# Identifying congeners and transformation products of organic contaminants within complex chemical mixtures in impacted surface waters with a top-down non-targeted screening workflow

Emmanuel Eysseric[1], Christian Gagnon[2], Pedro A. Segura[1, *]

* Tel: 1-(819) 821-7922. Fax: 1-(819) 821-8019. E-mail: pa.segura@usherbrooke.ca
[1]Department of Chemistry, Université de Sherbrooke, Sherbrooke, Canada
[2]Environment and Climate Change Canada, Montréal, Canada

The final version of this paper was published in *Science of the Total Environment*:
Eysseric E., Gagnon C., Segura P. A. (2022): Identifying congeners and transformation products of organic contaminants within complex chemical mixtures in impacted surface waters with a top-down non-targeted screening workflow. *Sci. Total Environ.* 822, 153540.
https://doi.org/10.1016/j.scitotenv.2022.153540

**Abstract**

Over 350 000 compounds are registered for production and use including a high number of congeners found in complex chemical mixtures (CCMs). With such a high number of chemicals being released in the environment and degraded into transformation products (TPs), the challenge of identifying contaminants by non-targeted screening (NTS) is massive. "Bottom-up" studies, where compounds are subjected to conditions simulating environmental degradation to identify new TPs, are time consuming and cannot be relied upon to study the TPs of hundreds of thousands of compounds. Therefore, the development of "top-down" workflows, where the structural elucidation of unknown compounds is carried directly on the sample, is of interest.

In this study, a top-down NTS workflow was developed using molecular networking and clustering (MNC). A total of 438 compounds were identified including 176 congeners of consumer product additives and 106 TPs. Reference standards were used to confirm the identification of 53 contaminants among them lesser-known pharmaceuticals (aliskiren, sitagliptin) and consumer product additives (lauramidopropyl betaine, 2,2,4-trimethyl-1,2-dihydroquinoline). The MNC tools allowed to group similar TPs and congeners together. As such, several previously unknown TPs of pesticides (metolachlor) and pharmaceuticals (gliclazide, irbesartan) were identified as tentative candidates or probable structures. Moreover, some congeners that had no entry on global repositories (PubChem, ChemSpider) were identified as probable structures. The workflow worked efficiently with oligomers containing ethylene oxide moieties, and with TPs structurally related to their parent compounds.

The top-down approach shown in this study addresses several issues with the identification of congeners of industrial compounds from CCMs. Furthermore, it allows elucidating the structure of TPs directly from samples without relying on bottom-up studies under conditions discussed herein. The top-down workflow and the MNC tools show great potential for data mining and retrospective analysis of previous NTS studies.

## 1. Introduction

Nowadays, over 350 000 compounds are registered for production and use and thus potentially released in environmental compartments (Wang et al., 2020). The study of human-made complex chemical mixtures (CCMs) is of interest to identify drivers of toxicity in surface waters (Altenburger et al., 2019). Rather than specific individual molecules, CCMs are often composed of multiple congeners of molecules of the same family. These mixtures are often ambiguous in their description or sometimes even confidential (Wang et al., 2020). Furthermore, these already ill-known molecules that make up CCMs are likely transformed through photolysis or biological processes. The resulting transformation products (TPs) constitute yet another layer of uncertainty. These CCMs and their TPs can then be accompanied by natural organic matter in surface waters which further complexifies the study of their occurrence. There is thus a concerning lack of knowledge regarding the occurrence, fate, and toxicity of these numerous compounds in the different environmental compartments.

Generally, the structural elucidation of unknown transformation products is carried out through "bottom-up" workflows. In these approaches, priority contaminants, consumer product additives (CPAs) such as polymer additives and surfactants, pesticides, and pharmaceuticals are submitted to various degradation pathways after which their transformation products are identified. Then, the environmental occurrence of the transformation products can be confirmed. As an example, 2-pyrrolidone, the hydrolysis transformation product of vinylpyrrolidone, was identified after being generated in laboratory experiments and then confirmed in retrospective analysis to be a widespread contaminant (Zahn et al., 2019). While "bottom-up" workflows allow to identify with high confidence unknown transformation products, it remains highly time- and resource-intensive. In the context hundreds of thousands of compounds being discharged in the environment and thousands of new ones being created each year, these "bottom-up" studies cannot keep up with the overwhelming burden of work that studying the fate and occurrence of all these compounds would entail. As such, only priority compounds like high production volume chemicals or with high concern for toxicity can realistically be selected in "bottom-up" studies.

However, there are online tools that can predict transformation products such as enviPath (Wicker et al., 2016), BioTransformer (Djoumbou-Feunang et al., 2019), and the Chemical Transformation Simulator (U.S. Environmental Protection Agency, 2022). The generated transformation products can then be added to suspect lists for suspect screening (Hollender et al., 2017). This combinatorial "bottom-up" approach has been incorporated into suspect and non-targeted screening (NTS) workflows of surface water in Europe (Bletsou et al., 2015; Li et al., 2017). However, this can lead to the omission of potentially important transformation products and to a "combinatorial explosion" where too many predicted transformation products are generated through multiple simulated degradation processes (Zahn et al., 2019).

These issues are considerable challenges that the scientific community is facing when conducting the identification of transformation products and chemical congeners mixtures of CPAs by NTS. Nevertheless, there are tools and software available that can help resolve these issues. The Global Natural Products Social Molecular Networking (GNPS) is an open-access web-based platform that groups compounds that share similar mass spectra into molecular networks (Wang et al., 2016). GNPS has been widely used in the study of natural products (Hebra et al., 2020; Olivon et al., 2017) and in metabolomics (Ernst et al., 2019; Quinn et al., 2017; Sedio et al., 2018). Still, molecular networking has seen little use for non-targeted analysis of surface waters. In one instance it was used to identify an unknown transformation product of the pharmaceutical telmisartan in a NTS assay and it also helped to identify in batch congeners of octylphenol ethoxylate (Eysseric et al., 2021). In another NTS study, it was used to identify several TPs of pharmaceutical compounds (Oberleitner et al., 2021).

There are other software and platforms that can help grouping compounds that share properties such as retention times and peak area. XCMS online (Gowda et al., 2014) and Compound Discoverer have been

used in these purposes in metabolomics (Hemmer et al., 2020) using hierarchical clustering analysis. They, however, only use $MS^1$ data. On the other hand, the open-source R package CluMSID generates distance matrixes for each precursor ion from $MS^2$ data and can thus help toward the identification of similar compounds from their spectra (Depke et al., 2017; Depke et al., 2019). As such, CluMSID and the molecular networking from GNPS can be used to assist the identification of transformation products and congeners. Both tools can then be incorporated in NTS workflows after the high resolution $MS^2$ spectra database search. Finally, the use of *in-silico* spectra matching algorithms such as MetFrag (Ruttkies et al., 2016) and SPS (Sweeney, 2014) allows employing large libraries of compounds. Those approaches have been recently used in surface water analysis (Eysseric et al., 2021; Ferrer et al., 2020; Gago-Ferrero et al., 2018; Lai et al., 2021).

The development of "top-down" approaches, where the structural elucidation of previously unknown TPs and congeners from CCMs is carried directly on the sample, would help to considerably alleviate the burden on "bottom-up" approaches to identify new TPs. Furthermore, top-down workflows could partly address the concerning issue of identifying congeners of CPAs from CCMs without relying on suspect lists.

The objective of this study was to evaluate the capacity of a "top-down" workflow to directly elucidate the structure of TPs and identify congeners of CPAs with a high level of confidence. To do so, NTS of water samples from a local river impacted with industrial, urban, and agricultural contamination sources was carried out. Then, molecular networking from GNPS and CluMSID were used to identify congeners and transformation products from CCMs.

## 2. Materials and Methods

### 2.1 Reagents and standards

Water, acetonitrile (ACN), methanol (MeOH), and formic acid were all LC-MS Optima grade and were obtained from Fisher Scientific (Waltham, MA, USA). Information about standards is shown in the Supplementary Material.

### 2.2 Collection and preparation of samples

Water samples (1000 mL) were collected from the Yamaska River upstream and downstream the wastewater treatment plants of Cowansville, Farnham and Saint-Hyacinthe (QC, Canada) on July 11, 2019; a satellite view the sampling points can be seen in Figure S-1 (Supplementary Material). Since the objective of this study was to evaluate the potential of the proposed top-down workflow to identify TPs and congeners of CPAs, representative sampling, which would have required a much larger sample size with multiple sampling replicates, was not necessary considering the scope of the article. Amber-coloured high-density polyethylene bottles were used for the sampling and kept in an ice cooler until arrival at the laboratory where they were immediately stored at -20°C. Prior to the extraction, the samples were thawed at room temperature, filtered through 1.2 µm glass fibre APFC prefilters and then through 0.45 µm mixed cellulose ester membranes, both from Millipore-Sigma (Oakville, ON, Canada). The samples were concentrated 250 mL on Strata-X polymer solid-phase extraction cartridges (200 mg, 6 mL) from Phenomenex (Torrance, CA, USA) and then eluted with 2 × 3 mL of a 1:1 (v/v) 2% formic acid solution of ACN-MeOH. The eluates were evaporated under a nitrogen stream and reconstituted to 625 µL, which amounts to a preconcentration factor of 400. While multi-layered SPE cartridges combining different sorbent chemistries have been used in the past (Gago-Ferrero et al., 2015; Köke et al., 2018; Moschet et al., 2013) and show high recoveries, especially for highly polar compounds, the Strata-X sorbent [poly(styrene-divinylbenzene) modified with *N*-vinylpyrrolidone] is able to obtain acceptable recoveries

(>75%) for a wide range of compounds in surface waters (Segura et al., 2019). The steps involved in the sampling and the preparation of the samples can be seen in Figure S-2 (Supplementary Material).

### 2.3 Quality control

A composite field blank made in all sampling points was prepared with Optima LC-MS water at each station and an instrumental blank was prepared prior to the injection of the sequence. No isotopically labeled standards nor spikes were used in this study. The composite field blank and the instrumental blanks were analyzed at the beginning and the end of the sequence to account for potential carry over. Both blanks were used for background signal subtraction to filter out possible sampling and laboratory contaminants. Features that were present in the blanks and whose peak areas were less than 4 times higher in the samples were removed from the peak list. The measures and steps used in this study for quality control are shown in Figure S-2 (Supplementary Material). To improve the transparency and reproducibility of this study, the *Nontargeted Analysis Study Reporting Tool* was used for this study based on the article by Peter et al. (2021). An Excel file downloaded from the website of Benchmarking and Publications for Non-Targeted Analysis (https://nontargetedanalysis.org/SRT) is available in the Supplementary Data (NTA_SRT_wPlot-and-ScoreTable.xlsx).

### 2.4 Instruments and methods

A Thermo Scientific Q-OrbitrapMS model Q Exactive Plus Orbitrap (San Jose, CA, USA) was interfaced with a Thermo Scientific UHPLC system using a pneumatic assisted heated electrospray ion source. The analytical settings used were the same as those used in a previous study in another municipality along the Yamaska River (Eysseric et al., 2021). MS detection was performed in the positive ion mode using Top 10 Data Dependent Acquisition (DDA). A DDA cycle entailed one $MS^1$ survey scan (*m/z* 100-1000) acquired at a full width at half maximum resolution ($R_{FWHM}$) of 35 000 and precursors ions meeting user defined criteria for monoisotopic precursor intensity (dynamic acquisition of $MS^2$ based Top 10 most intense ions with a $2\times10^5$ AGC target). The frequency of acquisition was

10 Hz. Precursor ions were isolated using the quadrupole (2 Da isolation width) and activated by higher-energy collision dissociation using stepped normalized energy (25, 35 and 45 units) and fragment ions were detected in the Orbitrap at $R_{FWHM}$=17 500. $R_{FWHM}$ parameters were selected to maximize the frequency of acquisition. With a Top10 DDA method and short chromatographic peak widths (10-20 s), high frequency acquisition is crucial. Dynamic exclusion was set to auto to filter out background signal, noise, and instrument contamination.

Instrument calibration was performed prior to all analyses and mass accuracy was notably below 1 ppm using the Thermo Pierce calibration solution and the automated instrument protocol. Source parameters were the following: capillary temperature was 300 °C; sheath gas was 50; auxiliary gas was 20; spray voltage was 4000 V. The liquid chromatographic column was a Waters Acquity UPLC HSS T3 (2.1 × 50 mm, 1.8 µm) and the mobile phase was composed of water with 0.1% (v/v) formic acid (solvent A) and MeOH-ACN (3:2, v/v) with 0.1% (v/v) formic acid (solvent B). The gradient elution program, according to volume percent of solvent B in the mobile phase, was the following: 0 min, 2%; 17 min, 100%; 21 min, 100%; 21.01 min, 2%; 25 min, 2%. Total run time was 25 min. Mobile phase flow rate was 350 µL $min^{-1}$ throughout the run and the injection volume was 2 µL. The instrumental parameters for the LC-MS acquisition in this study can be seen in Figure S-2 (Supplementary Material).

## 2.5 Software parameters

The identification was realized with a multi-tool approach recently developed using two *in-silico* high-resolution tandem mass spectrometry databases, MetFrag and the Similar Partition Algorithm (SPS) along with the Global natural products social networking (GNPS) (Eysseric et al., 2021). Settings for each of the tools are shown in the Supplementary Material as well as in Figure S-2 which summarizes the full workflow used for data treatment. The R (version 4.1.1) package CluMSID (Clustering of $MS^2$ Spectra for Metabolite Identification) version 1.6.0 was used to generate distance matrixes for data analysis along with dendrograms, clusters and other figures for data visualization (Depke et al., 2019). The R script containing the parameters for the figures is available in the Supplementary Material. CluMSID is complementary to GNPS since it operates offline and it builds a data matrix that allows to see the similarity of one spectrum with all the other ones unlike GNPS which is an online platform that only shows the degree of similarity between compounds that equals or exceeds the cosine score threshold. Throughout the text the term "clusters" will be used to refer to results associated to ClumSID and "molecular networks" to results obtained from GNPS.

## 2.6 Levels of confidence

Annotations carry different level of confidence in the identification. They were given based on Schymanski previous work on the matter (Schymanski et al., 2014). All structures in this study had a maximal deviation of 5 ppm for mass accuracy. For MetFrag and GNPS, a minimum of 4 matched peaks with the libraries was required to generate an identification. A score (the quality of the $MS^2$ match based on the difference between the reference and experimental spectra) of 70 was needed for SPS, a cosine score (a scalar product of two spectra represented as vectors where 1 is a complete similarity) of 0.7 was needed for GNPS and a score of 5 was needed for MetFrag. All spectra matches were manually inspected to reduce the number of false positives. Furthermore, the isotopic pattern was used to generate molecular formulas with GenForm on patRoon which was part of the score calculation performed by the patRoon tool. The criteria for identification can be seen in Figure S-2.

The "confirmed structure" level of confidence was given to compounds that were confirmed with reference standards. The "probable structure" level was given to compounds that either had an unambiguous match with an $MS^2$ library and/or enough diagnostic evidence such as experimental context, diagnostic $MS^2$ fragment ions. The third level, “tentative candidate", was given when there was a strong candidate structure either through a library match or diagnostic evidence for a compound, but not enough to unambiguously match a structure to a feature. This was seen when multiple library matches for a single feature had close scores.

All matches in the probable structure and tentative candidate levels of confidence had to carry environmental relevance or be likely to be found in the samples. For example, a pharmaceutical compound like rofecoxib that was withdrawn over 15 years ago would not be selected in either category despite a good library match. Similarly, a match for a compound like anthracene that would be unlikely to show affinity for the positive mode of electrospray ionization and that would have a drastically different chromatographic behavior would be filtered out.

# 3. Results

A total of 438 compounds in the 6 sampling sites along the Yamaska river were detected. Of those, 53 carry a confirmed structure level of confidence (Table 1), 258 carried a probable structure level of

confidence and 127 were tentative candidates. All compounds with their level of confidence, monoisotopic mass, super class, class, and their frequency of detection per sampling site can be seen in the Supplementary Data (IdentifiedCompounds.xlsx).

The chemicals were classified into five superclasses: consumer product additives (CPAs), illicit drugs, natural products, pesticides, and pharmaceuticals, based on the metadata in their PubChem and US-EPA Comptox Chemistry Dashboard profiles. CPAs were divided into cosmetics, food additives, polymer additives and surfactants. Natural products were divided into animal metabolites, plant metabolites, and toxins. Pesticides were subdivided into more specific classes: herbicides, insecticides, fungicides, and plant growth regulators. Pharmaceutical compounds were subdivided according to their Anatomical Therapeutic Chemical (ATC) Classification code (World Health Organization, 2021). Temperature, dissolved oxygen, conductivity and pH were also taken as water characteristics at each point (Table S-1, Supplementary Material).

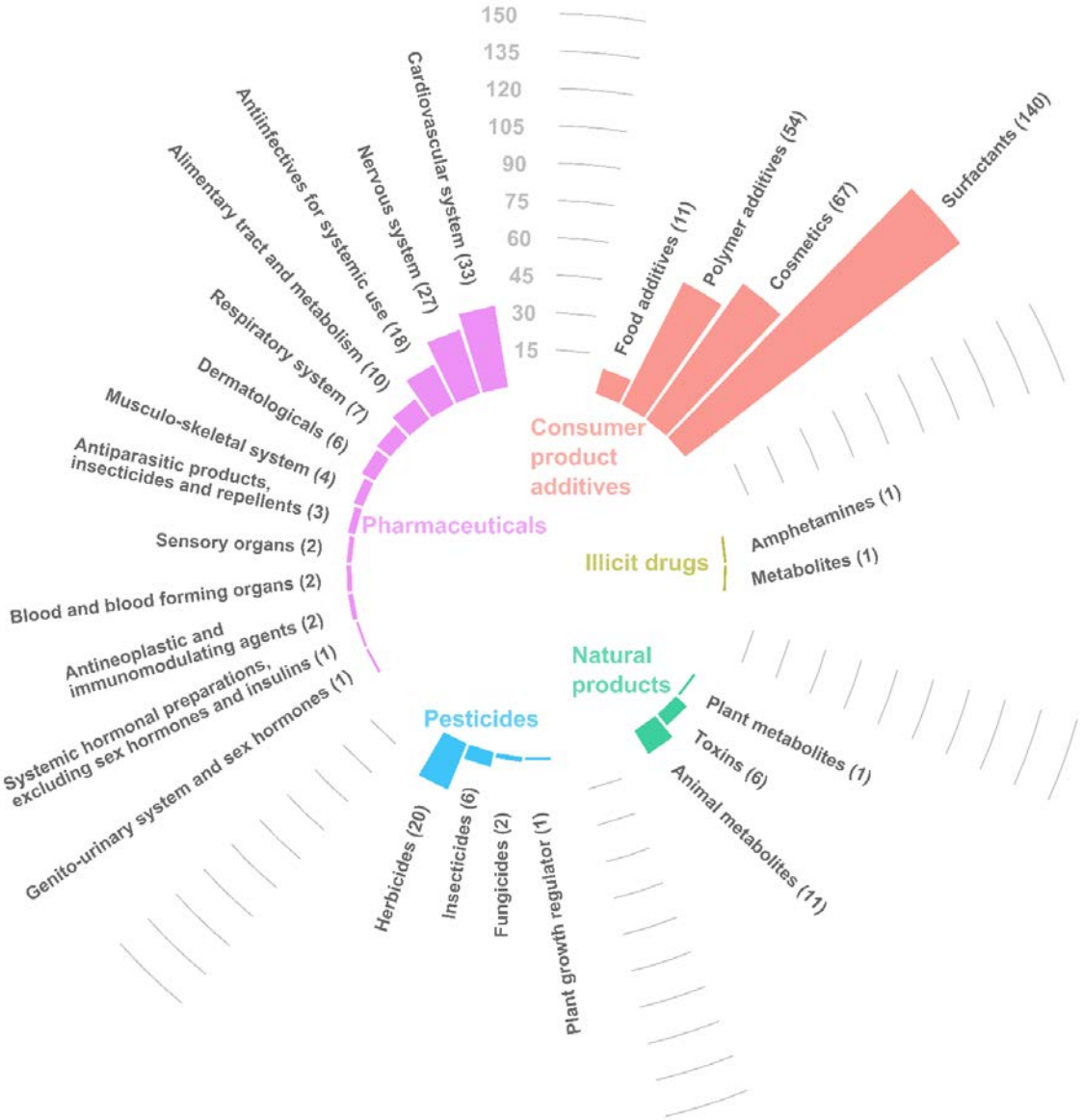



**Figure 1.** Total compounds tentatively identified and confirmed by superclass and class in all sampling points

3.1. Consumer Product Additives

Consumer product additives (CPAs) were the most numerous of the superclasses with 273 tentative identifications and 27 confirmed structures (Figure 1). CPA contamination was generalized to all points

ranging from 186 compounds tentatively identified or confirmed downstream of Cowansville to 244 downstream of Farnham (Figure 2).

**Table 1.** List of compounds confirmed with reference standards.

| Compound name | **Superclass,** class |
|---|---|
| 2,2,4-Trimethyl-1,2-dihydroquinoline (TMQ) | **CPA**, Polymer additive |
| 2,2,6,6-Tetramethyl- 4-piperidinol | **CPA**, Polymer additive |
| Acetaminophen | **Pharmaceutical**, Nervous system |
| Acetyltributyl citrate | **CPA**, Polymer additive |
| Aliskiren | **Pharmaceutical**, Cardiovascular system |
| Atrazine | **Pesticide**, Herbicide |
| Benzotriazole-1H | **CPA**, Polymer additive |
| Benzotriazole-5-methyl-1H | **CPA**, Polymer additive |
| Benzoylecgonine | **Illicit drugs**, Metabolite |
| Caffeine | **Pharmaceutical**, Nervous system |
| Carbamazepine | **Pharmaceutical**, Nervous system |
| Citalopram | **Pharmaceutical**, Nervous system |
| Desethylatrazine | **Pesticide**, Herbicide |
| Diethyltoluamide | **CPA**, Cosmetic |
| Diltiazem | **Pharmaceutical**, Cardiovascular system |
| Dimethenamid | **Pesticide**, Herbicide |
| Diphenhydramine | **Pharmaceutical**, Respiratory system |
| Diphenylguanidine | **CPA**, Polymer additive |
| Ditolylguanidine | **CPA**, Polymer additive |
| Erucamide | **CPA**, Polymer additive |
| Fexofenadine | **Pharmaceutical**, Respiratory system |
| Gliclazide | **Pharmaceutical**, Alimentary tract and metabolism |
| Irbesartan | **Pharmaceutical**, Cardiovascular system |
| Ketamine | **Pharmaceutical**, Nervous system |
| Lauramidopropyl betaine | **CPA**, Cosmetic |
| Lauryldiethanolamide | **CPA**, Cosmetic |
| Lauryldiethanolamine | **CPA**, Cosmetic |
| Losartan | **Pharmaceutical**, Cardiovascular system, |
| MDMA | I**llicit drugs**, Amphetamine |
| Metolachlor | **Pesticide**, Herbicide |
| Metribuzin | **Pesticide**, Herbicide |
| O-Desmethylvenlafaxine | **Pharmaceutical**, Nervous system |
| OPEO-3 to OPEO-15 | **CPA**, Surfactant |
| Oxazepam | **Pharmaceutical**, Nervous system |
| Oxybenzone | **CPA**, Cosmetic |
| Paraxanthine (caffeine metabolite) | **Pharmaceutical**, Nervous system |
| Rosuvastatin | **Pharmaceutical**, Cardiovascular system, |
| Sitagliptin | **Pharmaceutical**, Alimentary tract and metabolism |
| Trimethoprim | **Pharmaceutical**, Antiinfective for systemic use |
| Tris(2-butoxyethyl) phosphate | **CPA**, Polymer additive |
| Valsartan | **Pharmaceutical**, Cardiovascular system |

**Table 2.** List of networks or clusters of chemicals that were tentatively identified and confirmed with reference standards per class and superclass.

| Cluster family | Number of members identified | Superclass (class) | Molecular network or cluster |
|---|---|---|---|
| Fatty amides | 4 | CPA (polymer additives) | Figure S-3 |
| Polyethylene glycols | 24 | CPA (cosmetics) | Figure S-4 |
| Betaines and alkylamidopropyl dimethylamines | 11 | CPA (cosmetics) | Figure 3 |
| Diethanolamines and diethanolamides | 6 | CPA (cosmetics) | Figure S-5 |
| Polyoxyethylene alkyl ethers and esters | 64 | CPA (surfactants) | Figure S-6, Figure S-7 and Table S-2 |
| Octylphenol ethoxylates | 16 | CPA (surfactants) | Figure S-8 |
| Alkylphenols ethoxylates acids | 21 | CPA (surfactants) | Figure S-9 |
| Metolachlor transformation products | 3 | Pesticides (herbicides) | Figure S-11 |
| Beta-blockers and transformation products | 4 | Pharmaceuticals (cardiovascular system) | Figure S-13 and Figure S-14 |
| Irbesartan and transformation products | 5 | Pharmaceuticals (cardiovascular system) | Figure S-14 and Figure S-16 |
| Diltiazem and transformation products | 4 | Pharmaceuticals (cardiovascular system) | Figure S-14 and Figure S-17 |
| Gliclazide and transformation products | 2 | Pharmaceuticals (Alimentary tract and metabolism) | Figure S- 18 |

CPA: Consumer product additive.

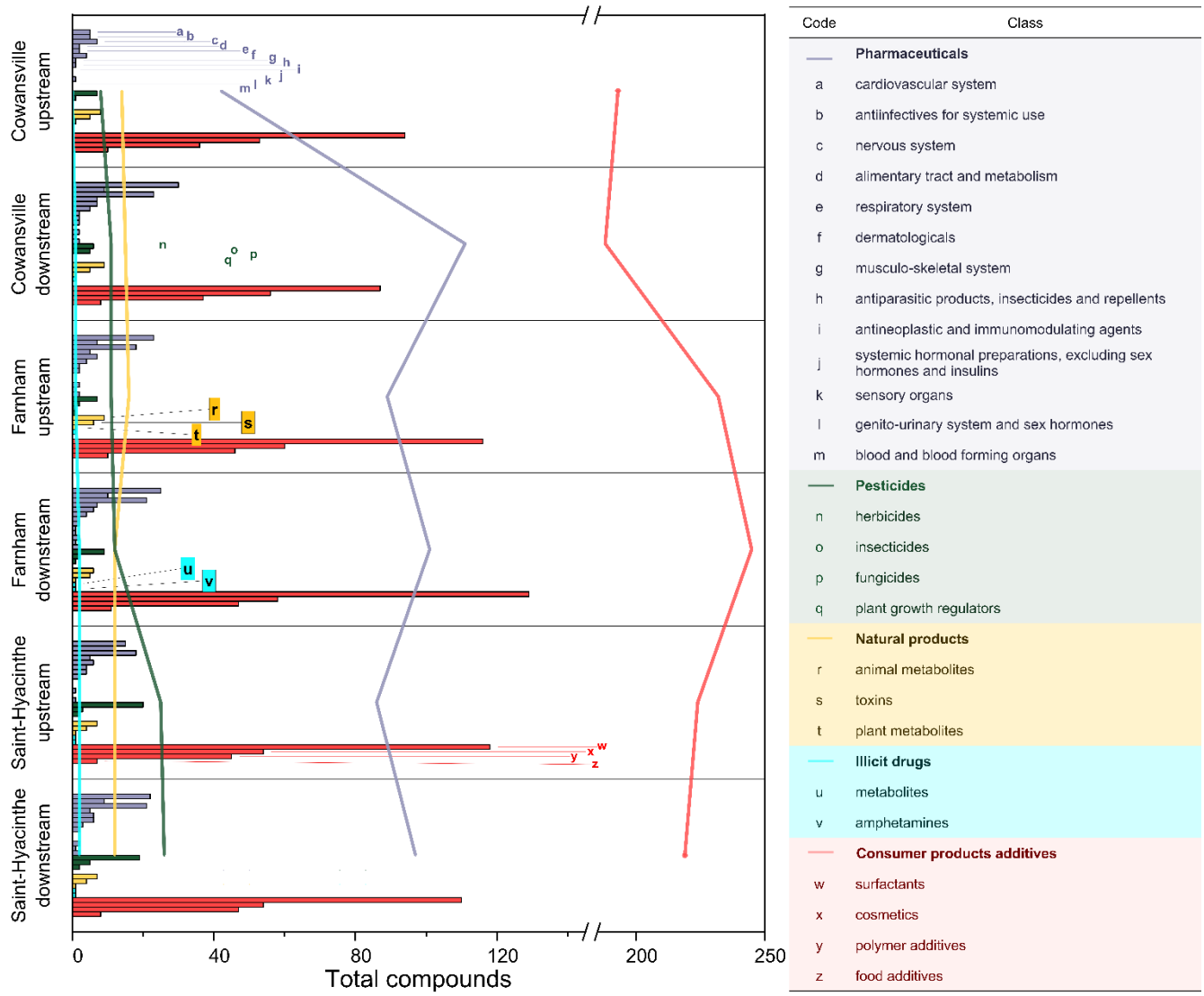


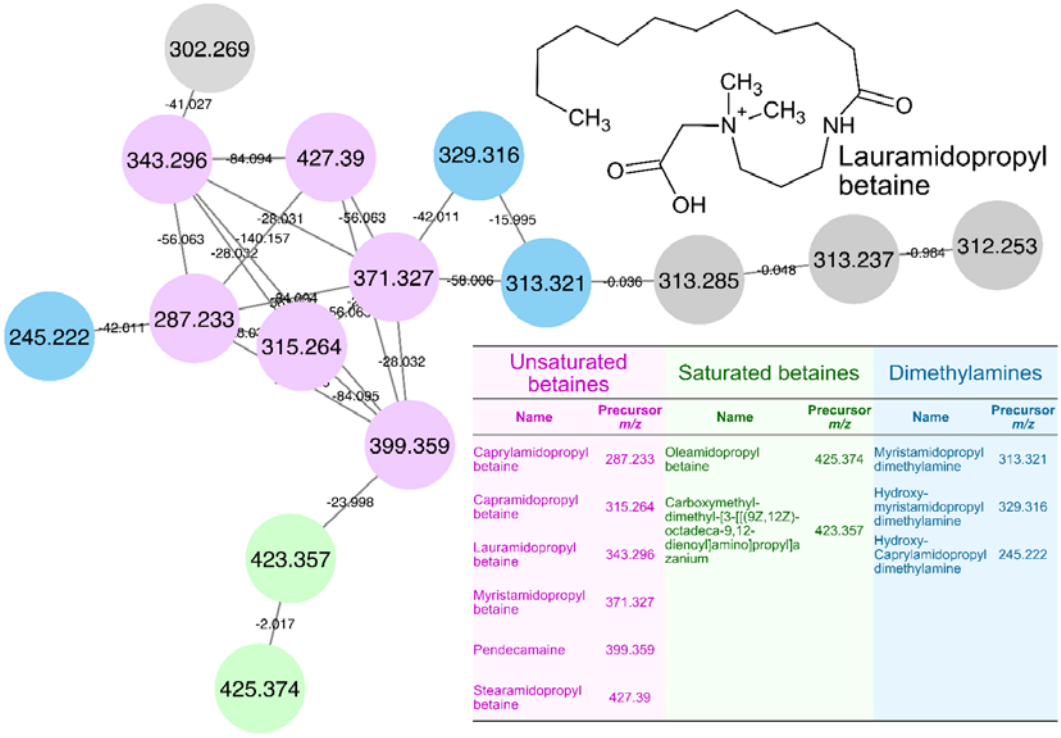

**Figure 2.** Frequency of detection of confirmed and tentatively identified compounds per station by superclass and class

Food additives included among others the emulsifiers sucrose palmitate and polysorbate 40, 60, and 80, along with the Maillard reaction product 5-hydroxymethylfurfural as tentative candidates.

Fifty-four polymer additives were identified or confirmed with reference standards (19 tentative candidates, 27 probable structures, 8 confirmed structures). The class included plasticizers, flame retardants, lubricants, heat and light stabilizers, antioxidants, antiozonants and vulcanization accelerators among others; several of these chemicals are high production volume (HPV) chemicals in Canada per the CompTox Chemistry Dashboard of the US EPA. The vulcanization accelerators diphenylguanidine and ditolylguanidine, that were detected in all points, have been found to originate from tire wear particle leachates (Sieira et al., 2020; Zahn et al., 2019). Another HPV compound related to tire wear was the antioxidant 2,2,4-trimethyl-1,2-dihydroquinoline, also known as TMQ. The transformation product 2,2,6,6-tetramethyl-4-piperidinol, resulting form the hydrolysis of the HPV light stabilizer 4-hydroxy-2,2,6,6-tetramethylpiperidine-1-ethanol, was also tentatively identified. Erucamide, an HPV fatty amide used as lubricant in polymers, was found in a molecular network which allowed to identify 3 other congeners (Figure S-3, Supplementary Material). Five transformation products of the flame retardant tris(2-butoxyethyl)phosphate (TBEP) as well as the parent compound itself were detected in all points. The TPs were manually searched after their structures were elucidated in a previous bottom-up study

(Eysseric et al., 2022). The TPs had been generated in laboratory and tentatively identified after a photolysis experiment of TBEP after which their environmental occurrence was confirmed in this current NTS.

Sixty-seven cosmetics were either tentatively identified or confirmed with reference standards. Whereas food additives and polymer additives were composed of diverse compounds and their TPs, the cosmetics are mostly composed of families of congeners. These compounds shared highly similar $MS^2$ spectra and were grouped in clusters and in networks which allowed to identify them in batch. In the first case, twenty-four polyethylene glycols (PEG) congeners were grouped in three separate molecular networks (Figure S-4, Supplementary Material). PEGs have a very wide variety of uses in cosmetics such as solvents in cologne, hair fixatives, and nails lacquers or as emulsifiers in shampoos and conditioners (Rieger, 2009). Congeners from PEG-3 to PEG-28 were detected in all points. Five polypropylene glycol (PPG) congeners from pentapropylene glycol to nonapropylene glycol were also identified with a level of confidence of 2a. Additionally, nine congeners of alkylamidopropyl betaines and two alkylamidopropyl dimethylamines (by-products from the manufacturing of betaines) were clustered in a molecular network (Figure 3). The compound lauramidopropyl betaine was confirmed with a reference standard (Table 1); the other compounds in the network are probable identifications considering the high degree of similarity they shared with it in terms of retention time and $MS^2$ spectra. These compounds are used in shampoos, conditioners, skin moisturizers, and skin cleansers. To the knowledge of the authors, only lauramidopropyl betaine and myristamidopropyl betaine have been reported in surface and waste waters before (Beckers et al., 2020; Peng et al., 2018) which means that we report 9 new betaine related compounds in this paper. Another network of diethanolamines and diethanolamides, also used in shampoos as foam boosters was also found (Figure S-5, Supplementary Material). Once again to the knowledge of the authors, only lauryldiethanolamide has been reported in waste and surface waters in the literature (Beckers et al., 2020; Peng et al., 2018).




| Unsaturated betaines | | Saturated betaines | | Dimethylamines | |
|---|---|---|---|---|---|
| Name | Precursor *m/z* | Name | Precursor *m/z* | Name | Precursor *m/z* |
| Caprylamidopropyl betaine | 287.233 | Oleamidopropyl betaine | 425.374 | Myristamidopropyl dimethylamine | 313.321 |
| Capramidopropyl betaine | 315.264 | Carboxymethyl-dimethyl-[3-[[(9Z,12Z)-octadeca-9,12-dienoyl]amino]propyl]azanium | 423.357 | Hydroxy-myristamidopropyl dimethylamine | 329.316 |
| Lauramidopropyl betaine | 343.296 | | | Hydroxy-Caprylamidopropyl dimethylamine | 245.222 |
| Myristamidopropyl betaine | 371.327 | | | | |
| Pendecamaine | 399.359 | | | | |
| Stearamidopropyl betaine | 427.39 | | | | |

**Figure 3.** Molecular network of cosmetics betaines and betaine related compounds. In grey are the unannotated precursors that are interference isobars of *m/z* 313.321 that were selected to the quadrupole at the same time. The structure of lauramidopropyl betaine, which was confirmed with a reference standard, is shown.

Surfactants represented by far the biggest class with 140 compounds that were either tentatively identified or confirmed with reference standards. The occurrence of these species was observed in nearly all sampling points (Figure 2). These surfactants are part of multiple very large families of PEG based congeners. A molecular network (Figure S-6, Figure S-7, Supplementary Material) and a cluster (Table S-2, Supplementary Material) gathered what amounted to 64 PEG alkyl ethers and esters allowing the tentative identification of 6 subgroups of alkyl PEG ethers with aliphatic chains length of 10, 11, 12, 13, 14, and 15 carbon atoms and 5 subgroups of alkyl esters with aliphatic chains length of 11, 12, 13, 14, and 15 carbon atoms. PEG alkyl ethers and esters are widely used as lubricants in textile processing, as emulsifiers in metal working fluids and as solvent cleaners (Pfaendner, 2019). This particular network really highlights the power of the molecular networking tool to identify congeners and especially congeners of PEG which share highly similar $MS^2$ spectra. A widespread contamination of alkylphenol ethoxylates, non-ionic surfactants that were identified in the Yamaska river in a recent work (Eysseric et al., 2021) was again identified. Octylphenol ethoxylates (OPEOs) congeners ranging from OPEO-3 to OPEO-19 were found in a molecular network (Figure S-8, Supplementary Material) and confirmed with reference standards. Additionally, 27 carboxylic acid TPs of OPEOs and closely related nonylphenol ethoxylates (NPEO) in another molecular network (Figure S-9, Supplementary Material) were tentatively identified. OPEOs and NPEOs have been known to biodegrade into carboxylic acid TPs under aerobic conditions (Komori et al., 2006). These compounds showed an inversed linear relationship between the number of ethylene oxide units and retention time which further strengthen the level of confidence in their identification (Figure S-10, Supplementary Material). Transformation experiments under controlled laboratory conditions as well as molecular modelling are necessary to elucidate the mechanisms leading to the formation of these TPs.

### 3.2. Illicit drugs, Pesticides and Natural Products

Of the 29 pesticides that were either tentatively identified or confirmed, herbicides were the most numerous with multiple common compounds such as atrazine (confirmed) and two of its metabolites desethylatrazine (confirmed) and 2-hydroxyatrazine. Metolachlor was confirmed with a reference standard while its TPs, metolachlor-ESA, metolachlor-OA, and metolachlor morpholinone, were annotated with an empirical library. Two additional TPs of metolachlor were tentatively identified, metolachlor_TP250 and metolachlor_TP266, because they were grouped in a molecular network with metolachlor-OA (Figure S-11, Supplementary Material). The number of pesticides identified went up sharply at the stations upstream and downstream of Saint-Hyacinthe (Figure 2). This was to be expected considering the intense agricultural activity around the river upstream both these stations that can be appreciated with the satellite images of the sampling points (Figure S-1, Supplementary Material).

The amphetamine MDMA and the main metabolite of cocaine, benzoylecgonine, were confirmed. The natural products tentatively identified included 6 toxins (Figure 1). Among them are the cyanotoxin lyngbiatoxin-C and the couple of mycotoxins zearalenone and zearalenol that are tentative candidates. Lingbyatoxin 1 and lingbyatoxin-6 were observed previously in benthic Lyngbya wollei algae samples collected in the St. Lawrence River (Lajeunesse et al., 2012). Zearalenone could be a source of concern because of its estrogenic activity (Rogowska et al., 2019). The number of natural products detected stayed relatively similar across all points (Figure 2).

### 3.3. Pharmaceuticals

A total of 116 pharmaceuticals subdivided into 13 classes (Figure 1) were either tentatively identified or confirmed. The fluctuation in the detections between upstream and downstream the wastewater treatment plants of each sampling site can be appreciated in Figure 2. The effect was most marked in Cowansville,

which is to be expected since it is the first sizeable city in this stretch of the river with a population of over 11 thousand inhabitants (Statistics Canada, 2017) and a regional hospital of 96 beds (Fondation de L'Hôpital Brome-Missisquoi-Perkins, 2022). Since all the following points are downstream the city of Cowansville, more pharmaceuticals were detected. Still the discrepancy between upstream and downstream could be observed but in a lesser manner.

The largest class of pharmaceuticals was the drugs for the treatment of the cardiovascular system with 33 compounds (Figure 1). Several transformation products that were never reported before to the knowledge of the authors were tentatively identified with the help of molecular networking and clustering tools (MNC). The transformation product metoprolol_TP282 results from hydroxylation followed by oxidation (Figure S-12, Supplementary Material) of metoprolol, a beta-blocker. Metoprolol_TP282 was located in a molecular network with 3 other beta-blockers including metoprolol which made its identification possible (Figure S-13, Supplementary Material) as well as next to metoprolol in the dendrogram of all precursors from downstream Cowansville, where it was detected (Figure S-14, Supplementary Material). Furthermore, one previously unknown TP of irbesartan was tentatively identified: irbesartan_TP445 is the result of hydroxylation at the end of the aliphatic chain (Figure S-15, Supplementary Material). Three other TPs that were found in the same network, irbesartan_TP443, the result of oxidation of the newly formed alcohol in irbesartan_TP445, irbesartan_TP459 the acid resulting from another hydroxylation on the same carbon, and irbesartan_TP387, that results from the loss of a propyl group (Figure S-15, Supplementary Material), had been tentatively identified in the past (Boix et al., 2016). These compounds were all grouped with MNC tools which made their identification possible despite them being absent in databases and irbesartan_TP445 being previously unknown (Figure S-16). Irbesartan is a widely consumed pharmaceutical partly removed during wastewater treatment (Boix et al., 2016). Finally, another molecular network of diltiazem and three of its TPs: desmethyldiltiazem, deacetyldiltiazem, and desmethyldeacetyldiltizaem (Figure S-17, Supplementary Material) were found. These TPs had been tentatively identified in a previous study in the Yamaska River (Eysseric et al., 2021).

The TP hydroxyatorvastatin-lactone was tentatively identified. This compound was also identified in the bottom-up study realized by Eysseric et al. (2022) as the parent compound atorvastatin had also been submitted to photolysis in laboratory settings. Atorvastatin, despite its common use, was not detected in any sample. This finding illustrates the importance of bottom-up studies as it would not have been identified with current top-down tools. Rosuvastatin, another statin pharmaceutical, along with the transformation product rosuvastatin lactone were tentatively identified. Only rosuvastatin was initially annotated following a $MS^2$ library match. It is upon looking at the dendrogram (Figure S-14, Supplementary Material) and the distance matrix for downstream Cowansville, where both features were detected, that a mass corresponding to the net loss of $H_2O$ in the formula was observed. A manual inspection of the $MS^2$ spectrum of the compound in addition to contextual evidence from past studies (Lee et al., 2009; Machado et al., 2015; Sulaiman et al., 2015) allowed assigning a probable structure.

The second largest class was the drugs classified as affecting the nervous system with 27 compounds (8 confirmed structures, 14 probable structures, and 4 tentative candidates). It includes several contaminants commonly found contaminants in surface waters such as acetaminophen, citalopram, oxazepam, carbamazepine, and caffeine, all which were confirmed. The TPs O-desmethylvenlafaxine and paraxanthine were also confirmed along with ketamine which also has recreative use. Gamma-aminobutyric acid (GABA), lidocaine and its TP N-desethyllidocaine, lamotrigine, methocarbamol, and three carbamazepine TPs were also tentatively identified as probable structures.

The 18 anti-infectives for systemic use that were tentatively identified and confirmed (Figure 1) were from five different antibiotic classes: aminoglycosides, cephalosporins, lincosamides, macrolides, and sulfonamides. Trimethoprim was confirmed with a reference standard, seven compounds were probable structures, and ten other compounds were tentative candidates. This can be concerning when considering that the distribution of antibiotic resistance genes is related to riverine inflows of antibiotics (Liang et al.,

2020). The list of all compounds confirmed and tentatively identified is shown in the Supplementary Data (IdentifiedCompounds.xlsx).

The alimentary tract and metabolism drug gliclazide which is used in the treatment of diabetes was confirmed with a reference standard. It was grouped in a small cluster of two compounds with a transformation product resulting from the formation of an acid (Figure S-18, Supplementary Material). To the authors' knowledge, this TP was previously unknown and unreported in the literature.

## 4. Discussion

The networking and clustering tools proved to be highly efficient when it came to identifying unannotated or even completely unknown TPs. An important caveat was that the parent compound had to be detected as well as linked in the network or cluster to realize the identification of the TPs. This was seen with irbesartan, diltiazem, rosuvastatin, gliclazide, metoprolol, metolachlor and citalopram where 16 TPs were identified among these compounds including several unknown ones, as can be seen in Table 2. However, in the cases where there was a single TP while the parent compound was not detected, networking and clustering tools could not assist toward the identification. This is because the TPs could not be connected to a similar compound. As such, the single TPs were identified either through an annotation from a high-resolution tandem mass spectra database, like celecoxib carboxylic acid, benzoylecgonine and clindamycin sulfoxide, or because they were previously identified in a bottom-up study realized by the authors, like hydroxy-atorvastatin lactone and the five transformation products of TBEP.

Furthermore, when the structure of the TPs was similar to the respective parent compound, they could be efficiently linked in a network or a cluster. In the cases where the transformation products were the result of reactions on the parent compound such as hydroxylation (irbesartan, metoprolol), oxidation (metoprolol, irbesartan, alkylphenol ehthoxylates acids) or dealkylation (irbesartan, diltiazem, citalopram), they could still be grouped by the networking and clustering tools. A mass bias affecting the formation of clusters and networks was also observed. Congeners over 400 Da containing polyoxyethylene units were generally in molecular networks. However, in instances where the molecular weight of a compound and its transformation product were lower than around 400 Da, they were less likely to be grouped together. For example, neither were atrazine and desethylatrazine nor caffeine and paraxanthine clustered despite being closely related structurally. Indeed, precursors with higher *m/z* were more likely to be part of a molecular network than ones with a lower *m/z* value as shown in Figure 4 where 47% of the precursors with a *m/z* equal and under 400 were in a network whereas 63% of the compounds over 400 *m/z* were in a molecular network. This could be explained by the fact that, generally, larger compounds can be fragmented into more product ions than smaller compounds thus they can be more easily grouped together by the algorithms that had a minimum of matching product ions of 4. There is also a high number of polyoxylethylene homologues in the samples which are at large over 400 Da and may skew the trend. Furthermore, we hypothesize that higher collision energies (stepped energy of 25, 35 and 45 units were used in the present workflow) could generate more fragments and thus reduce this threshold. However, such hypothesis was not tested since it was outside the scope of this study.

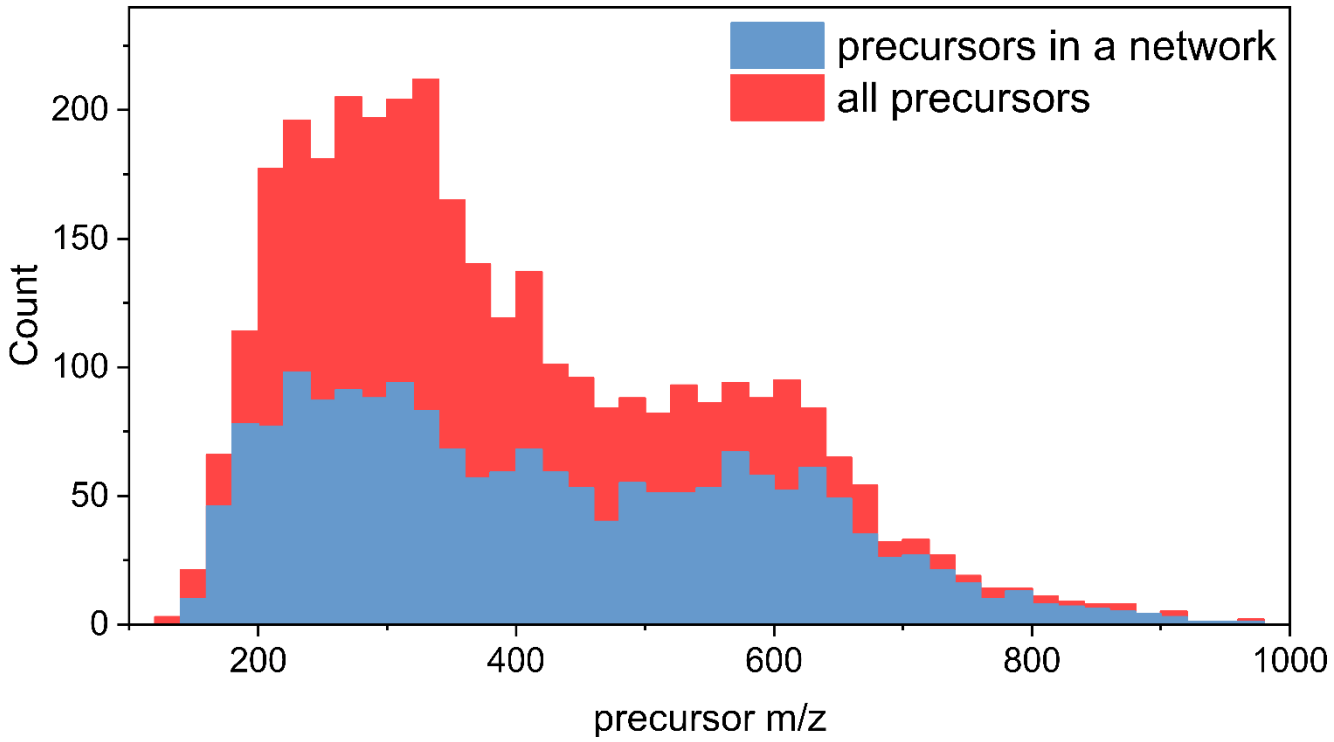



**Figure 4.** Repartition of the number of precursors in a network per precursor *m/z* on the total number of precursors across all sampling points. The background signal, noise and blank features were removed and are not presented in this figure.

Networking and clustering tools showed powerful capacities when it came to uncovering the structures of oligomers such as alkylphenol ethoxylates, polyoxyethylenes, polyoxyethylene alkyl ethers, and polyoxyethylene alkyl esters. The tools proved to be particularly useful since only a portion of the compounds in the networks were registered in a chemical repository. This was especially the case for polyoxyethylene alkyl ethers and esters that for a significant part did not have an entry on PubChem nor Chemspider. This means that even *in silico* tools such as MetFrag or SPS (both used in this study) could not have supplied an identification since they use these large chemical repositories as a source. A total of 125 compounds containing multiple polyoxyethylene units were tentatively identified or confirmed in the multiple networks and clusters that can be seen in Table 2. There was little to no ambiguity when it came to the assessment of a structure for these compounds because of the numerous product ions from the high resolution $MS^2$ spectra, network information, and other diagnostical evidence such as the linear retention time pattern in a family of congener seen in Figure S-10. All the polyoxyethylene congeners can be seen in an Excel File in the Supplementary Data (IdentifiedCompounds.xlsx)

Still, a much higher number of precursors, which includes the annotated ones, shared polyoxyethylene units and thus highly similar $MS^2$ spectra. The dendrogram and heatmap of all the features detected in downstream Farnham in Figure S-19 (Supplementary Material) illustrates the number of precursors sharing multiple product ions. Over 30% of all precursors (1240 on 4023 total unique precursor ions) had at least two product ions originating from polyoxyethylene units while over 40% (1717 on 4023 total unique precursor ions) had at least one. The product ions shared by most of these compounds corresponded to the protonated molecules of two (89.060 ± 1mDa), three (133.086 ± 1mDa), four (177.112 ± 1mDa), five (177.112 ± 1mDa), and six (265.164 ± 1mDa) polyoxyethylene units. It should be noted, however, that several compounds had more than one adduct, notably in the form of protonated molecules and ammonium adducts as can be seen in Figure S-4, Figure S-7, Table S-2, and Figure S-8 (Supplementary Material). Componentization performed by the CAMERA package and GNPS showed 76 instances of overlapping ammonium adducts in the ethylene oxide cluster. As such, the total number containing polyoxyethylene units is between 1641 and 1717.

Isobaric interferences proved to negatively impact the performance of MNC tools. There were several instances in which these interferences were observed. In some cases, a coeluting compound or a background contamination whose precursor was within the quadrupole window selection range (2 Da) was selected for a $MS^2$ experiment. Isobaric precursors were observed in the molecular networks of betaines (Figure 3), alkyldiethanolamines and alkyldiethanolamides (Figure S-5, Supplementary Material), and metolachlor (Figure S-11, Supplementary Material). While these were cases of false positives relatively simple to assess, there might also have been cases of false negatives where the spectral interference caused precursors that should have been linked to be separated which can be much more challenging to address. A better chromatographic separation could be a solution to minimize the impact of coeluting species which could be achieved with a longer column, e.g., 150 mm. Matrix effects could also prove to be a problem if severe ionization suppression were to happen.

Regarding the individual performances of the MNC tool, GNPS offered a simpler user experience with the website interface and allowed to treat the networks with the Cytoscape software. However, it required four different interfaces. CluMSID required a steeper learning curve and longer calculation time while only allowing to analyze one file at a time which is a very significant drawback when working on large sequences. On the other hand, it operated offline and as such did not require an internet connection. The main inconvenience of both tools was how time-consuming they are and cannot be applied to routine analysis as of now.

## 5. Conclusion

A “top-down” workflow consisting of a non-targeted water analysis of a local river in Southern Canada allowed to identify 126 tentative candidates, 258 probable structures and confirm 53 compounds with reference standards for a total of 438 compounds. While this method is limited by a Top 10 DDA acquisition technique that detects only the 10 most abundant with a frequency of acquisition of 10 Hz, it does not necessarily detect the most toxic compounds in the samples. No single method can detect all relevant compounds in a sample, but the proposed method is a useful tool to improve current knowledge about the occurrence of nontargeted contaminants. By using different ion sources such as atmospheric pressure ionization or dielectric barrier discharge ionization (Lara-Ortega et al., 2018) as well as hydrophilic interaction liquid chromatography, the analytical capabilities of the method can be further expanded. Thus, the obtained data, combined to toxicity prediction based on quantitative structure-activity relationships such as Ecosar (U.S. Environmental Protection Agency, 2019) or deep learning (Tang et al., 2018) can be employed to sort out the most toxic compounds.

The use of molecular networking and clustering tools permitted to group together similar compounds and thus made possible the structural elucidation of multiple previously unknown TPs of pharmaceuticals and pesticides. The tools also helped to identify 176 congeners of compounds units originating from complex chemical mixtures found in consumer product additives where only 47 were annotated with the empirical and in silico $MS^2$ matching tools. A total of 37 alkylphenol ethoxylates and their carboxylic acid transformation products, known for their estrogenicity, were thus identified all at once. A very powerful use of the MNC was showcased while allowing to tentatively identify multiple congeners of polyoxyethylene ethers and esters that in multiple instances did not figure on PubChem. Expanding the reach of identification further is of great value in NTS assays as the number of commercially available compounds continues to increase. Similarly, facilitating the structural elucidation of unknown transformation products directly in environmental samples without having to rely on “bottom-up” studies in controlled laboratory settings helps to alleviate the pressure on the scientific community and even speed up the identification of unknown contaminants. Still, these studies remain crucial as there were numerous instances where MNC tools did not group transformation products and parent compounds despite them all being present. In those cases, $MS^2$ data resulting from bottom-up studies, like in the case of hydroxylated atorvastatin lactone, and $MS^2$ databases had to be relied upon for tentative identifications.

Since MNC tools are used at the end of a non-targeted analysis workflow, they offer impressive possibilities with regards to data mining and retrospective analysis of data-dependent experiments while working on all file formats. While time investment and level of specialization required to use these tools can be a barrier for now in routine analysis, they showed how powerful they can be in multiple applications and should be implemented in "top-down" workflows and non-targeted analysis for more comprehensive contaminant monitoring.

## 6. Acknowledgements

This study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery grant awarded to P.A.S and by the St. Lawrence Action Plan. The authors would like to thank the Centre de recherche en écotoxicologie du Québec (EcotoQ) for a scholarship to E. Eysseric. Funding sources did not have any involvement in the design, experiments, data interpretation, writing, revision, or submission of this study. We would like to thank Audrey Roy-Lachapelle from Environment and Climate change Canada, Lou Paris from OBV Yamaska, and Clémence Soulié for assistance with the sample collection, extraction, and analysis.

## 7. References

Altenburger R, Brack W, Burgess RM, Busch W, Escher BI, Focks A, Hewitt LM, Jacobsen BN, de Alda ML, Ait-Aissa S. Future water quality monitoring: improving the balance between exposure and toxicity assessments of real-world pollutant mixtures. Environmental Sciences Europe 2019; 31: 1-17.

Beckers L-M, Brack W, Dann JP, Krauss M, Müller E, Schulze T. Unraveling longitudinal pollution patterns of organic micropollutants in a river by non-target screening and cluster analysis. Science of The Total Environment 2020; 727: 138388.

Bletsou AA, Jeon J, Hollender J, Archontaki E, Thomaidis NS. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. TrAC, Trends in Analytical Chemistry 2015; 66: 32-44.

Boix C, Ibáñez M, Sancho JV, Parsons JR, de Voogt P, Hernández F. Biotransformation of pharmaceuticals in surface water and during waste water treatment: Identification and occurrence of transformation products. Journal of Hazardous Materials 2016; 302: 175-187.

Depke T, Franke R, Brönstrup M. Clustering of $MS^2$ spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*. Journal of Chromatography B 2017; 1071: 19-28.

Depke T, Franke R, Brönstrup M. CluMSID: An R package for similarity-based clustering of tandem mass spectra to aid feature annotation in metabolomics. Bioinformatics 2019; 35: 3196-3198.

Djoumbou-Feunang Y, Fiamoncini J, Gil-de-la-Fuente A, Greiner R, Manach C, Wishart DS. BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification. Journal of cheminformatics 2019; 11: 1-25.

Ernst M, Kang KB, Caraballo-Rodríguez AM, Nothias L-F, Wandy J, Chen C, Wang M, Rogers S, Medema MH, Dorrestein PC. MolNetEnhancer: Enhanced molecular networks by integrating metabolome mining and annotation tools. Metabolites 2019; 9: 144.

Eysseric E, Beaudry F, Gagnon C, Segura PA. Non-targeted screening of trace organic contaminants in surface waters by a multi-tool approach based on combinatorial analysis of tandem mass spectra and open access databases. Talanta 2021; 230: 122293.

Eysseric E, Gagnon C, Segura PA. Uncovering new transformation products of concerning organic contaminants by photodegradation experiments and analysis of real samples from a local river. Chemosphere 2022; 293: 133408.

Ferrer I, Sweeney DL, Thurman EM, Zweigenbaum JA. Non-Targeted Screening of Water Samples Using Data Dependent Acquisition with Similar Partition Searching. Journal of the American Society for Mass Spectrometry 2020; 31: 1189-1204.

Fondation de L'Hôpital Brome-Missisquoi-Perkins. Our Hospital. Fondation de L'Hôpital Brome-Missisquoi-Perkins, Ottawa, 2022.

Gago-Ferrero P, Krettek A, Fischer S, Wiberg K, Ahrens L. Suspect screening and regulatory databases: a powerful combination to identify emerging micropollutants. Environmental science & technology 2018; 52: 6881-6894.

Gago-Ferrero P, Schymanski EL, Bletsou AA, Aalizadeh R, Hollender J, Thomaidis NS. Extended suspect and non-target strategies to characterize emerging polar organic contaminants in raw wastewater with LC-HRMS/MS. Environmental science & technology 2015; 49: 12333-12341.

Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, Rinehart D, Nguyen T, Ray J, Kuehl J, Arevalo B. Interactive XCMS Online: simplifying advanced metabolomic data processing and subsequent statistical analyses. Analytical chemistry 2014; 86: 6931-6939.

Hebra T, Eparvier V, Touboul D. Atmospheric pressure photoionization versus electrospray for the dereplication of highly conjugated natural products using molecular networks. Journal of Chromatography A 2020; 1630: 461533.

Hemmer S, Manier SK, Fischmann S, Westphal F, Wagmann L, Meyer MR. Comparison of three untargeted data processing workflows for evaluating LC-HRMS metabolomics data. Metabolites 2020; 10: 378.

Hollender J, Schymanski EL, Singer HP, Ferguson PL. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? Environmental Science & Technology 2017; 51: 11505-11512.

Köke N, Zahn D, Knepper TP, Frömel T. Multi-layer solid-phase extraction and evaporation—enrichment methods for polar organic chemicals from aqueous matrices. Analytical and Bioanalytical Chemistry 2018; 410: 2403-2411.

Komori K, Okayasu Y, Yasojima M, Suzuki Y, Tanaka H. Occurrence of nonylphenol, nonylphenol ethoxylate surfactants and nonylphenol carboxylic acids in wastewater in Japan. Water Science and Technology 2006; 53: 27-33.

Lai A, Singh RR, Kovalova L, Jaeggi O, Kondić T, Schymanski EL. Retrospective non-target analysis to support regulatory water monitoring: from masses of interest to recommendations via in silico workflows. Environmental Sciences Europe 2021; 33: 1-21.

Lajeunesse A, Segura PA, Gélinas M, Hudon C, Thomas K, Quilliam MA, Gagnon C. Detection and confirmation of saxitoxin analogues in freshwater benthic *Lyngbya wollei* algae collected in the St. Lawrence River (Canada) by liquid chromatography-tandem mass spectrometry. Journal of Chromatography A 2012; 1219: 93-103.

Lara-Ortega FJ, Robles-Molina J, Brandt S, Schütz A, Gilbert-López B, Molina-Díaz A, García-Reyes JF, Franzke J. Use of dielectric barrier discharge ionization to minimize matrix effects and expand coverage in pesticide residue analysis by liquid chromatography-mass spectrometry. Analytica Chimica Acta 2018; 1020: 76-85.

Lee H-B, Peart TE, Svoboda ML, Backus S. Occurrence and fate of rosuvastatin, rosuvastatin lactone, and atorvastatin in Canadian sewage and surface water samples. Chemosphere 2009; 77: 1285-1291.

Li Z, Kaserzon SL, Plassmann MM, Sobek A, Ramos MJG, Radke M. A strategic screening approach to identify transformation products of organic micropollutants formed in natural waters. Environmental Science: Processes & Impacts 2017; 19: 488-498.

Liang X, Guan F, Chen B, Luo P, Guo C, Wu G, Ye Y, Zhou Q, Fang H. Spatial and seasonal variations of antibiotic resistance genes and antibiotics in the surface waters of Poyang Lake in China. Ecotoxicology and environmental safety 2020; 196: 110543.

Machado TC, Pizzolato TM, Arenzon A, Segalin J, Lansarin MA. Photocatalytic degradation of rosuvastatin: Analytical studies and toxicity evaluations. Science of the Total Environment 2015; 502: 571-577.

Moschet C, Piazzoli A, Singer H, Hollender J. Alleviating the reference standard dilemma using a systematic exact mass suspect screening approach with liquid chromatography-high resolution mass spectrometry. Analytical Chemistry 2013; 85: 10312-10320.

Oberleitner D, Schmid R, Schulz W, Bergmann A, Achten C. Feature-based molecular networking for identification of organic micropollutants including metabolites by non-target analysis applied to riverbank filtration. Analytical and bioanalytical chemistry 2021; 413: 5291-5300.

Olivon F, Allard P-M, Koval A, Righi D, Genta-Jouve G, Neyts J, Apel C, Pannecouque C, Nothias L-F, Cachet X. Bioactive natural products prioritization using massive multi-informational molecular networks. ACS chemical biology 2017; 12: 2644-2651.

Peng Y, Fang W, Krauss M, Brack W, Wang Z, Li F, Zhang X. Screening hundreds of emerging organic pollutants (EOPs) in surface water from the Yangtze River Delta (YRD): occurrence, distribution, ecological risk. Environmental Pollution 2018; 241: 484-493.

Peter KT, Phillips AL, Knolhoff AM, Gardinali PR, Manzano CA, Miller KE, Pristner M, Sabourin L, Sumarah MW, Warth B. Nontargeted Analysis Study Reporting Tool: A Framework to Improve Research Transparency and Reproducibility. Analytical Chemistry 2021; 93: 13870-13879.

Pfaendner R. Pastic Additives. Kirk-Othmer Encyclopedia of Chemical Technology. John Wiley & Sons, Inc., Hoboken, NJ, 2019, pp. 1-53.

Quinn RA, Nothias L-F, Vining O, Meehan M, Esquenazi E, Dorrestein PC. Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. Trends in Pharmacological Sciences 2017; 38: 143-154.

Rieger MM. Cosmetics. Kirk-Othmer Encyclopedia of Chemical Technology. John Wiley & Sons, Inc., Hoboken, NJ, 2009, pp. 1-53.

Rogowska A, Pomastowski P, Sagandykova G, Buszewski B. Zearalenone and its metabolites: Effect on human health, metabolism and neutralisation methods. Toxicon 2019; 162: 46-56.

Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. Journal of cheminformatics 2016; 8: 3.

Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environmental science & technology 2014; 48: 2097-2098.

Sedio BE, Boya P CA, Rojas Echeverri JC. A protocol for high-throughput, untargeted forest community metabolomics using mass spectrometry molecular networks. Applications in Plant Sciences 2018; 6: e1033.

Segura PA, Racine M, Gravel A, Eysseric E, Grégoire A-M, Rawach D, Teysseire F-X. Impact of method parameters on the performance of suspect screening for the identification of trace organic contaminants in surface waters. Canadian Journal of Chemistry 2019; 97: 197–211.

Sieira BJ, Montes R, Touffet A, Rodil R, Cela R, Gallard H, Quintana JB. Chlorination and bromination of 1, 3-diphenylguanidine and 1, 3-di-o-tolylguanidine: Kinetics, transformation products and toxicity assessment. Journal of Hazardous Materials 2020; 385: 121590.

Statistics Canada. Cowansville [Population centre], Quebec and Quebec [Province] (table). Census Profile. 2016 Census. . Statistics Canada Catalogue no. 98-316-X2016001, Ottawa, 2017.

Sulaiman S, Khamis M, Nir S, Lelario F, Scrano L, Bufo SA, Mecca G, Karaman R. Stability and removal of atorvastatin, rosuvastatin and simvastatin from wastewater. Environmental Technology 2015; 36: 3232-3242.

Sweeney DL. A data structure for rapid mass spectral searching. Mass Spectrometry 2014; 3: S0035-S0035.

Tang W, Chen J, Wang Z, Xie H, Hong H. Deep learning for predicting toxicity of chemicals: A mini review. Journal of Environmental Science and Health, Part C 2018; 36: 252-271.

U.S. Environmental Protection Agency. Ecological Structure Activity Relationships (ECOSAR) Predictive Model, 2019.

U.S. Environmental Protection Agency. CTS: Chemical Transformation Simulator, 2022.
Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nature biotechnology 2016; 34: 828.
Wang Z, Walker GW, Muir DC, Nagatani-Yoshida K. Toward a global understanding of chemical pollution: a first comprehensive analysis of national and regional chemical inventories. Environmental Science & Technology 2020; 54: 2575-2584.
Wicker J, Lorsbach T, Gütlein M, Schmid E, Latino D, Kramer S, Fenner K. enviPath–The environmental contaminant biotransformation pathway resource. Nucleic acids research 2016; 44: D502-D508.
World Health Organization. Anatomical Therapeutic Chemical (ATC) Classification. 2021, Geneva, Switzerland, 2021.
Zahn D, Mucha P, Zilles V, Touffet A, Gallard H, Knepper T, Frömel T. Identification of potentially mobile and persistent transformation products of REACH-registered chemicals and their occurrence in surface waters. Water research 2019; 150: 86-96.