Title: Report on HRMS fingerprinting tool

Deliverable 2.5

WP 2 SENSORS – Innovative monitoring systems



Author: Francisco Zafra-Navarro, María José Farré, Sofia Semitsoglou Tsiapou, Luigi Marino, Paolo Roccaro, Mira Petrovič, Wolfgang Gernjak.

Date: 30/11/2025

Funded by the European Union

GRANT AGREEMENT NUMBER	**101081728**		
FULL TITLE / ACRONYM	**INNOVATIVE TOOLS TO CONTROL ORGANIC MATTER AND DISINFECTION BYPRODUCTS IN DRINKING WATER (intoDBP)**		
START DATE	01 December 2022	**DURATION**	48 months
END DATE	30 November 2026		
PROJECT URL	https://intodbp.eu/		
DELIVERABLE	**D2.5: Report on HRMS fingerprinting tool**		
WORK PACKAGE	**WP2**: SENSORS - Innovative monitoring systems		
CONTRACTUAL DATE OF DELIVERY	**30 November 2025**		
ACTUAL DATE OF DELIVERY	**25 August 2026**		
NATURE	Report	**DISSEMINATION LEVEL**	PUBLIC
LEAD BENEFICIARY	**ICRA**		
RESPONSIBLE AUTHOR	Francisco Zafra-Navarro, María José Farré, Sofia Semitsoglou Tsiapou, Luigi Marino, Paolo Roccaro, Mira Petrovič, Wolfgang Gernjak.		
CONTRIBUTIONS FROM			
ABSTRACT	Disinfection byproducts (DBPs) formed during chemical disinfection remain a major challenge for drinking water treatment. Although regulations target a limited set of DBPs, chlorination of dissolved organic matter (DOM) produces a much broader and largely unregulated pool of compounds. This study assessed the performance of a full-scale drinking water treatment plant of Barcelona case study using an integrated analytical approach combining targeted DBP quantification by GC–MS, AOX measurement, and molecular-level DOM characterization by high-resolution mass spectrometry. Despite achieving ~60% DOC removal, molecular fingerprinting revealed a highly selective removal pattern: aromatic compounds were efficiently eliminated, whereas aliphatic and oxygen-containing unsaturated compounds persisted and showed strong associations with DBP formation. These findings demonstrate that DOM composition, rather than bulk DOC concentration, drives DBP formation potential and highlight the need for treatment strategies that specifically target aliphatic and unsaturated DOM fractions to improve drinking water safety.		

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

© intoDBP Consortium, 2022
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both. Reproduction is authorised provided the source is acknowledged.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS 4

LIST OF FIGURES 5

LIST OF TABLES 5

EXECUTIVE SUMMARY 6

1.0 INTRODUCTION AND RATIONALE 7

2.0 METHODOLOGY 9

2.1. Drinking Water Treatment Plant (DWTP) and Sampling 9
2.2. Formation Potential (FP) Experiments 10
2.3 DOM Characterization via LC-HRMS 10
2.3.1 SPE Extraction 10
2.3.2 LC-HRMS Instrumental Analysis 10
2.3.3 Data Processing and Formula Assignment 10
2.4 Targeted DBP Analysis (GC-MS) 10
2.5 AOX Analysis 11
2.6 UV absorbance measurement 11
2.7 Fluorescence measurement 11

3.0 RESULTS AND DISCUSSION 12

3.1 DOM Characterization Across the DWTP 12
3.1.1 Overall DOM Changes 12
3.1.2 Chlorinated Dissolved Organic Matter (Cl-DOM) 13
3.1.3 Brominated Dissolved Organic Matter (Br-DOM) 13
3.2 Targeted DBPs and AOX analysis 14

3.3 Correlations Between DOM and DBPs 16
3.4 Fluorescence and UV absorbance analysis and their correlations with HRMS analysis 17
3.4.1 UV-Vis and Fluorescence measurements of the DWTP 17
3.4.2 Correlation between UV-Vis and Fluorescence with HRMS molecular features 18
4.0 CONCLUSIONS 21
5.0 FUTURE WORK 22
6.0 REFERENCES 23
7.0 ANNEX 25

LIST OF ABBREVIATIONS

AOBr – Adsorbable Organic Bromide
AOCl – Adsorbable Organic Chloride
AOI – Adsorbable Organic Iodide
AOX – Adsorbable Organic Halides
APHA – American Public Health Association
Br-DOM – Brominated Dissolved Organic Matter
DBPs – Disinfection Byproducts
DOM – Dissolved Organic Matter
DOC – Dissolved Organic Carbon
DWTP – Drinking Water Treatment Plant
ESI – Electrospray Ionization
FP – Formation Potential
GAC – Granular Activated Carbon
GC-MS – Gas Chromatography–Mass Spectrometry
HAAs – Haloacetic Acids
HANs – Haloacetonitriles
HRT – Hydraulic Retention Time
HRMS – High-Resolution Mass Spectrometry
KMD – Kendrick Mass Defect
LC-HRMS – Liquid Chromatography–High-Resolution Mass Spectrometry

MTBE – Methyl tert-butyl ether

PPL – Bond Elut PPL cartridge

SRNOM – Suwannee River Natural Organic Matter

THMs – Trihalomethanes

UHPLC – Ultra-High-Performance Liquid Chromatography

LIST OF FIGURES

Figure 1. Mean relative changes in the molecular features detected across all samplings along the DWTP (A) and FP samples (B), highlighting depletion and enrichment trends based on features showing at least a tenfold change in abundance compared to the previous treatment stage (A and B). Van Krevelen diagrams showing the distribution of molecular features at each DWTP (C) and FP (D) sampling points, distinguishing halogenated from non-halogenated compounds. 13
Figure 2. Concentration profiles of halogenated species using multiple analytical approaches. (A) Target DBP analysis in actual and FP samples in µg/L; (B) Adsorbable organic halides, chlorine (AOCl), bromine (AOBr) and iodine (AOI) in actual and FP samples in mg/L; (C) Fingerprinting analysis of halogenated features count across van Krevelen regions. Sampling points: R = Raw, C = Clarified, PAC = Post-Activated Carbon filtration, PW = Produced Water, D1 = Distribution system after 24 h, D2 = Distribution system after 48 h and D3 = Distribution system after 72 h. 15
Figure 3. Strongly correlated (ρ > 0.8) DBP precursors in raw water. 16
Figure 4. Response of fluorescence and absorbance-based optical indicators to DOM transformation during drinking water treatment. 18
Figure 5. Van Krevelen diagram representation of the features associated (ρ > 0.5) with the maximum intensity at the region V (humic-like compounds). 19
Figure S1. Example of a fluorescence spectrum divided into five characteristic regions, tyrosine-like (Region I), tryptophan-like (Region II), fulvic-like (Region III), soluble microbial product-like (Region IV), and humic-like (Region V) compounds. Adapted from Chen et al. (2003) 25

LIST OF TABLES

Table 1. Drinking water treatment plant conditions during the samplings. 9
Table S1. % Recovery of DOM extraction 25
Table S2. Criteria for formula assignment according to their regions in the Van Krevelen diagram 26
Table S3. First 10 molecular formulae presented in at least 4 of 6 samples and employed for all data analysis and representations 27
Table S4. Number of high correlated features and classified according to their region in the van Krevelen presented in each sample 28
Table S5. Codification of Fluorescence and Absorbance parameters measured 29
Table S6. Fluorescence and Adsorbance measurements 30

EXECUTIVE SUMMARY

Disinfection byproducts (DBPs) are unintended chemical compounds formed when disinfectants, primarily chlorine, react with dissolved organic matter (DOM) in drinking water. While regulatory frameworks typically monitor only a narrow group of DBPs, such as trihalomethanes (THMs) and haloacetic acids (HAAs), hundreds of other halogenated compounds remain unregulated. Conventional water treatment processes remove DOM in a largely non-selective manner and often fail to target the most reactive or persistent fractions responsible of DBP formation. Therefore, understanding the molecular characteristics of DOM and identifying specific precursor fractions is essential to optimizing treatment operations and reducing the formation of harmful DBPs.

This study highlights the critical **role of high-resolution mass spectrometry (HRMS) in characterizing DOM at the molecular level and linking specific fractions to DBP formation pathways. The resulting molecular-level information provides a strong foundation for integrating optical measurements, such as fluorescence and absorbance, into robust DBP monitoring and prediction strategies.** HRMS enables comprehensive non-targeted profiling of complex DOM mixtures, capturing both reactive and recalcitrant fractions. By combining HRMS with targeted DBP analysis via GC-MS, measurements of adsorbable organic halides (AOX), and optical characterization through fluorescence and absorbance, this study provides an integrated assessment of DOM transformation, halogenated byproduct formation, and DBP prediction in a full-scale drinking water treatment plant (DWTP) in Barcelona, serving more than 4.5 million inhabitants. Results demonstrated selective DOM removal: aromatic and condensed aromatic fractions were efficiently eliminated, whereas aliphatic and low-oxygen unsaturated fractions persisted and showed strong correlations with DBP formation. Chlorinated DOM was formed primarily during pre-oxidation and partially removed by GAC filters, whereas brominated DOM appeared after final chlorination following the incorporation of desalinated water. Targeted DBPs accounted for less than 31% of total AOX, indicating the presence of substantial fraction of uncharacterized halogenated organic matter. HRMS fingerprinting identified 97 DOM features associated with DBP formation, predominantly within low-oxygen unsaturated and aliphatic regions. Correlations analyses between optical indices and HRMS features revealed that fluorescence signals, including the maximum intensity peak in the humic-like region (region V), were associated with molecular features located in the low-oxygen unsaturated region of the van Krevelen diagram, which were previously linked to DBP formation. The next step will be to configure the sensors by integrating the knowledge generated in this work package with the sensor framework established in work package 2, enabling online monitoring and prediction of DBPs in real time. Overall, HRMS provides a detailed molecular-level view, emerging as a **valuable tool for evidence-based water quality management and public health protection**, particularly in conventional DWTPs where technological capacity may be limited.

1.0 Introduction and rationale

Disinfection byproducts (DBPs) are unintended chemical compounds formed during the disinfection of drinking water, primarily resulting from reactions between disinfectants, such as chlorine, and dissolved organic matter (DOM) or other precursors present in source water[1]. While regulatory frameworks, including the European Drinking Water Directive (2020/2184)[2], establish limits for specific, commonly monitored DBPs such as trihalomethanes (THMs) and haloacetic acids (HAAs), a large portion of DBPs remains unregulated. To date, more than 700 distinct DBPs have been identified, encompassing a wide variety of chemical classes including haloacetonitriles, haloketones, haloaldehydes, halonitromethanes, and haloacetamides, many of which exhibit diverse physicochemical properties and toxicological profiles[3].

The presence of DBPs in drinking water raises significant concerns for public health, as many of these compounds are associated with genotoxic, mutagenic, and potentially carcinogenic effects[4–9]. Conventional water treatment strategies focus on the removal of DOM prior to disinfection to limit DBP formation; however, these processes are generally non-selective and often fail to target the specific DOM fractions that are most reactive with disinfectants or those that persist through treatment[10]. Consequently, certain highly reactive or recalcitrant DOM fractions can survive treatment and contribute disproportionately to DBP formation.

A detailed understanding of which molecular fractions of DOM act as DBP precursors and whether these fractions overlap or differ between reactive and persistent components, is therefore essential. Such knowledge can inform the optimization of treatment strategies, enabling selective removal of critical DOM fractions, minimizing the formation of harmful DBPs, and ultimately enhancing the safety and quality of drinking water delivered to consumers.

High-resolution mass spectrometry (HRMS) has become an essential tool for characterizing DOM at the molecular level and identifying both known and previously unrecognized DBPs[11,12]. In contrast to targeted methods, such as GC-MS for volatile DBPs, HRMS enables comprehensive non-targeted profiling of complex organic matter, capturing a wide range of DBP precursors[13,14]. When integrated with another techniques, including targeted DBP analysis and adsorbable organic halide (AOX) measurements, HRMS provides a holistic view of DOM transformation and DBP formation throughout the drinking water treatment process. This approach is particularly valuable for:

- Identifying DOM fractions that are highly reactive with disinfectants.
- Detecting recalcitrant fractions that persist through treatment.
- Establishing correlations between DOM molecular features and DBP formation potential.
- Informing targeted treatment strategies that minimize DBP precursors while maximizing water quality.

By generating detailed molecular “fingerprints” of DOM, HRMS enables the identification of reactive and recalcitrant fractions, provides insights into their transformation during

treatment, and helps link specific DOM components to the formation of particular DBPs. When integrated with conventional approaches such as targeted DBP analysis (e.g., GC-MS) and bulk halogen measurements (e.g., AOX), HRMS offers a comprehensive framework to evaluate treatment performance, assess risks, and guide the development of targeted strategies for DBP mitigation.

Despite the valuable insights provided by this multi-analytical approach, the high cost and time requirements associated with these measurements limit their direct implementation for online monitoring in full-scale DWTPs. Therefore, faster and more cost-effective techniques, such as fluorescence and absorbance measurements, are increasingly being explored for real-time monitoring applications. By correlating optical signals with the molecular information obtained from the multi-analytical approach, relevant optical indices and predictive metrics can be established to support DBP formation monitoring and prediction.

This report (Deliverable D2.5) focuses on the application of an integrated analytical framework combining HRMS-based DOM fingerprinting, target GC-MS analysis of DBPs, AOX quantification, and optical measurements to assess drinking water treatment performance. This approach provides new insights into DBP formation pathways, supports improved water quality management, and contributes to evidence-based strategies for public health protection.

2.0 Methodology

2.1. Drinking Water Treatment Plant (DWTP) and Sampling

The studied DWTP is located in Barcelona, Spain, serving >4.5 million people with an average flow of 2.85 m^3/s. Source water originates from interconnected reservoirs (Sau-Susqueda-Pasteral)[15]. The treatment train includes:

- Primary disinfection (chlorine dioxide and sodium hypochlorite)
- Coagulation-flocculation (poly-aluminum chloride, modified starch) and sedimentation
- Granular activated carbon (GAC) filtration
- Blending with bromide-rich desalinated water
- Secondary disinfection (sodium hypochlorite)

To capture seasonal variability and reduce sampling bias, six sampling campaigns were performed during 2023 and 2024 in three seasons, spring (April 2023), autumn (November 2023), and winter (January 2024), with two sampling events per season.

Conditions during the samplings are shown in **Table 1**. Seven sampling points were selected: Raw water, Clarified, Post-GAC, Produced water, and Distribution at 24h, 48h, and 72h HRT. Routine physicochemical parameters (DOC, pH, chlorine, bromide, nitrate, etc.) were measured at each point.

Table 1. Drinking water treatment plant conditions during the samplings.

Campaign	Sampling	Date	Flow (m^3/s)	%Desalinated water in the tanks	Discharge of Chlorine dioxide (Preoxidation-Clarified samples) (mg/L)	Discharge of Chlorine (Disinfection process – Produced water samples) (mg/L)
1	1	17/04/2023	2.87	13.6	1.00	0.80
1	2	24/04/2023	2.87	13.6	1.00	0.80
2	3	06/11/2023	3.38	9.7	0.45	1.65
2	4	20/11/2023	3.89	10.8	0.28	1.30
3	5	08/01/2024	2.19	14.6	0.25	1.10
3	6	29/01/2024	1.87	9.8	0.45	1.82

2.2. Formation Potential (FP) Experiments

DBP formation potential was assessed following a modified APHA (2005) method[4–7]. Chlorine was added to 2 L glass bottles (no headspace) to achieve 1–3 mg/L residual after 72 h at 25°C in the dark. Post-incubation, residual chlorine was measured, and samples were quenched using ascorbic acid (for DBPs) or sodium sulfite (for AOX and DOM analyses).

2.3 DOM Characterization via LC-HRMS

2.3.1 SPE Extraction

Water samples (2 L) were filtered (0.7 µm GF/F), acidified to pH 2, and extracted using Bond Elut PPL cartridges. Cartridges were conditioned with methanol and 0.1% formic acid. DOM was eluted with 2 mL methanol, dried under nitrogen, and reconstituted in ultrapure water to standardize total carbon at 1000 mg/L. DOC was measured to assess extraction recovery, results are summarized in the Annex **Table S1**.

2.3.2 LC-HRMS Instrumental Analysis

DOM extracts were analyzed using an Orbitrap Exploris 120 coupled to a Vanquish UHPLC system with a Hypersil GOLD™ C18 column (50 × 2.1 mm, 1.9 µm). ESI in negative mode, full scan m/z 200–1000, 120,000 FWHM. 1000 spectra per sample were averaged; methanol blanks were run between samples. Fluoranthene (m/z 201.07825) was used for internal calibration.

2.3.3 Data Processing and Formula Assignment

Mass lists were generated from 4–8 min retention time, exported to Excel, and processed in RStudio (v4.4.0) using MFAssignR[16]. Noise reduction used Kendrick Mass Defect (KMD) plots; peaks >6× noise were assigned formulas (C, H, O, N, S, Cl, Br). Halogenated compounds were validated via isotopic patterns. Only features which are presented at the samples at least in 4 of the 6 sampling events are employed in the Figures, discussion and data treatment; an example of the features employed are summarized in the Annex **Table S2**. Features were visualized using van Krevelen diagrams and classified as aliphatic, aromatic, condensed aromatic, high-oxygen unsaturated, and low-oxygen unsaturated compounds[17], the classification parameters are summarized in the Annex **Table S3**. SRNOM standard was used for quality control, achieving 85% match with reference datasets[17].

2.4 Targeted DBP Analysis (GC-MS)

DBPs were extracted via liquid-liquid salted microextraction, adjusting pH to 3.5, using MTBE and internal standards. Extracts were analyzed using a TSQ Quantum triple quadrupole GC-MS system with a TG-1701MS column. Oven program: 35°C for 5 min → 100°C at 10°C/min → 200°C at 20°C/min. Data were processed with TraceFinder EFS 3.1. THMs, HAAs, and HANs were quantified.

2.5 AOX Analysis

Water samples were acidified (pH 2) and loaded onto activated carbon cartridges using a Mitsubishi TX-3AA adsorption module. After washing, cartridges were combusted at 1000°C, and pyrolysis gases containing halides were collected in an absorption solution. Halides were quantified using a Dionex ICS-5000 ion chromatography system.

2.6 UV absorbance measurement

Before the UV analysis, samples were stored at room temperature overnight. Once the spectrophotometer (Agilent Cary 4000 UV) was switched on, a 30-minute warm-up period was allowed for the lamp to stabilize. The excitation–emission scan was performed from 800 nm to 200 nm. Milli-Q water was used for baseline correction, and the Inner Filter Effect (IFE) was avoided. In cases where the absorbance at 254 nm exceeded 0.3 cm^{-1}, samples were diluted and the IFE correction was applied.

2.7 Fluorescence measurement

Before the fluorescence analysis, samples were stored at room temperature overnight. Once the spectrophotometer (Hitachi Fluorescence Spectrophotometer F-7000) was switched on, a 45-minute warm-up period was allowed for the lamp to stabilize. The excitation range was set from 200 nm to 450 nm with a voltage of 700 V, and the emission range was set from 250 nm to 600 nm. All scans were performed at a speed of 12,000 $nm \cdot min^{-1}$. Once the fluorescence excitation–emission matrix (EEM) is obtained, it is divided into five characteristic regions based on the type of fluorophores associated with each region. Fluorescence signals are subsequently interpreted according to this regional classification, which links each region to specific DOM fluorophore groups: tyrosine-like (Region I), tryptophan-like (Region II), fulvic-like (Region III), soluble microbial product-like (Region IV), and humic-like (Region V) compounds. A schematic representation of the fluorescence spectrum and the corresponding regional classification is provided in **Figure S1**.

3.0 Results and Discussion

All data presented in this section is stored in CORA database and is available under request.

To gain a comprehensive understanding of halogenated organic matter formation during drinking water treatment, three complementary analytical approaches were applied to samples collected along the DWTP and their corresponding formation potential (FP) tests:

1. Non-targeted HRMS fingerprinting of DOM.
2. Targeted analysis of 17 DBPs via GC-MS and the determination of adsorbable organic halides (AOX), differentiating chlorinated (AOCl), brominated (AOBr), and iodinated (AOI) organic matter.
3. Correlations between DOM and DBP formation potential
4. Optical characterization and their correlations with HRMS.

This integrated approach provides insights into DOM transformation, highlights which DOM fractions are reactive or recalcitrant, and identifies fractions persisting through treatment that may contribute to DBP formation, as the monitoring of DBP formation.

3.1 DOM Characterization Across the DWTP

3.1.1 Overall DOM Changes

Van Krevelen diagrams (**Figure 1**) were used to visualize compositional changes in DOM throughout the DWTP. **Figure 1A** shows a general depletion of DOM features, with the greatest reduction in aromatic and condensed aromatic-like compounds, reflecting their effective removal during pre-oxidation and coagulation-flocculation. Conversely, low-oxygen unsaturated-like compounds increased, indicating formation of new molecular features, particularly after GAC filtration. This likely results from oxidative breakdown of high-molecular-weight DOM.

FP samples (**Figure 1B**) exhibited more pronounced reductions, especially in unsaturated regions, suggesting controlled removal of DBP precursors. DOC measurements confirmed overall removal of 41.6–75.2 %, highlighting selective targeting of aromatic structures while aliphatic and unsaturated compounds largely persisted. Removal efficiencies across van Krevelen regions ranged from 0–0.8 % (aliphatics), 51–59 % (aromatics), 32–58 % (condensed aromatics), 13–17 % (high-oxygen unsaturated), and 5–11 % (low-oxygen unsaturated).

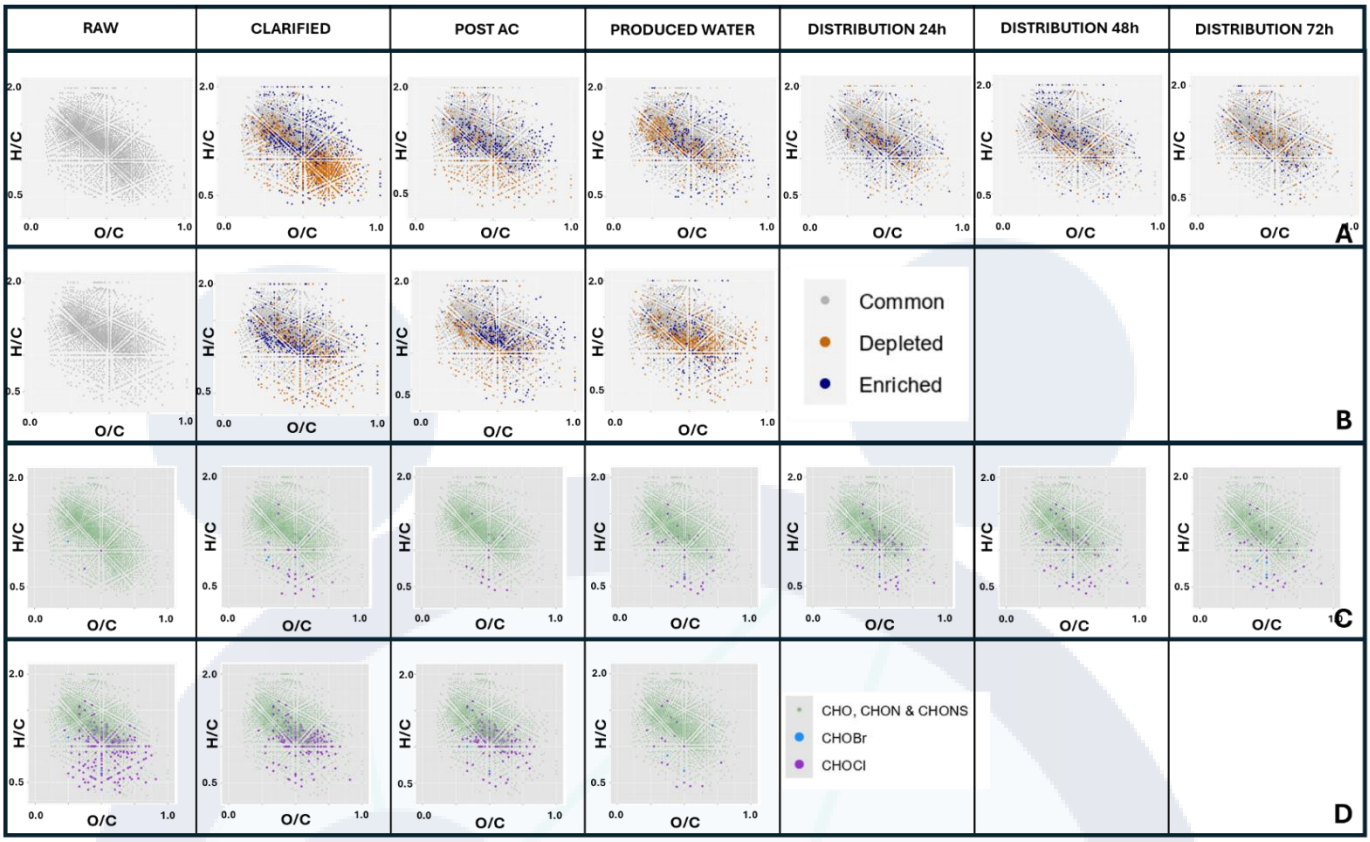


Figure 1. Mean relative changes in the molecular features detected across all samplings along the DWTP (A) and FP samples (B), highlighting depletion and enrichment trends based on features showing at least a tenfold change in abundance compared to the previous treatment stage (A and B). Van Krevelen diagrams showing the distribution of molecular features at each DWTP (C) and FP (D) sampling points, distinguishing halogenated from non-halogenated compounds.

3.1.2 Chlorinated Dissolved Organic Matter (Cl-DOM)

Cl-DOM only represents a maximum number of 4.7 % of the detected features. As is shown in **Figure 1C** it first appeared after pre-oxidation and increased after main disinfection. Condensed aromatic and high-oxygen unsaturated regions contained the highest Cl-DOM abundance, whereas aliphatic and aromatic regions were less populated. The persistence of Cl-DOM in unsaturated regions (<17 % removal) is concerning due to their high reactivity and DBP formation potential. Aromatic halogenated species formed primarily during pre-oxidation, benefiting from faster reaction kinetics with chlorine, and were partially removed by GAC filtration, though removal was incomplete, as could also see in the FP samples (**Figure 1D**).

3.1.3 Brominated Dissolved Organic Matter (Br-DOM)

Br-DOM fraction in number of features terms represents less than the 0.3 % of the total but remains toxicologically relevant. Br-DOM was absent in the aliphatic region and predominantly appeared in the condensed aromatic-like region. Two intensity peaks were observed (**Figure 1C**): after pre-oxidation and after main chlorination, the latter

amplified by the addition of bromide-rich desalinated water (**Figure 1D**). Activated carbon filtration partially mitigated Br-DOM formation. Peak intensities correlated with sodium hypochlorite doses and proportion of desalinated water, confirming the role of bromide availability and disinfectant concentration in Br-DOM, also showed in the increase of Br-DOM in produced water FP sample (**Figure 1D**).

3.2 Targeted DBPs and AOX analysis

Targeted DBP analysis (**Figure 2A**) detected 12 out of 17 DBPs (4 THMs, 3 HANs, 5 HAAs). Produced water contained:

- THMs: 13–42 µg/L
- HANs: 1–4 µg/L
- HAAs: 6–13 µg/L

FP samples exhibited higher concentrations:

- THMs: 51–64 µg/L
- HANs: 7–9 µg/L
- HAAs: 56–63 µg/L

DBP mitigation ranged from 34–75 % for THMs, 69–83 % for HANs, and 79–89 % for HAAs, with THMs showing the least reduction.

AOX measurements (**Figure 2B**) revealed that AOCl increased during distribution, consistent with post-disinfection THM formation, while AOBr peaked in produced water due to bromide input. Targeted DBPs accounted for <31 % of AOX in DWTP samples and <51 % in FP samples, highlighting the presence of uncharacterized halogenated organic matter.

HRMS analysis (**Figure 2C**) showed biphasic halogenated feature formation: an initial increase after pre-oxidation, partial removal during GAC filtration, and a second increase after final chlorination. Most regions exhibited decreasing halogenated feature counts post-disinfection, except aliphatic-like compounds. In FP samples, low- and high-oxygen unsaturated regions retained the highest number of halogenated features, reflecting their persistence and reactivity as DBP precursors.

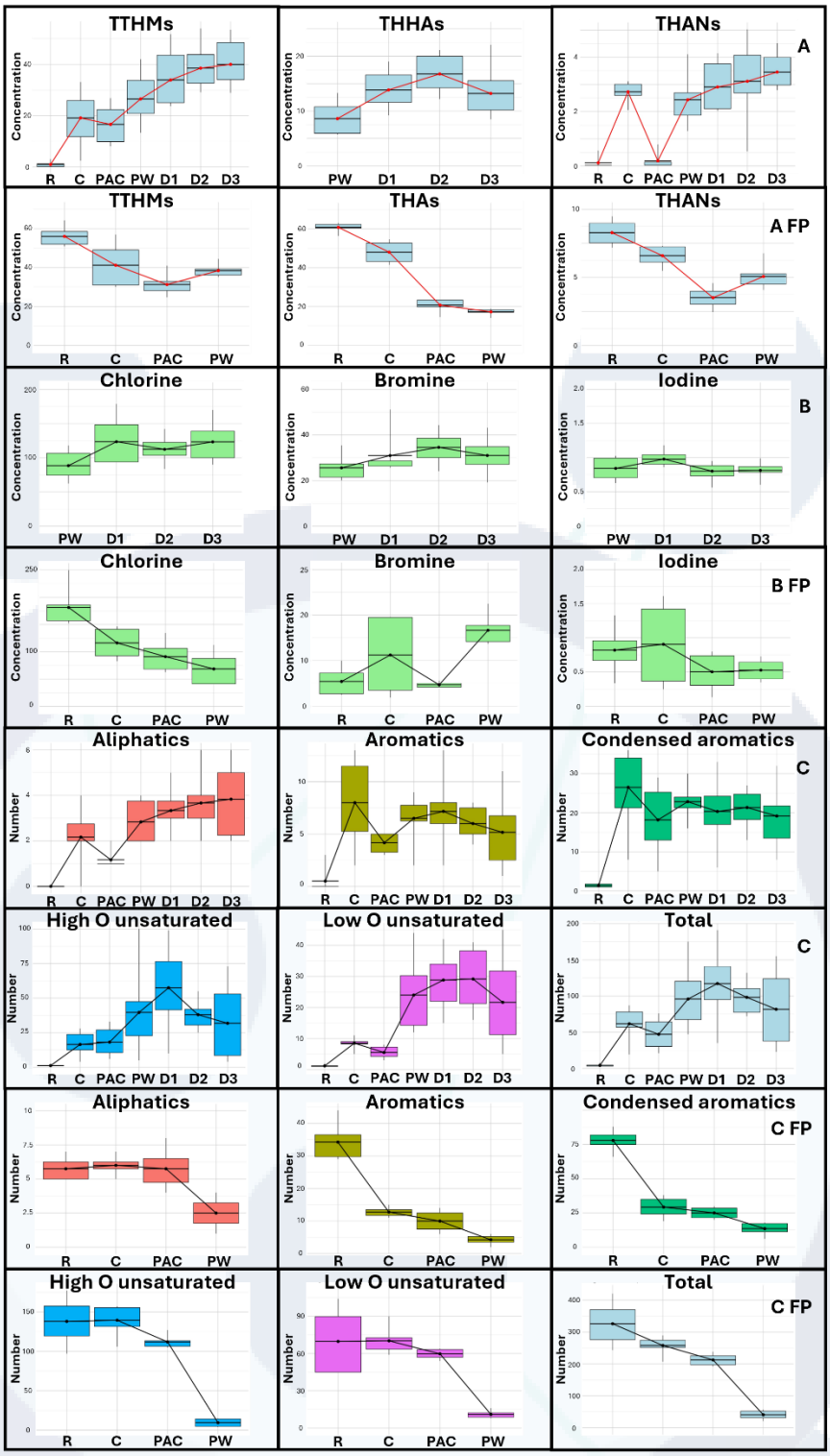


Figure 2. Concentration profiles of halogenated species using multiple analytical approaches. (A) Target DBP analysis in actual and FP samples in µg/L; (B) Adsorbable organic halides, chlorine (AOCl), bromine (AOBr) and iodine (AOI) in actual and FP samples in mg/L; (C) Fingerprinting analysis of halogenated features count across van Krevelen regions. Sampling points: R = Raw, C = Clarified, PAC = Post-Activated Carbon filtration, PW = Produced Water, D1 = Distribution system after 24 h, D2 = Distribution system after 48 h and D3 = Distribution system after 72 h.

3.3 Correlations Between DOM and DBPs

While the DWTP removed up to 75.2 % of DOC, HRMS fingerprints revealed selective removal: aromatic and condensed aromatic regions were largely eliminated, whereas aliphatic and unsaturated regions persisted, contributing to DBP formation.

To investigate the relationship between DOM composition and DBP formation, and to determine whether the DOM fractions remaining after treatment contribute to DBP formation, Spearman rank correlation coefficients (ρ) were calculated between HRMS-derived molecular features of raw water samples and FP DBP concentrations. Only features detected in at least four of the six samples were included. These features were correlated with DBP concentrations measured under FP conditions, and those showing strong positive correlations (ρ > 0.8) were retained and are shown in **Figure 3**.

Analysis of individual features (ρ > 0.8) identified 97 potential DBP precursors (**Figure 3**):

- 38 aliphatics
- 3 aromatics
- 18 high-O unsaturated
- 38 low-O unsaturated

Most precursors were present in low-oxygen unsaturated (39 %) and aliphatic (39 %) regions, consistent with their persistence and reactivity. Aromatic region accounted for only 3 % and there are not condensed aromatics. As summarized in **Table S4**, 61 of the 97 tentative precursors remained present after activated carbon filtration, the final step before disinfection. Thus, the DWTP removed only 37% of these indicative DBP precursors.

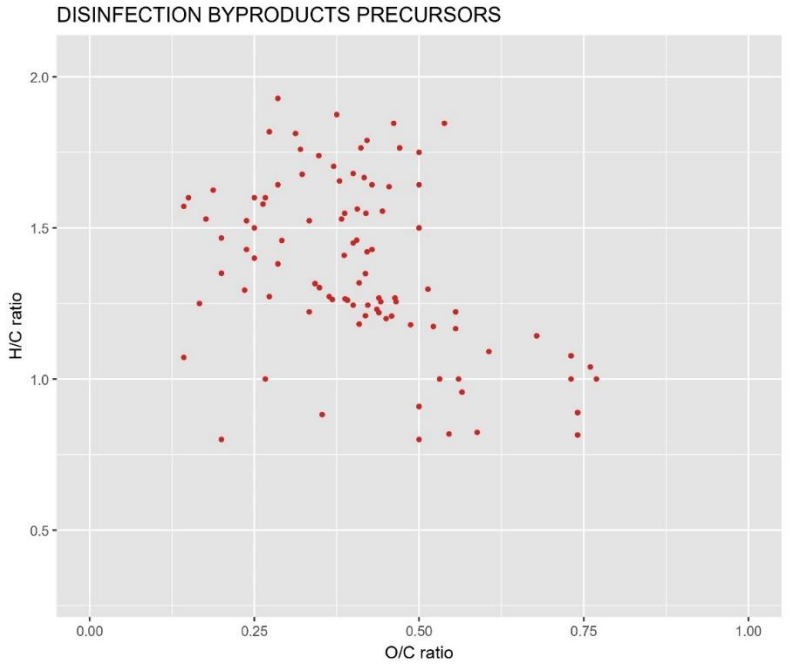


Figure 3. Strongly correlated (ρ > 0.8) DBP precursors in raw water.

3.4 Fluorescence and UV absorbance analysis and their correlations with HRMS analysis.

The fluorescence excitation–emission matrices were divided into five characteristic regions (**Figure S1**), each associated with different classes of dissolved organic matter (DOM), tyrosine-like (Region I), tryptophan-like (Region II), fulvic-like (Region III), soluble microbial product-like (Region IV), and humic-like (Region V) compounds. Several fluorescence descriptors were extracted from these regions, including maximum intensity, integrated intensity, raw intensity, and Raman-normalized intensity. In parallel, absorbance values at selected UV–Vis wavelengths were used as optical indicators. The complete set of fluorescence and UV–Vis parameters is summarized in **Table S5**, while the values of the indices measured and subsequently correlated with the HRMS-derived molecular features are presented in **Table S6**.

3.4.1 UV-Vis and Fluorescence measurements of the DWTP

Firstly, the behaviour of representative fluorescence and absorbance indicators across the treatment train was evaluated to assess their response to DOM removal and transformation processes. **Figure 4** shows the evolution of fluorescence at the humic-like compounds region (Region V) maximum intensity (I5MAX) and UV absorbance at 254 nm (UVA254) throughout the treatment train. Values were normalized relative to raw water (100%) using the mean values obtained from the six sampling campaigns.

Both optical indicators showed a progressive decrease along the treatment process, reflecting the reduction and transformation of chromophoric DOM fractions. The decrease was particularly pronounced for fluorescence intensity, which I5MAX decreased to 24% of its initial value in produced water, whereas UVA254 remained at 58% of the raw water value. This stronger reduction in fluorescence suggests that humic-like fluorescent components are highly sensitive to treatment-induced changes in DOM composition.

The decrease observed in optical indicators was consistent with the DOM removal trends defined through DOC measurements and HRMS molecular fingerprinting. Similar selective removal patterns were observed across the five fluorescence regions, with the most pronounced reductions occurring in fulvic-like (Region III) and humic-like (Region V) components. These trends agree with the HRMS results, where specific molecular regions of the van Krevelen diagram showed greater removal than others, particularly aromatic and condensed aromatic fractions.

The higher relative reduction observed for optical measurements compared with DOC and HRMS-based metrics can be explained by the preferential removal of highly aromatic and conjugated DOM fractions during treatment. These molecular classes contribute disproportionately to DOM optical properties, including UV absorbance and fluorescence signals. Therefore, fluorescence and absorbance measurements provide sensitive indicators of changes in DOM composition and can serve as rapid proxies for tracking treatment-induced transformations of DBP-relevant DOM fractions.

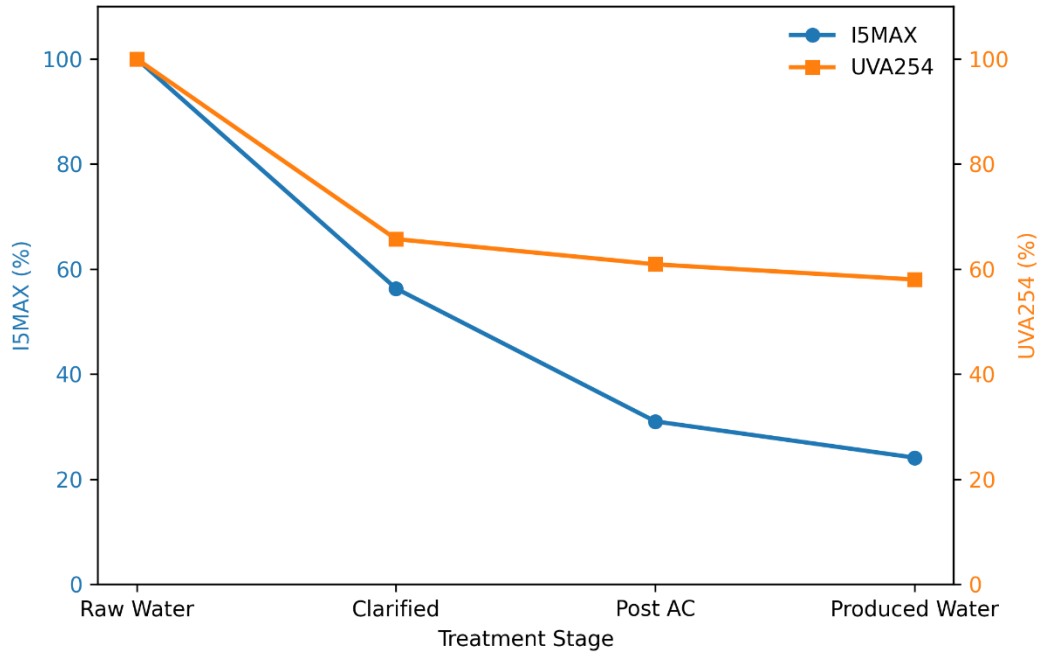


Figure 4. Response of fluorescence and absorbance-based optical indicators to DOM transformation during drinking water treatment.

3.4.2 Correlation between UV-Vis and Fluorescence with HRMS molecular features

To assess whether rapid optical measurements could indicate changes in DOM molecular composition, Spearman rank correlations were calculated between 28 fluorescence and UV–Vis parameters and the molecular features detected by HRMS. Spearman's coefficient was selected a priori because it evaluates monotonic associations without requiring normally distributed variables or a strictly linear response and is less sensitive to extreme values than Pearson's correlation. No alternative correlation coefficients were used in the primary analysis; therefore, Spearman was retained as the most appropriate metric for this exploratory analysis. The correlations assessed in this section were established between optical parameters and HRMS-derived molecular features, rather than directly between the optical measurements and DBP-related endpoints such as THMs, HAAs, HANs, AOX, or DBP formation potential. The intended predictive framework is therefore indirect: fluorescence and UV–Vis parameters measured in raw water are linked to molecular regions of DOM identified by HRMS, and these molecular regions are, in turn, associated with DBP formation. This approach could ultimately allow optical measurements in raw water to act as rapid proxies for the abundance of relevant DBP precursor fractions and, consequently, for the potential risk of DBP formation before disinfection. Direct correlations with DBP formation potential could also be evaluated using laboratory formation-potential tests performed on raw-water samples, whereas correlations with DBP concentrations measured in untreated raw water would not be meaningful because DBPs have not yet formed at that stage.

The initial dataset contained approximately 9,000 features. After retaining features detected in at least 50% of the samples and excluding the 20 % with the lowest variance,

approximately 3,000 features remained. Correlations with absolute Spearman coefficients above 0.5 were considered relevant. Raw water, treatment-train, and distribution-system samples were analyzed separately to prevent differences between treatment stages from dominating the observed relationships. Although particular attention was given to raw water because of its potential use for anticipating DBP formation before disinfection, the treatment train and distribution system were also evaluated to determine how optical-molecular relationships evolved during treatment and subsequent water distribution.

In raw water, the maximum intensity of fluorescence Region 5 (humic-like compounds) (I5MAX) showed the strongest relationship with the HRMS fingerprint, correlating with 2,076 molecular features (**Figure 5**). Most of these features were classified as low-oxygen unsaturated-like (1,193) or high-oxygen unsaturated-like compounds (698), followed by aliphatic (162), aromatic (21), and condensed aromatic-like features (2). Correlated formulas were mainly CHO (1,855) and CHON (221), with most associations occurring between m/z 400 and 800. These results indicate that fluorescence associated with humic-like compounds (Region V) responds to a broad and chemically heterogeneous fraction of unsaturated DOM. Its strong relationship with molecular fractions previously associated with DBP formation supports its potential use as a rapid optical proxy for assessing precursor abundance in raw water.

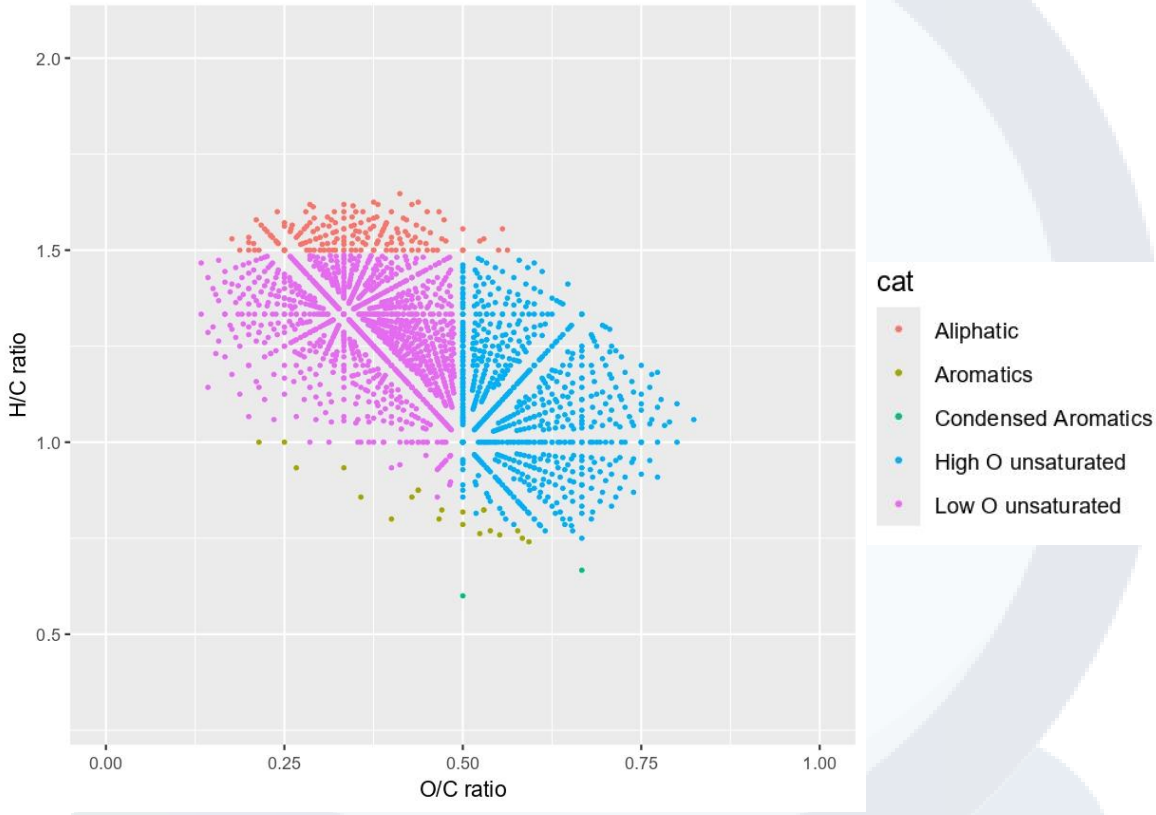


Figure 5. Van Krevelen diagram representation of the features associated (ρ > 0.5) with the maximum intensity at the region V (humic-like compounds).

Within the treatment train, fewer correlations were observed than in raw water, although the general patterns were consistent across the different fluorescence metrics. Regions 3, 4, and 5 showed the highest numbers of associated HRMS features, whereas region 1 showed only limited relationships. For Region 5, comparable patterns were obtained using the integral, maximum intensity, raw intensity, and Raman-normalized intensity. This consistency supports the robustness of Region 5 as an optical indicator not only of precursor fractions in raw water but also of DOM transformations occurring throughout treatment. Regions 3 and 4 were mainly associated with high-oxygen unsaturated-like compounds and, to a lesserregiont, with aliphatic and low-oxygen unsaturated-like fractions.

UV–Vis absorbance showed a more selective response. Wavelengths between 270 and 340 nm were mainly correlated with high-oxygen unsaturated-like features; for example, absorbance at 320 nm was associated with 49 formulas in this region. This suggests that UV–Vis preferentially reflects high-oxygenated unsaturated DOM, whereas fluorescence regions 3–5 integrate a broader molecular fraction. The two optical approaches therefore provide complementary rather than redundant information.

Following disinfection, the number of correlations decreased markedly. In post-disinfection, absorbance at 270 and 280 nm was associated with 12 features, while UVA254 correlated with 6. In the distribution system, UVA270, UVA280, and UVA340 correlated with 46, 44, and 32 features, respectively. This reduction may reflect lower sample variability, transformation of fluorescent compounds during chlorination, or the formation of products with weak optical responses. Nevertheless, analysing the distribution system provides insight into whether the selected optical indicators remain informative after treatment and during water residence in the network.

Overall, fluorescence, UV–Vis, and HRMS provide complementary but non-interchangeable information. Optical methods are rapid, low-cost, and suitable for frequent or potentially online monitoring, although they provide an aggregated response. HRMS provides detailed molecular-level characterization but requires more complex instrumentation, sample preparation, and data processing. The present results support the use of fluorescence regions 3–5 and selected UV–Vis wavelengths as rapid indicators of DOM fractions relevant to DBP formation. Their main predictive application would be the early assessment of DBP precursor fractions in raw water, while measurements across the treatment train and distribution system could additionally support process monitoring and the evaluation of DOM transformations over time.

4.0 Conclusions

Access to safe and high-quality drinking water is essential for human health. However, ensuring effective disinfection while minimizing the formation of DBPs requires optimized treatment processes. Therefore, comprehensive evaluation of drinking water treatment systems is critical. In this study, a conventional DWTP with limited technological complexity was assessed using a multi-analytical approach, including the quantification of 17 target DBPs, the measurement of AOX, the fingerprint of DOM by HRMS, and fluorescence, and UV–Vis absorbance measurements.

This combined approach enabled a detailed assessment of DOM removal efficacy and DBP formation dynamics in a full-scale DWTP. The results showed that although the plant achieved substantial DOC removal, this process was selective, primarily targeting aromatic and condensed aromatic-like compounds. In contrast, aliphatic and low- and high-oxygen unsaturated-like regions exhibited limited removal effectiveness. This insufficient elimination was strongly correlated with the formation of halogenated organic features and DBPs, particularly THMs. A Spearman correlation analysis further supported this relationship, linking the regions with the lowest removal rates to DBP formation.

In conclusion, several factors can influence DBP formation during disinfection, for example, the addition of desalinated water. However, the most significant factor is the chemical composition and partial removal of DOM. Even under relatively low DOC concentration, as observed in this DWTP, the presence of reactive DOM regions contributes substantially to DBP formation. If such treatment technologies are applied to water sources with higher DOC loads, their limited capacity to selectively remove key precursor regions may become a major critical concern in controlling DBP formation.

Correlations between optical parameters and HRMS features indicated that fluorescence regions 3–5, especially Region 5, reflected broad fractions of oxygenated and unsaturated DOM. UV–Vis absorbance between 270 and 340 nm showed a more selective association with high-oxygen unsaturated compounds. These techniques therefore provide complementary information and may be useful as rapid indicators of DOM changes relevant to DBP formation.

Despite progress in DBP research, the link between DOM composition and DBP formation remains only partially understood. This study highlights the importance of integrating targeted and non-targeted analytical tools (GC-MS, AOX, HRMS) to better characterize DOM reactivity and support the development of more selective treatment strategies to mitigate DBP formation.

5.0 Future work

Future work perspectives are to employ this multi-analytical approach to the remaining case studies. For the rest of the case studies, evaluating their capacity to remove specific fractions of dissolved organic matter (DOM), their disinfection byproducts (DBPs) formation, and to establish if their DOM fractions that react to form DBPs are consistent across all site cases or vary depending on the disinfection methodology used, for example, chloramination at the Canal drinking water plant or more traditional treatment systems such as those used in Limassol.

The Barcelona case study has already provided preliminary correlations between HRMS fingerprints and UV–Vis and fluorescence parameters. Therefore, the next step will be to validate the robustness and transferability of these relationships across different plants and treatment conditions, rather than to repeat the same analysis. Direct relationships between optical parameters and DBP-related endpoints – including THMs, HAAs, HANs, AOX, and DBP formation potential – will be investigated to assess their predictive performance and to develop cross-site calibration models. HRMS will serve as the molecular-level reference for interpreting these relationships and identifying the DOM fractions responsible for both the optical responses and DBP formation.

Particular attention will be given to the overlap between HRMS features associated with fluorescence Region 5 and those identified as potential DBP precursors. This analysis will determine whether Region 5, together with selected UV–Vis wavelengths, can provide a transferable optical proxy for reactive DOM and DBP formation risk. Following cross-site calibration and validation, these indicators could be integrated into online sensors for continuous monitoring and early warning of conditions associated with increased DBP formation. The resulting sensor outputs and predictive models will be linked to the monitoring and forecasting activities of WP3 and WP5, supporting treatment optimization and project-wide DBP risk management.

6.0 References

(1) Zhang, H.; Zhang, Y.; Shi, Q.; Zheng, H.; Yang, M. Characterization of Unknown Brominated Disinfection Byproducts during Chlorination Using Ultrahigh Resolution Mass Spectrometry. *Environ Sci Technol* 2014, *48* (6), 3112–3119. https://doi.org/10.1021/es4057399.

(2) *Drinking water - European Commission*. https://environment.ec.europa.eu/topics/water/drinking-water_en (accessed 2025-07-11).

(3) Richardson, S. D.; Plewa, M. J. To Regulate or Not to Regulate? What to Do with More Toxic Disinfection by-Products? *J Environ Chem Eng* 2020, *8* (4). https://doi.org/10.1016/j.jece.2020.103939.

(4) Rout, P. R.; Zhang, T. C.; Bhunia, P.; Surampalli, R. Y. Treatment Technologies for Emerging Contaminants in Wastewater Treatment Plants: A Review. *Science of the Total Environment* 2021, *753*. https://doi.org/10.1016/j.scitotenv.2020.141990.

(5) Font-Ribera, L.; Gràcia-Lavedan, E.; Aragonés, N.; Pérez-Gómez, B.; Pollán, M.; Amiano, P.; Jiménez-Zabala, A.; Castaño-Vinyals, G.; Roca-Barceló, A.; Ardanaz, E.; Burgui, R.; Molina, A. J.; Fernández-Villa, T.; Gómez-Acebo, I.; Dierssen-Sotos, T.; Moreno, V.; Fernandez-Tardon, G.; Peiró, R.; Kogevinas, M.; Villanueva, C. M. Long-Term Exposure to Trihalomethanes in Drinking Water and Breast Cancer in the Spanish Multicase-Control Study on Cancer (MCC-SPAIN). *Environ Int* 2018, *112*, 227–234. https://doi.org/10.1016/j.envint.2017.12.031.

(6) Villanueva, C. M.; Cordier, S.; Font-Ribera, L.; Salas, L. A.; Levallois, P. Overview of Disinfection By-Products and Associated Health Effects. *Curr Environ Health Rep* 2015, *2* (1), 107–115. https://doi.org/10.1007/s40572-014-0032-x.

(7) Li, J.; Aziz, M. T.; Granger, C. O.; Richardson, S. D. Halocyclopentadienes: An Emerging Class of Toxic DBPs in Chlor(Am)Inated Drinking Water. *Environ Sci Technol* 2022, *56* (16), 11387–11397. https://doi.org/10.1021/acs.est.2c02490.

(8) He, H.; Sun, N.; Li, L.; Ai, J.; Zhou, H.; Yang, X.; Yang, X.; Wang, D.; Zhang, W. Effects of Dissolved Organic Matter Removal and Molecular Transformation in Different Water Treatment Processes on Formation of Disinfection Byproducts. *Water Res* 2023, *245*. https://doi.org/10.1016/j.watres.2023.120626.

(9) Richardson, S. D.; Plewa, M. J.; Wagner, E. D.; Schoeny, R.; DeMarini, D. M. Occurrence, Genotoxicity, and Carcinogenicity of Regulated and Emerging Disinfection by-Products in Drinking Water: A Review and Roadmap for Research. *Mutation Research - Reviews in Mutation Research*. November 2007, pp 178–242. https://doi.org/10.1016/j.mrrev.2007.09.001.

(10) Andersson, A.; Lavonen, E.; Harir, M.; Gonsior, M.; Hertkorn, N.; Schmitt-Kopplin, P.; Kylin, H.; Bastviken, D. Selective Removal of Natural Organic Matter during Drinking Water Production Changes the Composition of Disinfection By-Products. *Environ Sci (Camb)* 2020, *6* (3), 779–794. https://doi.org/10.1039/c9ew00931k.

(11) Phungsai, P.; Kurisu, F.; Kasuga, I.; Furumai, H. Changes in Dissolved Organic Matter Composition and Disinfection Byproduct Precursors in Advanced Drinking Water Treatment Processes. *Environ Sci Technol* 2018, *52* (6), 3392–3401. https://doi.org/10.1021/acs.est.7b04765.

(12) Gonsior, M.; Schmitt-Kopplin, P.; Stavklint, H.; Richardson, S. D.; Hertkorn, N.; Bastviken, D. Changes in Dissolved Organic Matter during the Treatment Processes of a Drinking Water Plant in Sweden and Formation of Previously Unknown Disinfection Byproducts. *Environ Sci Technol* 2014, *48* (21), 12714–12722. https://doi.org/10.1021/es504349p.

(13) Kimura, S. Y.; Cuthbertson, A. A.; Byer, J. D.; Richardson, S. D. The DBP Exposome: Development of a New Method to Simultaneously Quantify Priority Disinfection by-Products and Comprehensively Identify Unknowns. *Water Res* 2019, *148*, 324–333. https://doi.org/10.1016/j.watres.2018.10.057.

(14) Zhou, R.; Xu, Z.; Wu, Y.; Yang, L.; Li, X.; Meng, Y.; Zhu, P.; Kong, L.; Ding, X. Analysis of 13 Haloacetamide DBPs in Drinking Water Using a Highly Sensitive LC-MS/MS Method. *Environ. Sci.: Water Res. Technol.* 2023, *9* (4), 1231–1237. https://doi.org/10.1039/D2EW00781A.

(15) Munthali, E.; Marcé, R.; Farré, M. J. Drivers of Variability in Disinfection By-Product Formation Potential in a Chain of Thermally Stratified Drinking Water Reservoirs. *Environ Sci (Camb)* 2022, *8* (5), 968–980. https://doi.org/10.1039/d1ew00788b.

(16) Schum, S. K.; Brown, L. E.; Mazzoleni, L. R. MFAssignR: Molecular Formula Assignment Software for Ultrahigh Resolution Mass Spectrometry Analysis of Environmental Complex Mixtures. *Environ Res* 2020, *191*. https://doi.org/10.1016/j.envres.2020.110114.

(17) Hawkes, J. A.; D'Andrilli, J.; Agar, J. N.; Barrow, M. P.; Berg, S. M.; Catalán, N.; Chen, H.; Chu, R. K.; Cole, R. B.; Dittmar, T.; Gavard, R.; Gleixner, G.; Hatcher, P. G.; He, C.; Hess, N. J.; Hutchins, R. H. S.; Ijaz, A.; Jones, H. E.; Kew, W.; Khaksari, M.; Palacio Lozano, D. C.; Lv, J.; Mazzoleni, L. R.; Noriega-Ortega, B. E.; Osterholz, H.; Radoman, N.; Remucal, C. K.; Schmitt, N. D.; Schum, S. K.; Shi, Q.; Simon, C.; Singer, G.; Sleighter, R. L.; Stubbins, A.; Thomas, M. J.; Tolic, N.; Zhang, S.; Zito, P.; Podgorski, D. C. An International Laboratory Comparison of Dissolved Organic Matter Composition by High Resolution Mass Spectrometry: Are We Getting the Same Answer? *Limnol Oceanogr Methods* 2020, *18* (6), 235–258. https://doi.org/10.1002/lom3.10364.

(18) Chen, W.; Westerhoff, P.; Leenheer, J. A.; Booksh, K. Fluorescence Excitation-Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter. Environ. Sci. Technol. 2003, 37(24), 5701-5710. https://doi.org/10.1021/es034354c.

7.0 Annex

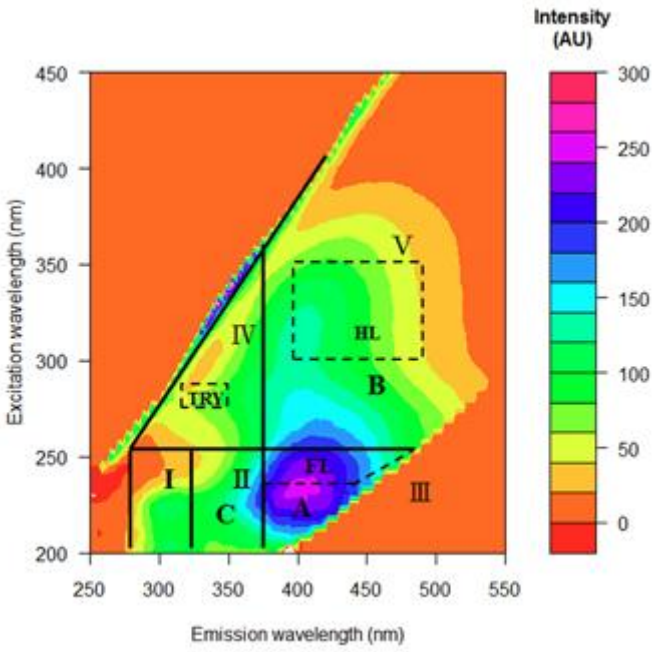


Figure S1. Example of a fluorescence spectrum divided into five characteristic regions, tyrosine-like (Region I), tryptophan-like (Region II), fulvic-like (Region III), soluble microbial product-like (Region IV), and humic-like (Region V) compounds. Adapted from Chen et al. (2003).

Table S1. % Recovery of DOM extraction.

Sample	Sampling 1	Sampling 2	Sampling 3	Sampling 4	Sampling 5	Sampling 6
Raw	72.26	50.00	82.78	45.18	61.75	54.83
Clarified	74.61	60.33	83.73	71.33	65.50	49.33
Post AC	36.49	45.86	30.66	48.35	42.30	43.38
Produced water	47.60	54.09	49.23	57.45	45.98	50.73
Distribution 24h	65.13	38.34	74.20	32.83	56.05	43.85
Distribution 48h	52.90	37.61	64.98	7.41	40.83	39.60
Distribution 72h	47.01	56.88	52.13	74.43	41.90	39.33
Raw FP	51.75	50.78	-	-	52.00	49.05
Clarified FP	40.32	38.45	-	-	44.20	33.43
Post AC FP	45.87	51.23	-	-	36.38	58.73
Produced water FP	50.22	47.96	-	-	48.60	49.43

Table S2. Criteria for formula assignment according to their regions in the Van Krevelen diagram.

Class	H/C ratio	O/C ratio	AImod value
Aliphatics-like	≥1.5	-	-
Aromatics-like	-	-	≥0.67
Condensed aromatics-like	-	-	0.5 < AImod < 0.67
High O unsaturated-like	<1.5	≥0.5	<0.5
Low O unsaturated-like	<1.5	<0.5	<0.67

Note. Hawkes, J. A., d'Andrilli, J., Agar, J. N., Barrow, M. P., Berg, S. M., Catalán, N., ... & Podgorski, D. C. (2020). An international laboratory comparison of dissolved organic matter composition by high resolution mass spectrometry: Are we getting the same answer?. Limnology and Oceanography: Methods, 18(6), 235-258.

Table S3. First 10 molecular formulae presented in at least 4 of 6 samples and employed for all data analysis and representations.

Raw			Clarified			Post AC			Produced water			Distribution 24h			Distribution 48h			Distribution 72h		
F	G	C	F	G	C	F	G	C	F	G	C	F	G	C	F	G	C	F	G	C
C9H9O2Br	O2Br	CHOBr	C10H10N2O3	N2O3	CHNO	C10H10N2O3	N2O3	CHNO	C10H10N2O3	N2O3	CHNO	C10H10N2O3	N2O3	CHNO	C10H10N2O3	N2O3	CHNO	C10H10N2O3	N2O3	CHNO
C17H20N2O10	N2O10	N2O10	C10H10N2O4	N2O4	CHNO	C10H10N2O4	N2O4	CHNO	C10H10N2O4	N2O4	CHNO	C10H10N2O4	N2O4	CHNO	C10H10N2O4	N2O4	CHNO	C10H10N2O4	N2O4	CHNO
C15H12N2O10	N2O10	CHNO	C10H10N2O5	N2O5	CHNO	C10H10N2O5	N2O5	CHNO	C10H10N2O5	N2O5	CHNO	C10H10N2O5	N2O5	CHNO	C10H10N2O5	N2O5	CHNO	C10H10N2O5	N2O5	CHNO
C22H26N2O10	N2O10	CHNO	C10H10N2O6	N2O6	CHNO	C10H10N2O6	N2O6	CHNO	C10H10N2O6	N2O6	CHNO	C10H10N2O6	N2O6	CHNO	C10H10N2O6	N2O6	CHNO	C10H10N2O6	N2O6	CHNO
C15H14N2O10	N2O10	CHNO	C10H10O3S	O3S	CHOS	C10H10O3S	O3S	CHOS	C10H10O3S	O3S	CHOS	C10H10O3S	O3S	CHOS	C10H10O3S	O3S	CHOS	C10H10O3S	O3S	CHOS
C16H10N2O10	N2O10	CHNO	C10H10O4S	O4S	CHOS	C10H10O4S	O4S	CHOS	C10H10O4S	O4S	CHOS	C10H10O4S	O4S	CHOS	C10H10O4S	O4S	CHOS	C10H10O4S	O4S	CHOS
C16H12N2O10	N2O10	CHNO	C10H10O5	O5	CHO	C10H10O5	O5	CHO	C10H10O5	O5	CHO	C10H10O5	O5	CHO	C10H10O5	O5	CHO	C10H10O5	O5	CHO
C16H14N2O10	N2O10	CHNO	C10H10O5S	O5S	CHOS	C10H10O5S	O5S	CHOS	C10H10O5S	O5S	CHOS	C10H10O5S	O5S	CHOS	C10H10O5S	O5S	CHOS	C10H10O5S	O5S	CHOS
C16H16N2O10	N2O10	CHNO	C10H10O6	O6	CHO	C10H10O6	O6	CHO	C10H10O6	O6	CHO	C10H10O6	O6	CHO	C10H10O6	O6	CHO	C10H10O6	O6	CHO
C17H12N2O10	N2O10	CHNO	C10H10O6S	O6S	CHOS	C10H10O6S	O6S	CHOS	C10H10O6S	O6S	CHOS	C10H10O6S	O6S	CHOS	C10H10O6S	O6S	CHOS	C10H10O6S	O6S	CHOS

Note. F: Formulae; G: Group; C: Class. The rest of the dataset is provided under request.

Table S4. Number of highly correlated features and classified according to their region in the van Krevelen presented in each sample.

Sample	Aliphatic	Aromatic	Condensed aromatic	Low-Oxygen unsaturated	High-Oxygen unsaturated	Total
Raw	38	3	0	18	38	97
Clarified	33	2	0	4	19	58
Post AC	35	1	0	6	19	61
Produced water	31	0	0	5	17	53
Distribution 24h	29	0	0	2	15	46
Distribution 48h	32	0	0	2	17	51
Distribution 72h	30	0	0	3	15	48

Table S5. Codification of Fluorescence and Absorbance parameters measured.

Codification	Parameter
RUI1	I1 (224/289) RAMAN UNITS
RUI2	I2 (224/340) RAMAN UNITS
RUI3	I3 (245/439) RAMAN UNITS
RUI4	I4 (275/346) RAMAN UNITS
RUI5	I5 (344/439) RAMAN UNITS
RAWI1	I1 (224/289) RAW DATA
RAWI2	I2 (224/340) RAW DATA
RAWI3	I3 (245/439) RAW DATA
RAWI4	I4 (275/346) RAW DATA
RAWI5	I5 (344/439) RAW DATA
I1MAX	MAXIMUN INTENSITY REGION 1 (224/289)
I2MAX	MAXIMUN INTENSITY REGION 2 (224/340)
I3MAX	MAXIMUN INTENSITY REGION 3 (245/439)
I4MAX	MAXIMUN INTENSITY REGION 4 (275/346)
I5MAX	MAXIMUN INTENSITY REGION 5 (344/439)
Φ1	FLUORESCENCE INTENSITY AT REGION 1
Φ2	FLUORESCENCE INTENSITY AT REGION 2
Φ3	FLUORESCENCE INTENSITY AT REGION 3
Φ4	FLUORESCENCE INTENSITY AT REGION 4
Φ5	FLUORESCENCE INTENSITY AT REGION 5
ΦTOT	TOTAL FLUORESCENCE INTENSITY
UVA254	ABSORBANCE AT λ 254 nm
UVA270	ABSORBANCE AT λ 270 nm
UVA280	ABSORBANCE AT λ 280 nm
UVA300	ABSORBANCE AT λ 300 nm
UVA320	ABSORBANCE AT λ 320 nm
UVA330	ABSORBANCE AT λ 330 nm
UVA340	ABSORBANCE AT λ 340 nm

Table S6. Fluorescence and Absorbance measurements.

Sample name	RU I1	RU I2	RU I3	RU I4	RU I5	RA WI1	RA WI2	RA WI3	RA WI4	RA WI5	I1M AX	I2M AX	I3M AX	I4M AX	I5M AX	Φ1	Φ2	Φ3	Φ4	Φ5	ΦT OT	UVA2 54	UVA2 70	UVA2 80	UVA3 00	UVA3 20	UVA3 30	UVA3 40
RAW S1	0.24	0.83	1.81	0.40	0.64	119	334	698	147	240	0.49	1.66	2.14	1.14	0.87	941	4588	5538	1967	1490	14524	0.035	0.027	0.023	0.016	0.010	0.008	0.007
CLARIFIED S1	0.20	0.61	1.16	0.27	0.42	101	224	463	102	159	0.30	1.23	1.40	0.82	0.58	651	3556	3443	1403	902	9955	0.021	0.016	0.013	0.009	0.006	0.004	0.004
POST AC S1	0 16	0.29	0.55	0.12	0.19	109	119	216	45	70	0.19	0.60	0.74	0.38	0.26	405	1582	1689	605	401	4862	0.019	0.013	0.011	0.007	0.004	0.003	0.002
PW S1	0 08	0.25	0.45	0.10	0.15	70	119	171	38	57	0.16	0.47	0.61	0.32	0.22	343	1362	1429	525	321	3980	0.018	0.012	0.10	0.007	0.004	0.003	0.002
DIST 1 S1	0.14	0.34	0.45	0.11	0.14	69	168	186	45	55	0.19	0.51	0.58	0.32	0.21	382	1536	1315	552	297	4082	0.012	0.007	0.006	0.004	0.002	0.002	0.001
DIST 2 S1	0.09	0.20	0.34	0.07	0.10	66	111	149	30	38	0.12	0.39	0.44	0.26	0.15	280	1076	1020	405	217	2998	0.013	0.008	0.006	0.004	0.002	0.002	0.001
DIST 3 S1	0.12	0.22	0.39	0.09	0.12	100	121	157	36	46	0.13	0.42	0.48	0.30	0.17	283	1225	1133	460	251	3352	000	0.000	0.000	0.000	0.000	0.000	0.000
RAW S2	0.28	0.80	1.88	0.38	0.66	128	339	717	146	255	0.53	1.68	2.24	1.14	0.88	991	4677	5660	1929	1528	14785	0.033	0.025	0.021	0.014	0.008	0.006	0.005
CLARIFIED S2	0.25	0.62	1.16	0.27	0.38	115	241	470	97	146	0.36	1.19	1.45	0.77	0.52	724	3503	3438	1324	842	9831	0.018	0.013	0.011	0.007	0.004	0.003	0.002
POST AC S2	0.18	0.28	0.48	0.09	0.16	85	120	200	36	61	0.22	0.56	0.66	0.33	0.23	403	1420	1474	515	345	4157	0.018	0.013	0.010	0.007	0.004	0.002	0.002
PW S2	0.2	0.24	0.38	0.08	0.12	60	119	155	37	48	0.18	0.47	0.56	0.30	0.18	361	1290	1216	454	2777	3598	0.017	0.012	0.010	0.006	0.003	0.002	0.002
DIST 1 S2	0.16	0.30	0.45	0.10	0.15	56	128	184	39	59	0.21	0.54	0.61	0.34	0.21	385	1540	1403	547	320	4195	0.017	0.011	0.009	0.006	0.003	0.002	0.002
DIST 2 S2	0.14	0.24	0.4	0.10	0.14	88	128	177	39	54	0.17	0.48	0.56	0.33	0.20	335	1339	1296	514	298	3781	0.016	0.011	0.009	0.006	0.003	0.002	0.002
DIST 3 S2	0.11	0.23	0.42	0.09	0.13	74	119	171	36	53	0.14	0.49	0.53	0.32	0.20	285	1294	1249	475	280	3582	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RAW S3	0.26	0.80	1.98	0.47	0.71	99	263	699	156	246	0.56	1.77	2.23	1.25	0.98	1063	5057	5982	2300	1632	16033	0.032	0.024	0.020	0.014	0.008	0.006	0.005
CLARIFIED S3	0.24	0.62	1.18	0.29	0.42	91	237	423	102	148	0.32	1.17	1.44	0.82	0.58	682	3679	3624	1468	912	10366	0.017	0.012	0.010	0.006	0.003	0.002	0.002
POST AC S3	0.11	0.21	0.53	0.10	0.17	61	102	191	38	60	0.15	0.50	0.99	0.37	0.25	318	1398	1701	553	379	4350	0.014	0.009	0.008	0.005	0.002	0.001	0.001
PW S3	0.11	0.13	0.35	0.07	0.12	32	54	131	25	42	0.14	0.36	0.44	0.24	0.16	249	1009	1065	382	250	2955	0.013	0.009	0.007	0.005	0.002	0.001	0.001

DIST 1 S3	0.09	0.19	0.44	0.09	0.13	55	97	160	35	48	0.13	0.41	0.51	0.30	0.19	280	1189	1223	467	081	3440	0.013	0.008	0.007	0.004	0.002	0.01	0.001
DIST 2 S3	0.10	0.18	0.40	0.09	0.12	57	68	143	30	43	0.14	0.41	0.48	0.28	0.18	270	1149	1146	443	257	3265	0.012	0.008	0.006	0.004	0.001	0.001	0.000
DIST 3 S3	0.08	0.16	0.39	0.08	0.12	32	52	145	31	44	0.15	0.39	0.47	0.29	0.18	268	1151	1149	442	260	3270	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RAW S4	0.25	0.76	1.81	0.44	0.64	106	288	703	159	245	0.49	1.67	2.09	1.14	0.88	984	4713	5467	2096	1481	14741	0.031	0.024	0.020	0.014	0.008	0.006	0.005
CLARIFIED S4	0.20	0.59	1.17	0.28	0.40	105	241	424	103	154	0.33	1.11	1.35	0.77	0.53	639	3486	3411	1365	865	9765	0.017	0.012	0.010	0.007	0.004	0.003	0.002
POST AC S4	0.11	0.14	0.46	0.09	0.16	63	82	194	34	60	0.13	0.42	0.56	0.30	0.21	282	117	1364	480	339	3642	0.014	0.010	0.008	0.006	0.003	0.002	0.002
PW S4	0.12	0.11	0.29	0.06	0.09	73	54	116	25	35	0.14	0.30	0.36	0.20	0.13	220	798	843	308	199	2368	0.014	0.010	0.008	0.005	0.003	0.002	0.002
DIST 1 S4	0.10	0.19	0.36	0.07	0.11	94	96	155	32	43	0.12	0.37	0.46	0.25	0.17	237	1010	1049	399	242	2938	0.012	0.008	0.005	0.002	0.002	0.002	0.001
DIST 2 S4	0.09	0.17	0.34	0.07	0.10	46	93	133	30	40	0.13	0.37	0.45	0.23	0.17	228	977	1014	385	227	2833	0.013	0.009	0.005	0.003	0.003	0.002	0.001
DIST 3 S4	0.11	0.14	0.34	0.07	0.11	82	68	131	27	41	0.11	0.37	0.44	0.24	0.15	241	952	988	375	226	2783	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RAW S5	0.17	0.56	2.52	0.32	1.05	79	223	959	115	390	0.34	1.83	2.85	1.30	1.21	688	4207	7751	1868	2510	17024	0.046	0.034	0.029	0.020	0.012	0.012	0.007
CLARIFIED S5	0.19	0.46	1.37	0.24	0.50	89	172	535	91	189	0.29	1.25	1.66	0.84	0.63	561	3115	4021	1286	1116	10099	0.037	0.027	0.023	0.015	0.010	0.010	0.005
POST AC S5	0.15	0.38	1.04	0.31	0.39	94	148	402	106	147	0.25	0.98	1.26	0.66	0.49	491	2499	3133	1161	849	8133	0.034	0.024	0.0179	0.013	0.008	0.008	0.005
PW S5	0.12	0.33	0.90	0.16	0.34	79	92	360	59	123	0.17	0.90	1.10	0.59	0.43	390	2242	2791	896	725	7045	0.030	0.021	0.017	0.012	0.007	0.007	0.004
DIST 1 S5	0.14	0.33	0.94	0.16	0.32	85	143	373	60	119	0.17	0.81	1.06	0.59	0.41	370	2211	2647	2647	885	682	6796	0.027	0.019	0.016	0.007	0.005	0.004
DIST 2 S5	0.14	0.37	0.83	0.15	0.28	89	149	330	55	107	0.19	0.83	0.98	0.56	0.38	377	2196	2405	2405	874	602	6454	0.026	0.018	0.015	0.007	0.005	0.004
DIST 3 S5	0.14	0.32	0.75	0.15	0.26	68	82	301	58	98	0.18	0.74	0.92	0.54	0.36	408	2050	2050	2277	850	554	6140	0.000	0.000	0.000	0.000	0.000	0.000
RAW S6	0.15	0.60	2.05	0.32	0.76	88	271	811	128	300	0.40	1.65	2.34	1.17	0.95	764	4184	6186	1752	1780	14667	0.033	0.025	0.021	0.014	0.008	0.008	0.004
CLARIFIED S6	0.19	0.40	0.93	0.18	0.32	102	180	378	76	129	0.27	0.91	1.14	0.58	0.41	563	2461	2739	948	699	7410	0.028	0.021	0.018	0.012	0.007	0.007	0.004
POST AC S6	0.16	0.37	0.78	0.16	0.26	120	183	318	65	104	0.30	0.74	0.98	0.50	0.35	537	2108	2339	813	582	6379	0.029	0.021	0.017	0.011	0.006	0.006	0.003
PW S6	0.23	0.43	0.58	0.16	0.20	1116	190	255	64	81	0.36	0.61	0.78	0.38	0.27	644	1946	1749	770	429	5538	0.030	0.022	0.017	0.011	0.006	0.006	0.003

DIST 1 S6	0.20	0.50	0.78	0.21	0.36	129	266	325	86	106	0.33	0.80	0.94	0.52	0.35	679	2323	2317	1006	577	6902	0.031	0.023	0.018	0.011	0.007	0.005	0.003
DIST 2 S6	0.18	0.32	0.84	0.18	0.30	120	189	346	74	119	0.25	0.85	1.02	0.57	0.39	542	2259	2482	959	626	6868	0.031	0.022	0.018	0.012	0.007	0.005	0.003
DIST 3 S6	0.16	0.35	0.80	0.16	0.29	93	165	335	70	118	0.30	0.78	1.06	0.54	0.38	614	2271	2468	905	612	6869	0.000	0.000	0.000	0.000	0.000	0.000	0.000

intoDBP will create innovative tools and strategies to improve water quality management for safe human use and a healthy environment. It focuses on catchment protection and forecasting, transformative drinking water treatment, and real-time monitoring to combat the effects of climate and global change. In particular, intoDBP focuses on pollution and risks related to disinfection byproducts (DBPs). By developing and applying advanced, integrated, and cost-effective sensors and analytical methods, intoDBP will expand knowledge on water quality and DBP precursors to better understand its formation and human exposure in Europe. intoDBP monitoring results will feed into numerical forecasting tools to predict source water changes and formulate climate change adaptation pathways at catchment and treatment scale. intoDBP also develops transformative options for advanced and cost-effective upgrade of water treatment and disinfection.

In the intoDBP consortium researchers, small and large enterprises, communication experts and public services join forces to generate interdisciplinary solutions, that will generate a renewed perspective of drinking water surveillance, support decision-making and governance, and increase system resilience. intoDBP will implement and validate its cross-cutting products in four complementary case studies from three European countries where compliance with DBP regulation currently is an acknowledged challenge. The direct and visible positive impact of intoDBP in the case studies will foster rapid product adoption at a European and global scale, thus strengthening Europe's position and role in the global water market.

Reaching out beyond the water sector itself, intoDBP will directly engage society through surveys to analyze exposure to DBPs, collect data about catchment protection initiatives, create awareness and promote sustainable consumer behavior such as reducing bottled water consumption.



Funded by the European Union