A., Tognazzi, A. and Rossi, C. 2010. Spatial and seasonal changes in optical properties
 of autochthonous and allochthonous chromophoric dissolved organic matter in a stratified
 mountain lake. Photochemical & Photobiological Sciences 9(3), 304-314.
- Casas-Ruiz, J.P., Spencer, R.G.M., Guillemette, F., von Schiller, D., Obrador, B., Podgorski, D.C.,
 Kellerman, A.M., Hartmann, J., Gomez-Gener, L., Sabater, S. and Marce, R. 2020. Delineating
 the Continuum of Dissolved Organic Matter in Temperate River Networks. Global
 Biogeochem. Cycles 34(8).
- 596 Cox, D.R. and Stuart, A. 1955. SOME QUICK SIGN TESTS FOR TREND IN LOCATION AND DISPERSION.
 597 Biometrika 42(1-2), 80-95.
- Dadi, T., Harir, M., Hertkorn, N., Koschorreck, M., Schmitt-Kopplin, P. and Herzsprung, P. 2017.
 Redox Conditions Affect Dissolved Organic Carbon Quality in Stratified Freshwaters.
 Environmental Science & Technology 51(23), 13705-13713.
- Dittmar, T., Koch, B., Hertkorn, N. and Kattner, G. 2008. A simple and efficient method for the solid phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnology and
 Oceanography-Methods 6, 230-235.
- Flerus, R., Lechtenfeld, O.J., Koch, B.P., McCallister, S.L., Schmitt-Kopplin, P., Benner, R., Kaiser, K. and
 Kattner, G. 2012. A molecular perspective on the ageing of marine dissolved organic matter.
 Biogeosciences 9(6), 1935-1955.
- Fudyma, J.D., Chu, R.K., Graf Grachet, N., Stegen, J.C. and Tfaily, M.M. 2021. Coupled Biotic-Abiotic
 Processes Control Biogeochemical Cycling of Dissolved Organic Matter in the Columbia River
 Hyporheic Zone. Frontiers in Water 2.
- Gonsior, M., Schmitt-Kopplin, P. and Bastviken, D. 2013. Depth-dependent molecular composition
 and photo-reactivity of dissolved organic matter in a boreal lake under winter and summer
 conditions. Biogeosciences 10(11), 6945-6956.
- Hawkes, E.A., Patrlarca, A., Sjoberg, P.J.R., Tranvik, L.J. and Bergquist, J. 2018. Extreme isomeric
 complexity of dissolved organic matter found across aquatic environments. Limnology and
 Oceanography Letters 3(2), 21-30.

- Herlemann, D.P.R., Manecki, M., Meeske, C., Pollehne, F., Labrenz, M., Schulz-Bull, D., Dittmar, T. and
 Jurgens, K. 2014. Uncoupling of Bacterial and Terrigenous Dissolved Organic Matter
 Dynamics in Decomposition Experiments. Plos One 9(4).
- Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A. and
 Hedges, J.I. 2006. Characterization of a major refractory component of marine dissolved
 organic matter. Geochim. Cosmochim. Acta 70(12), 2990-3010.
- Herzsprung, P., von Tumpling, W., Hertkorn, N., Harir, M., Buttner, O., Bravidor, J., Friese, K. and
 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.
 Environmental Science & Technology 46(10), 5511-5518.
- Herzsprung, P., Wentzky, V., Kamjunke, N., von Tümpling, W., Wilske, C., Friese, K., Boehrer, B.,
 Reemtsma, T., Rinke, K. and Lechtenfeld, O.J. 2020. Improved Understanding of Dissolved
 Organic Matter Processing in Freshwater Using Complementary Experimental and Machine
 Learning Approaches. Environmental science & technology 54(21), 13556-13565.
- Holt, A.D., Kellerman, A.M., Li, W.B., Stubbins, A., Wagner, S., McKenna, A., Fellman, J., Hood, E. and
 Spencer, R.G.M. 2021. Assessing the Role of Photochemistry in Driving the Composition of
 Dissolved Organic Matter in Glacier Runoff. Journal of Geophysical Research-Biogeosciences
 126(12).
- Kamjunke, N., Hertkorn, N., Harir, M., Schmitt-Kopplin, P., Griebler, C., Brauns, M., von Tumpling, W.,
 Weitere, M. and Herzsprung, P. 2019. Molecular change of dissolved organic matter and
 patterns of bacterial activity in a stream along a land-use gradient. Water Res. 164.
- Kellerman, A.M., Dittmar, T., Kothawala, D.N. and Tranvik, L.J. 2014. Chemodiversity of dissolved
 organic matter in lakes driven by climate and hydrology. Nature Communications 5.
- Kellerman, A.M., Kothawala, D.N., Dittmar, T. and Tranvik, L.J. 2015. Persistence of dissolved organic
 matter in lakes related to its molecular characteristics. Nature Geoscience 8(6), 454-U452.
- Kida, M., Fujitake, N., Kojima, T., Tanabe, Y., Hayashi, K., Kudoh, S. and Dittmar, T. 2021. Dissolved
 Organic Matter Processing in Pristine Antarctic Streams. Environmental Science &
 Technology 55(14), 10175-10185.
- Kuzyakov, Y. 2010. Priming effects: Interactions between living and dead organic matter. Soil
 Biology & Biochemistry 42(9), 1363-1371.
- Lavonen, E.E., Gonsior, M., Tranvik, L.J., Schmitt-Kopplin, P. and Kohler, S.J. 2013. Selective
 Chlorination of Natural Organic Matter: Identification of Previously Unknown Disinfection
 Byproducts. Environmental Science & Technology 47(5), 2264-2271.
- Lavonen, E.E., Kothawala, D.N., Tranvik, L.J., Gonsior, M., Schmitt-Kopplin, P. and Kohler, S.J. 2015.
 Tracking changes in the optical properties and molecular composition of dissolved organic
 matter during drinking water production. Water Res. 85, 286-294.
- Lechtenfeld, O.J., Hertkorn, N., Shen, Y., Witt, M. and Benner, R. 2015. Marine sequestration of
 carbon in bacterial metabolites. Nature Communications 6.
- Maizel, A.C., Li, J. and Remucal, C.K. 2017. Relationships Between Dissolved Organic Matter
 Composition and Photochemistry in Lakes of Diverse Trophic Status. Environmental Science
 & Technology 51(17), 9624-9632.
- Martinez-Perez, A.M., Nieto-Cid, M., Osterholz, H., Catala, T.S., Reche, I., Dittmar, T. and Alvarez Salgado, X.A. 2017. Linking optical and molecular signatures of dissolved organic matter in
 the Mediterranean Sea. Scientific Reports 7.

660 Ohno, T., Parr, T.B., Gruselle, M.C.I., Fernandez, I.J., Sleighter, R.L. and Hatcher, P.G. 2014.
661 Molecular Composition and Biodegradability of Soil Organic Matter: A Case Study Comparing
662 Two New England Forest Types. Environmental Science & Technology 48(13), 7229-7236.
663 Olefeldt, D., Turetsky, M.R. and Blodau, C. 2013. Altered Composition and Microbial versus UV-

Mediated Degradation of Dissolved Organic Matter in Boreal Soils Following Wildfire.
Ecosystems 16(8), 1396-1412.

- 666Osterholz, H., Kirchman, D.L., Niggemann, J. and Dittmar, T. 2018. Diversity of bacterial communities667and dissolved organic matter in a temperate estuary. FEMS Microbiol. Ecol. 94(8).
- Phungsai, P., Kurisu, F., Kasuga, I. and 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.
- Phungsai, P., Kurisu, F., Kasuga, I. and Furumai, H. 2018. Changes in Dissolved Organic Matter
 Composition and Disinfection Byproduct Precursors in Advanced Drinking Water Treatment
 Processes. Environmental Science & Technology 52(6), 3392-3401.
- Raeke, J., Lechtenfeld, O.J., Wagner, M., Herzsprung, P. and Reemtsma, T. 2016. Selectivity of solid
 phase extraction of freshwater dissolved organic matter and its effect on ultrahigh resolution
 mass spectra. Environmental Science-Processes & Impacts 18(7), 918-927.
- Reemtsma, T. 2009. Determination of molecular formulas of natural organic matter molecules by
 (ultra-) high-resolution mass spectrometry Status and needs. J. Chromatogr. A 1216(18),
 3687-3701.
- Riedel, T., Zak, D., Biester, H. and Dittmar, T. 2013. Iron traps terrestrially derived dissolved organic
 matter at redox interfaces. Proceedings of the National Academy of Sciences of the United
 States of America 110(25), 10101-10105.
- Riedel, T., Zark, M., Vahatalo, A.V., Niggemann, J., Spencer, R.G.M., Hernes, P.J. and Dittmar, T. 2016.
 Molecular Signatures of Biogeochemical Transformations in Dissolved Organic Matter from
 Ten World Rivers. Frontiers in Earth Science 4.
- Stubbins, A., Lapierre, J.F., Berggren, M., Prairie, Y.T., Dittmar, T. and del Giorgio, P.A. 2014. What's
 in an EEM? Molecular Signatures Associated with Dissolved Organic Fluorescence in Boreal
 Canada. Environmental Science & Technology 48(18), 10598-10606.
- Tang, G., Zheng, X., Hu, S., Li, B., Chen, S., Liu, T., Zhang, B. and Liu, C. 2022. Microbial metabolism
 changes molecular compositions of riverine dissolved organic matter as regulated by
 temperature. Environ. Pollut. 306, 119416.
- Tanuwidjaja, I., Vogel, C., Pronk, G.J., Scholer, A., Kublik, S., Vestergaard, G., Kogel-Knabner, I., Fuka,
 M.M., Schloter, M. and Schulz, S. 2021. Microbial Key Players Involved in P Turnover Differ
 in Artificial Soil Mixtures Depending on Clay Mineral Composition. Microbial Ecology 81(4),
 897-907.
- Wagner, S., Jaffe, R., Cawley, K., Dittmar, T. and Stubbins, A. 2015. Associations Between the
 Molecular and Optical Properties of Dissolved Organic Matter in the Florida Everglades, a
 Model Coastal Wetland System. Frontiers in Chemistry 3.
- Ward, J.H. 1963. HIERARCHICAL GROUPING TO OPTIMIZE AN OBJECTIVE FUNCTION. Journal of the
 American Statistical Association 58(301), 236-&.
- Wilcoxon, F. 1945. INDIVIDUAL COMPARISONS BY RANKING METHODS. Biometrics Bulletin 1(6), 80 83.

703	Wilske, C., Herzsprung, P., Lechtenfeld, O.J., Kamjunke, N., Einax, J.W. and von Tumpling, W. 2021.
704	New Insights into the Seasonal Variation of DOM Quality of a Humic-Rich Drinking-Water
705	Reservoir-Coupling 2D-Fluorescence and FTICR MS Measurements. Water 13(12).
706	Wilske, C., Herzsprung, P., Lechtenfeld, O.J., Kamjunke, N. and von Tumpling, W. 2020.
707	Photochemically Induced Changes of Dissolved Organic Matter in a Humic-Rich and Forested
708	Stream. Water 12(2).
709	Yuan, Z.W., He, C., Shi, Q., Xu, C.M., Li, Z.S., Wang, C.Z., Zhao, H.Z. and Ni, J.R. 2017. Molecular
710	Insights into the Transformation of Dissolved Organic Matter in Landfill Leachate Concentrate
711	during Biodegradation and Coagulation Processes Using ESI FT-ICR MS. Environmental
712	Science & Technology 51(14), 8110-8118.
713	Zhang, H.F., Zhang, Y.H., Shi, Q., Ren, S.Y., Yu, J.W., Ji, F., Luo, W.B. and Yang, M. 2012.
714	Characterization of low molecular weight dissolved natural organic matter along the
715	treatment trait of a waterworks using Fourier transform ion cyclotron resonance mass
716	spectrometry. Water Res. 46(16), 5197-5204.
717	Zheng, Y.X., He, W., Li, B.H., Hur, J., Guo, H.M. and Li, X.M. 2020. Refractory Humic-like Substances:
718	Tracking Environmental Impacts of Anthropogenic Groundwater Recharge. Environmental
719	Science & Technology 54(24), 15778-15788.
720	