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

Emmanuel Eysseric[1], Francis Beaudry[2], Christian Gagnon[3], 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

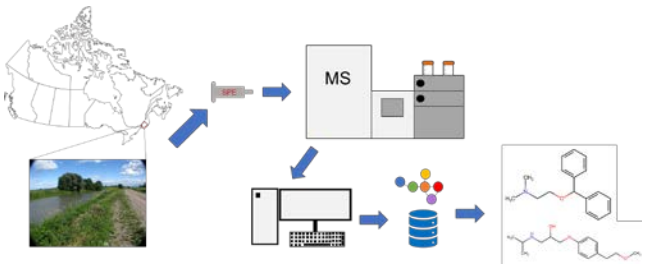
[2] Département de Biomédecine Vétérinaire, Faculté de Médecine Vétérinaire Université de Montréal, Saint-Hyacinthe, QC, Canada.

[3]Environment and Climate Change Canada, Montréal, Canada

The final version of this paper was published in *Talanta*:

Eysseric E., Beaudry F., Gagnon C., Segura P. A. (2021) 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*:122293. 10.1016/j.talanta.2021.122293

Graphical abstract



Keywords: data-dependent acquisition, Global Natural Products Social Networking, MetFrag, ultra-trace concentration, pharmaceuticals and personal care products, consumer product additives

Abstract

Non-targeted screening (NTS) in mass spectrometry (MS) helps alleviate the shortcoming of targeted analysis such as missing the presence of concerning compounds that are not monitored and its lack of retrospective analysis to subsequently look for new contaminants. Most NTS workflows include high resolution tandem mass spectrometry ($HRMS^2$) and structure annotation with libraries which are still limited. However, combinatorial fragmentation tools that simulate MS^2 spectra are available to help close the gap of missing compounds in empirical libraries. Three NTS tools were combined and used to detect and identify unknown contaminants at ultra-trace levels in surface waters in real samples in this qualitative study. Two of them were based on combinatorial fragmentation databases, MetFrag and the Similar Partition Searching algorithm (SPS), and the third, the Global Natural Products Social Networking (GNPS), was an ensemble of empirical databases. The three NTS tools were applied to the analysis of real samples from a local river. A total of 253 contaminants were identified by combining all three tools: 209 were assigned a probable structure and 44 were confirmed using reference standards. The two major classes of contaminants observed were pharmaceuticals and consumer product additives. Among the confirmed compounds, octylphenol ethoxylates, denatonium, irbesartan and telmisartan are reported for the first time in surface waters in Canada. The workflow presented in this work uses three highly complementary NTS tools and it is a powerful approach to help identify and strategically select contaminants and their transformation products for subsequent targeted analysis and uncover new trends in surface water contamination.

1 Introduction

Recent advances in mass spectrometry such as higher sensitivity [1], resolution, mass accuracy and computing (i.e., faster processors, increased memory, etc.) are making non-targeted screening (NTS) of trace organic contaminants more applicable to environmental applications [2-4]. Still, NTS entails a heavy multistep data analysis method from peak selection, to the prioritization, and the eventual identification of prioritized compounds [4]. Compared with targeted analysis, NTS remains inconclusive, or tentative, in its identification with no reference standards used and the identification confidence level of a given compound may vary [5]. There are several methods to improve the confidence in a molecular formula such as the determination of the spectral accuracy of precursor ions [6], the open-source tool SIRIUS [7] and other algorithms used by manufacturers of mass spectrometers [8]. Suspect screening of specific classes of contaminants can also help narrow down the scope of research by performing a pseudo-NTS. As an example, analysis of illicit drugs has seen recent advances in regard with online shared depositories [9] and MS software [10]. However, such methods can be difficult for compounds at lower intensities as it is often the case in trace and ultra-trace analysis. Compounds may also be identified with MS^2 data. A probable structure can be proposed by using an MS^2 library spectrum match [5]. Modern high-resolution tandem mass spectrometers can automatically gather structural information using functions such as data-dependent acquisition (DDA) [11] and data-independent acquisition (DIA)[12]. There have been successful cases of application of both DDA[13, 14] and DIA[15] to environmental analysis.

Once structural information for compounds of interest is obtained, library spectrum matching is a convenient and powerful tool for identification of unknowns but is severely limited by the low number of MS^2 spectra in libraries. Unlike very comprehensive electron ionization-mass spectrometry libraries for gas chromatography-mass spectrometry like the NIST Standard Reference Database, most online high-resolution electrospray ionization-MS^2 libraries for UHPLC-MS analysis such as mzCloud, MassBank, Metlin or Riken contain small numbers of MS^2 spectra (<200 000) representing, in most cases, a small number of molecules (<20 000) [16]. The general heterogenicity of spectral data due to the oftentimes ultra-trace level concentration of analytes and matrix effects is a major hindrance against the effective usage of these libraries [17]. Additionally, the small number of compounds in electrospray ionization-MS^2, databases, represents only a small fraction of the total number of known chemical compounds. For example, the CAS Registry contains about 162 000 000 unique compounds as of July 2020 [18]. To date one of the most comprehensive freely available chemical compound databases is PubChem which contains over 103 000 000 compound entries [19].

A way to solve the problem of limited MS^2 spectra databases is to integrate computational techniques into NTS methods. There are several computational approaches to assess the structure of known unknowns. Among them are rule-based fragmentation and combinatorial fragmentation. On the first approach, the spectrum is compared to simulated fragmentation spectra using a set of fragmentation rules that are applied to a proposed structure. Mass Frontier (HighChem, Slovakia) and Fragmenter (ACD/Labs, Canada) software packages use this approach. These rules predict hundreds of possible fragments but only a fraction is actually observed. Also of note is that bond cleavage rates are rarely considered which makes the relative abundance of product ions

unavailable or inaccurate at best [20]. The combinatorial approach is used to explain the peaks found in an observed spectrum. Peaks are matched to a substructure and candidates are ranked by annotation score [20]. Possible fragments at cleavable links of a candidate compound's structure are enumerated and compared with peaks present in the MS^2 spectrum. There are some drawbacks to the combinatorial approach such as the inability to account for structural rearrangements and a lower accuracy because some predicted fragments are highly unlikely which can lead to a higher false identification rate [20-22]. Among the multiple NTS tools available at the moment, there are three are of interest given the different on they way that they function and thus they are highly complementary: Similar Partition Searching (SPS), the Global Natural Products Social Networking (GNPS), and MetFrag.

The Similar Partition Searching (SPS) algorithm, developed by Sweeney [21, 23] and used in the present study, is based on a combinatorial approach. The SPS database is a subset of about 240 000 common compounds from the PubChem Compound database that have been divided into mathematical partitions of their molecular mass, i.e., as masses of complementary substructures that when put together contain all the atoms of a given molecule. The database was formed by systematic bond disconnection using only a few very basic fragmentation rules. The SPS software first compares the selected precursor ion in each MS^2 spectrum to the corresponding MS^1 spectrum to determine adduct ion assignment (e.g. $[M+H]^+$, $[M+Na]^+$, $[2M+H]^+$). The accurate mass of the precursor ion is then adjusted, based on this adduct assignment, to calculate the accurate mass of the analyte molecule. The SPS algorithm then compares the accurate-mass fragmentation data from the MS^2 spectrum to the partitions of all compounds in the database that have molecular weights that are within 4 mDa of the analyte. Each partition is scored mainly by the number and intensity of the neutralized product ions matched by virtue of being within 4 mDa of masses in the partition. There is also a small score adjustment for mass accuracy and the number of rings and double bonds that were disconnected in generating the partition. For many MS^2 spectra, multiple partitions of an individual molecule will generate different scores; these scores are then combined into one final score. The SPS scoring does not use or consider the isotope ratio data or the number of synonyms in the Pubchem database for each compound. SPS has been recently applied to the identification of up to 200 contaminants in wastewater and surface water in the US [24].

To illustrate how SPS works, one of the eleven 3-substructure partitions of dicyclohexylurea is shown in Figure 1. There are 248 3-substructure partitions in the SPS database of eight compounds with an exact mass within 4 mDa of the calculated accurate mass of dicyclohexylurea from the spectrum in Figure 1. The SPS algorithm will check each of these 248 partitions (plus 56 2-substructure partitions and 524 4-substructure partitions) against the MS^2 fragmentation data and then generate a combined score for each compound.

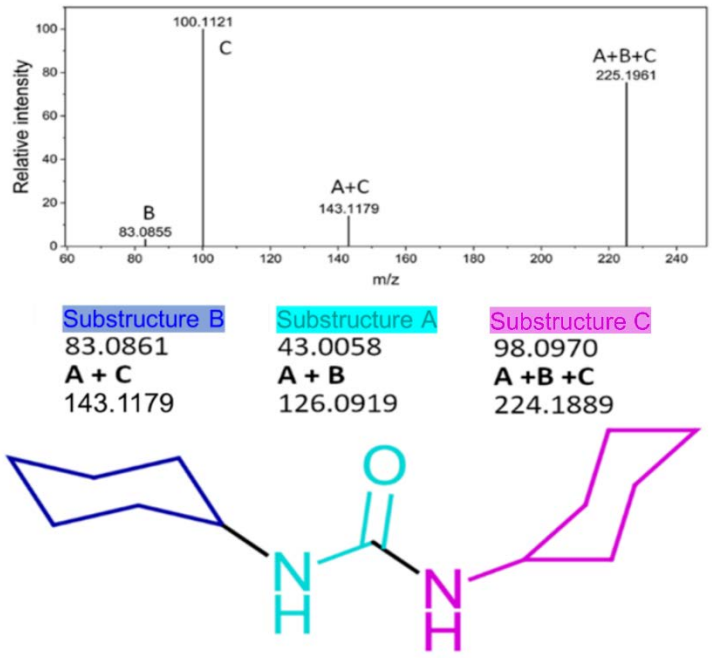


Figure 1. Four ions are found in the MS^2 spectrum of dicyclohexylurea on mzCloud. All four ions can be explained by substructures in the figure with the addition/subtraction of one or two hydrogen atoms. These complementary substructures were generated by systematically disconnecting the breakable bonds of dicyclohexylurea and then placing the masses of the complementary substructures in the SPS database with masses of over 240 000 other common compounds. Searching is then done by pattern matching.

The Global Natural Products Social Networking (GNPS) (https://gnps.ucsd.edu/) is a freely available online platform that works twofold by performing empirical library search from a bundle of online databases, including MassBank, MoNA, the Human Metabolome Database [25]. It also creates networks from data-dependent acquisition (DDA) files by converting the selected precursor ions into multidimensional vectors where each product ion from the MS^2 spectrum is a dimension. It then calculates the scalar product of each combination of vectors. When the scalar product between two vectors is closer to a value of one, the more similar they are and thus the more similar are their respective MS^2 to each other. This is very useful to mark which compounds are structurally related like natural products of the same family or, in organic contaminants analysis, transformation products [26].

MetFrag, an *in silico* combinatorial fragmenter, initially released in 2010 [27],works by searching candidates for a given *m/z* from a compound database. The candidates' molecular structures are then split into smaller units by bond dissociation [28]. These *in silico* generated fragments are then compared to the experimentally obtained MS^2 spectra. Scoring is based on the number of matched exact mass product ions between the combinatorial and experimental spectrum, the intensity of the product ions and the bond dissociation energy of the matched fragments. Several databases are available for MetFrag like PubChem, ChemSpider and KEGG. MetFrag was updated in 2016 to include better identification by using parameters such as the number of PubChem data sources for

a candidate, the number of PubMed articles referencing it, its presence on lists of relevant candidates for the identification, the presence of a substructure and or specific elements in a candidate and information about the retention time [28]. MetFrag has been used in non-targeted screening previously, notably in the Rhine River in Basel, Switzerland [4].

Still, despite the use of empirical and computational MS^2 database searching, NTS remains a challenging exercise. Significant shortcomings like handling large sets of data and the identification of unreported transformation products prevent it from being routinely applied in monitoring programs yet [4].

The objective of this paper was to use three complementary qualitative NTS tools (SPS, MetFrag and GNPS) to clearly identify organic contaminants at ultra-trace to trace levels in real samples of surface waters for qualitative non-targeted analysis purposes and highlight the use of *in silico* databases. These tools were used in conjunction for the analysis of samples from a local river collected near a municipal wastewater treatment plant. Reference standards were then used to confirm some of the matches made by the tools.

2 Experimental section

2.1 Reagents and standards

Water, acetonitrile (ACN), methanol (MeOH) and 0.1% formic acid (FA) in ACN were all HPLC-MS Optima grade and were obtained from Fisher Scientific (Waltham, MA, USA). Analytical standards for the confirmation of suspects in the case study were of high purity (in most cases ≥98 %) and are shown in the Supporting Information, section SI-1.1.

2.2 Sample collection and preparation of River samples

River water samples (one amber high-density polyethylene bottle of 1000 mL per sampling point) were collected from the Yamaska River upstream and downstream the wastewater treatment plant of Granby (QC, Canada) on July 11, 2019 (Figure SI-1). Granby is a town in southern Quebec with around 60 000 inhabitants, and it has a strong industrial sector and some agricultural activity upstream [29]. Samples were conserved in an ice cooler until arrival to the laboratory and were immediately stored at -20 °C. Before extraction, samples were thawed at room temperature and buffered to pH 7 with phosphoric acid monobasic and phosphoric acid dibasic. Then samples were extracted by solid-phase extraction according to a previous published method [30]. Briefly, samples of 250 mL were concentrated on polymeric Strata-X solid-phase extraction cartridges (200 mg, 6 mL) from Phenomenex (USA) and then eluted with 2 × 3 mL of an ACN-MeOH 1:1 (v/v) solution. Eluates were evaporated under a nitrogen flow and reconstituted to 625 µL with MeOH to obtain a concentration factor of 400. While using MeOH as reconstitution solvent leads

to peak distortion for early eluting peaks, the effect was minor and the benefits of solubilizing a large range of compounds outweighed the peak distortion effects observed in the early stages of the separation (Figure SI-2 in the Supporting information). Three extraction replicates per sample were carried for each sample. This improved the number of identifications and accounted for potential extraction and instrumental variability.

2.3 Instruments and methods

A Thermo Scientific Vanquish Flex ultra-high performance liquid chromatography system was coupled to a Thermo Scientific Q-OrbitrapMS model Q Exactive Plus Orbitrap (San Jose, CA, USA) using a pneumatic assisted heated electrospray ion source. 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, 5%; 8 min, 18%; 22 min, 80%; 32 min, 100%; 40 min, 100%; 40.01 min, 5%; 45 min, 5%. Total run time was 45 min. Mobile phase flow rate was 250 µL min^{-1} throughout the run and the injection volume was 2 µL.

For mass spectrometry, ion source parameters were the following: polarity was positive, capillary temperature was 300 °C; sheath gas was 50; auxiliary gas was 20; spray voltage was 4000 V. A data dependent acquisition (DDA) experiment was used for detection. A DDA cycle entailed one MS^1 survey scan (*m/z* 100-1000) acquired at 35 000 mass resolution (FWHM) and precursor ions meeting user defined criteria for monoisotopic precursor intensity (dynamic acquisition of MS^2 based Top 10 most intense ions with at least 2×10^5 intensity threshold). 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 17 500 mass resolution (FWHM). Instrument calibration was performed prior to all analyses and mass accuracy was notably below 1 ppm using Thermo Pierce calibration solution and automated instrument protocol. The calibration mixture was composed of caffeine, n-butylamine, the tetrapeptide MRFA, and Ultramark 1621, a mixture of flourinated phosphazenes, in an acetonitrile/methanol/acetic acid solution.

2.4 Data conversion and processing

For the Similar partition searching (SPS) workflow, MSConvertGUI from the ProteoWizard tool Suite [31] was used to convert data files from vendor format to universal formats which were then compressed and uploaded to an Amazon Web Service S3 folder for the SPS algorithm from MathSpec Inc. (USA) to process. After blank subtraction, the results were imported into a Microsoft Access database. Tentative matches were then evaluated based on the match score and the number of synonyms for that compound. The latter reflects the popularity of a chemical and it

is analogous to the number of different literature references for that compound. Using the number of references as filter has been found to be a useful in the identification of unknowns [32]. For more details, consult the Supporting Information (Section 1.4 and Figure SI-3).

The Global Natural Products Social Networking (GNPS) was used to perform embedded empirical library searches. GNPS also generates networks of related MS^2 spectra (molecular networks) which is a powerful and efficient way to visualize DDA data. Vendor files were converted with MSConvert into a readable format and uploaded to the GNPS server where the search and networks were conducted. Once the networking was performed, the network files were treated with Cytoscape software [33]. Information regarding the search and network parameters is given in a schematized workflow in the Supporting Information (Section SI-1.5 and Figure SI-4).

For the MetFrag workflow, PatRoon, a package from the R programming language that functions as a common interface for different NTS tools currently available was used [34]. PatRoon has been used in NTS studies in the past [35]. Vendor files were first converted with MSConvert into the mzML format before the data treatment. Peak picking and feature selection were conducted by XCMS, background subtraction and sample metadata were done with PatRoon itself. Formulas were generated with GenForm and detection of adduct ions was performed with CAMERA. Computational MS^2 database search was performed by MetFrag on CompTox Chemicals Dashboard from the US EPA using metadata files according to McEachran, Mansouri, Grulke, Schymanski, Ruttkies and Williams [36]. For more details on the parameters of the tools used, see the Supporting Information (section SI-1.6 and Figure SI-5).

2.5 Quality control

A composite field blank of LC-MS Optima grade water was collected in the two sampling points and it was stored and then extracted in the same way as the samples. The field blank as well as an additional MeOH instrumental blank were injected for background subtraction and to control for potential laboratory and instrument contamination. Details about how the background subtraction was applied with SPS are shown in section SI-1.4 and Figure SI-3; details about how the field and instrumental blanks were used to look for contaminants with GNPS are given in section SI-1.5 and Figure SI-4; details about how the background subtraction were applied with patRoon and MetFrag are given in section SI-1.6 and Figure SI-5.

2.6 Levels of identification confidence

Only matches with a level of confidence of probable structure (level 2) and confirmed structure (level 1), according to the scheme proposed by Schymanski, Jeon, Gulde, Fenner, Ruff, Singer and Hollender [5] are reported in this article. The probable structure level was attained using either library (MS^2 database match) or diagnostic evidence (e.g., possible ionization by electrospray in the positive mode and environmental relevance of the annotated chemicals on suspect lists). The structure confirmation level was attained using reference standards. However, all probable

structure matches do not carry the same level of certainty and informed judgement based on the chemistry and environmental context of a potential match must be considered. For instance, parameters like consistency between the retention time and the structure are all considered to filter out "bogus" matches. The quality of the match is, of course, a major factor; it considers the number of matched exact *m/z* for each fragment as well as the number of unexplained *m/z*. Finally, if the match originates from an empirical (GNPS) or combinatory (MetFrag, SPS) library, it does not carry the same level of confidence; the former being more reliable. As such, annotations not made by GNPS were cross-checked with the online MS^2 database mzCloud. Where a feature's match given by either SPS and/or Metfrag was not made by GNPS, its MS^2 spectrum was submitted to mzCloud.

Additionally, spectral accuracy and formula ranking were determined using Mass Works software from Cerno Bioscience (Las Vegas, NV) according to a method published previously [6]. Briefly, molecular formulas were generated according to the following parameters: search mode was sCLIPS; allowed elements were C, H, N, O, P, F, S, Cl, Br, Na; mass tolerance was 5 ppm; charge was chosen depending on the ion's deconvolution state, even electron state; double bond equivalent range was 0.5 to 25; interference rejection was 0.001.

3 Results and discussion

3.1 Non-targeted screening of river water samples collected near a wastewater treatment plant

An example of a match with the "probable structure" level of confidence for the pharmaceutical compound metoprolol is shown in Table 1 for SPS and in Figure 2 for GNPS and MetFrag. In Table 1, the "EPA DashBd" link redirects to a monoisotopic mass search for the calculated molecular weight (MW) ± 0.04 Da on the EPA CompTox Chemistry Dashboard. The links in the PubChemLink column open the PubChem Compound Summary for the tentatively identified candidate. In this example, the first candidate (PubChem Link: 4171) is metoprolol, the second (PubChem Link: 441308) one is metoprolol tartrate and the third one (PubChem Link: 62937) is metoprolol succinate. The next four hits with the same score of 82 are other salts of metoprolol with different counterions. These salts were not detected by the instrument, but since metoprolol is a component, these hits refer to the same compound. This match carries a probable structure level of confidence since the main *m/z* from the experimental and combinatorial spectra match and metoprolol is on several lists of suspects in surface waters. The matches with lower scores are from other compounds.

Table 1. Example of a SPS match for a feature.

Analyte	RT	MW	Intensity	Adduct	EPA DashBd	Score	Num Syn	PubChem Link	Class
267185	5.23	267.18	694693	H+	267.1844	82	151	4171	pharmaceutical
267185	5.23	267.18	694693	H+	267.1844	82	93	441308	pharmaceutical
267185	5.23	267.18	694693	H+	267.1844	82	60	62937	pharmaceutical
267185	5.23	267.18	694693	H+	267.1844	82	43	5702086	pharmaceutical
267185	5.23	267.18	694693	H+	267.1844	82	18	6440651	not classified
267185	5.23	267.18	694693	H+	267.1844	82	10	6446646	not classified
267185	5.23	267.18	694693	H+	267.1844	82	4	16219665	not classified
267185	5.23	267.18	694693	H+	267.1844	79	16	162812	xenobiotc metab
267185	5.23	267.18	694693	H+	267.1844	45	20	3151271	not classified
267185	6.04	267.18	439562	H+	267.1842	43	16	162812	xenobiotc metab

In Figure 2a, the experimental spectrum acquired with the Q-OrbitrapMS is compared to the empirical database spectrum of metoprolol from GNPS. As it can be seen, the MS^2 mass spectrum of the unknown compound found in the river sample matches well with the library spectrum of metoprolol. In both spectra the most abundant peak is the $[M+H]^+$ ion (*m/z* 268) and characteristic product ions frequently used for MRM experiments such as *m/z* 98, *m/z* 116, *m/z* 133 and *m/z* 159 can be clearly seen the spectrum [37-39]. In Figure 2b, the bottom part of the graph is automatically generated by a report-making script embedded in MetFrag. On the spectrum view, the color match shows which algorithm annotated the *m/z* of specific fragments. The bar graph indicates the match score for the different criteria. These scores are normalized and 1 is the highest score. This figure shows that Metfrag identified multiple product that can be explained by the metoprolol structure such as $C_{15}H_{24}NO_2^+$ (*m/z* 250, loss of H_2O) and $C_{12}H_{20}NO_3^+$ (*m/z* 226, loss of isopropyl), among others.

A total of 253 compounds were identified by the multi-tool method in both sampling points, the complete list is found in the Microsoft Excel file IdentifiedCompounds.xslx (Supporting Information). These compounds were classified in five generic classes (Figure 3a): consumer product additives and other synthetic compounds (116 compounds), pharmaceuticals (87), natural products (28), illicit drugs (14) and pesticides (8). Out of the 253 identified compounds, 209 were assigned a probable structure and 44 compounds were confirmed with reference standards (Table 2). A more detailed account of the matched product ions can be found in Table SI-1. All identified compounds found by the combinatorial tools (SPS and/or Metfrag) were cross-checked with the online MS^2 spectra library mzCloud. In the file IdentifiedCompounds.xlsx, matches are listed as "not on mzCloud" if the compound is not present in the online library. A score from mzCloud is given if the annotation is the same as the one given by SPS and/or MetFrag and it is the best match. As many as 45 compounds were confirmed by mzCloud while 82 were not present on mzCloud's database. Most of those absent compounds were chemical congeners related to consumer product additives.

Additionally, the file IdentifiedCompounds.xlsx indicates in which sampling points the compounds were detected. All compounds that were detected upstream the wastewater treatment plant were also detected downstream while fewer compounds, especially pharmaceuticals, were detected upstream compared to downstream. This is expected since Granby is the first sizeable city in this branch of the river. The number of compounds of each category detected upstream the wastewater treatment plant is shown in Figure SI-6 (Supporting Information).

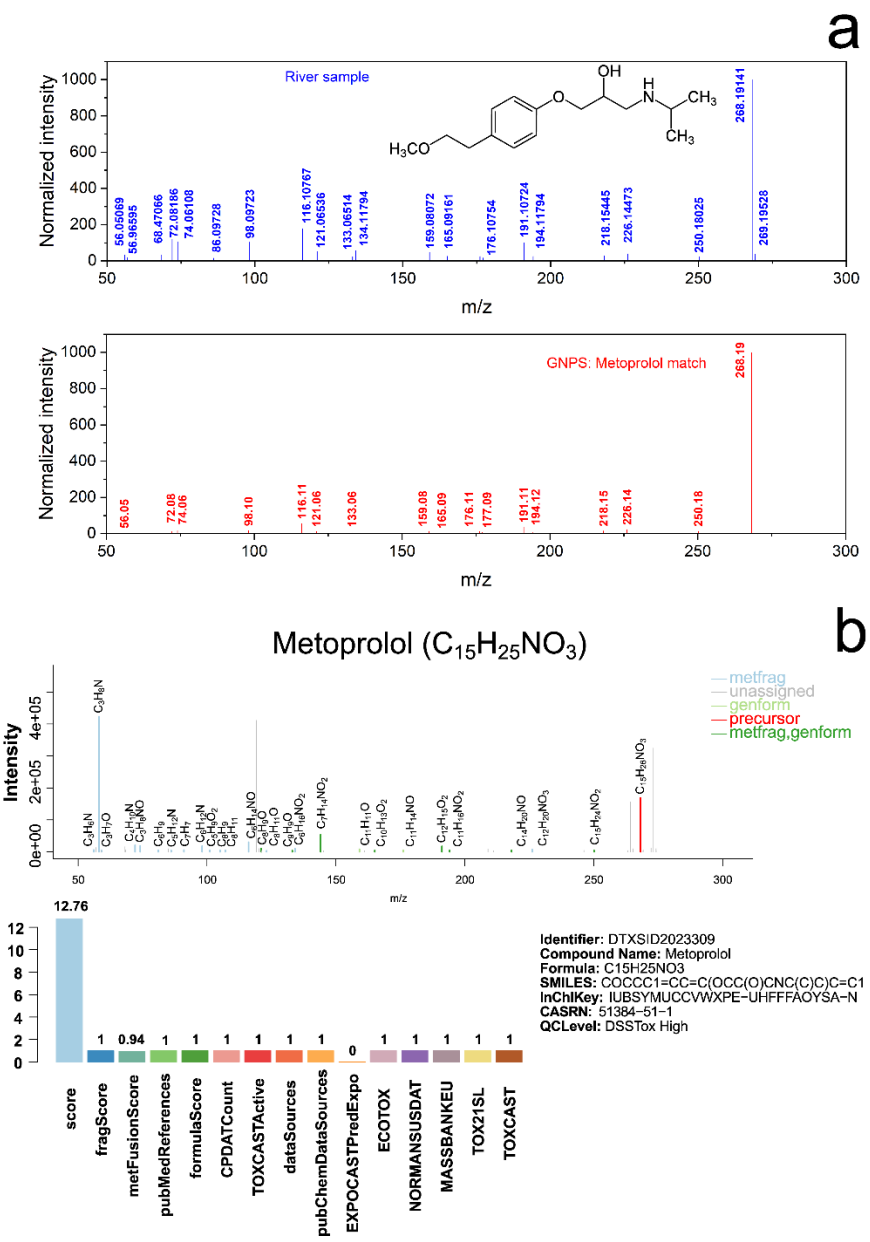


Figure 2. Example of probable structure match for metoprolol a) from GNPS and b) from MetFrag. FragScore is the MetFrag score; metFusion score combines MetFrag score with MassBank when applicable i.e., when MassBank has an entry for the compound; pubMedReference is the number of times a compound is referenced in PubMed; formulaScore is a score based on the number of

explained molecular formulas in the spectrum; CPDATCount is from the CPDAT list (Chemical and Products Database) that categorize chemicals functions; TOXCASTActive is the list of compounds screened by the US EPA; dataSources is the number of synonyms a compound has in the CompTox database, pubChemDataSources refers to the number of synonyms in PubChem, EXPOCASTPredExpo is a US EPA exposition prediction program; ECOTOX is a US EPA curated database that gives ecotoxicology data; NORMANSUSDAT, MASSBANKEU, TOX21SL and TOXCAST are databases of contaminants of emerging concern.

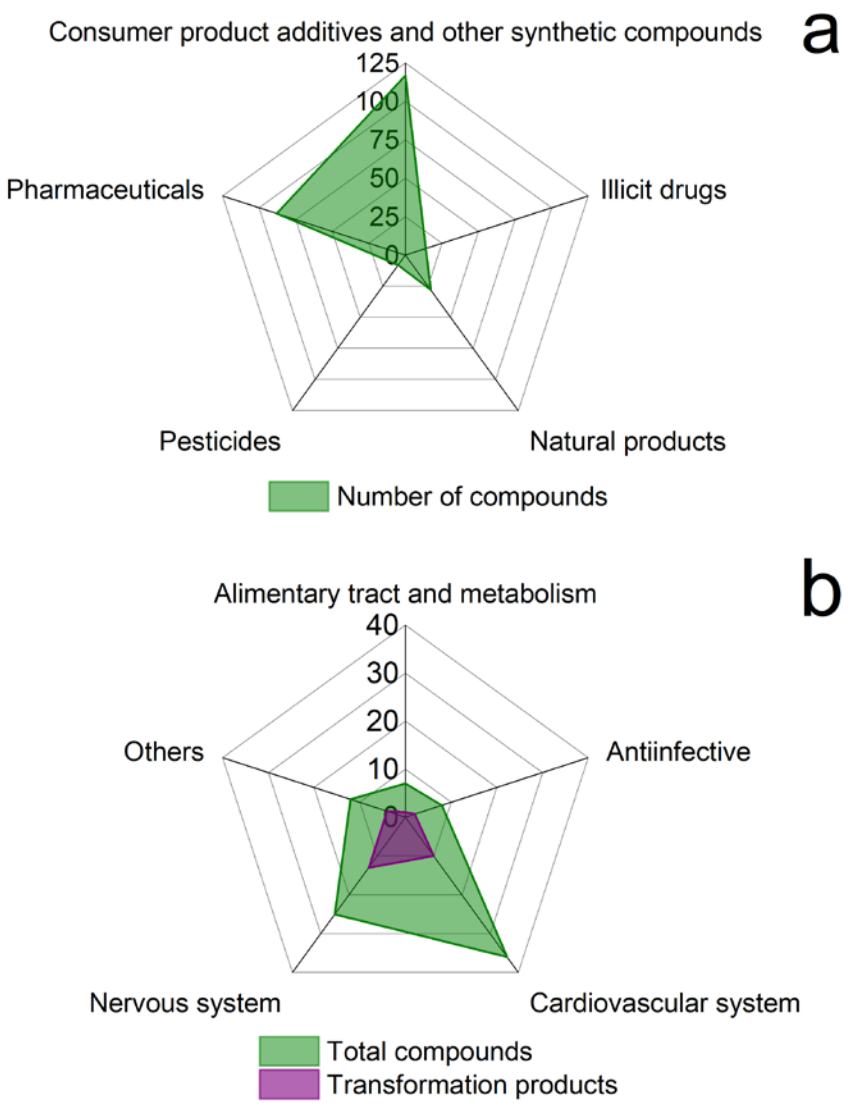


Figure 3. a: Generic classes of the compounds identified as probable or confirmed structures by the multi-tool method. **b:** Anatomical Therapeutic Chemical (ATC) classes of pharmaceutical compounds identified as probable or confirmed structures. All the identified chemicals are found in the "IdentifiedCompounds.xslx" file (Supporting Information).

In Table 2, the precursor ions of all confirmed compounds (except for azithromycin) mass accuracies < 2.5 mDa as well as their most intense product ion. Spectral accuracy, a measure of the similarity between experimental and theoretical isotopic patterns [6], is also reported to further confirm the experimental data. In most cases, spectral accuracy was higher than 90% and the molecular formula was ranked among the top five possible formulas. While low values of spectral accuracy and low formula rankings were observed (e.g., cetirizine, valsartan) this was due generally to co-eluting isobars that lowered the match between theoretical and calibrated isotopic patterns in Mass Works. Such effect was already observed, especially for compounds at low concentrations in environmental matrices [6]. Another factor that affects ranking according to spectral accuracy is that molecules with masses > 400 Da have a higher number of potential matches than molecules with lower masses.

The number of pharmaceuticals among the identified compounds (Figure 3b) is extensive: 57 parent compounds and 30 transformation products were detected. Among these, four anatomical therapeutic chemical classes had the highest number of compounds: cardiovascular system (26 parent compounds, 10 transformation products), nervous system (12 parent compounds, 13 transformation products), antiinfectives (5 parent compounds, 2 transformation products) and alimentary tract and metabolism (6 parent compounds, 1 transformation product). While some of the confirmed compounds are frequently occurring pharmaceuticals such as carbamazepine and venlafaxine, others less commonly reported compounds in surface waters were also found. For example, to the authors' knowledge the angiotensin II receptor antagonists irbesartan and telmisartan and one of telmisartan's transformation products were detected for the first time in surface waters in Canada but were found widely in other parts of the world [40]. Telmisartan spectra from the reference standard, the river sample and its transformation product can be seen in Figure SI-7 (Supporting Information).

Table 2. Compounds confirmed using reference standards.

Confirmed structure	Precursor (river sample) (*m/z*)	Mass accuracy (mDa)	Spectral accuracy* (%)	Product ion (river sample) (*m/z*)	Mass accuracy (mDa)	Usage**
Atenolol	267.1708	0.46	89.6 (1)	190.0867	0.34	Beta-blocker
Atorvastatin	559.2628	1.91	92.0 (4)	440.2248	0.87	Statin
Azithromycin	375.2635	4.78	94.2 (2)	591.4237	2.34	Antibiotic
Benzoylecgonine	290.1398	1.16	92.0 (1)	168.1024	0.43	Opioid (M)
Caffeine	195.0883	0.43	96.7 (1)	138.0667	2.28	Stimulant
Carbamazepine	237.1028	1.19	86.8 (2)	194.0970	-0.51	Antiepileptic
Cetirizine	389.1637	1.69	73.9 (15)	201.0472	-0.50	Antihistamine
Citalopram	325.1721	0.70	93.7 (2)	109.0456	1.03	Antidepressant
Cocaine	304.1555	0.88	83.7 (1)	182.1181	0.34	Opioid
N,N-Diethyl-meta-toluamide (DEET)	192.1388	0.49	99.3 (1)	119.0498	0.82	Insect repellant
Denatonium	325.2280	0.31	92.6 (1)	86.0974	0.18	Bittering agent
O-Desmethylvenlafaxine	264.1966	0.37	97.8 (1)	246.1863	0.77	Antidepressant (M)
Diltiazem	415.1695	-1.54	87.0 (2)	178.0326	-0.42	Calcium channel blocker

Diphenhydramine	256.1703	0.06	95.8 (1)	167.0860	-0.12	Antihistamine
Fexofenadine	502.2967	3.39	94.7 (2)	466.2756	0.98	Antihistamine
Irbesartan	429.2405	-1.16	83.3 (2)	207.0924	0.02	Angiotensin II receptor antagonist
3,4-Methylenedioxy methamphetamine (MDMA)	194.1181	0.79	77.3 (1)	109.9594	0.50	Amphetamine
Methadone	310.2163	0.46	87.6 (1)	265.1595	1.37	Synthetic opiod
Octaethylene glycol (PEG-8)	371.2280	0.40	97.3 (2)	133.0860	-0.10	Ethylene glycol oligomer
Octylphenol ethoxylate-9 (OPEO-9)†	625.3925	-1.20	96.1 (80)	347.1677	0.60	Nonionic Surfactant
Oxazepam	287.0588	1.28	97.5 (2)	269.0470	1.25	Tranquilizer, antidepressant and illicit drug
Pentaethylene glycol (PEG-5)	239.1497	0.80	95.2 (1)	151.0964	0.10	Ethylene glycol oligomer
Quetiapine	384.1763	2.05	76.0 (13)	253.0800	0.38	Antipsychotic
Tris(2-butoxyethyl) phosphate	399.2506	0.67	96.3 (13)	299.1634	-0.48	Flame retardant
Telmisartan	515.2459	0.64	95.3 (7)	276.13711371	-1.13	Angiotensin II receptor antagonist
Temazepam	301.0749	1.59	88.5 (1)	228.0577	0.81	Tranquilizer, antidepressant and illicit drug
Valsartan	436.2356	0.35	79.8 (2)	235.0990	1.02	Angiotensin II receptor antagonist
Venlafaxine	278.2133	0.59	94.5 (1)	260.2016	0.22	Antidepressant

*Number in parentheses indicates the rank among possible formulas according to spectral accuracy. ** (M) indicates metabolite or transformation product. † Other OPEOs, from OPEO-1 to OPEO-17 were also observed and confirmed in the samples.

Among the pharmaceuticals used to treat cardiovascular system disorders, the antihypertensive diltiazem, is interesting since it showcases the use of the molecular networks as can be seen in Figure 4. Only diltiazem and desmethyldiltiazem were identified with the databases originally. However, the other transformation products were sharing a single network since their MS^2 spectra were highly similar. From the information available in the network such as the *m/z* difference between each precursor ion and the structure of diltiazem, it was possible to deduce the structure of the other transformations products even though they were not initially annotated by the databases. As mentioned earlier, irbesartan and telmisartan were confirmed with standards. Additionally, a transformation product that could correspond to of hydroxy-telmisartan was also found in the samples. The high number of pharmaceuticals from the sartan family observed (8 in total) could be explained by high ionization efficiency. According to studies on the relationship of ionization efficiency in electrospray and molecular properties, three physico-chemical parameters appear to have a significant influence: molecular volume, pK_a and log D [41, 42]. The sartans identified have all relatively high molecular volumes and, at the pH of the mobile phase, they are cationic (except valsartan) and have log D values that would favor their transfer from droplets to the gas phase. Since these compounds appear to extensively degrade into multiple transformation products, it could be interesting to monitor their fate and occurrence in Canadian WWTP effluents and surface waters.

The presence of several pharmaceuticals used to treat nervous system disorders such as oxazepam, temazepam, quetiapine and venlafaxine were confirmed in the samples with reference standards. For venlafaxine, four of its transformation products: O-desmethylvenlafacine (confirmed), N-desmethylvenlafaxine, N-oxide venlafaxine and oxo-venlafaxine, were also observed in the samples. Carbamazepine (confirmed), its metabolite 10,11-dihydro-10,11-dihydroxycarbamazepine, citalopram (confirmed) along its transformation product desmethylcitalopram were identified as well in the river extracts. Finally, several antiinfectives were identified including azithromycin (confirmed), cefprozil as well as oxopterin-sulfamethoxazole, an algal metabolite of sulfamethoxazole [43, 44] and 4-desmethoxy-4-ethoxy trimethoprim, a metabolite of trimethoprim.

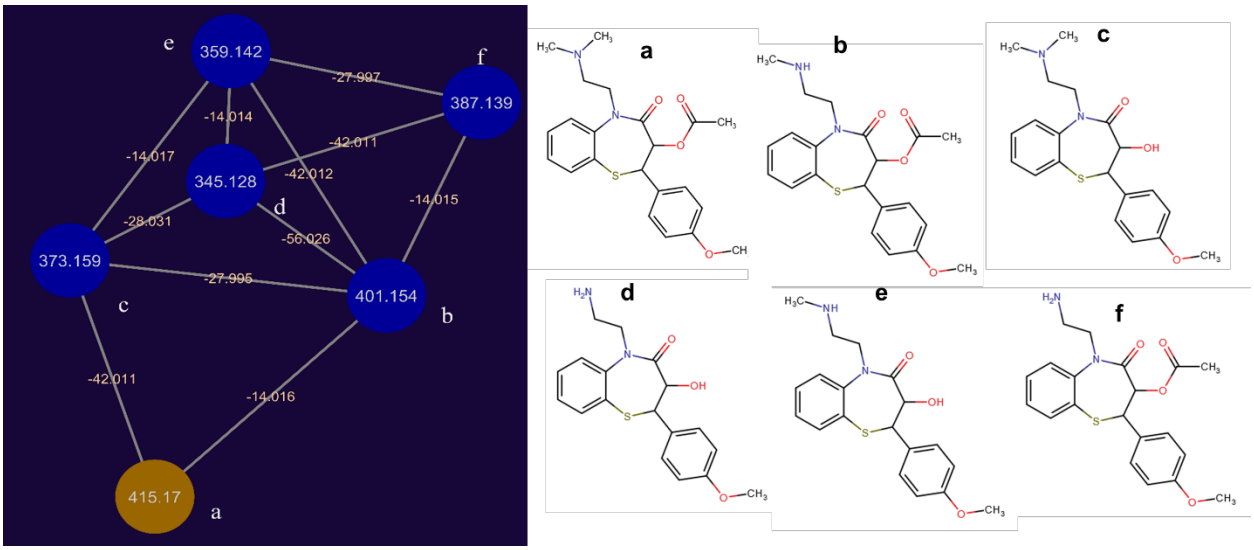


Figure 4. Molecular network of the calcium channel blocker diltiazem (a) and its transformation products (b to f) with proposed structures to the left. Each detected precursor is a node linked in the network with precursors that have similar MS^2 spectra. In the figure, **a** is diltiazem, **b** is desmethyldiltiazem, **c** is deacethyldiltiazem, **d** is didesmethyldeacethyldiltiazem, **e** is desmethyldeacethyldiltizaem and **f** didesmethyldiltiazem.

Among the illicit drugs, cocaine, and its metabolites benzoylecgonine, tropine and tropinone along with other opioids such as dezocine and methadone were observed in the samples. Cocaine, benzoylecgonine and methadone were confirmed with reference standards. Several amphetamines were also identified like methamphetamine, 3,4-methylenedioxymethamphetamine (confirmed), mephedrone, norephedrine and lefetamine. These results are consistent with a previous study where cocaine, benzoylecgonine and methamphetamine were found in surface waters near the town of Granby [45].

As for the consumer product additives and synthetic compounds, more than half of them (65 compounds) were congeners containing repeating ethylene oxide or propylene oxide units such as octylphenol ethoxylates (OPEOs), alcohol ethoxylates, polyethylene glycols (PEGs) and alkyl PEG ethers. OPEOs were notably absent in the mzCloud database but the GNPS molecular

networks proved to be particularly effective in their identification as these very similar congeners were linked in a network as shown in Figure SI-8 (Supporting Information). These compounds are used as non-ionic surfactants in multiple industries. However, OPEOs have shown estrogenic activity [46-48] and have been used to replace nonylphenol ethoxylates which were found in neighboring rivers previously [49]. In total, 18 congeners of the OPEO family (OPEO-1, with one ethoxylate monomer to OPEO-18) were tentatively identified then confirmed with reference standards as can be seen in the case of OPEO-9 in Table 2. For the other congeners' confirmation, please see Figure SI-9 (Supporting Information). While this study is fully qualitative, signal intensity for OPEO-3 to OPEO-15 was saturated in samples preconcentrated by solid-phase extraction (SPE) with a factor of 400. This speaks of a potentially very concerning level of contamination. Following identification by the proposed screening investigation, quantitation of these compounds could be planned in future studies to assess the level of contamination in the river and further estimate the risk that OPEOs and the other concerning contaminants pose. Congeners such as the OPEOs however represent a very complex challenge in term of quantification as these compounds are not commercially available as pure individual reference standards but rather as a mixture of congeners of various polyethylene oxide chain lengths.

Other consumer additives that were confirmed were the flame-retardant tris(2-butoxyethyl) phosphate and the bittering agent denatonium. The spectra match of denatonium can be seen in Figure SI-10 (Supporting Information). This compound has been detected in several WWTPs in Germany [50], but to the authors' knowledge, this is the first time it is reported in Canadian surface waters. Other industrial compounds were identified with a confidence level of 2 (probable structure) such as three members of the phthalate family (dibutylphthalate, dioctylphthalate, and diisodecylphthalate), the surfactant dimethyldioctadecylammonium, used in detergents, fabric softeners and flocculating agent in WWTPs and the multipurpose chemical 2-(2-(2-(2-phenoxyethoxy) ethoxy)ethoxy)ethanol.

The number of compounds identified as probable structures (212) and confirmed structures (44) with the three NTS tools compares well with recent a work using non-targeted methods and an empirical library where a number of 68 compounds were tentatively identified in a Mediterranean River basin [51]. The strength of the multi-tool method partly rests on the comprehensive size of the databases that counted over 240 000 compounds compared to 2000 compounds for the referenced article. A work where the *in silico* database SPS was used indeed showed a higher number of tentatively identified compounds with 200 compounds [24] which would suggest a higher rate of identification with larger databases. SPS, MetFrag and GNPS are complementary: MetFrag (with patRoon) offers peak picking, formula generation and access to custom databases and international suspects lists while being open source; SPS is simple to operate, and its data can be efficiently managed with Microsoft Access databases; GNPS gives access to empirical databases and generates molecular networks. The annotations made by the two combinatorial tools can in turn help gather more information on the GNPS networks can be useful to identify transformation products.

4 Conclusion

The hybrid method presented in this paper showcased its efficiency for identifying new trends in contamination as was the case for denatonium and the hypertension medications irbesartan and telmisartan. It also showed its potency to uncover yet unknown transformation products. The complementarity of SPS, GNPS and MetFrag allowed for increased confidence in the tentative identification made by only one tool.

Consumer product additives and other synthetic compounds required more treatment time since there is often less information about them on their PubChem page compared to pesticides and pharmaceuticals as was also the case with mzCloud where few of these compounds were present in the library. Additional research must be conducted to make sure it is a likely match. For that reason, databases need to provide data and metadata more readily accessible to address this issue in the future. While at this point extensive NTS analysis is still too time intensive for frequent monitoring, it remains crucial to detect new forms of contamination (OPEOs, alkyl PEG ethers) and identify pharmaceutical metabolites or transformation products, it shows invaluable ability to guide water quality programs to include new target compounds in monitoring programs, thus acting like an analytical compass for quantitative target-oriented approaches. Quantitation remains the end-goal as concentrations are needed to properly evaluate risk and the extent of contamination and it presents its own significant challenges. Following the identification step, the quantitation of the most concerning contaminants will be then tackled in a future study. Still, since no laboratory can afford to buy and keep every single reference standard likely to be present in environmental samples, and even less so their stable labeled isotopes as internal standards, NTS is set to become a cornerstone for the analysis of trace organic contaminants in surface waters and it will continue to improve its performance in the next years.

5 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 Q-OrbitrapMS was funded by the *Fonds de Recherche du Québec* (FRQ) and the Canadian Foundation for Innovation (CFI) through a John R. Evans Leaders grant (no. 36706) awarded to F. B. 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 Daniel L. Sweeney from MathSpec Inc. for his help with the Similar Partition Searching (SPS) algorithm.

6 References

[1] A.-H.M. Emwas, The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research, in: J.T. Bjerrum (Ed.), Metabonomics, Humana Press, New York, NY, 2015, pp. 161-193.
[2] M. Krauss, H. Singer, J. Hollender, LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns, Anal. Bioanal. Chem. 397 (2010) 943-951.

[3] M. Zedda, C. Zwiener, Is nontarget screening of emerging contaminants by LC-HRMS successful? A plea for compound libraries and computer tools, Anal. Bioanal. Chem. 403(9) (2012) 2493-2502.
[4] J. Hollender, E.L. Schymanski, H.P. Singer, P.L. Ferguson, Nontarget screening with high resolution mass spectrometry in the environment: ready to go?, Environ. Sci. Technol. 51(20) (2017) 11505-11512.
[5] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: communicating confidence, Environ. Sci. Technol. 48(4) (2014) 2097-2098.
[6] E. Eysseric, K. Barry, F. Beaudry, M. Houde, C. Gagnon, P.A. Segura, Application of spectral accuracy to improve the identification of organic compounds in environmental analysis, Anal. Chem. 89(18) (2017) 9805–9813.
[7] K. Dührkop, M. Fleischauer, M. Ludwig, A.A. Aksenov, A.V. Melnik, M. Meusel, P.C. Dorrestein, J. Rousu, S. Böcker, SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information, Nature Methods 16(4) (2019) 299-302.
[8] J.C. Erve, M. Gu, Y. Wang, W. DeMaio, R.E. Talaat, Spectral accuracy of molecular ions in an LTQ/Orbitrap mass spectrometer and implications for elemental composition determination, J. Am. Soc. Mass Spectrom. 20(11) (2009) 2058-2069.
[9] M. Mardal, M.F. Andreasen, C.B. Mollerup, P. Stockham, R. Telving, N.S. Thomaidis, K.S. Diamanti, K. Linnet, P.W. Dalsgaard, HighResNPS. com: an online crowd-sourced HR-MS database for suspect and non-targeted screening of new psychoactive substances, J. Anal. Toxicol. 43(7) (2019) 520-527.
[10] R. Bade, B.J. Tscharke, J.M. White, S. Grant, J.F. Mueller, J. O'Brien, K.V. Thomas, C. Gerber, LC-HRMS suspect screening to show spatial patterns of new psychoactive substances use in Australia, Sci. Total Environ. 650 (2019) 2181-2187.
[11] Y. Picó, D. Barceló, Transformation products of emerging contaminants in the environment and high-resolution mass spectrometry: a new horizon, Anal. Bioanal. Chem. 407(21) (2015) 6257-6273.
[12] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, Nature methods 12(6) (2015) 523.
[13] S. Broecker, S. Herre, B. Wüst, J. Zweigenbaum, F. Pragst, Development and practical application of a library of CID accurate mass spectra of more than 2,500 toxic compounds for systematic toxicological analysis by LC–QTOF-MS with data-dependent acquisition, Anal. Bioanal. Chem. 400(1) (2011) 101-117.
[14] H.P. Singer, A.E. Wössner, C.S. McArdell, K. Fenner, Rapid screening for exposure to “non-target” pharmaceuticals from wastewater effluents by combining HRMS-based suspect screening and exposure modeling, Environ. Sci. Technol. 50(13) (2016) 6698-6707.
[15] S. Samanipour, M.J. Reid, K. Bæk, K.V. Thomas, Combining a deconvolution and a universal library search algorithm for the nontarget analysis of data-independent acquisition mode liquid chromatography− high-resolution mass spectrometry results, Environ. Sci. Technol. 52(8) (2018) 4694-4701.
[16] R.R. da Silva, P.C. Dorrestein, R.A. Quinn, Illuminating the dark matter in metabolomics, P Natl Acad Sci USA 112(41) (2015) 12549-12550.

[17] S. Herrera-Lopez, M. Hernando, E. García-Calvo, A. Fernández-Alba, M. Ulaszewska, Simultaneous screening of targeted and non-targeted contaminants using an LC-QTOF-MS system and automated MS/MS library searching, J. Mass Spectrom. 49(9) (2014) 878-893.
[18] Chemical Abstracts Service, CAS registry - The gold standard for chemical substance information, 2020. https://www.cas.org/support/documentation/chemical-substances, (access: 2020/07/07).
[19] PubChem, PubChem Data Counts, 2020. https://pubchemdocs.ncbi.nlm.nih.gov/statistics, (access: 2020/01/14).
[20] F. Hufsky, S. Böcker, Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data, Mass Spectrom. Rev. 36(5) (2017) 624-633.
[21] D.L. Sweeney, Small molecules as mathematical partitions, Anal. Chem. 75(20) (2003) 5362-5373.
[22] K. Scheubert, F. Hufsky, D. Petras, M. Wang, L.-F. Nothias, K. Dührkop, N. Bandeira, P.C. Dorrestein, S. Böcker, Significance estimation for large scale metabolomics annotations by spectral matching, Nat. Commun. 8(1) (2017) 1494.
[23] D.L. Sweeney, A data structure for rapid mass spectral searching, Mass Spectrometry 3(Special Issue 2) (2014) S0035-S0035.
[24] I. Ferrer, D.L. Sweeney, E.M. Thurman, J.A. Zweigenbaum, Non-Targeted Screening of Water Samples Using Data Dependent Acquisition with Similar Partition Searching, J. Am. Soc. Mass Spectrom. 31 (2020) 1189-1204.
[25] M. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kapono, T. Luzzatto-Knaan, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, Nat Biotechnol 34(8) (2016) 828.
[26] R.A. Quinn, L.-F. Nothias, O. Vining, M. Meehan, E. Esquenazi, P.C. Dorrestein, Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy, Trends Pharmacol. Sci. 38(2) (2017) 143-154.
[27] S. Wolf, S. Schmidt, M. Müller-Hannemann, S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra, BMC Bioinformatics 11(1) (2010) 148.
[28] C. Ruttkies, E.L. Schymanski, S. Wolf, J. Hollender, S. Neumann, MetFrag relaunched: incorporating strategies beyond in silico fragmentation, Journal of cheminformatics 8(1) (2016) 3.
[29] Statistics Canada, Granby [Population centre], Quebec and Quebec [Province] (table). Census Profile. 2016 Census., 2017. http://geodepot.statcan.ca/Diss/GeoSearch/index.cfm?lang=E (access: 2019/11/29).
[30] E. Eysseric, X. Bellerose, J.-M. Lavoie, P.A. Segura, Post-column hydrogen-deuterium exchange technique to assist in the identification of small organic molecules by mass spectrometry, Can. J. Chem. 94 (2016) 781-787.
[31] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, A cross-platform toolkit for mass spectrometry and proteomics, Nat Biotechnol 30(10) (2012) 918-920.
[32] B.L. Milman, A procedure for decreasing uncertainty in the identification of chemical compounds based on their literature citation and cocitation. Two case Studies, Anal. Chem. 74(7) (2002) 1484-1492.

[33] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome research 13(11) (2003) 2498-2504.
[34] R. Helmus, patRoon handbook, 2019. https://rickhelmus.github.io/patRoon/articles/handbook.html, (access: 2020/01/18).
[35] V. Albergamo, J.E. Schollée, E.L. Schymanski, R. Helmus, H. Timmer, J. Hollender, P. De Voogt, Nontarget screening reveals time trends of polar micropollutants in a riverbank filtration system, Environ. Sci. Technol. 53(13) (2019) 7584-7594.
[36] A.D. McEachran, K. Mansouri, C. Grulke, E.L. Schymanski, C. Ruttkies, A.J. Williams, "MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies, Journal of Cheminformatics 10(1) (2018) 45.
[37] R.M. Borkar, B. Raju, R. Srinivas, P. Patel, S.K. Shetty, Identification and characterization of stressed degradation products of metoprolol using LC/Q-TOF-ESI-MS/MS and MSn experiments, Biomed. Chromatogr. 26(6) (2012) 720-736.
[38] B. Kasprzyk-Hordern, R. Dinsdale, A. Guwy, Multi-residue method for the determination of basic/neutral pharmaceuticals and illicit drugs in surface water by solid-phase extraction and ultra performance liquid chromatography–positive electrospray ionisation tandem mass spectrometry, J. Chromatogr. A 1161(1-2) (2007) 132-145.
[39] M. Gros, M. Petrović, D. Barceló, Development of a multi-residue analytical methodology based on liquid chromatography- tandem mass spectrometry(LC-MS/MS) for screening and trace level determination of pharmaceuticals in surface and wastewaters, Talanta 70(4) (2006) 678-690.
[40] K. Zhang, Y. Zhao, K. Fent, Cardiovascular drugs and lipid regulating agents in surface waters at global scale: Occurrence, ecotoxicity and risk assessment, Sci. Total Environ. (2020) 138770.
[41] M. Oss, A. Kruve, K. Herodes, I. Leito, Electrospray ionization efficiency scale of organic compounds, Anal. Chem. 82(7) (2010) 2865-2872.
[42] A. Kiontke, A. Oliveira-Birkmeier, A. Opitz, C. Birkemeyer, Electrospray ionization efficiency is dependent on different molecular descriptors with respect to solvent pH and instrumental configuration, PLoS One 11(12) (2016) e0167502.
[43] Q. Xiong, Y.-S. Liu, L.-X. Hu, Z.-Q. Shi, W.-W. Cai, L.-Y. He, G.-G. Ying, Co-metabolism of sulfamethoxazole by a freshwater microalga Chlorella pyrenoidosa, Water Res. (2020) 115656.
[44] M.A. Stravs, F. Pomati, J. Hollender, Exploring micropollutant biotransformation in three freshwater phytoplankton species, Environ. Sci.: Processes Impacts 19(6) (2017) 822-832.
[45] Y. Ryu, D. Barceló, L.P. Barron, L. Bijlsma, S. Castiglioni, P. de Voogt, E. Emke, F. Hernández, F.Y. Lai, A. Lopes, M. Lopez de Alda, N. Mastoianni, K. Munro, J. O'Brien, C. Ort, B.G. Plósz, M.J. Reid, V. Yargeau, K.V. Thomas, Comparative measurement and quantitative risk assessment of alcohol consumption through wastewater-based epidemiology: An international study in 20 cities, Sci. Total Environ. 565 (2016) 977-983.
[46] E.J. Routledge, J.P. Sumpter, Structural features of alkylphenolic chemicals associated with estrogenic activity, J Bio Chem 272(6) (1997) 3280-3288.
[47] M. Seki, H. Yokota, M. Maeda, H. Tadokoro, K. Kobayashi, Effects of 4-nonylphenol and 4-tert-octylphenol on sex differentiation and vitellogenin induction in medaka (Oryzias latipes), Environmental Toxicology and Chemistry: An International Journal 22(7) (2003) 1507-1516.
[48] S. Gronen, N. Denslow, S. Manning, S. Barnes, D. Barnes, M. Brouwer, Serum vitellogenin levels and reproductive impairment of male Japanese Medaka (Oryzias latipes) exposed to 4-tert-octylphenol, Environ. Health Perspect. 107(5) (1999) 385-390.

[49] D. Berryman, Un suivi des nonylphénols éthoxylés dans sept cours d'eau recevant des eaux usées traitées d’entreprises de textiles, in: Direction du suivi de l’état de l’environnement (Ed.) Ministère du Développement durable de l’Environnement et des Parcs, Québec, QC, 2005, p. 41.
[50] S. Lege, G. Guillet, S. Merel, J.E.Y. Heras, C. Zwiener, Denatonium–A so far unrecognized but ubiquitous water contaminant?, Water Res. 112 (2017) 254-260.
[51] A. Ccanccapa-Cartagena, Y. Pico, X. Ortiz, E.J. Reiner, Suspect, non-target and target screening of emerging pollutants using data independent acquisition: Assessment of a Mediterranean River basin, Sci. Total Environ. 687 (2019) 355-368.