Supporting Information

Formation of carbonyl compounds during ozonation of lake water and wastewater: development of a non-target screening method and quantification of target compounds

TAREK MANASFI[1,†,], JOANNA HOUSKA[1,†,§], ISABELLE GEBHARDT[†] AND URS VON GUNTEN[†,‡,§*]

[†]Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

[‡]Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zurich, Switzerland

[§]School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Submitted to *Water Research*

*Corresponding author:
vongunten@eawag.ch

[1] equal contributions from T.M. and J.H.

Contents

Section S1. Chemicals, reagents, and stock solutions

Chemicals and Reagents. Information on the target carbonyl compounds including their CAS numbers, purities, and suppliers are provided in Table S1. Benzaldehyde (>99.5%), benzaldehyde-d_6 (98% deuterated), benzaldehyde tosylhydrazone (benzaldehyde-TSH, 98%), *p*-toluenesulfonylhydrazide (TSH, 97% purity), sodium hydroxide solution (≥98%), sodium phosphate monobasic monohydrate (99%), and sodium phosphate dibasic dihydrate (≥98%) were obtained from Sigma-Aldrich (Switzerland). rac-tramadol-d_6 hydrochloride (tramadol-d_6, chemical purity 98%, isotopic purity 99.8%) was obtained from TRC Canada. Methanol (99.99%) and formic acid (LCMS grade) were obtained from Fisher Scientific (Switzerland). The Suwannee River II Standard Fulvic Acid (SRFA, 2S101F) isolate was obtained from the International Humic Substances Society (IHSS), St. Paul, Minnesota, USA). Ultrapure water (resistivity > 18.2 MΩ·cm, ASTM type 1) used for the preparation of aqueous solutions was produced using Arium® pro Ultrapure Laboratory Water Systems, Sartorius.

Stock solutions. Carbonyl compound stock solutions at concentrations in the range of 1 to 50 mM were prepared in acetonitrile (ACN) with the exception of glyoxylic acid, which was prepared in methanol due to its low solubility in acetonitrile. The carbonyl compound stock solutions were stored in amber vials with no or little headspace and stored in the freezer at -20 °C for up to 4 months. Intermediate solutions were prepared freshly in ultrapure water prior to the experiments. The stock solutions of the derivatising agent TSH were prepared freshly in ACN at a concentration of 10 mM and kept at room temperature for up to 7 days. Caps of vials containing stock solutions with solvents were wrapped with parafilm (Parafilm M, Bemis) to provide a better sealing during storage. Stock solutions of SRFA (c_{DOC} ~ 50 mgC·L^{-1}) were prepared in ultrapure water.

Generation of ozone stock solutions. An ozone (O_3) generator (BMT 803 BT, BMT Messtechnik, Berlin) producing ozone-containing gas from pure oxygen (Carbagas, 99.995%) was used. The ozone stock solutions were prepared by bubbling the generated ozone gas into

ice-cooled ultrapure water (Bader and Hoigné 1981). The concentrations of ozone stock solutions were determined with a spectrophotometer (Cary 100, Varian, USA) at 260 nm, with a molar absorption coefficient of $\varepsilon = 3200\ M^{-1}\ cm^{-1}$ (von Sonntag and von Gunten 2012).

Table S1. List of the selected target carbonyl compounds with information about their purity, CAS number, molecular formulas, compound class, and supplier

Chemicals	**CAS number**	**Molecular formula**	**Compound class**	**Supplier**
Formaldehyde, 37 wt.%	50-00-0	CH_2O	Aldehyde	Sigma-Aldrich
Acetaldehyde, 99.5%	75-07-0	C_2H_4O	Aldehyde	Acros organics
Decanal, 95%	112-31-2	$C_{10}H_{20}O$	Aldehyde	Sigma-Aldrich
Benzaldehyde, 99%	100-52-7	C_7H_6O	Aldehyde	Fluka
Indole-3-carboxaldehyde, 97%	487-89-8	C_9H_7NO	Aldehyde	Sigma-Aldrich
4-Hydroxy-2-methoxybenzaldehyde, 98%	673-22-3	$C_8H_8O_3$	Aldehyde	Sigma-Aldrich
Crotonaldehyde (cis/tans), 99.5%	4170-30-3	C_4H_6O	Aldehyde, unsaturated	Sigma-Aldrich
Cinnamaldehyde, 95%	104-55-2	C_9H_8O	Aldehyde, unsaturated	Sigma-Aldrich
Methacrolein, 95%	78-85-3	C_4H_6O	Aldehyde, unsaturated	Sigma-Aldrich
Cyclopentanone, 99%	120-92-3	C_5H_8O	Ketone, cyclic	Fluka
1-Acetyl-1-cyclohexene, 97%	932-66-1	$C_8H_{12}O$	Ketone, unsaturated	Sigma-Aldrich
2,3-Butanedione, 97%	431-03-8	$C_4H_6O_2$	Diketone	Sigma-Aldrich
3,5-Heptanedione, 97%	7424-54-6	$C_7H_{12}O_2$	Diketone	Sigma-Aldrich
Glyoxal, 40 wt.%	107-22-2	$C_2H_2O_2$	Diketone	Sigma-Aldrich
Glutaraldehyde, 50 wt.%	111-30-8	$C_5H_8O_2$	Diketone	Sigma-Aldrich
Glyoxylic acid, 98%	563-96-2	$C_2H_2O_3$	Ketoacid	Sigma-Aldrich
Pyruvic acid, 98%	127-17-3	$C_3H_4O_3$	Ketoacid	Sigma-Aldrich

Section S2. Surface water and wastewater samples

Table S2. Water quality parameters of the collected wastewater (WW) and lake water samples

Matrix	**pH**	**DOC (mg/L)**	**Alkalinity (mmol/L)**	**NH_4^+ (µg/L)**	**NO_2^- (µg/L)**	**Conductivity (µS/cm at 20 °C)**
WW Neugut*	8.15	5.06	5.14	42.4	5.8	1014
WW Werdhölzli September 2020*	7.90	5.50	3.59	109.6	43.0	ND
WW Werdhölzli March 2021 a[#]	7.94	4.70	3.68	390.0	110.0	670
WW Werdhölzli March 2021 b[#]	7.85	4.70	3.68	613.0	179.0	693
WW Werdhölzli March 2021 c[#]	7.80	5.20	3.46	529.0	240.0	711
WW Werdhölzli October 2021	8.07	6.00	3.77	428.0	153.0	656
O_3 Werdhölzli March 2021[1]	7.77	6.50	3.68	439.0	43.1	673
O_3 Werdhölzli March 2021[2]	7.76	6.40	3.68	722.0	38.8	692
O_3 Werdhölzli March 2021[3]	7.70	6.50	3.43	561.0	28.6	708
SF Werdhölzli March 2021[1]	7.82	4.40	3.55	16.8	1.5	769
SF Werdhölzli March 2021[2]	7.92	4.20	3.44	25.1	2.4	799
SF Werdhölzli March 2021[3]	7.65	3.80	3.46	18.9	2.5	814
WW Glarnerland*	8.73	10.97	1.76	ND	227.5	ND
Lake Greifensee water*	8.42	3.40	3.36	18.4	26.9	273
Lac de Bret water	8.40	3.10	3.45	29.5	19.6	352

[*]samples underwent pH adjustment to pH 7 using phosphate buffer prior to ozonation [#] grab samples of secondary wastewater effluent collected on the same day; [1,2,3]full-scale ozonation at the WWTP at specific ozone doses of 0.27, 0.42 and 0.90 mgO_3/mgC, respectively.

Section S3. Optimisation of the derivatisation method

Samples used for optimising the derivatisation procedure were prepared by spiking carbonyl compounds to ultrapure water, secondary wastewater effluent, ozonated wastewater, and SRFA-containing water (see below for concentrations). The derivatisation efficiency for benzaldehyde and cinnamaldehyde could be calculated with standards of derivatised benzaldehyde-TSH and cinnamaldehyde-TSH. For the other carbonyl compounds, the derivatisation efficiency was assessed using peak areas. All samples were spiked with benzaldehyde-d_6 (100 nM) and tramadol-d_6 (32.7 nM) as derivatisation surrogate and internal

standard, respectively (Figure S1). An ultrapure water sample was derivatised according to the same procedure and was used as a blank for quality control.

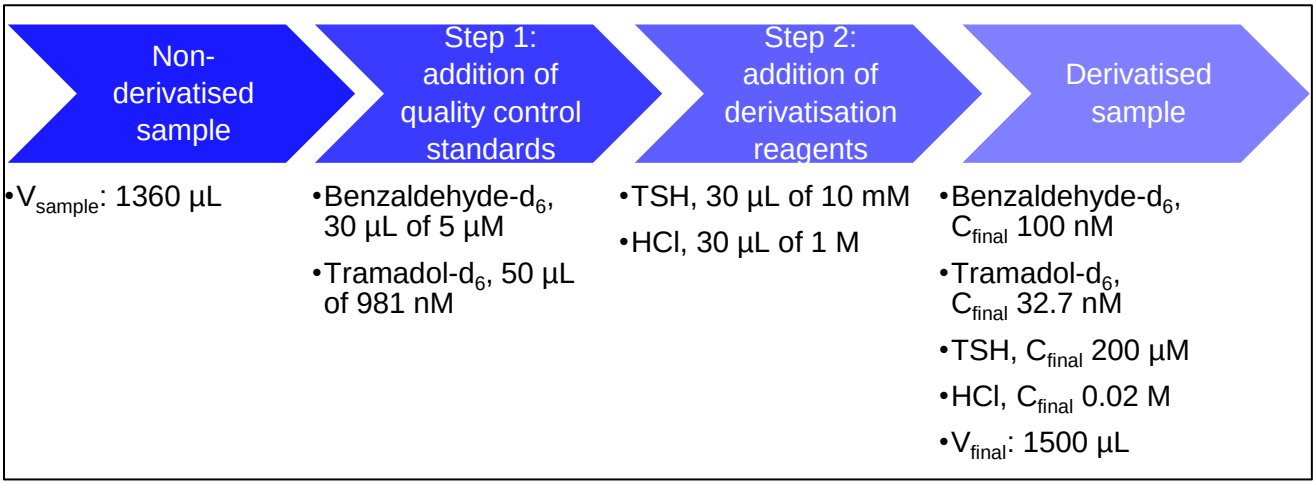

Figure S1. Flow scheme of the steps of the derivatisation procedure.

To optimise the derivatisation procedure, the influence of adding varying TSH (Figures S2 and S3) and HCl concentrations (Figure S4a), and modification of the reaction times (Figure S4b) on the derivatisation efficiency were evaluated in multiple matrices (ultrapure water, ozonated and non-ozonated SRFA-containing water and wastewater). Samples were left at room temperature for different reaction times (10-120 min) after addition of TSH and HCl and were loaded on an autosampler (4 °C) for analysis by LC-ESI-HRMS.

Effect of TSH concentration

Increasing the applied TSH concentrations from 12.5 - 200 µM (corresponds to a molar excess of 7.5- to 125-fold with respect to the total carbonyl compounds content) enhanced the derivatisation efficiency of the tested carbonyl compounds in the different tested water types (Figures S2).

Benzaldehyde-TSH and cinnamaldehyde-TSH concentrations were determined by external calibration based on the response ratios. The latter corresponds to the ratio of the peak area of the analyte to the peak area of the TSH-derivatised benzaldehyde-d_6 in the same sample. To calculate the derivatisation efficiencies, the measured concentrations ($C_{measured}$) were divided

by the theoretical (spiked) concentration (C_{spiked} = 200 nM), and multiplied by 100 (equation 1).

$$\text{Derivatisation efficiency (\%)} = \frac{C_{measured}}{C_{spiked}} \cdot 100 \quad (1)$$

At a TSH concentration of 200 µM, the derivatisation efficiency of benzaldehyde and cinnamaldehyde reached 82–98% and 91–105%, respectively (Figure S3). In most water types, the use of higher TSH concentration (300 µM) did not significantly enhance the derivatisation efficiency (Figure S3). The disadvantages of potential impurities and carry-over of TSH in LC-ESI-HRMS due to higher TSH concentrations were considered to outweigh the benefit of slightly higher derivatisation efficiencies. Therefore, 200 µM TSH was selected as an optimum concentration for derivatisation, which is higher than in a previous study (~135 µM) (Marron et al. (2020)).

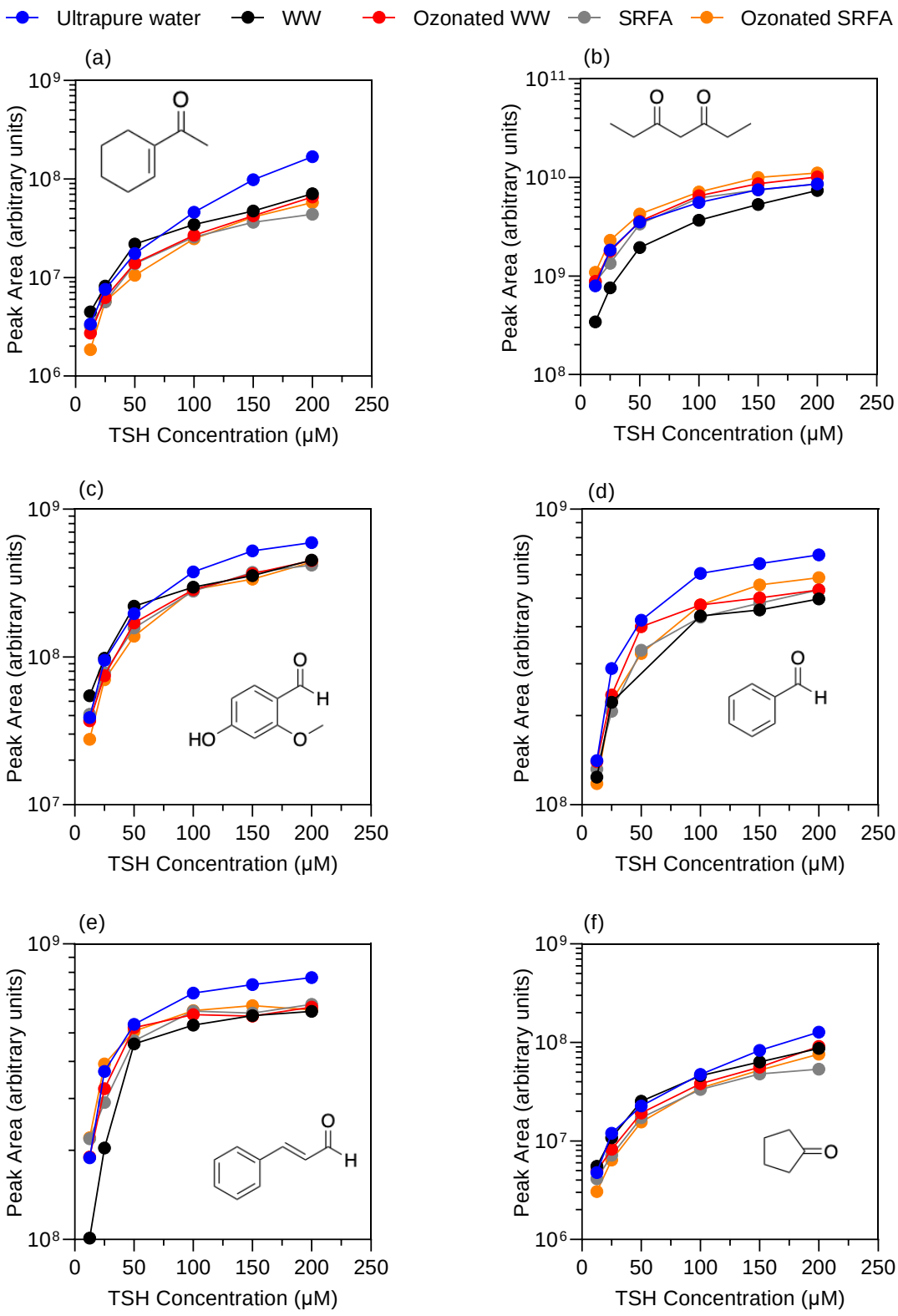

Figure S2. Influence of the TSH concentration on the derivatisation efficiency of 6 carbonyl compounds (100 nM each) in the different (non)-ozonated water types (ultrapure water, (non)-ozonated wastewater from Neugut WWTP (WW), and (non-)ozonated SRFA-containing water (SRFA)). (a) 1-Acetyl-1-cyclohexene, (b) 3,5-heptanedione, (c) 4-hydroxy-2-methoxybenzaldehyde, (d) benzaldehyde, (e) cinnamaldehyde, and (f) cyclopentanone.

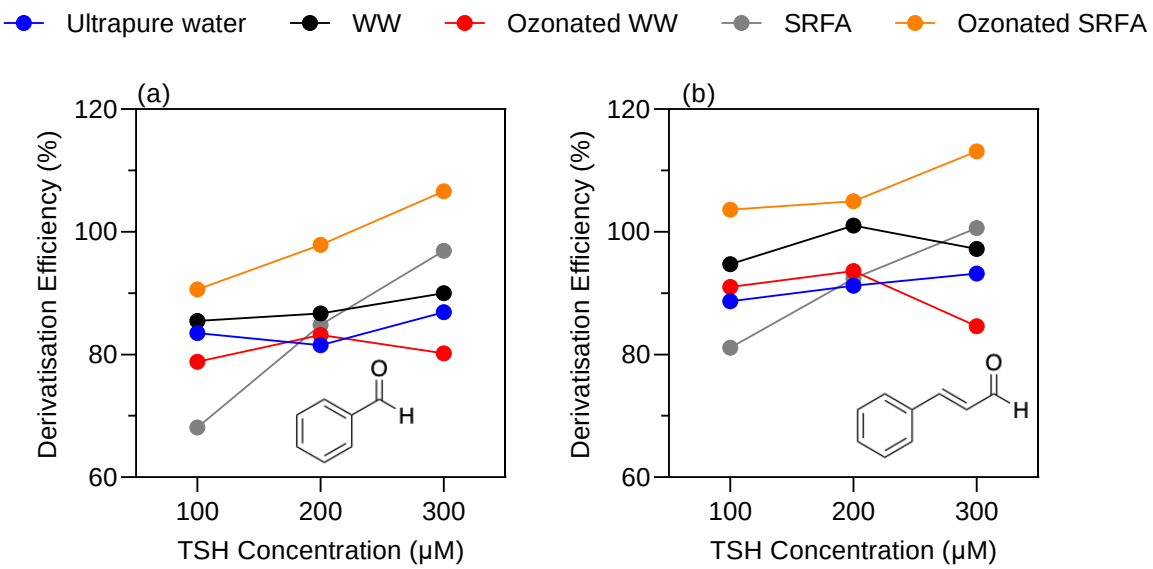

Figure S3. Derivatisation efficiency of (a) benzaldehyde and (b) cinnamaldehyde at 200 nM using varying TSH concentrations in different (non)-ozonated water types (ultrapure water, (ozonated) wastewater from Neugut WWTP (WW), and SRFA-containing water (SRFA)).

Effect of HCl concentration

Ultrapure water was spiked with a mix containing 100 nM of each target carbonyl compound (except 2,3-butanedione which was acquired later). TSH (30 µM, which is equivalent to about 18-fold molar excess to the total carbonyl compounds concentration) and different HCl concentrations (0-0.05 M) were added (Figure S4a). HCl catalyses the reaction between TSH and carbonyl compounds (Marron et al., 2020; Siegel et al., 2014). The derivatisation efficiencies (based on peak areas) of all target carbonyl compounds increased considerably with the addition of HCl up to 0.01 M, except for formaldehyde, acetaldehyde and decanal. Higher HCl concentrations either increased, slightly decreased or did not affect the derivatisation efficiency (Figure S4a). Based on this and to guarantee the same efficiency in all water types including wastewater, a concentration of 0.02 M HCl was sufficient to overcome a potential buffering capacity of a real water and to achieve maximum derivatisation yields. This is similar to Marron et al. (2020) in water and wastewater matrices, whereas in Siegel et al. (2014) a higher HCl concentration (0.04 M) was applied for the derivatisation of yeast cell sample extracts.

Effect of reaction time

Ultrapure water samples spiked with target carbonyl compounds (500 nM each) were derivatised with TSH (200 µM, about 25-fold molar excess with respect to the total carbonyl compounds content) and 0.02 M HCl. The samples were stored at room temperature for reaction times ranging from 10 to 120 min before their analyses (Figure S4b). For most compounds, the derivatisation efficiencies did not change when the reaction time for derivatisation increased from 10 to 120 min (Figure S4b). For one compound (glutaraldehyde), a higher intensity of the derivatised species was detected after 120 min compared to 10 min. In contrast, for a few compounds (glyoxal, methacrolein, pyruvic acid and crotonaldehyde) the intensity decreased slightly with time. Therefore, a reaction time of 10 min was selected to reach maximum derivatisation efficiencies for the majority of the compounds.

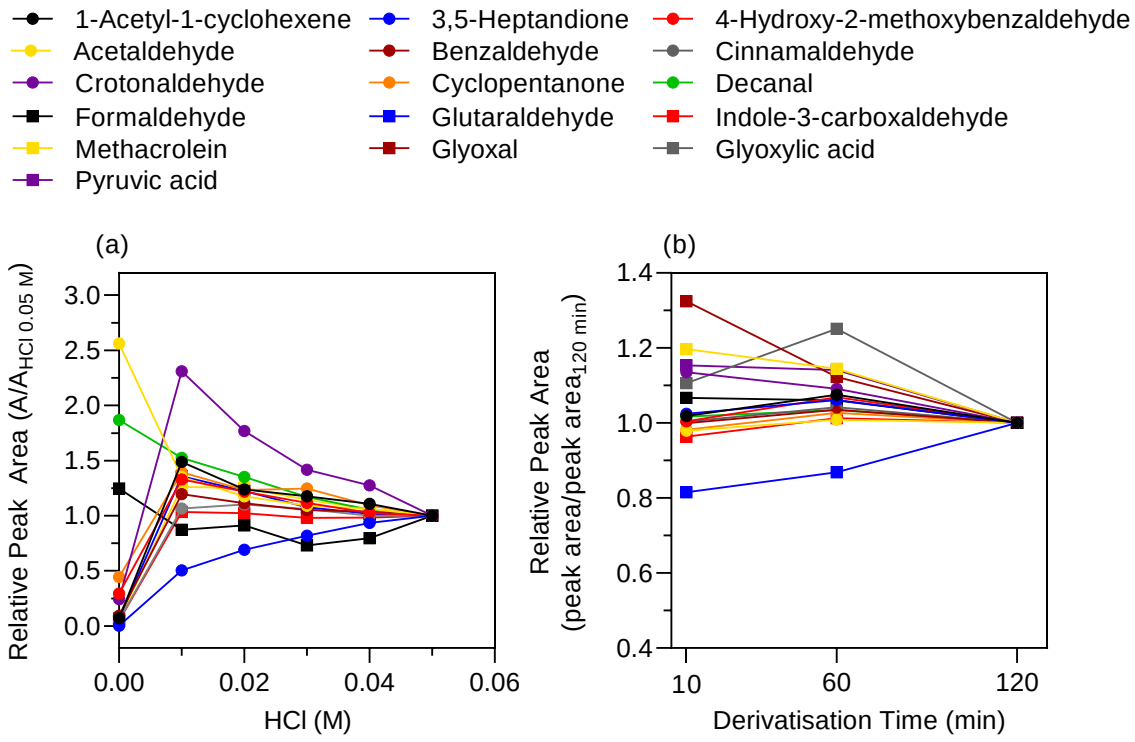

Figure S4. Derivatisation efficiency of 16 carbonyl compounds (see header of the figure) in ultrapure water at room temperature: (a) Effect of HCl concentration, spiked carbonyl compound concentration 100 nM each (glyoxal, glyoxylic acid and pyruvic acid were not detected), and (b) effect of reaction time, spiked carbonyl compound concentration 500 nM each.

Effect of derivatisation and ionization mode on the detectability of carbonyl compounds by LC-ESI-HRMS

A mixture of carbonyl compounds spiked to ultrapure water was analysed before and after derivatisation (200 µM TSH and 0.02 M HCl) in full scan mode with positive and negative polarity. The instrument mass range was set to either 55-550 Da (before derivatisation) or 100-1000 Da (after derivatisation). The mixture of carbonyl compounds contained 250 nM of each, crotonaldehyde, benzaldehyde, cinnamaldehyde, 4-hydroxy-2-methoxybenzaldehyde, cyclopentanone, decanal, indole-3-carboxaldehyde, methacrolein, 3,5-heptanedione and 1-acetyl-1-cyclohexene. It also contained the following compounds at different concentrations: glutaraldehyde (750 nM), glyoxal (3 µM), pyruvic acid (1.5 µM), acetaldehyde (3 µM), glyoxylic acid (4.5 µM) and formaldehyde (3 µM). The total carbonyl concentration in the sample was 18.25 µM. Under these conditions, the applied TSH concentration (200 µM) represents an 11-fold molar excess with respect to the total carbonyl compounds concentration.

Figure S5 shows the detectability of target carbonyl compounds in positive and negative ESI modes with and without derivatisation. The responses of the carbonyl compounds were significantly enhanced after derivatisation, with the exception of 1-acetyl-1-cyclohexene which had a two-fold higher peak area in the positive mode before derivatisation. Formaldehyde, acetaldehyde, crotonaldehyde and 3,5-heptanedione were only detected after derivatisation in the positive mode. Decanal, benzaldehyde, methacrolein, glyoxal and glutaraldehyde were only detected after derivatisation in both the positive and negative modes. Cinnamaldehyde, cyclopentanone, and 1-acetyl-1-cyclohexene were detected in both modes when derivatised, and only in the positive mode without derivatisation. Glyoxylic acid and pyruvic acid were detected in both modes when derivatised, and in the negative mode without derivatisation. Indole-3-carboxaldehyde and 4-hydroxy-2-methoxybenzaldehyde were detected for all four conditions.

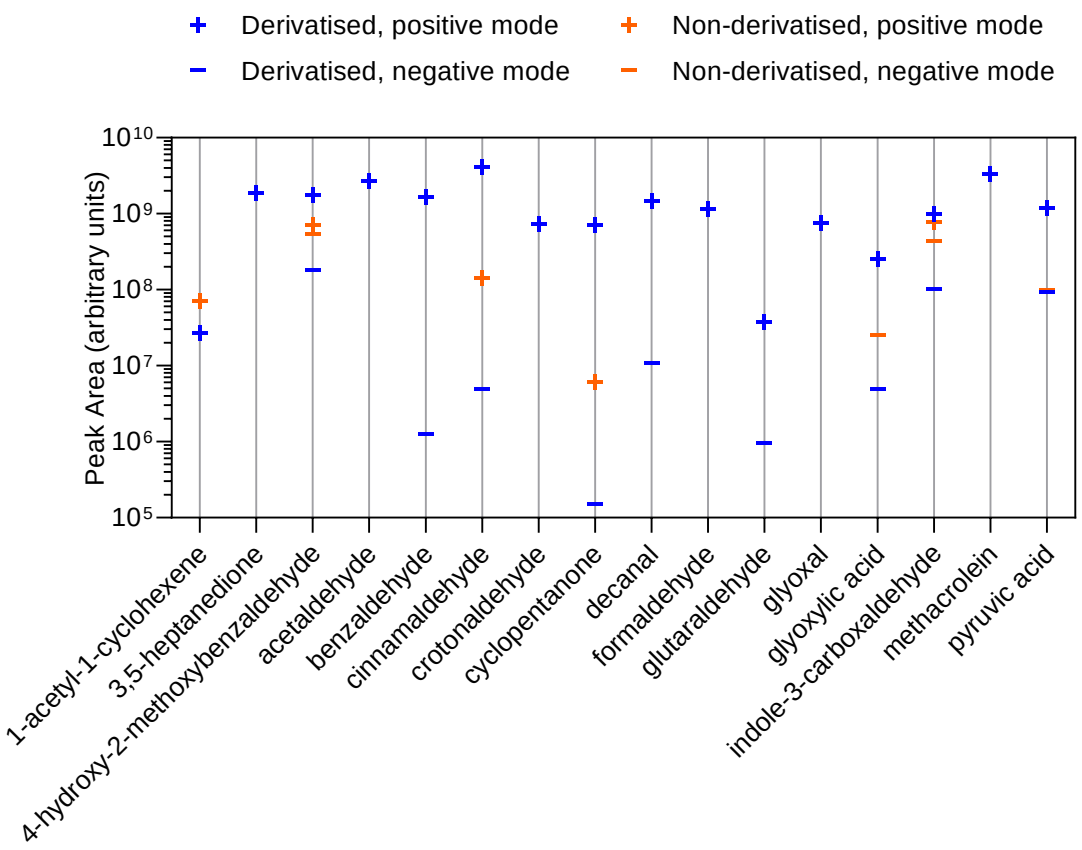

Figure S5. Detectability of 16 target carbonyl compounds before and after derivatisation in ultrapure water in both positive and negative ESI modes. For experimental conditions see text.

Section S4. Method performance characteristics

Limits of detection (LOD), limits of quantification (LOQ), and measurement ranges

Table S3. Limits of detection (LOD), limits of quantification (LOQ) and measurement ranges of the target carbonyl compounds

Compound	LOD (µg/L)	LOD (nM)	LOQ (µg/L)	LOQ (nM)	Measurement Range (µg/L)	Measurement Range (nM)
1-acetyl-1-cyclohexene	0.11	0.9	0.37	3	0.37-18.63	3-150
2,3-butanedione	0.26	3.0	0.86	10	0.86-21.52	10-250
3,5-heptanedione	0.04	0.3	0.13	1	0.13-12.82	1-100
4-hydroxy-2-methoxybenzaldehyde	0.05	0.3	0.15	1	0.15-22.82	1-150
Acetaldehyde	0.40	9.0	1.32	30	1.32-132.15	30-3000
Benzaldehyde	0.32	3.0	1.06	10	1.06-21.22	10-200
Cinnamaldehyde	0.12	0.9	0.40	3	0.40-33.04	3-250
Crotonaldehyde	0.004	0.06	0.01	0.2	0.01-3.50	0.2-50
Cyclopentanone	0.25	3.0	0.84	10	0.84-21.03	10-250
Decanal	0.47	3.0	1.56	10	1.56-23.44	10-150
Formaldehyde	4.50	150.2	15.01	500	15.01-180.16	500-6000
Glutaraldehyde	1.35	13.5	4.51	45	4.51-30.04	45-300
Glyoxal	0.17	3.0	0.58	10	0.58-58.04	10-1000
Glyoxylic acid	1.00	13.5	3.33	45	3.33-333.18	45-4500
Indole-3-carboxaldehyde	0.04	0.3	0.15	1	0.15-14.52	1-100
Methacrolein	0.06	0.9	0.21	3	0.21-7.01	3-100
Pyruvic acid	1.19	13.5	3.96	45	3.96-264.18	45-3000

Assessment of matrix effects

Analytical matrix effects on the measurement of derivatised carbonyl compounds using LC-ESI-HRMS were investigated in ozonated and non-ozonated secondary wastewater effluent and SRFA-containing water in comparison to ultrapure water. Ozonated (specific ozone dose 1 mgO_3/mgC) and non-ozonated SRFA and WW Neugut samples were spiked with benzaldehyde-TSH and cinnamaldehyde-TSH at a concentration of 200 nM, followed by derivatisation with 200 µM TSH and 0.02 M HCl (referred to as spiked samples in Equation 2). Non-spiked samples of ozonated (specific ozone dose 1 mgO_3/mgC) and non-ozonated SRFA and WW Neugut derivatised similarly to the spiked samples were used for blank correction, to account for any benzaldehyde or cinnamaldehyde that might have been present originally (not from the spiked amount of benzaldehyde-TSH and cinnamaldehyde-TSH). Ultrapure water

spiked with benzaldehyde-TSH and cinnamaldehyde-TSH at a concentration of 200 nM, followed by derivatisation with 200 µM TSH and 0.02 M HCl was used as a reference (spiked ultrapure water, equation 2) for the determination of matrix effects. All samples contained 32.7 nM tramadol-d_6 and 100 nM benzaldehyde-d_6 as internal standards. The analytical matrix effects were calculated according to equation 2:

$$\text{Matrix effect (\%)} = \frac{\text{Peak area}_{\text{spiked sample}} - \text{Peak area}_{\text{unspiked sample}}}{\text{Peak area}_{\text{spiked ultrapure water}}} \cdot 100 \quad (2)$$

After analysis of the benzaldehyde-TSH- and cinnamaldehyde-TSH-spiked ozonated and non-ozonated SRFA and WW Neugut, the analytical matrix effects were determined by comparing the response of a spiked sample to a spiked ultrapure water sample in which matrix effects are assumed to be absent. The determined matrix effects in the different water types ranged between 107 to 117% for benzaldehyde-TSH, and between 106 to 115% for cinnamaldehyde-TSH, indicating the absence of any significant analytical matrix effect. Based on these results for cinnamaldehyde and benzaldehyde and the similarity of intensities of derivatised carbonyls in the different matrices (see Section S3, influence of TSH concentration), the matrix effects are assumed to be negligible during the analysis of derivatised carbonyl compounds.

Table S4. Matrix effects (%) in the analysis of benzaldehyde-TSH and cinnamaldehyde-TSH in the different tested water matrices

	Matrix effects (%)* with respect to spiked ultrapure water	
Matrix	Benzaldehyde-TSH	Cinnamaldehyde-TSH
SRFA-containing water	113	112
Ozonated SRFA-containing water	107	106
WW Neugut	117	115
Ozonated WW Neugut	108	108

WW = wastewater; *the matrix effect % were calculated with respect to the response in spiked ultrapure water. A 100% matrix effect means that the response in the matrix was the same as in ultrapure water (equation 2)

Formaldehyde background contamination

Formaldehyde was always detected in derivatised blanks (of ultrapure water) to which no formaldehyde spiking was carried out. Several factors were tested to examine potential sources of the background formaldehyde contamination: (i) replacing ultrapure water with LC/MS grade water for sample preparation, (ii) enhanced rinsing of the glassware (four times with ultrapure water and then four times with acetonitrile), (iii) fresh preparation of the HCl stock solution in LC/MS grade water, (iv) fresh preparation of the TSH stock solution in different hoods than the stock solutions to avoid potential contamination, (v) fresh preparation of the TSH stock solution in ultrapure water or methanol instead of acetonitrile, (vi) replacing formic acid in the eluents with acetic acid, and (vii) replacing the column (already used with charged matrices) with a new clean column. All these measures did not lead to a reduction in blank background levels. The concentrations of formaldehyde in the blanks were mostly dependent on the applied TSH concentrations, with higher formaldehyde concentrations at higher TSH concentrations. Furthermore, the background formaldehyde concentration was detected even in derivatized blanks which were analysed immediately after derivatization and did not change substantially with increasing holding times. For all these reasons, the formaldehyde background contamination was attributed to mainly TSH rather than to a contamination from air or other sources. It is worth mentioning that a main difference to Marron et al. 2020, where increasing holding times resulted in higher background formaldehyde concentrations, is that in the latter the assessment was made based on samples held on the sampling tray at room temperature, while in the present study samples were placed on a sampling tray immediately after derivatisation and kept at 4 °C until injection. To avoid an overestimation of formaldehyde concentrations, the same concentration of TSH was used for derivatising calibration standard solutions and samples, and the background level coming from TSH was consistently accounted for.

Section S5. Stability of the target carbonyl compounds

Stability of the derivatised carbonyl compounds

An evaluation of the stability of derivatised carbonyl compounds was performed by spiking secondary wastewater effluent (from Werdhölzli WWTP) with carbonyl compounds, conducting derivatisation, and storing the samples for different times ≤ 6.5 days before analysis.

Significant differences in the stability of the different derivatised carbonyl compounds spiked into secondary wastewater effluent were observed during storage at 4 °C (Figure S6). Some derivatised compounds including glyoxal, glyoxylic acid, pyruvic acid and formaldehyde were very unstable (Figure S6a). After ~1 day of storage, glyoxal entirely disappeared, and losses of 72%, 60%, and 40% were observed for pyruvic acid, glyoxylic acid, and formaldehyde, respectively. All four compounds were below LOQ after 6.5 days. Moderately unstable derivatised carbonyl compounds included decanal (13% loss after 1 day), crotonaldehyde (39% loss after 1 day), and methacrolein (32% loss after 1 day). However, unlike the highly unstable compounds, these compounds were still quantifiable after 6.5 days with residual fractions of 34% for decanal, 36% for crotonaldehyde, and 47% for methacrolein (Figure S6a). The other target carbonyl compounds were stable during the 6.5 days period (Figure S6b).

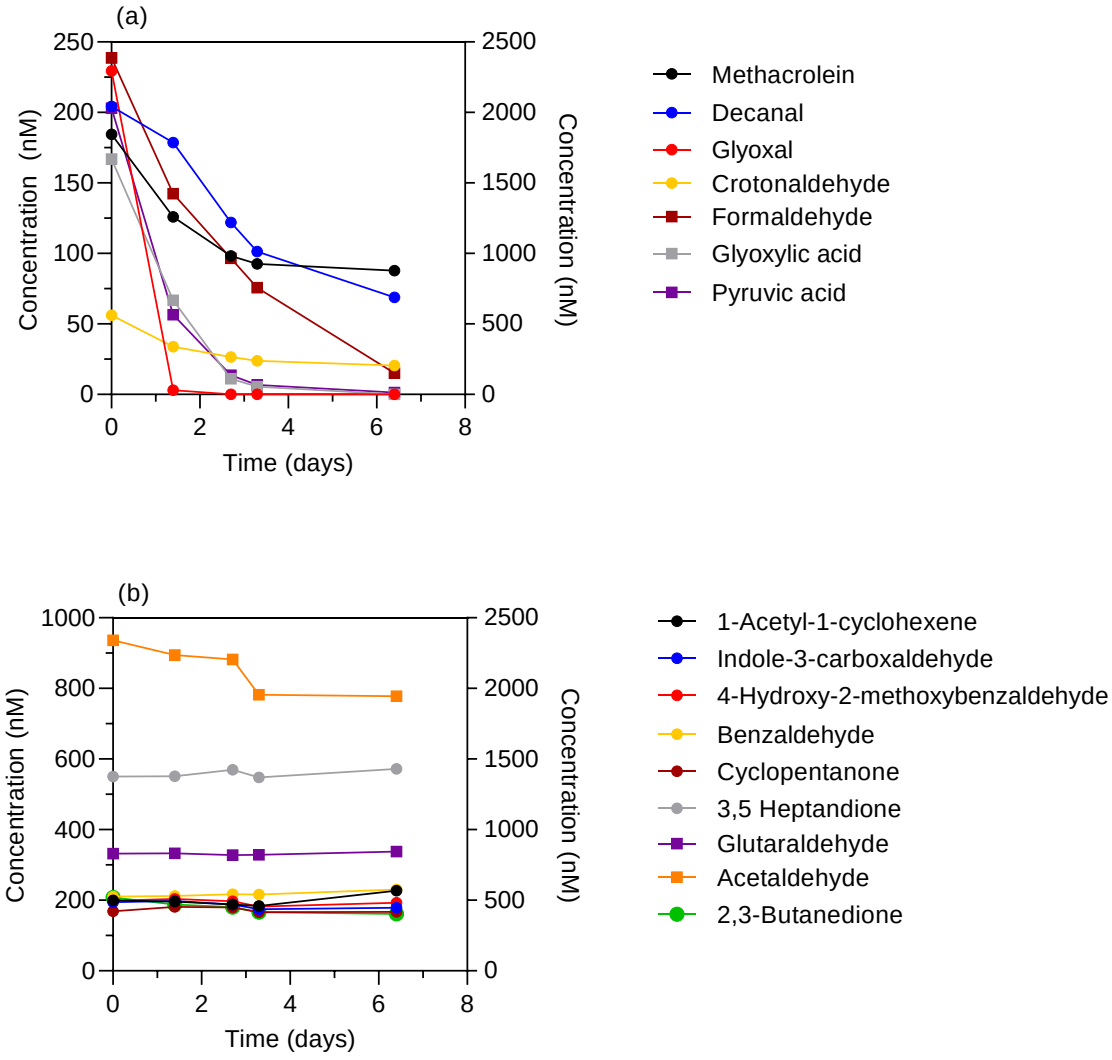

Figure S6. Stability of derivatised target carbonyl compounds in spiked secondary wastewater effluent over a period of 6.5 days; storage T = 4 °C. (a) Unstable carbonyl compounds; formaldehyde, pyruvic acid and glyoxylic acid concentrations are displayed on the secondary y-axis. (b) Stable carbonyl compounds; acetaldehyde and glutaraldehyde concentrations are displayed on the secondary y-axis.

Stability of the non-derivatised carbonyl compounds

To mimic sample storage effects and assess the stability of carbonyl compounds, ozonated wastewater samples (three different specific ozone doses in full-scale at the Werdhölzli WWTP) were derivatised after 1, 2 and 7 days and immediately analysed after derivatisation.

Figure S7 shows the variation of the concentrations of target carbonyls in ozonated wastewater for holding times of up to 7 days prior to derivatisation.

1-acetyl-1-cyclohexene increased during the first 2 days of storage but decreased significantly after 7 days. High concentrations of glyoxal were consistently (in all three ozonation experiments) formed during a storage period of 7 days. This delayed formation of glyoxal may be caused by the decomposition of a precursor formed during ozonation (cf. paragraph h, Section 3.4.2, main text).

2,3-butanedione and acetaldehyde only showed minor variations for the applied holding time. Glutaraldehyde decreased slightly within 2 days and a loss of around 50% was observed after 7 days. Formaldehyde showed inconsistent trends for different sample sets with a decrease of ca. 25% observed in sets O3-1 and O3-2 (Figure S7) and a decrease below < LOQ in set O3-3.

Pyruvic acid and glyoxylic acid were highly unstable and decreased to <LOQ within 2 days. However, low concentrations were measured again after 7 days indicating potential minor formation during prolonged storage, which may be caused by decomposition of a precursor.

Overall, the assessment of stability shows that the evolution of carbonyl compounds is dynamic whether derivatised or not. This means that, to guarantee highest accuracy, delays before and after derivatisation should be avoided. Moreover, once derivatised, samples should be analysed as soon as possible (optimally within 1 day). In the current study, sample analysis by LC-HRMS was performed within ≤ 2 days after derivatisation due to logistical boundary conditions. Accounting for the instability of some target compounds, the measured concentration could be

underestimated for samples which came late on the measurement sequence. However, given that the latter was established coherently in terms of having samples from the same (waste)water sample with increasing ozone doses measured consecutively with negligible time difference, this provides a high degree of certainty regarding the determined trends and interpretations. This is also confirmed by the redundancy and coherence of trends across multiple water types and in some cases multiple sample sets for the same water type (e.g., 5 sets for wastewater).

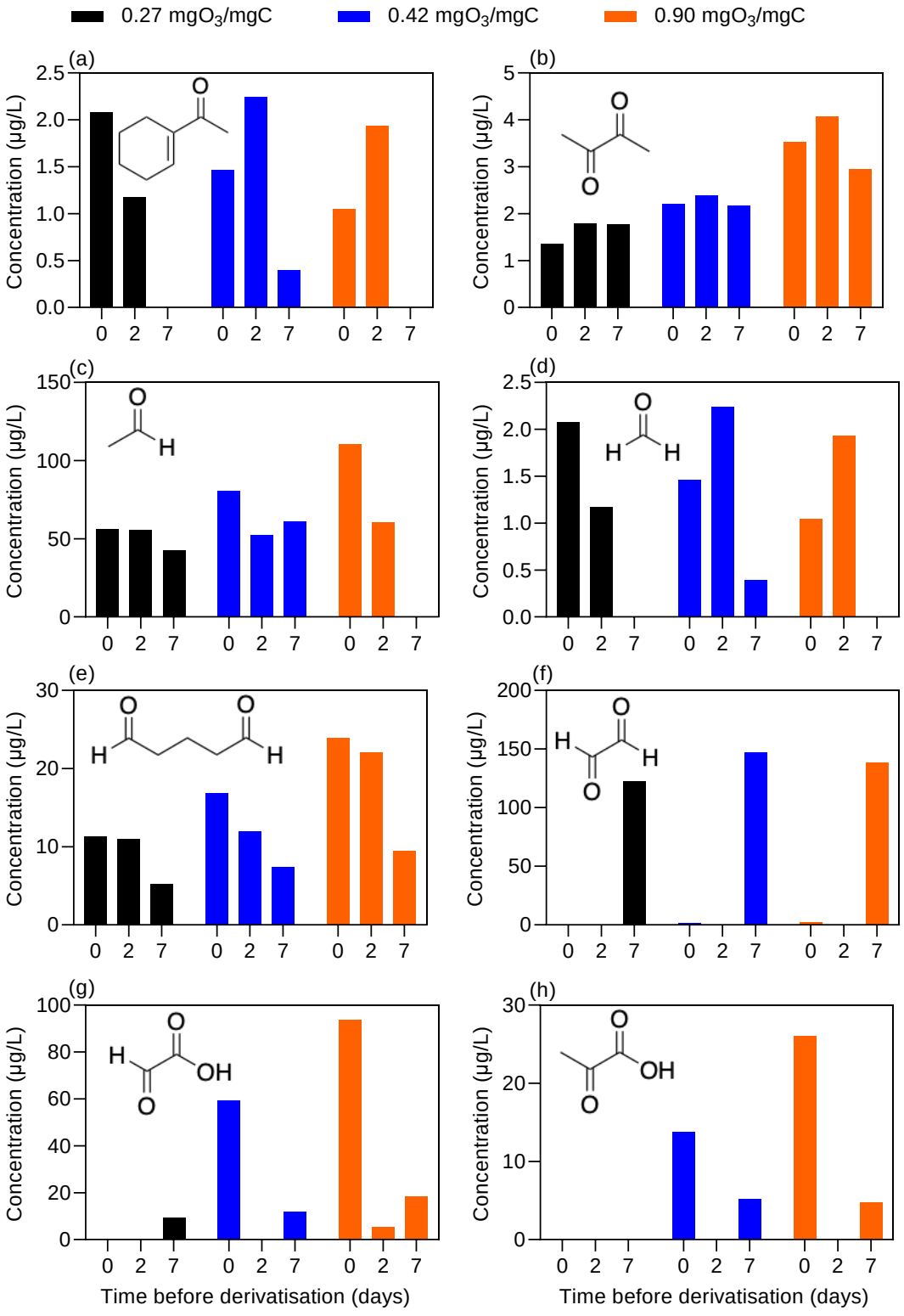

Figure S7. Stability of the target carbonyl compounds before derivatisation. Target carbonyl compounds concentrations in plant-ozonated wastewaters (at three specific ozone doses in Werdhölzli WWTP, see legend on top) derivatised within 24 hours (0 days), after 2 days, or after 7 days from sample collection; pH = 7.70 – 7.77, storage T = 4 °C; (a) 1-acetyl-1-cyclohexene, (b) 2,3-butanedione, (c) acetaldehyde, (d) formaldehyde, (e) glutaraldehyde, (f) glyoxal, (g) glyoxylic acid, and (h) pyruvic acid.

Section S6. Non-target screening and data processing workflow

The non-target screening workflow was based on the software Compound Discoverer 3.2 (CD3.2, Thermo Scientific, Germany) and consisted of nodes categorized into two main groups: processing and post-processing nodes (Figure S8).

Processing nodes included the selection of spectra, alignment of retention times across multiple LC-MS data files, detection and grouping of compounds by MW and retention time across different LC-MS data files, filling gaps, marking background compounds, and compound identification. The Fill Gaps node was included to provide an estimate of the background/noise intensity for peaks missing in certain data files. The Mark Background Compounds node was used to flag compounds that were also found in the blanks (TSH-derivatised ultrapure water). Compounds which occurred at a lower than five-fold peak area in the samples compared to the blanks were flagged as background compounds and were excluded from further processing. Compound identification nodes included predicting compositions with the minimum elemental composition of $C_8H_{10}N_2O_2S$ (at least one TSH with an additional carbon, which corresponds to formaldehyde-TSH, the smallest derivatized carbonyl compound) based on the measured accurate mass (error tolerance 5 ppm), isotope patterns and MS^2 fragmentation. Other compound identification nodes included searching ChemSpider, searching mzCloud spectral library, and searching an in-house mass list containing all the target carbonyl compounds. Since the derivatised carbonyl compounds were, expectedly, not found in neither chemical (ChemSpider) nor spectral databases (mzCloud), the Predicted Compositions node was set as the first data source for compound annotation, followed by the MassList Search node, ChemSpider Search node, and finally mzCloud Search node. Afterwards, the annotated compounds were processed in a Compound Class Scoring node where they were scored against a set of fragment ions originating from the TSH moiety and present in the fragmentation scans of the derivatised carbonyl compounds. The three TSH signature fragments had a m/z (in positive mode) of 139.0212, 157.0318, and 155.0161. This

Compound Class Scoring node annotated the centroids in the fragmentation scans of the detected compounds with the matching TSH fragments and provided a class coverage score with the percent coverage of the three TSH fragments. A minimal score of 33% (1 out of 3 fragments) was required to assign hits as carbonyl compounds. Other compound scoring nodes included the spectral distance node which calculated the spectral similarity score (between theoretical and measured isotope patterns) for the compound annotations, and the application of the mzLogic node (algorithm for spectral annotation) to score explanations from the ChemSpider node and the Search Mass List node.

The following post-processing nodes were used to provide additional information about the detected compounds in addition to the processing nodes described above: Differential Analysis node was used to run statistics for differential analysis (fold changes and ratios) across selected samples (e.g., before and after ozonation). Scripting nodes were run using R scripts which were either previously developed by Thermo Scientific and slightly customized (calculation of C:O ratio and C:H ratio for the identified molecular formulae) or developed internally specifically for the analysis of carbonyl compounds (subtraction of the TSH moiety). The latter was used to automatically generate the molecular formulas of carbonyl compounds (in the non-derivatised form) in Compound Discoverer by subtracting the TSH moiety from the predicted composition of the derivatised compound. The subtracted elemental composition was defined as $C_{(7 \times n)}H_{(8 \times n)}N_{(2 \times n)}S_{(1 \times n)}O_{(1 \times n)}$, with n = number of derivatisations in the same compound (n= 1 or 2 for monocarbonyl and dicarbonyl compounds, respectively).

Further processing and filtering of the data was performed by focusing on compounds with peak areas at least two-fold higher in an ozonated sample than a non-ozonated sample. The analysis outputs in Compound Discoverer were exported into excel files. After further processing of the data (see criteria below), the occurrence of the identified non-target carbonyl compounds in the different samples was determined. The software Trace Finder EFS 4.1 (Thermo Scientific, Germany) was used to extract the response ratios of the identified carbonyl

compounds (peak area of the compound relative to the peak area of the internal standard benzaldehyde-d_6-TSH in the same sample) from all the MS data files. The output was exported into excel files. The resulting excel files from Compound Discoverer and Trace Finder were further processed in R.

In summary, the following criteria were applied for the screening of unknown carbonyl compounds:

i) accurate mass within a mass error of 5 ppm,

ii) matching of isotope patterns based on S and N atoms originating from TSH and potentially from the carbonyl compounds

iii) presence of at least one of the TSH signature fragments in the MS^2 spectrum of the candidate compound

iv) peak areas absent or at least five-fold lower in the derivatised blanks than in the derivatised samples

v) peak areas at least two-fold higher in ozonated than in non-ozonated samples

The combination of these criteria allowed the identification of carbonyl compounds with low risk of false positive results. The list of identified carbonyl compounds (confidence levels 2 and 3) was processed further to achieve higher levels of identification confidence (Schymanski et al. 2014). Structural elucidation of selected carbonyl compounds was carried out by integrating the data regarding: MS^2 fragments and retention time, formation trends upon ozonation, influence of ozonation conditions (formation in the presence/absence of •OH), ozone and •OH reaction kinetics with the considered compound, and control experiments with model carbonyl compounds and ozonated DOM model compounds. Details and results of this structure elucidation approach are presented in Houska et al. (in press).

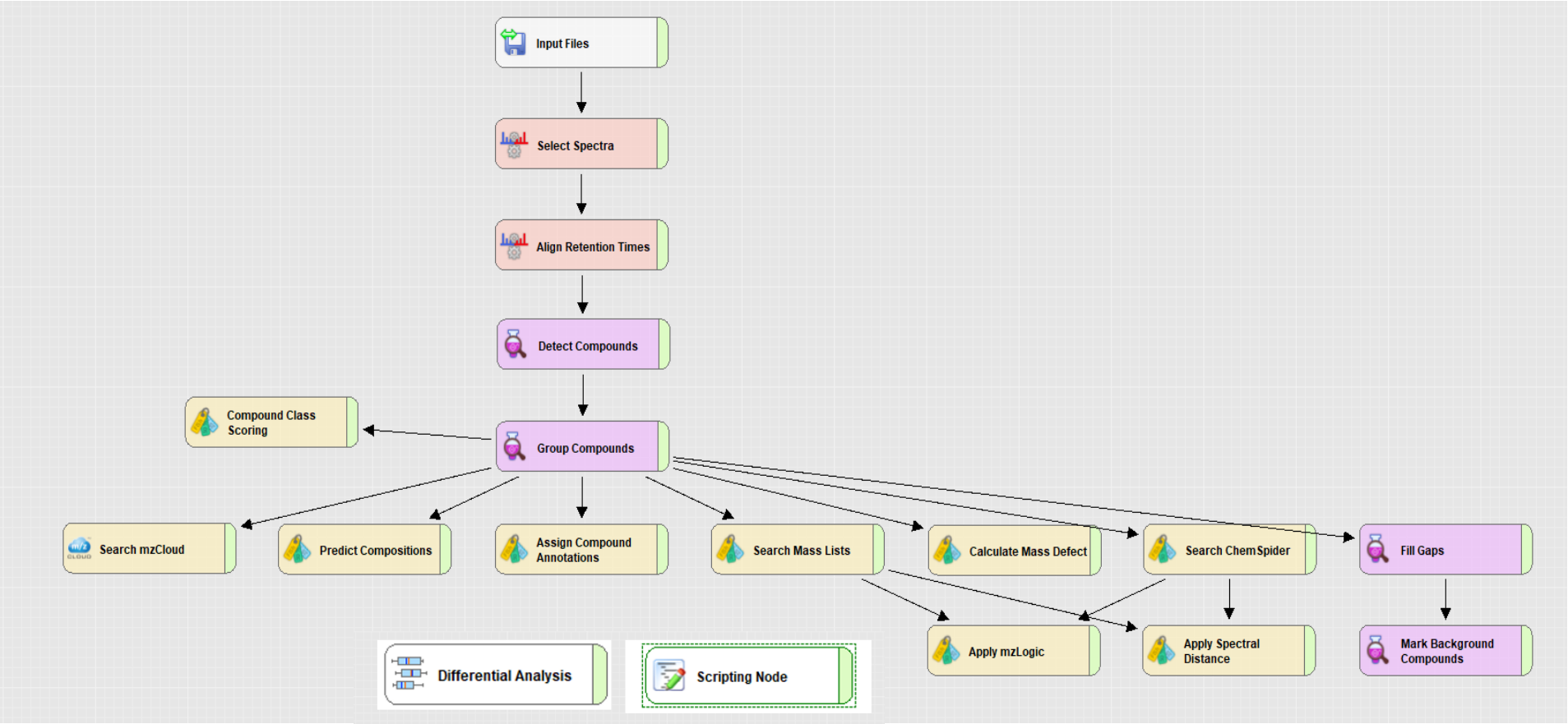

Figure S8. Non-target screening workflow used in Compound Discoverer 3.2 for the identification of unknown carbonyl compounds.

Section S7. Formation of target carbonyl compounds during ozonation

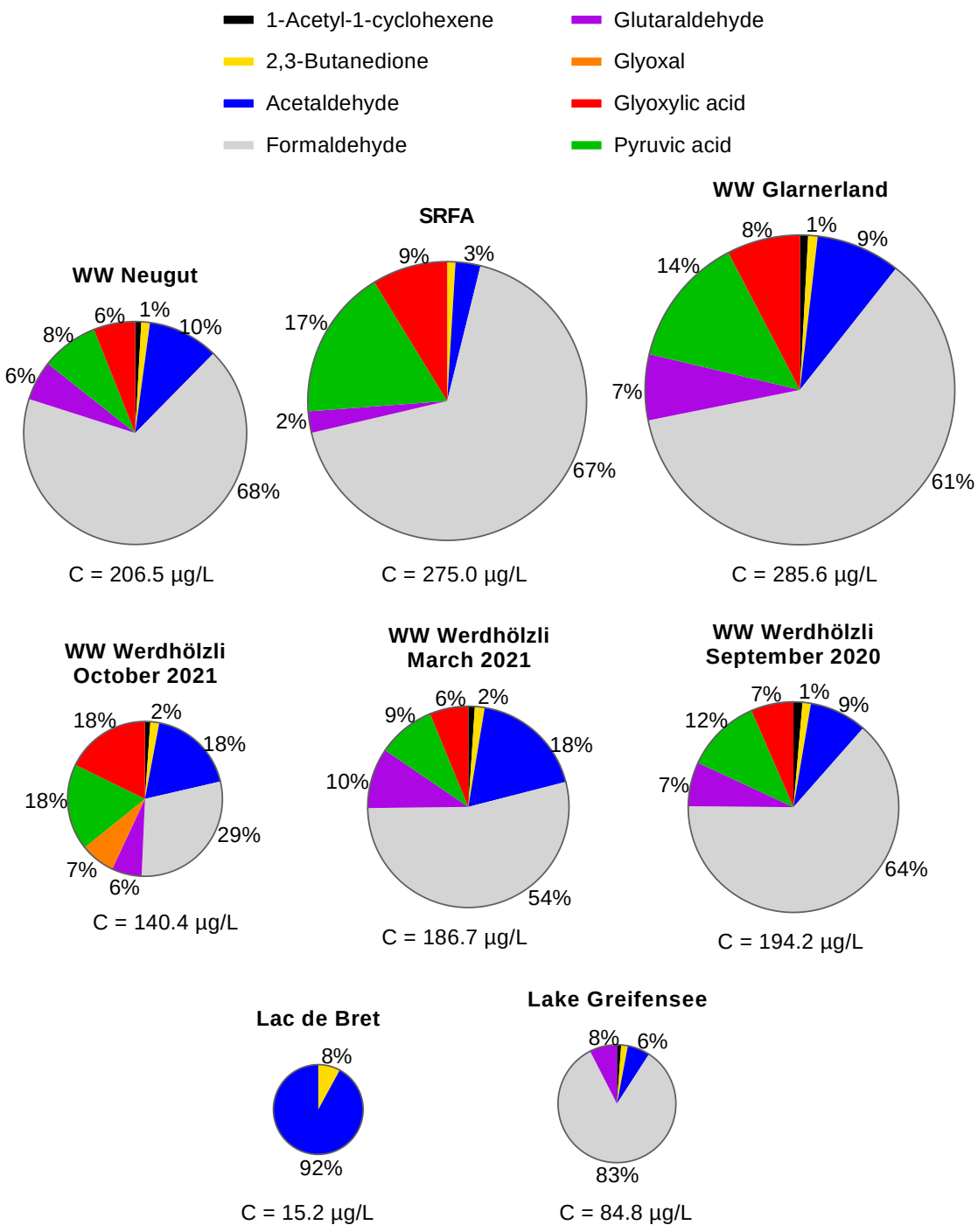

Figure S9. Fractional distribution of the quantified target carbonyl compounds based on mass concentrations in ozonated water and wastewater (WW) samples at a specific ozone dose of 0.5 mgO_3/mgC. Percentage values correspond to the mass concentrations of individual target carbonyl compounds relative to the sum of 8 target carbonyl compounds concentrations detected in the ozonated samples. The size of pie charts is proportional to the sum of the carbonyl mass concentrations in the different samples (C in µg/L).

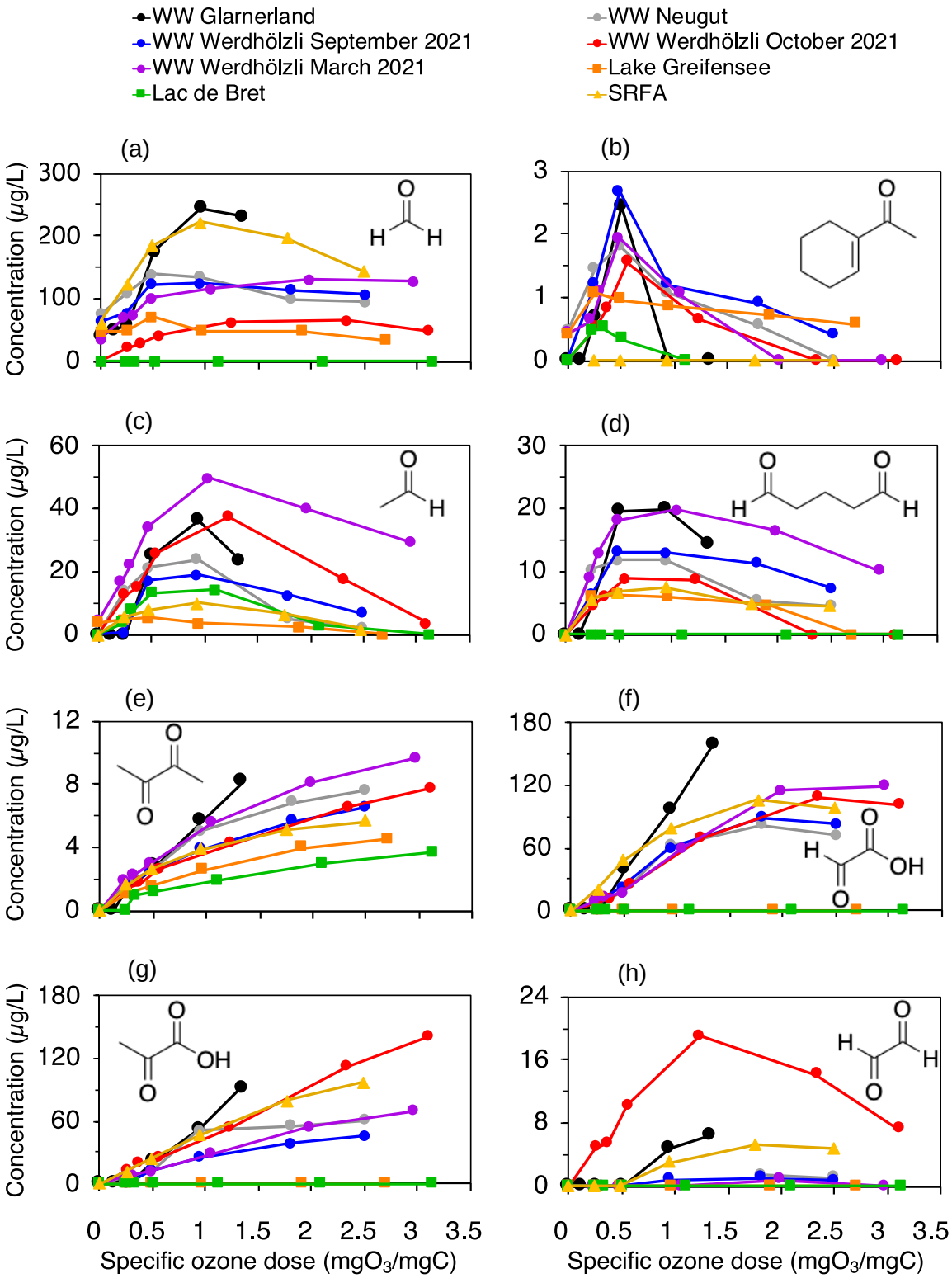

Figure S10. Mass concentrations of 8 target carbonyl compounds as a function of the specific ozone doses in the different water types (see legend); (a) formaldehyde, (b) 1-acetyl-1-cyclohexene, (c) acetaldehyde, (d) glutaraldehyde, (e) 2,3-butanedione, (f) glyoxylic acid, (g) pyruvic acid, and (h) glyoxal. The structure of each compound is shown in the corresponding panel.

Figure S11 and S12 show the concentrations in μM and the DOC concentration-normalised molar concentration (in μM/mgC), respectively, of the eight carbonyl compounds and their sum (total carbonyl compounds concentrations) before and after ozonation at various specific ozone doses from 0.2 to 2 mgO_3/mgC. Before ozonation, formaldehyde and in some samples acetaldehyde were detected above LOQs. During ozonation, the concentration of carbonyl compounds increased at different magnitudes in the different water types. The highest total carbonyl compound molar concentrations were measured in SRFA (9.4 μM) and WW Glarnerland (11.2 μM) at 0-9-1.2 mgO_3/mgC. In wastewater samples from Werdhölzli WWTP, the concentrations were in a comparable range for WW Werdhölzli March 2021 and WW Werdhölzli September 2020, while they were slightly lower in WW Werdhölzli October 2021. In comparison to SRFA-containing water and wastewater samples, the formation of carbonyl compounds was lower in lake waters, with highest total molar concentrations of 0.34 and 1.79 μM in Lac de Bret and Lake Greifensee samples, respectively, for a specific ozone dose of around 1 mgO_3/mgC (Figure S11). The lower formation of carbonyl compounds in lake waters compared to SRFA-containing water and wastewaters remained after normalizing the concentrations with respect to DOC concentrations (Figure S12). The fractional distribution of carbonyl compounds in samples ozonated with a specific ozone dose of 0.5 mgO_3/mgC (close to the doses applied in practice) is shown in Figure 4 in the main text based on molar concentrations and in Figure S9 based on mass concentrations.

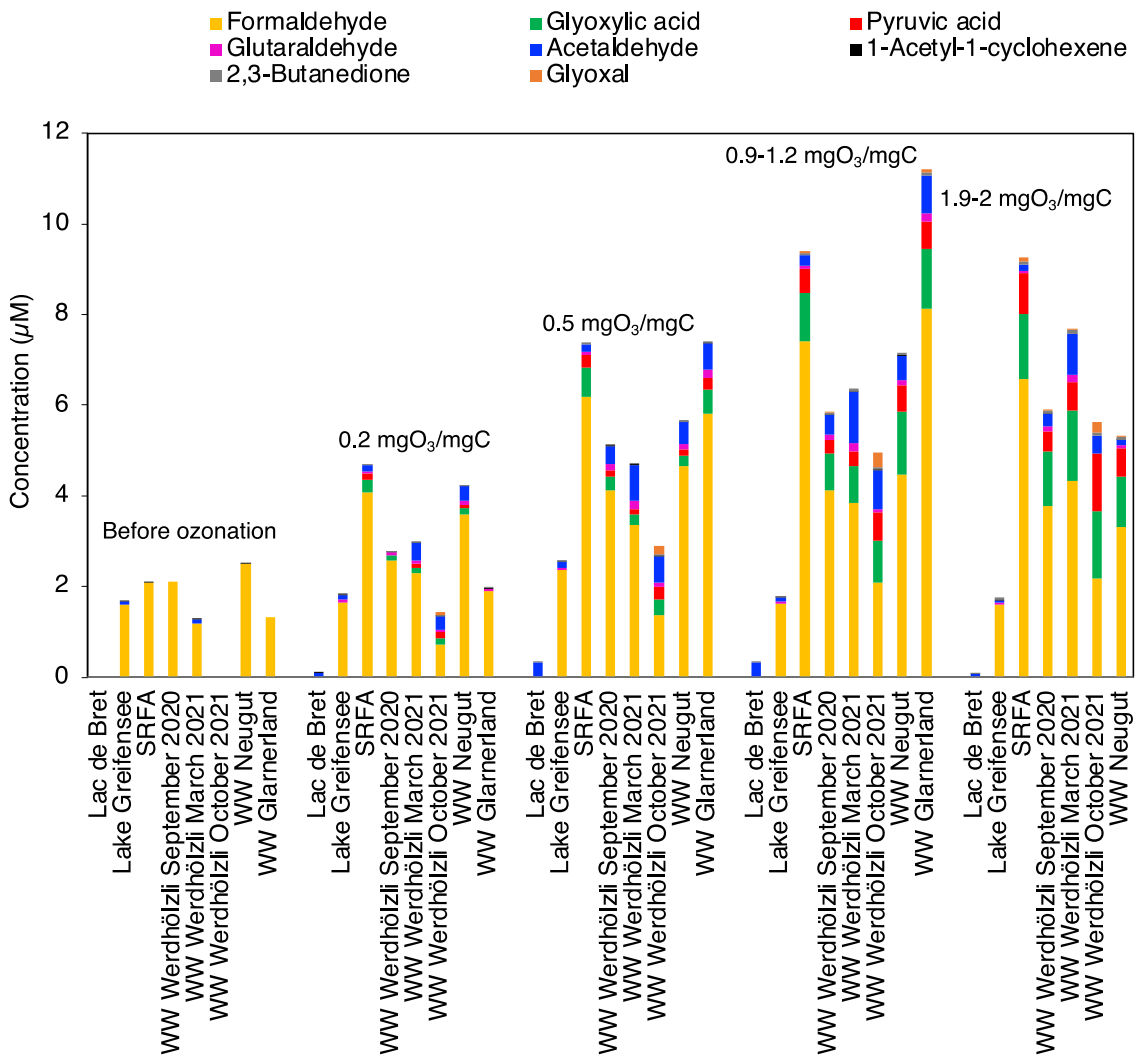

Figure S11. Molar concentrations of the target carbonyl compounds in the different water matrices as functions of increasing specific ozone doses (laboratory-scale ozonation); T: 22 °C, pH (cf. Table S2).

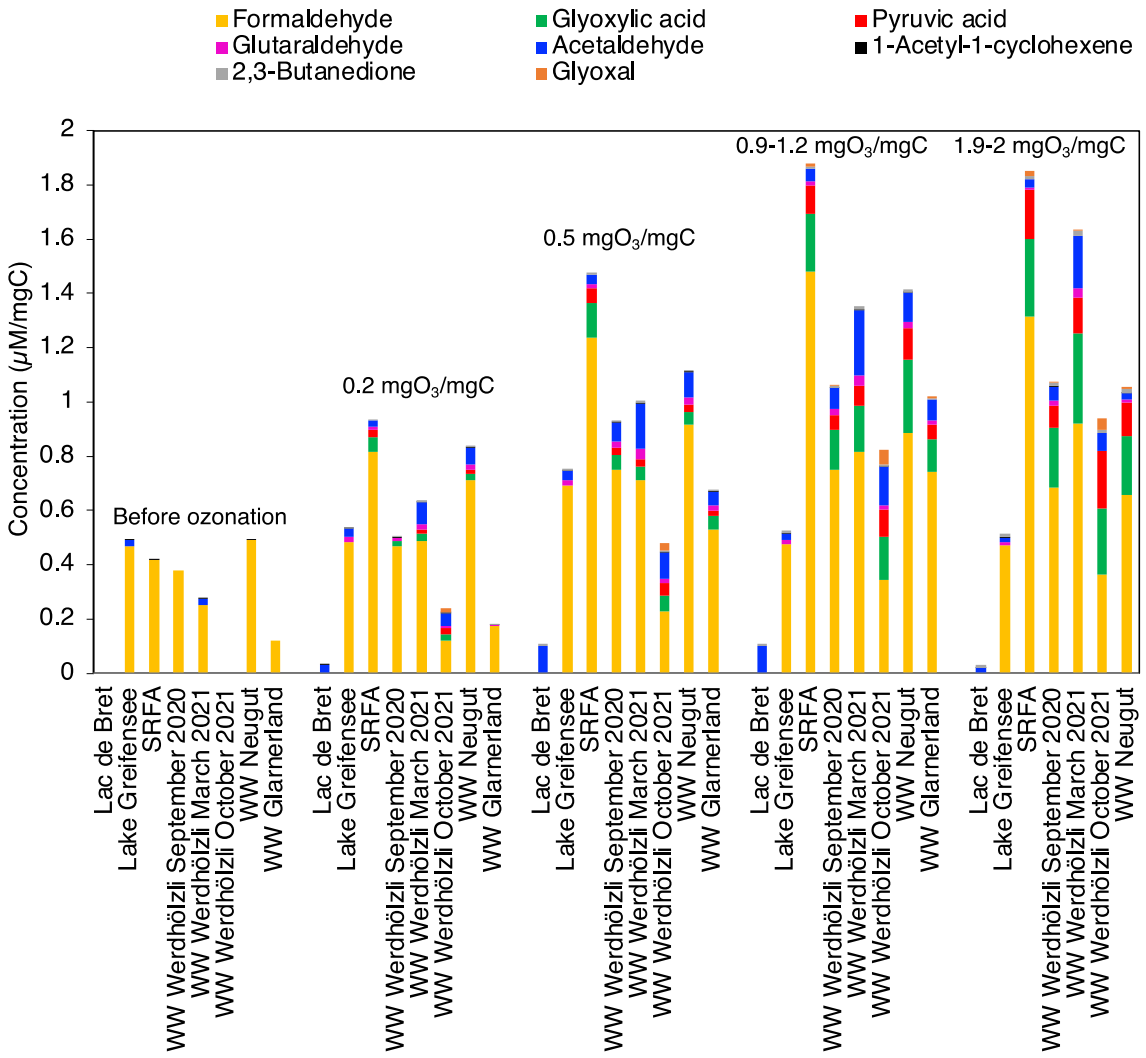

Figure S12. DOC concentration-normalised molar concentrations of carbonyl compounds in the different water matrices as functions of increasing specific ozone doses (laboratory-scale ozonation); T: 22 °C, pH (cf. Table S2).

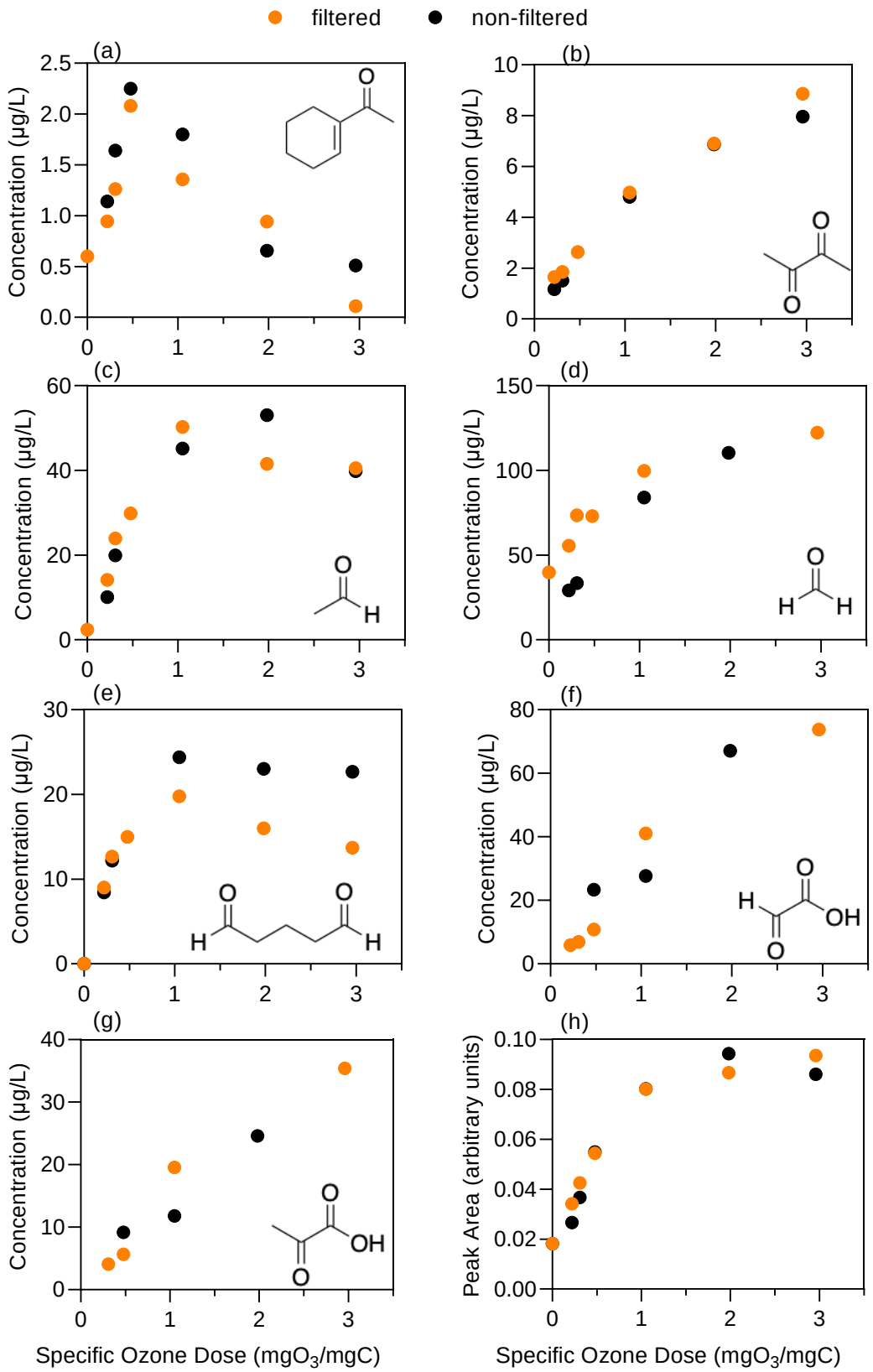

Figure S13. Laboratory ozonation at various specific ozone doses of secondary wastewater effluent (Werdhölzli, March 2021) with and without filtration (0.45 µm), pH = 7.80, T = 22° C. (a) 1-acetyl-1-cyclohexene, (b) 2,3-butanedione, (c) acetaldehyde, (d) formaldehyde, (e) glutaraldehyde, (f) glyoxylic acid, (g) pyruvic acid , and (h) non-target carbonyl compounds based on compounds identified in Houska et al. (in press).

Section S8. Fate of carbonyl compounds during full-scale wastewater treatment

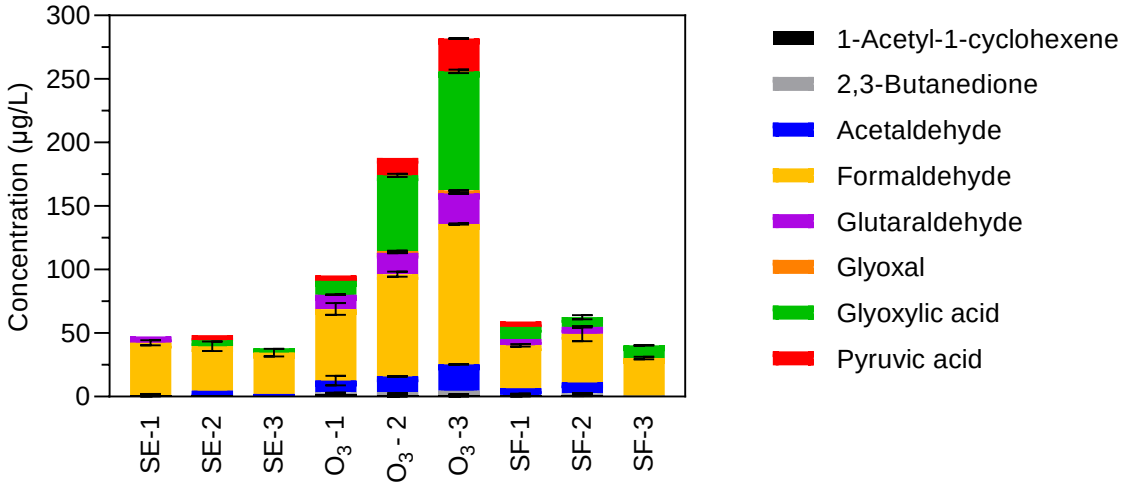

Figure S14. Fate of carbonyl compounds during full-scale wastewater treatment: Concentrations of target carbonyl compounds during different treatment steps. Samples SE-1, SE-2- and SE-3 represent secondary wastewater effluent collected at different times. O_3-1, O_3-2 and O_3-3 correspond to the plant-ozonated samples at specific ozone doses 0.27, 0.42, and 0.90 mgO_3/mgC, respectively. SF-1, SF-2, and SF-3 correspond to the sand filtration effluents from the water packages originating from O_3-1, O_3-2, and O_3-3, respectively. pH = 6.98-6.99, T = 15.4 – 15.7 °C (in the ozone reactor). Error bars represent the range of carbonyl concentrations based on two replicates.

References

Bader, H. and Hoigné, J. (1981) Determination of ozone in water by the indigo method. Water Research 15(4), 449-456.

Houska, J., Manasfi, T., Gebhardt, I. and von Gunten, U. (in press) Ozonation of lake water and wastewater: Identification of carbonous and nitrogenous carbonyl-containing oxidation byproducts by non-target screening. Water Research. https://doi.org/10.1016/j.watres.2022.119484

Marron, E.L., Prasse, C., Buren, J.V. and Sedlak, D.L. (2020) Formation and Fate of Carbonyls in Potable Water Reuse Systems. Environ Sci Technol 54(17), 10895-10903.

Siegel, D., Meinema, A.C., Permentier, H., Hopfgartner, G. and Bischoff, R. (2014) Integrated quantification and identification of aldehydes and ketones in biological samples. Anal Chem 86(10), 5089-5100.

von Sonntag, C. and von Gunten, U. (2012) Chemistry Of Ozone In Water and Wastewater Treatment. IWA Publishing.