Amélioration des techniques d'identification en spectrométrie de masse et étude de la transformation de contaminants organiques

par

Emmanuel Eysseric

Thèse présentée au Département de chimie en vue
de l'obtention du grade de Philosophia doctor (Ph.D.)
FACULTÉ DES SCIENCES, UNIVERSITÉ DE SHERBROOKE

Juillet 2022

Le 20 juillet 2022

le jury a accepté la thèse de monsieur Emmanuel Eysseric
dans sa version finale.

Membres du jury

Professeur Pedro A. Segura
Directeur de recherche
Département de chimie

Monsieur Christian Gagnon
Codirecteur de recherche
Environnement et changement climatique Canada

Professeur Jean-Philippe Bellenger
Évaluateur interne
Département de chimie

Professeure Violaine Ponsin
Évaluatrice externe
Département des sciences de la Terre et de l'atmosphère
Université du Québec à Montréal

Professeure Céline Guéguen
Présidente-rapporteuse
Département de chimie

SOMMAIRE

On comptait en 2020 plus de 350 000 composés organiques en circulation, dont un grand nombre mal défini et dans des mélanges chimiques complexes. Il y a un manque de connaissance majeur sur la présence environnementale et la stabilité des contaminants organiques dans les milieux aquatiques et encore plus pour leurs produits de transformation (PTs) qui y sont formés. Les méthodes de monitoring de contaminants sont spécifiques à quelques dizaines de composés et ne fournissent pas un tableau suffisant de la contamination. Les analyses non ciblées cherchent à pallier ce problème, mais obtenir une identification conclusive est complexe en raison de la taille limitée des banques de données et des faibles concentrations des contaminants. Les PTs sont particulièrement difficiles à élucider comme leur identité est généralement inconnue.

Cette thèse en trois axes tente d'abord d'améliorer et développer des techniques et méthodes d'identification de contaminants organiques et de leurs PTs dans les eaux de surface. Puis, elle se penche sur l'identification des PTs stables de quatre contaminants organiques préoccupants après une exposition simulant la photolyse solaire pour ensuite les retrouver dans des échantillons environnementaux. Finalement, dans le troisième axe, on cherche à identifier des PTs et des composés issus de mélanges chimiques complexes directement depuis des échantillons d'eaux de surface sans avoir recours à des études de dégradation de contaminants en laboratoire.

Dans le premier axe, une méthode d'échange hydrogène-deutérium permettant de distinguer des composés similaires a d'abord été développée. Puis, un outil de génération de formules moléculaires a été évalué puis adapté à des conditions environnementales ; il s'est avéré supérieur aux autres méthodes avec lesquelles il a été comparé. Finalement, un flux de travail non ciblé a été développé avec des outils de correspondances de spectres de masse *in silico* et de réseaux moléculaires. Plus de 250 contaminants ont été identifiés dans la rivière Yamaska à Granby dont plusieurs jamais reportés au Canada.

Dans le deuxième axe, des PTs de quatre contaminants préoccupants ont été générés puis identifiés en laboratoire avec les méthodes développées dans l'axe 1. Ces PTs ont ensuite

été recherchés et détectés dans des échantillons réels. Dans le troisième axe, le flux de travail développé dans l'axe 1 a été amélioré en incorporant des outils regroupant des spectres de masse en tandem similaires en réseaux. Cela a permis d'identifier plus de 400 contaminants. Plusieurs PTs de composés pharmaceutiques et de pesticides auparavant inconnus ont été identifiés ainsi que plus d'une centaine de congénères d'additifs de produits de consommation. Ce flux répond à des besoins importants en analyse environnementale en ce qui a trait à l'identification d'homologues et de produits de transformation divers. Cependant, les PTs des composés étudiés en laboratoires n'ont pas pu être identifiés directement par le flux de travail, soulignant l'importance des méthodes de contrôle en laboratoire.

Cette thèse joint les approches ascendantes où l'on génère en laboratoire des produits de transformation et les approches descendantes où ceux-ci sont identifiés directement à même l'échantillon avec le flux de travail. D'une part, elle a accru notre compréhension sur le devenir de plusieurs contaminants particulièrement préoccupants avec les études en laboratoire puis a poussé le cheminement en les retrouvant dans échantillons réels et en simulant leur toxicité pour une approche ascendante totale. Parallèlement, elle consiste une preuve de principe sur comment mener une analyse non ciblée selon l'état de l'art. Les outils ont été validés en fournissant des informations précieuses aux élaborateurs de programmes de monitoring et de suivi de toxicité sur l'occurrence locale de nombreux contaminants organiques et leurs PTs.

Mots clés : Spectrométrie de masse à haute résolution, produits de transformation, conformité spectrale, analyse donnée-dépendante, banques de données combinatoires

REMERCIEMENTS

Je tiens à remercier mon superviseur Pedro Segura et mon co-superviseur Christian Gagnon de m'avoir conseillé, encouragé et guidé pendant cette thèse. Ils ont su m'instiller leur passion de la recherche et leur rigueur scientifique. Je suis très reconnaissant à Francis Beaudry pour ses conseils et les opportunités qu'il a offerts ; travaillé avec lui a été un véritable plaisir. J'aimerais aussi remercier Audrey Roy-Lachapelle et Sylvie Poirier Larabie pour leur aide, leur soutien et leurs conseils. Je tiens aussi à remercier les professionnels du département Philippe Venne, René Gagnon et Michel Trottier pour leurs conseils et leur disponibilité. Je tiens aussi à remercier tous les étudiants du labo Segura avec qui j'ai eu la chance de travailler lors de ma thèse : Marc-André Lecours, Killian Barry, Lisa Lahens, François-Xavier Teysseire, Xavier Bellerose, Laurie Fréchette Viens, Joey Isabelle, Mathieu Racine, Cassandra Guérette, Valérie Boucher, Hugo Alarie, Mégane Moreau, Judith Boudrias. Je suis aussi reconnaissant à Lou Paris de l'Organisme du Bassin Versant de la rivière Yamaska pour la belle opportunité qu'a été cette campagne d'échantillonnage. J'aimerais aussi remercier Yongdong Wang de Cerno Bioscience et Daniel Sweeney de Mathspec avec lesquels des partenariats fructueux ont pu être établis. Je tiens finalement à remercier ma famille, ma conjointe Mary-Jane, mes amis et mes trois chats Titanic, Daisy et Taïga qui ont su animer la rédaction de ma thèse comme de mes articles.

Je dédie cette thèse à mon grand-père Serge qui aurait tant aimé assister à ma soutenance et qui m'a transmis son amour de la chimie.

TABLE DES MATIERES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

MS : spectrométrie de masse

LC-MS : chromatographie liquide couplée à la spectrométrie de masse

QqTOFMS: Quadrupole time of flight

HDX: échange hydrogène deuterium

MS^2 : spectrométrie de masse en tandem

ATV : atorvastatine

BEZ : bézafibrate

OXZ : oxybenzone

TBEP : phosphate de tris(2-butoxyethyle)

DDA : analyse donnée dépendante (data dependent analysis)

APEOs : éthoxylates d'alkylphénol

OPEOs : éthoxylates d'octylphénol

NTS : dépistage non ciblé (non targeted screening)

SPS : similar partition searching

GNPS: global natural products social networking

HRMS: spectrométrie de masse à haute résolution

$HRMS^2$: spectrométrie de masse en tandem à haute résolution

UHPLC : chromatographie liquide à ultra-haute performance

Ppm : partie par million

Ppb : partie par milliard

Ppt : partie par trillion

MeOH : méthanol

ACN : acétonitrile

NMR : résonnance magnétique nucléaire

SPE : extraction en phase solide

INTRODUCTION

En date d'octobre 2020, il existait plus de 350 000 composés et mélanges de composés enregistrés pour la production et l'utilisation dans 19 pays.[1] La mondialisation impliquant une grande circulation des biens produits, beaucoup de ces composés peuvent ainsi se retrouver potentiellement dans les différents compartiments environnementaux.[1] Les agences et institution gouvernementales peinent à inventorier tous les nouveaux composés et à prédire leurs impacts potentiels alors que l'enjeu défraye de plus en plus la chronique.[2–6]

Les contaminants d'origine anthropique peuvent prendre la forme de composés pharmaceutiques, de pesticides et d'additifs de produits de consommation ; une importante partie de ces derniers est par ailleurs largement inconnue alors que les structures de 50 000 composés étaient non divulguées par secret industriel et 70 000 autres étaient mal définies.[1] La pollution chimique est ainsi complexe à caractériser considérant la diversité de contaminants et la méconnaissance de leur occurrence et leur toxicité.[7]

Ces contaminants se retrouvent dans les cours d'eau à travers des voies différentes selon leur usage.[8] Leur présence dans les milieux aquatiques cause des problèmes d'exposition pour les organiques aquatiques. Dans le cas de composés constamment réintroduits dans l'eau comme les produits pharmaceutiques et les cosmétiques par exemple, les effets peuvent se faire ressentir de façon multigénérationnelle même si les composés relâchés ont une demi-vie de l'ordre du mois en milieu environnemental.[9]

Les programmes de surveillance (ou *monitoring*) de contaminants dont le Plan de Gestion des produits chimiques canadien fait partie surveillent exclusivement la présence de contaminants organiques jugés préoccupants selon leur niveau de dangerosité et d'exposition sans chercher à identifier d'autres composés.[10,11] Au Canada ce sont ainsi environ 400 composés qui sont suivis; un tel nombre n'est pas suffisant pour procurer un portrait de la contamination en comparaison aux centaines de milliers de contaminants potentiellement présents. Aux lacunes des programmes de

monitoring vient s'ajouter la méconnaissance de l'identité des produits de transformation de la très grande majorité des contaminants organiques.[14,15] Les contaminants dans les eaux de surface peuvent être transformés de multiples manières; si la photolyse est la voie de dégradation majeure pour de nombreux pesticides et composés pharmaceutiques, [16–23] la biotransformation génère aussi des produits de transformation.[14,24–27] Il en résulte ainsi un grand manque de connaissances sur la présence environnementale des contaminants organiques et leur devenir dans les milieux aquatiques.

L'analyse non ciblée est universelle dans son approche et tente de recueillir le maximum de données possible en restreignant le moins possible les composés détectables.[28–37] L'analyse non ciblée est toutefois non quantitative; elle a un but qualitatif où l'on cherche à identifier des composés inconnus et est en ce sens complémentaire à l'analyse ciblée.[14] L'élucidation structurale des composés inconnus est conduite avec les multiples informations obtenues suite à leur analyse par spectrométrie de masse en tandem à haute résolution couplée aux chromatographies liquide ou en phase gazeuse.[14,32,38] Malgré tout, une identification avec un haut niveau de certitude ou de confiance reste ardue.[38] Les incertitudes reliées à la formule moléculaire d'un inconnu, à ses potentiels isomères ou à la taille limitée des banques de données de composés sont autant de freins à son identification. Les produits de transformation et les homologues, ou congénères, issus de mélanges chimiques complexes sont encore plus difficiles à identifier en raison du manque de données partiel ou total sur leur structure.

Cette thèse cherche à répondre à trois questions : i) comment améliorer les outils et méthodes d'identification de composés inconnus par spectrométrie de masse à haute résolution en analyse environnementale ; ii) quelle est l'identité des produits de transformation stables et principaux de quatre contaminants organiques préoccupants générés en laboratoire par photolyse en imitant le soleil ; et iii) peut-on identifier les produits générés en laboratoire en plus d'autres produits de transformation et d'homologues issus de mélanges chimiques complexes dans une analyse non ciblée d'échantillons réels ? Ainsi, la thèse s'articule autour de l'identification des composés pour

lesquels les données d'occurrence sont les plus incomplètes : les produits de transformation et les homologues d'additifs de produits de consommation.

Le chapitre 1 de la présente thèse est un sommaire de l'état de l'art sur l'occurrence et le devenir des contaminants organiques dans les eaux de surface avec un intérêt particulier pour les composés pharmaceutiques ainsi que les méthodes analytiques utilisées pour identifier ces composés. Le deuxième chapitre détaille les techniques, méthodes et logiciels utilisés pour obtenir les résultats de cette thèse. Les chapitres 3, 4, 5, 6 et 7 sont les résultats de la thèse sous forme d'articles insérés. Dans le troisième chapitre, il est question d'une méthode d'échange hydrogène-deutérium pour augmenter la qualité de l'identification. Le quatrième chapitre porte sur le développement est l'application d'un outil de détermination de formule moléculaire à opérer à des concentrations environnementales avec une matrice d'eau de surface. Le cinquième chapitre quant à lui présente un flux de travail utilisant des outils de spectres combinatoires pour identifier des composés inconnus. Puis, le sixième chapitre se penche sur l'identification de produits de transformation de contaminants préoccupants générés en laboratoire. Le septième chapitre traite du flux de travail développé dans le cinquième chapitre amélioré puis utilisé pour identifier des produits de transformation et des congénères d'additifs de produits de consommation. Finalement, le huitième chapitre est la conclusion générale alors que le neuvième liste les références.

CHAPITRE 1. ÉTAT DE L'ART

Les connaissances sur la présence des contaminants dans les différents compartiments environnementaux, aussi appelée l'occurrence environnementale, et sur leur dégradation et les composés en lesquels ils sont transformés sont étroitement liées aux avancées en analyse environnementale.[39]

1.1 Occurrence et devenir des contaminants organiques à l'état de trace et de leurs produits de transformation

Les contaminants organiques peuvent entrer dans l'environnement de multiples façons selon leur nature et leur utilisation comme on peut le voir en **Figure 1**. Ils y sont exposés à des conditions pouvant mener leur dégradation en produits de transformation.[9]

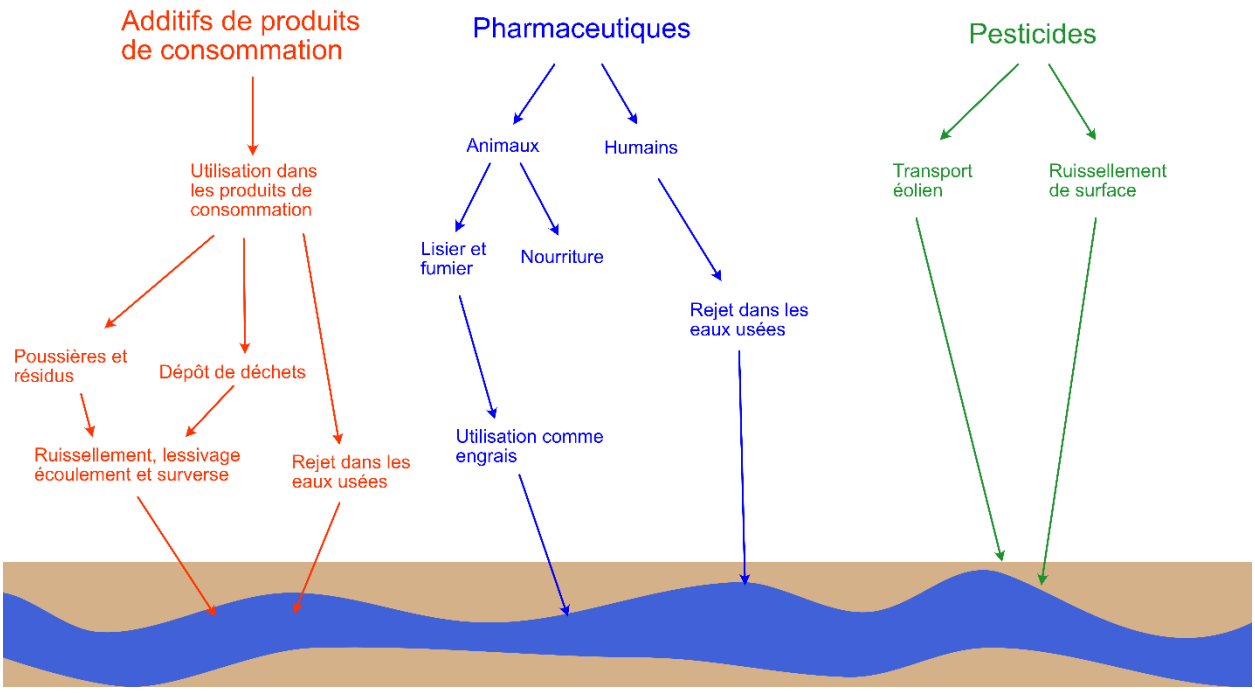


Figure 1. Voies d'entrées potentielles de contaminants organiques dans les milieux aquatiques. Inspiré de Suman et al., 2022.[8]

1.1.1 Occurrence des contaminants organiques à l'état de trace

Les contaminants organiques retrouvés dans les milieux aquatiques peuvent être classifiés en différents groupes. Dans cette thèse, la classification utilisée les séparera comme les additifs de produits de consommation, les composés pharmaceutiques et les pesticides.

1.1.1.1 Les additifs de produits de consommation

Les additifs de produits de consommation forment une grande classe de composés regroupant notamment les additifs de polymères, les tensioactifs, les produits cosmétiques et de soin personnel et les additifs alimentaires. Les additifs de polymères sont ajoutés aux plastiques pour modifier leurs propriétés; ils incluent les stabilisateurs de chaleurs et de froid, les stabilisateurs UV, les produits ignifuges, les plastifiants et les accélérateurs de polymérisation. Un volume annuel estimé de 25 milliards de kilogrammes était projeté pour 2021.[40] Certains additifs ont été règlementés comme plusieurs produits ignifuges bromés lors de la convention de Stockholm[41] ou le bisphénol A dans les biberons[42]. La communauté scientifique avait également étudié les tensioactifs nonylphénols éthoxylates compte tenu de leur très forte présence dans les eaux de surface et leur activité oestrogénique.[43]

Un problème majeur dans la compréhension de l'occurrence environnementale des additifs de produits de consommation réside dans l'absence d'information sur le nombre croissant de ces composés. Même dans les cas où les formulations des composés sont connues, leur concentration l'est rarement des consommateurs.[44] Les structures de plus de 50 000 additifs sont inconnues car leurs producteurs invoquent le secret de fabrication pour justifier leur non-divulgation.[1] Un tel angle mort affaiblit les modèles d'analyse de risques en écotoxicologie comme en exposomique. De plus, les structures de 70 000 autres composés sont décrites de manières ambiguë dans les cas de ce qu'on appelle les mélanges chimiques complexes, des formulations chimiques multi composantes comme des séries d'oligomères ou des combinaisons de composés uniques.[1] Dans le cas d'oligomères, les propriétés de bioaccumulation, de toxicités et de solubilité peuvent varier fortement selon la longueur de la chaine.[45–47] Les produits de transformation des additifs sont donc

particulièrement hasardeux à investiguer compte-tenu du manque de clarté sur leurs composés parents. Un produit de transformation d'un anti-ozonant retrouvé dans les particules de pneus le long des routes a été mis en cause pour toxicité aiguë chez le saumon coho en 2021.[48]

1.1.1.2 Les composés pharmaceutiques

La question de la présence des produits pharmaceutiques en milieu aquatique a pris de l'ampleur à la fin 20e siècle[9,49–54] et la « pollution pharmaceutique » a été reconnue formellement comme un enjeu par les acteurs industriels et non gouvernementaux en 2015.[55] Un cas médiatisé de pollution pharmaceutique est celui de l'antiinflammatoire non stéroïdien diclofénac qui avait été lié à la quasi extinction plusieurs espèces de vautours au Pakistan et dans le sous-continent indien.[56] Un autre est celui de la féminisation des poissons par les hormones contraceptives[57–60]. La réintroduction continuelle de composés pharmaceutiques dans les milieux aquatiques est une préoccupation importante pour les organismes qui y vivent. Une demi-vie plus courte d'un composé dans le milieu est compensée par son insertion constante[9] et les effets sous-létaux multigénérationnels peuvent se faire ressentir.[61]

Les antiinflammatoires non stéroïdiens forment un groupe de composés pharmaceutiques très présents dans les rejets urbains. Des composés comme le naproxène[62], le diclofénac[51,63–65] et l'ibuprofène[62,66–68] ont été retrouvés dans les rejets urbains dans plusieurs villes du monde. À ces composés s'ajoute l'acétaminophène qui est fortement investigué, suivi et trouvé dans toutes les régions du monde.[69–74] Les médicaments utilisés dans le traitement et la prévention des troubles relatifs au système cardiovasculaire sont également investigué et détecté en importante quantité. Ainsi des composés de multiples familles comme les bêta bloquants[22,49,50,53,61,75–77], statines[49,53,76–82], fibrates[17,49,51,67,76–78,83–89], antagonistes des récepteurs de l'angiotensine 2[24,53,72,77,90] ainsi que certains de leurs métabolites et produits de transformation connus ont été étudiés et retrouvés dans des eaux de surface à travers le monde. Les médicaments utilisés dans le traitement et la prévention des troubles du système nerveux comme les antidépresseurs et les antiépileptiques sont une autre classe de composés fréquemment retrouvés dans les rejets urbains. Des composés comme la carbamazépine[51,90–93], la venlafaxine[94,95] et leurs métabolites ont été retrouvés dans des cours d'eau

de manière presque omniprésente. Finalement les antibiotiques forment une catégorie de composés très préoccupantes en raison de la formation de gènes de biorésistance dans l'environnement.[96,97] Des antibiotiques de plusieurs familles ont été trouvés dans les rejets urbains et les eaux de surface à travers le monde.[97]

1.1.1.3 Les pesticides

La présence des pesticides dans l'environnement et leurs impacts est connue depuis des décennies. Dans *Silent Spring* en 1962, Rachel Carson lève l'alarme sur le DDT et les pesticides organochlorés en relatant les effets néfastes du DDT sur les oiseaux.[98] Cet ouvrage s'avèrera majeur et contribuera à ouvrir la voie pour des règlementations environnementales sur les pesticides organochlorés.[99] L'usage de ces derniers deviendra restreint dans les décennies suivantes à travers le monde; ce sont neuf pesticides qui figureront sur la liste initiale des douze polluants organiques persistants de la convention de Stockholm en 2001: l'aldrin, le chlordane, le DDT, le dieldrin, l'endrin, l'heptachlor, l'hexachlorobenzene, le mirex et le toxaphene.[41]

Plus récemment, d'autres pesticides sont des sources de préoccupations en ce qui a trait à leur occurrence, leur persistance, leur potentiel de bioaccumulation, leur toxicité. L'atrazine est un herbicide triazine disponible sur le marché depuis des décennies. Bien que son usage ait été banni dans l'Union européenne 2004 en raison de son potentiel de contaminer les aquifères, l'atrazine et ses produits de transformation continuent d'être détectés dans les eaux côtières européennes de même que dans l'urine de femmes enceintes et d'ouvriers agricoles, ce qui témoigne de leur persistance.[100] Parmi les produits de transformation de l'atrazine, on compte l'hydroxyatrazine, la dééthyltrazine et la déisopropylatrazine.[100] L'atrazine a des impacts négatifs sur de nombreux organismes dont les crustacés, les insectes, les mollusques, les poissons, les amphibiens et les reptiles.[100] Le glyphosate est un herbicide en circulation depuis 1974 utilisé en formulation avec les tensioactifs polyethoxyéthylène amine (POEA) pour former le Roundup. Le glyphosate se transforme en acide phosphonique d'aminométhyle (AMPA). On a trouvé du glyphosate et de l'AMPA dans des eaux de surface, des eaux souterraines et des sédiments à travers le monde à des concentrations atteignant jusqu'à 430 µg/L.[101] Le glyphosate en formulation avec les POEA peut

avoir des impacts négatifs sur des organismes aquatiques dont les moules, crustacés, poissons et amphibiens.[101] Le métolachlore est un des pesticides les plus utilisés dans le monde et équivalait à 4.2% de l'utilisation globale de pesticides en 2015.[102] On compte parmi ses produits de transformation les acides éthanesulfonique et oxalique.[103] Le métolachlore et l'acide oxalique du métolachlore ont notamment des impacts négatifs sur l'écrevisse marbrée dans l'échelle du microgramme par litre, concentrations dans lesquelles on les mesurés dans des eaux de surface.[104]

D'autres pesticides sont également surveillés dans les compartiments environnementaux dont l'eau; on compte ainsi 22 pesticides organiques suivis dans les *recommandations pour la qualité de l'eau potable au Canada*.[105]

1.1.2 Devenir des contaminants organiques à l'état de trace en milieu aquatique

Une fois introduits en milieu aquatique, les contaminants organiques sont soumis à de multiples voies par lesquelles ils peuvent être dégradés telles la photolyse, l'hydrolyse et la transformation biotique.[106] Les produits de transformation sont généralement plus mobiles et polaires que les composés parents. Des réactions fréquemment observées dans les composés organiques, incluant notamment les produits pharmaceutiques, comme l'hydroxylation, la déalkylation, l'hydrolyse et l'oxydation sont montrées dans la **Table 1**.

Table 1. Transformations observées dans divers composés pharmaceutiques

Transformation	**Composé initial**	**Référence**
Hydroxylation	Diclofénac	[27,107]
Déalkylation	Ibuprofène	[107]
	Irbesartan	[24]
Dihydroxylation	Carbamazépine	[108]
Formation d'un acide carboxylique depuis un alcool terminal	Métoprolol	[109]

Hydrolyse d'un éther	Naproxène	[107]
	Venlafaxine	[94]
	Diltiazem	[110,111]
Déalkylation d'une amine	Lidocaïne	[94]
	Venlafaxine	[94]
	Diltiazem	[110,111]
Oxydation	Ibuprofène	[107]
Déacétylation	Diltiazem	[110,111]

1.1.2.1 Métabolisme de phase I, II et III

Le métabolisme des composés pharmaceutiques et des pesticides ingérés dans les aliments est un cas particulier parmi les contaminants observés en milieu environnemental en raison de leur métabolisme par les êtres humains ou les animaux qui les consomment avant qu'ils ne soient relâchés dans l'environnement.[112,113]

Les composés sont métabolisés en plusieurs phases; dans la première des enzymes vont généralement introduire de groupements fonctionnels polaires.[114] Puis, en phase II, ils vont être conjugués à des molécules polaires pour faciliter leur métabolisme subséquent et leur excrétion en phase III.[115,116] Il a été observé pour plusieurs composés pharmaceutiques et leurs métabolites conjugués avec des composés en phase II comme les glucuronides un clivage du lien avec le conjugué et un retour vers le composé original.[11] Or, les conjugués de phase II peuvent avoir des concentrations supérieures à celles des composés initiaux dans les eaux de surface et ainsi agir comme un réservoir de composés pharmaceutiques.[117] Ainsi, un mélange de composés pharmaceutiques non métabolisés, métabolisés partiellement ou totalement ou métabolisés puis retransformés est déversé dans les eaux de surface. Les propesticides organothiophosphates et organodithiophosphate, comme le chlorpyrifos et le malathion respectivement, sont sujets à des transformations similaires à celles des composés pharmaceutiques en phase I. Les produits de transformation oxon comme le chlorpyrifos-oxon et le malaoxon, plus toxiques que les composés parents, sont ainsi relâchés dans l'environnement.[118]

1.1.2.2 Photolyse

La photolyse est le mécanisme de transformation majeur pour plusieurs composés qui résistent aux transformations biotiques dans les eaux de surface.[19,20,119] C'est notamment le cas de plusieurs composés pharmaceutiques comme le diclofénac, le kétoprofène, la sulfaméthoxazole et l'hydrochlorothiazide mais aussi pour des pesticides comme l'atrazine et le métolachlore.[16,22,27,102] Les transformations peuvent avoir lieu par photolyse directe ou indirecte. Dans le premier cas, un composé va directement absorber la lumière et subir une transformation. Dans le deuxième, des espèces réactives comme des radicaux hydroxy ou des oxygènes singulets vont réagir avec le composé; la présence dans l'eau de substances fulviques, de fer, de bicarbonate, de sulfate ou de nitrate va grandement accélérer la photolyse indirecte.[106] L'atorvastatine est principalement dégradée par photolyse indirecte; les nitrates et les acides fulviques sont les facteurs principaux influençant sa transformation, ces derniers permettant la formation d'oxygène singulets.[120]

Les acides fulviques peuvent aussi ralentir la photolyse des composés selon leur concentration et leur nature en diminuant la photolyse directe des molécules par rapport à si elles étaient en solution pure.[23]

1.2 L'analyse des contaminants organiques à l'état de trace dans les eaux de surface

La spectrométrie de masse en tandem à ionisation par électro-nébulisation couplée à la chromatographie liquide à ultra-haute performance (UHPLC-MS/MS) est la technique principale pour l'analyse des contaminants organiques dans les eaux de surface en raison de la sensibilité, la sélectivité, la grande gamme de composés analysables, la possibilité d' analyser des composés qui coéluent et l'information générée par des spectres MS^2.[14,38,118] Le type d'instrument varie selon si l'analyse est à but qualitatif ou quantitatif; ainsi les spectromètres de masse à triple-quadripôle sont utilisés pour les analyses quantitatives, ou ciblées, alors que les spectromètres de masse à haute résolution (HRMS) sont utilisés pour les analyses pour l'identification de contaminants, ce qu'on appelle une analyse non ciblée.[38,118,121]

1.2.1 L'analyse ciblée

En analyse ciblée, on sélectionne spécifiquement les transitions MS^2 des composés dans une fenêtre de temps de rétention chromatographique. Des étalons de référence sont nécessaires pour déterminer la fenêtre de sélection de temps de rétention, les paramètres de fragmentation et les paramètres quantitatifs de la méthode pour chaque molécule. [14,121] Seul le signal correspondant aux transitions des composés ciblés est acquis; l'appareil est ainsi « aveugle » à tout autre signal hors des gammes de *m/z* programmées dans la fenêtre de temps, peu importe son intensité. Il en découle que ce type d'analyse, nommé ciblé, est lourdement limité dans ses capacités de dépistage ou d'analyse à l'aveugle. De plus, il arrive fréquemment en analyse ciblée que les composés ciblés soient choisis selon leur présence dans des études précédentes et la facilité de les analyser et non selon la vraisemblance de les retrouver dans l'échantillon ou le danger qu'ils représentent. C'est ce qu'on appelle l'effet de Mathieu et cela causé que des contaminants préoccupants et potentiellement présents en importante concentration aient été délaissés pendant des années.[13]

1.2.2 L'analyse non ciblée

L'analyse non-ciblée a gagné en importance de la dernière décennie devant les besoins de la communauté scientifique confrontée d'une part à la quantité toujours croissante de composés produits, importés et relâchés dans l'environnement et l'autre à leurs produits de transformation.[8,32,38,122] L'analyse non ciblée est réalisée par spectrométrie de masse en tandem avec des instruments munis d'un quadripôle et d'un analyseur de masse à haute résolution. Les analyseurs de masse à temps de vol, les orbitrap, et les analyseurs de masse à résonance cyclonique ionique sont tous utilisés dans les méthodes d'analyses non ciblées. Ces analyseurs de masse sont non seulement plus à même de séparer les *m/z* qui seraient isobares avec un analyseur de plus faible résolution et de distinguer pleinement le patron isotopique d'un ion suffisamment intense dans les spectres plus chargés, mais ils peuvent également être calibrés pour garder une haute exactitude de masse, généralement inférieure à 1 mDa.[122] Ces masses à haute précision et exactitude en plus du patron isotopique sont utiles pour attribuer une formule chimique potentielle à un ion. À ces informations obtenues lors du balayage des ions s'ajoutent celles des expériences

d'acquisitions de MS^2. Les ions produits renseignent sur la structure de leur ion précurseur. Il existe des modes d'acquisition qui permettent d'obtenir des spectres MS^2 de manière automatique; dans l'acquisition donnée-dépendante (data dependent acquisition en Anglais), les ions dont l'intensité dépasse une valeur seuil sont sélectionnés et filtrés automatiquement au premier quadripôle puis acheminés à la cellule de collision où ils sont fragmentés et leurs ions produits sont balayés par l'analyseur de masse à haute résolution. Les spectres MS^2 générés automatiquement peuvent ensuite être comparés des banques de données de spectres pour identifier des inconnus. Ce type d'analyse a été utilisé avec succès en analyse environnementale comme dans d'autres disciplines comme la protéomique et la métabolomique.[15,30,32,36,48,123–126]

1.2.2.1 Les niveaux de confiance en identification

La qualité d'une identification réalisée en LC-MS est variable selon les informations à disposition, allant d'un simple ratio masse sur charge sans information sur l'adduit à une confirmation du temps de rétention et spectre MS^2 avec un étalon de référence. En 2014, Schymanski *et al.* propose le concept de niveaux de confiance dans une identification potentielle (Figure 2).[38] Une masse précise et exacte y figure comme plus bas niveau de confiance comme plusieurs formules moléculaires sont possibles. Le niveau suivi est celui dans lequel une formule sans ambiguïté est obtenue grâce à notamment des informations sur le patron isotopique. Grâce à des informations déduites dans le contexte analytique comme le temps de rétention ou des données de fragmentation issues de spectres MS^2, des structures potentielles sont possibles dans le troisième niveau de confiance, mais il n'y a pas suffisamment d'information pour pouvoir déterminer une structure en particulier. Dans le deuxième niveau de confiance, une correspondance de haute qualité avec une banque de données de spectre est trouvée sans ambigüité. Il s'agit du plus haut niveau de confiance qui peut être obtenu pour une structure sans avoir besoin d'un étalon de référence. C'est le cas pour un grand nombre de composés comme la majorité des produits de transformation et un grand nombre de congénères d'additifs de produits de consommation qui ne sont disponibles commercialement que sous forme de formulation complexe.[1,32,38] Le premier niveau de confiance est finalement obtenu

quand on utilise un étalon de référence pour valider la correspondance avec le spectre MS^2 et le temps de rétention.[38]

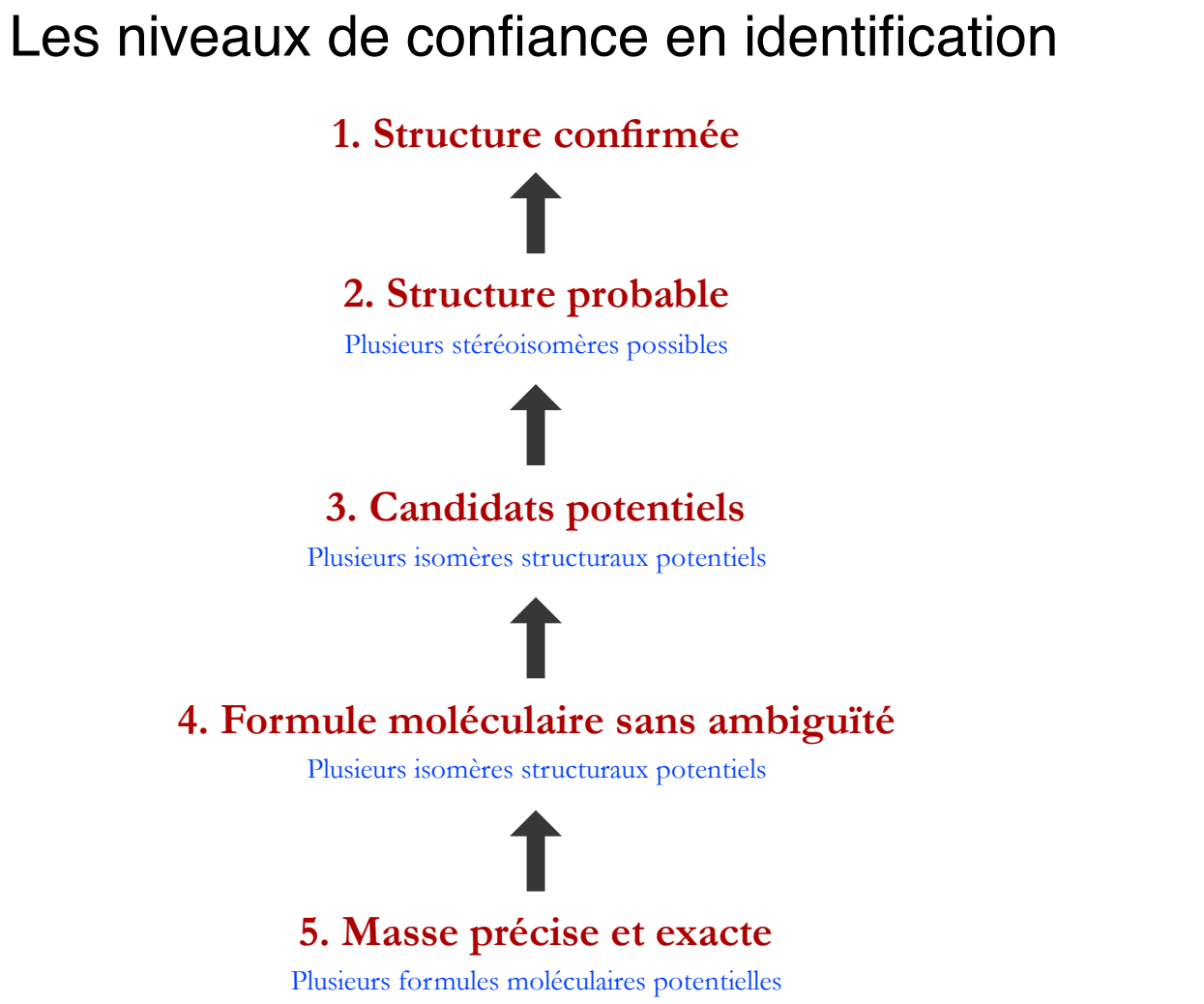


Figure 2. Les niveaux de confiance en identification pour les petites molécules en spectrométrie de masse. Adapté de Schymanski et al.[38]

1.2.2.2 Les types d'analyse non ciblée

L'analyse non ciblée se divise en deux catégories selon les objectifs initiaux de l'étude: le dépistage des suspects et le dépistage non-ciblé.

1.2.2.2.1 Le dépistage des suspects

Dans le dépistage des suspects, appelé *suspect screening* en Anglais, on va chercher à identifier des contaminants suspectés d'être présents dans l'échantillon après leur analyse par LC-HRMS et LC-HRMS^2. Le processus d'identification des inconnus se base sur les niveaux de confiance montrés en **Figure 2**. Le ou les *m/z* d'un suspect sont recherchés, sa formule moléculaire est ensuite déterminée et, si elle la même que celle du suspect, on continue à monter les niveaux de confiance pour le suspect avec des informations supplémentaires comme la correspondance du temps de rétention de l'inconnu avec le suspect selon un modèle prédictif puis celle des spectres MS^2.[127] Finalement, le suspect peut être confirmé avec un étalon de référence.

Le dépistage des suspects est une méthode efficace pour prioriser des composés préoccupants dans l'environnement parmi les milliers aux dizaines de milliers présents dans un échantillon. De surcroit, les étalons de référence ne sont pas nécessaires contrairement à l'analyse ciblée. Des composés non disponibles commercialement comme des produits de transformation récemment identifiés ou des congénères d'additifs de produits de consommation peuvent ainsi être ajoutés à aux listes de suspects. Il existe également des listes de contaminants potentiels suspects de se trouver dans diverses matrices comme celle du réseau NORMAN et les listes multiples disponibles sur le *Comptox Chemistry Dashboard* de l'US EPA.[128–130] Le dépistage des suspects est ainsi une forme d'analyse ciblée dans des données acquises de façon non-ciblée. En ce sens, il n'est pas exploratoire et des composés qui ne figurent pas sur la liste des suspects ne seront pas identifiés. L'effet de Mathieu peut aussi se manifester à travers des listes de suspects si les hypothèses sont limitées sur les contaminants potentiels. Il reste en revanche beaucoup moins prononcé comme les listes internationales contiennent plusieurs milliers de composés.[128]

Le dépistage des suspects reste cependant rétrospectif, ce qui signifie qu'il est possible de revenir en arrière dans d'anciennes analyses non ciblées pour confirmer la présence d'un contaminant identifié subséquemment. Cela a été notamment fait pour retracer la présence d'un contaminant

industriel dans le Rhin après qu'il ait été identifié par dépistage non ciblé.[32] Le dépistage des suspects a été utilisé pour identifier des contaminants issus de retours d'écoulement de fracturation hydraulique[131], des tensioactifs[132,133], des pharmaceutiques et leurs produits de transformation[22,25,134,135], des additifs illicites[31] et des toxines[136]. À terme, le dépistage des suspects se destine à suivre l'analyse ciblée pour la surveillance de contaminants dans les analyses de routine et comme moyen de préserver les échantillons numériquement.

1.2.2.2.2 Le dépistage non ciblé

Le dépistage non ciblé ou *non targeted screening* (NTS) est l'autre branche de l'analyse non ciblée. Le début du processus est essentiellement le même que pour le dépistage des suspects : une analyse LC-HRMS^2 est encore réalisée. Le but de l'expérience diffère en revanche en ce sens qu'il n'est pas d'identifier des composés choisis à l'avance dans une liste, mais plutôt d'identifier des inconnus sans a priori ou restrictions dans la recherche. Ainsi, le processus ascendant en **Figure 2** par lequel on gravit en confiance en passant d'un niveau à l'autre n'est pas utilisé de façon aussi rigide puisque la priorisation diffère. Tous les ions ayant un pic chromatographique sont considérés comme des inconnus. Leurs spectres MS^2, s'ils sont disponibles, sont extraits puis une recherche dans une banque de données est effectuée. C'est ici que la taille limitée des banques de spectres LC-HRMS^2 devient un frein sérieux à la capacité d'identification des analyses non ciblées. Il existe des outils computationnels de génération de spectres MS^2 qui permettent de construire des pseudo banques de données et ainsi d'élargir considérablement la quantité de composés identifiables par correspondance des spectres.[137,138] D'autres outils peuvent venir se greffer aux flux de travail; des algorithmes et logiciels permettent de regrouper les ions ayant des spectres MS^2 similaires, mais leur utilisation se limite surtout à d'autres disciplines comme l'étude des produits naturels et la métabolomique .[139–141]

Le NTS est ainsi véritablement non ciblé dans son approche. On tente d'identifier des contaminants en limitant les biais potentiels pour diminuer l'effet de Mathieu. Ce type d'approche limitant les hypothèses initiales pour les générer après l'analyse des résultats est dit agnostique.[142,143] Le NTS

est donc l'approche la mieux adaptée pour détecter des sources de contamination émergeantes ou jusque-là inconnues. La découverte du produit de transformation 6-PPD quinone, qui induit une toxicité aiguë chez le saumon coho du nord-ouest du Pacifique, dans des cours d'eau a été rendue possible grâce à une analyse non ciblée.[48] Suite à son identification, la présence du contaminant avait pu être confirmée rétrospectivement dans des études non ciblées antérieures. Les approches purement non ciblées dont fait partie le NTS sont importantes d'un point de vue exploratoire pour guider les programmes de monitoring et de toxicité. Ces méthodes sont cependant longues à réaliser et la constellation d'outils disponibles n'est pas unie sous une interface, ce qui ralentit le traitement des données et force une certaine redondance dans les analyses. Un haut niveau d'expertise et de familiarité est également nécessaire pour déployer les outils et modifier les paramètres de traitement des données pour un jeu de données spécifique.[144] De ce fait, la grande majorité des études non ciblées prennent la forme de dépistage des suspects en raison de la complexité de réaliser un dépistage non ciblé.

1.2.3 L'identification des produits de transformation

Considérant les limitations du dépistage non ciblé pour identifier des contaminants préalablement inconnus comme c'est le cas pour de nombreux produits de transformation, l'identification des produits de transformation est le plus souvent menée dans des études où les conditions de dégradation environnementales sont simulées en laboratoire.[14,18,118]

Ces études suivent un processus analytique ascendant, ou *bottom-up*, que l'on peut voir dans la **Figure 3**, dans lequel les composés sont choisis selon leur niveau et de toxicité et leur occurrence environnementale connue ou prédite. Les composés sont ensuite soumis à des conditions imitant ce qu'on trouverait dans l'environnement comme la photodégradation, la biodégradation et les réactions abiotiques oxiques ou anoxiques.[8,14,18,106] Si certaines études s'arrêtent à l'identification des produits de transformation générés en laboratoire, la confirmation de leur présence environnementale est en revanche cruciale pour pouvoir déterminer le risque qu'ils posent et la validité du montage de dégradation.[145] Les composés pharmaceutiques sont un cas particulier dans

ces études comme ils entrent fréquemment dans l'environnement partiellement ou complètement métabolisés; l'étude de transformation du composé parent peut en ce sens être moins pertinente.[14] Après la confirmation de leur présence environnementale, une évaluation de la toxicité des produits identifiés permet de déterminer s'ils doivent être ajoutés à des listes de suspects ou même surveillés dans le cadre de programme de monitoring.

L'identification de produits de transformation inconnus par processus descendant, c'est-à-dire directement dans l'échantillon naturel reste particulièrement complexe et peu appliquée.

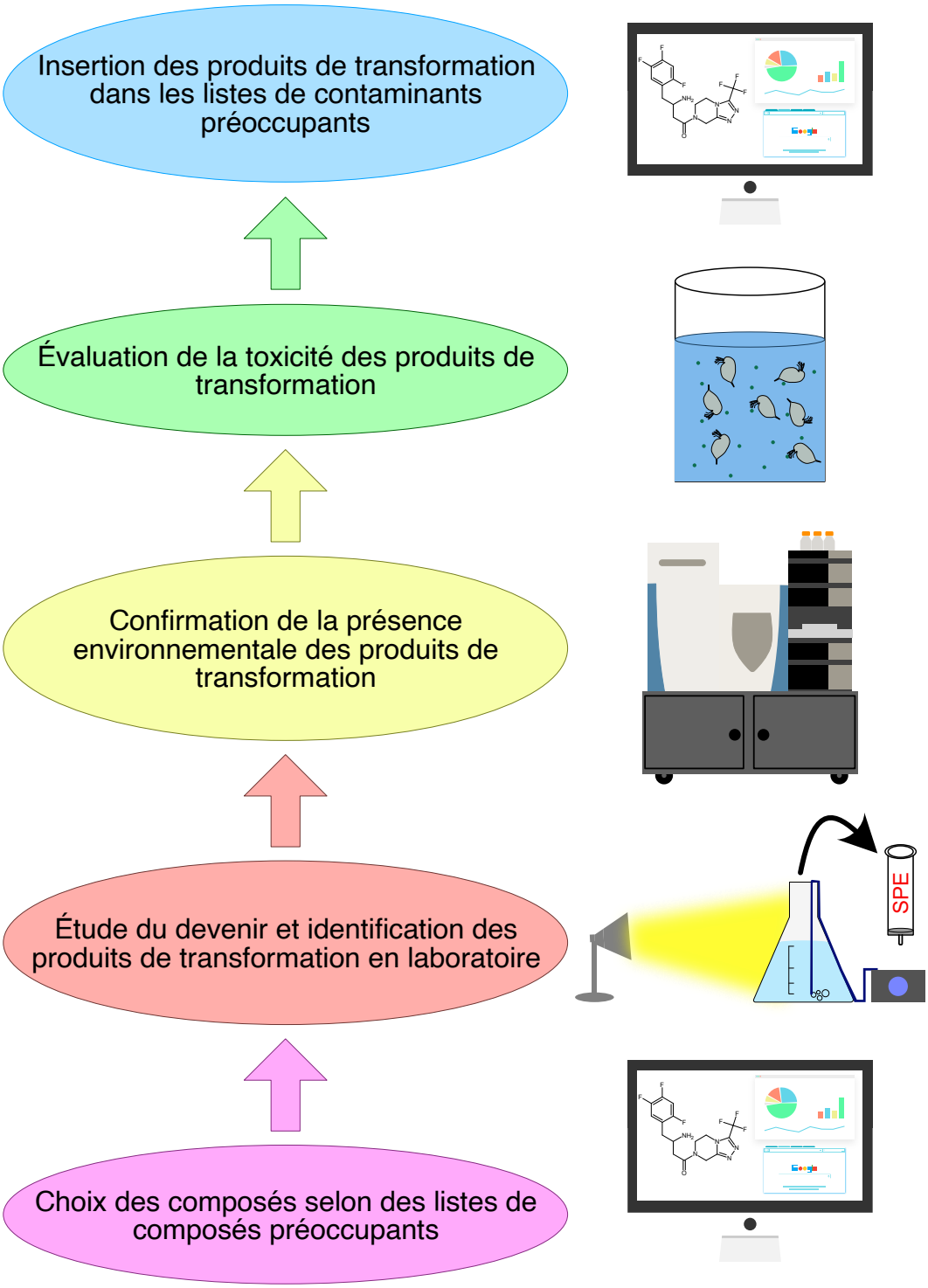


Figure 3. Processus ascendant d'étude du devenir de contaminants prioritaires et de l'identification de leurs produits de transformation

1.3 Objectifs du projet et définition des axes de recherche

Cette thèse est axée sur le développement, l'amélioration et l'application d'outils, méthodes et techniques d'identification de composés inconnus en contexte d'analyse environnementale, l'étude de produits de transformation et l'analyse d'échantillons d'eau de surface. Ces trois thèmes forment chacun un objectif de cette thèse.

Objectif 1 : Il y a des besoins dans l'augmentation du niveau de confiance pour l'identification de composés inconnus en analyse environnementale. La génération de formule moléculaire avec le moins d'ambiguïté possible à de faibles concentrations reste un problème comme la capacité de pouvoir distinguer des isomères structuraux pour limiter le nombre de structures potentielles pour un inconnu. Le dépistage non ciblé est aussi sévèrement limité par le nombre de spectres $HRMS^2$ dans les banques de spectres. Le premier objectif est de développer et optimiser des outils analytiques efficaces par spectrométrie de masse à haute résolution dans des conditions d'analyse environnementale de contaminants organiques à l'état de trace, c'est-à-dire avec des concentrations potentiellement très faibles et avec un certain effet de matrice. Ils pourront ainsi permettre d'améliorer le processus d'identification des suspects en consolidant le passage d'un niveau de confiance à l'autre. Parallèlement, l'objectif est de consolider l'identification des composés inconnus en dépistage non ciblé avec des outils computationnels de génération de spectres $HRMS^2$. Ainsi, l'identification dans les deux types d'analyse non ciblée sera améliorée.

Objectif 2 : Les informations sur le devenir et l'identité des produits de transformation de beaucoup de contaminants préoccupants sont très largement inconnus. Le deuxième objectif est d'identifier les produits de transformation générés en laboratoire de quatre contaminants organiques préoccupants suite à une expérience de dégradation en laboratoire. Les outils, techniques et méthodes pour augmenter la confiance dans l'identification développés à l'objectif précédent seront employés.

Objectif 3 : Peu d'études de dégradation de contaminants en laboratoire viennent confirmer la présence environnementale des produits de transformation. L'identification de produits de transformation inconnus directement dans les échantillons est également très complexe. Le troisième l'objectif est de réaliser un flux de travail de dépistage non-ciblé descendant pour identifier des contaminants inconnus incluant des produits de transformation et des homologues de produits de consommation.

CHAPITRE 2. MATÉRIEL ET MÉTHODE

Ce chapitre fait état des méthodes et paramètres méthodologiques utilisés dans la présente thèse. Pour éviter la répétition, des redirections vers les chapitres de résultats publiés sous forme d'article sont présentes pour plusieurs paramètres et informations sur le développement méthodologique.

2.1 Instrumentation

Les instruments d'analyse utilisés dans le cadre de cette thèse étaient un triple quadripôle Quattro Premier relié à un système de chromatographie liquide Acquity de Waters, un spectromètre de masse en temps de vol en tandem avec un quadripôle de modèle Maxis (Bruker) relié à un système de chromatographie liquide Nexera de Shimatzu et deux spectromètres de masse quadripôle-Orbitrap de modèle Q-Exactive Plus de Thermo reliés à des systèmes de chromatographie liquide Vanquish de Thermo.

2.2 Outils d'analyse combinatoire

2.2.1 Le Similar Partition Searching algorithm – SPS

La capacité d'un outil d'analyse combinatoire de recherche des composantes similaires (SPS) pour générer des spectres $HRMS^2$ computationnellement en analyse environnementale de contaminants organiques à l'état de trace a été évaluée puis une méthode employant cet outil a été développée; la méthode a ensuite été appliquée dans un dépistage non ciblé de contaminants organiques dans la rivière Yamaska. Le fonctionnement de SPS est expliqué en détail dans le CHAPITRE 5, plus spécifiquement dans la **Figure 14** et la section 5.3. Les paramètres relatifs au développement méthodologique sont également montrés et expliqués en détail dans le CHAPITRE 5. Finalement, les paramètres relatifs au contrôle qualité, à l'échantillonnage, à la purification et à l'extraction, à la séparation chromatographique, à l'analyse MS et au traitement de données sont montrés en détail dans le CHAPITRE 5.

2.2.2 MetFrag

MetFrag est un logiciel à accès libre qui génère des spectres $HRMS^2$ computationnellement pour identifier des composés inconnus.[137] Pour ce faire, MetFrag commence par assembler une liste de composés potentiels dont le *m/z* de la molécule protonnée ou des adduits choisis se situent dans une fenêtre de masse déterminée préalablement. Ces composés sont recherchés dans des grands registres de composés comme PubChem et ChemSpider ou une banque de données locale. Les composés sélectionnés selon les critères de masse sont ensuite fragmentés avec une approche basée sur la dissociation des liens moléculaires et ces fragments sont comparés aux ions produits du composé inconnu. Le score de la correspondance est basé sur la différence des *m/z* des fragments de l'inconnu par rapport à ceux des candidats, l'intensité des fragments et le nombre de fragments similaires.[146] La taille croissante de grands registres de composés comme PubChem et ChemSpider augmente le nombre de candidats potentiels pour un inconnu; cela augmente le nombre de composés qui peuvent être identifiés mais aussi le risque de faire une annotation erronée.[137,147] Considérant la taille croissante des registres de composés, une nouvelle version de MetFrag lancée en 2016 permet de consolider les identifications en ajoutant des outils comme souligner la présence de candidats potentiels sur des listes de contaminants suspects, l'utilisation de la banque de données de spectres $HRMS^2$ empirique MassBank et l'approximation des temps de rétention en fonction du coefficient de partition octanol-eau (LogP) quoique d'importants écarts sont observés avec cette méthode.

MetFrag est utilisé dans le CHAPITRE 5 et le CHAPITRE 7 pour l'identification de contaminants organiques dans le dépistage non ciblé de contaminants de la rivière Yamaska. MetFrag était une composante du processus dans les dépistages non ciblés et commandé depuis le langage R. Les paramètres relatifs à l'utilisation de MetFrag y sont fournis, plus spécifiquement à la section 9.3.1.6.

2.3 **Outils de réseaux moléculaires et d'agglomération des spectres de masse en tandem**

Deux outils d'organisation de données de spectres MS^2 ont été appliqués et évalués à des fins d'analyse non ciblée d'eaux de surface, l'écosystème GNPS (https://gnps.ucsd.edu) qui permet la formation de réseaux moléculaires et l'ensemble de fonctions (appelé *package*, un terme qui est conservé tel quel en Français) du langage de programmation R CluMSID qui permet l'agglomération de spectres MS^2.

2.3.1 Fonctionnement et paramètres de GNPS

GNPS est un outil de réseaux moléculaire qui permet de regrouper sous forme de réseaux les composés ayant des spectres MS^2 similaires tel que montré dans la **Figure 4**.

Les données sont d'abord simplifiées par la formation de spectres consensus regroupant les ions précurseurs similaires en termes de ratio masse sur charge et de temps de rétention et de spectres ensemble pour diminuer la charge de calcul.[148] Les spectres MS^2 simplifiés, appelés spectres consensus, sont ensuite utilisés pour générer l'analyse des réseaux moléculaires. Les similarités vectorielles sont calculées pour chaque paire possible de spectres avec un minimum d'ions fragmentés correspondants défini. La similarité est déterminée par le produit scalaire entre chaque paire d'ions possible; elle prend notamment en compte les intensités relatives des ions produits ainsi que les différences de *m/z* entre les ions produits dans les spectres appariés. Les scores varient donc entre 0 (totalement dissemblable) et 1 (totalement identique).[141,149] Les spectres MS^2 sont également comparés aux banques de données de spectres $HRMS^2$ de GNPS, ce qui permet d'établir en parallèle d'identifier les spectres.[139,150] Les réseaux peuvent être visualisés en ligne à même le site ou exportés vers des logiciels de visualisation tiers comme Cytoscape.[151,152]

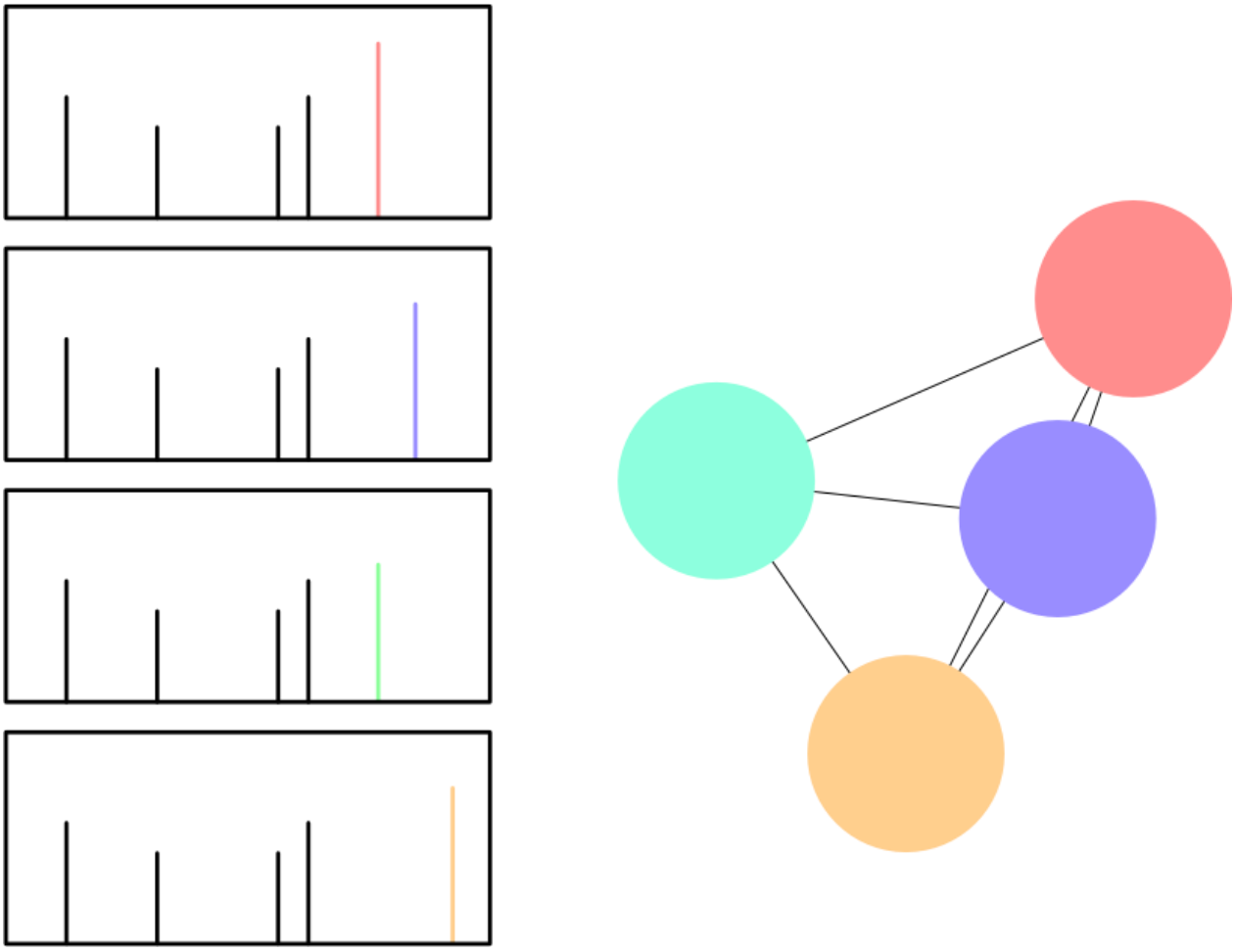


Figure 4. Exemple simplifié de réseau moléculaire de GNPS. Chaque nœud correspond à un spectre MS^2. Les ions précurseurs, représentés par des couleurs, sont distincts pour chaque spectre consensus. Ces quatre nœuds ont des spectres hautement similaires et forment un réseau moléculaire selon les critères définis.

Les détails des paramètres relatifs à la fenêtre de sélection des m/z pour la formation spectres consensus, de score minimal pour former un réseau, de nombre d'ions produits en commun pour former un lien, de liens maximums pour un nœuds sont montrés dans la section 9.3.1.5. Le mode opératoire schématisé de l'utilisation des réseaux moléculaires par GNPS est montré en section 9.5.1.

2.3.2 Fonctionnement et paramètres de CluMSID

CluMSID, pour *Clustering of MS2 spectra for metabolite identification*, est un package du langage de programmation R qui permet de regrouper les spectres $HRMS^2$ similaires pour identifier des métabolites.[140,153] Pour ce faire, une matrice de dissimilarité créée où chaque ion précurseur est comparé à tous les autres ions précurseurs selon leurs ions produits. La similarité entre deux spectres a et b est calculée en utilisant une forme dérivée de la similarité cosinus; les spectres sont considérés comme deux vecteurs à *n* dimensions et leur similarité est calculée en déterminant le cosinus de leur angle en divisant leur produit scalaire par le produit de leur norme.[153] Dans le cadre de cette thèse, CluMSID a été utilisé pour regrouper les spectres $HRMS^2$ non pas de métabolites, mais de produits de transformation et de congénères d'additifs de produits de consommation dans le CHAPITRE 7. Les paramètres spécifiques à l'utilisation de CluMSID sont affichés dans la section 9.5.1.

2.4 Logiciels

Les outils informatiques employés dans cette thèse pour la manipulation des instruments, le traitement de données, la visualisation des données, l'organisation des données et la création de graphiques pour leur publication sont affichés dans la **Table 2**. Le logiciel R était utilisé en conjonction avec de multiples packages pour le traitement, l'organisation et la visualisation des données et la conception graphique. Plus d'informations et de détails sur l'utilisation et les paramètres des outils informatiques sont montrées dans les sections 9.1.1 pour la méthode d'échange hydrogène-deutérium, 4.4.4 et 9.2 pour l'outil de conformité spectrale, 5.4.4 pour le dépistage non ciblé avec les outils combinatoires de génération de spectres, 6.4.5 pour l'identification des produits de transformation générés en laboratoire et leur dépistage dans un échantillon réel et 7.4.5 et 9.5 pour le dépistage non ciblé descendant avec les outils de regroupement des spectres.

Table 2. Liste des logiciels utilisés dans la thèse

Manipulation des instruments	Traitement des données	Organisation des données	Visualisation et conception graphique
Bruker Compass	Bruker DataAnalysis	R	Microsoft Word
Waters MassLynx	Thermo Xcalibur	Microsoft Excel	Microsoft PowerPoint
Thermo Xcalibur	Waters MassLynx	MSP Cloudberry	Affinity Designer
Bruker CompassHystar	Cerno MassWorks	Microsoft Access	Affinity Publisher
Thermo Tune	Mathspec SPS	Cytoscape	Affinity Photo
	R	R Studio	ACD Chemsketch
	MetFrag CL		R
	Proteowizard		R Studio
			Marvin Sketch

CHAPITRE 3. DÉVELOPPEMENT D'UNE MÉTHODE D'ÉCHANGE HYDROGÈNE-DEUTÉRIUM POST COLONNE EN TEMPS RÉEL DANS DES CONDITIONS ENVIRONNEMENTALES

3.1 Avant-propos

Ce chapitre a été publié dans le journal « Canadian Journal of Chemistry » sous les références Eysseric et *al*. 10.1139/cjc-2016-0281

« Post-column hydrogen-deuterium exchange technique to assist in the identification of small organic molecules by mass spectrometry »

3.1.1 Auteurs et affiliation

Emmanuel Eysseric[1], Xavier Bellerose[1], Jean-Michel Lavoie[2], Pedro A. Segura[1, *]

[1] Department of Chemistry, Université de Sherbrooke, 2500 Boulevard de l'Université, Sherbrooke, QC, Canada J1K 2R1

[2] Department of Chemical and Biotechnical Engineering, Université de Sherbrooke

[*] Corresponding author: e-mail: pa.segura@usherbrooke.ca, Tel: 1-819-7922, Fax: 1-819-821-8017

3.1.2 Présentation de l'article

L'identification de composés inconnus par spectrométrie de masse comporte un certain degré d'incertitude selon les informations à disposition. Il arrive fréquemment qu'il demeure une ambiguïté sur la structure exacte d'un inconnu comme notamment dans le cas d'isomères structuraux. Il existe peu de méthodes pour remédier à ce problème et même les informations MS^2 peuvent ne pas suffire pour une identification plus conclusive. De plus il n'est pas toujours possible d'obtenir de spectre MS^2 de qualité pour un composé donné. Il y a ainsi un besoin en identification par spectrométrie de masse pour diminuer le nombre de candidats potentiels pour une structure.

L’échange hydrogène deutérium (HDX) est une technique qui, en introduisant un échantillon dans un milieu en excès d’atomes de deutérium labiles, permet de déterminer le nombre d’hydrogènes échangeables des molécules ionisées qui le composent en observant un gain de masse. Le D_2O utilisé pour le HDX remplace l’eau dans les phases mobiles aqueuses pour la séparation chromatographique. Or, cette manière de faire est très coûteuse et requiert l’utilisation d’une importante quantité de D_2O est nécessaire pour une seule analyse.

Cette étude a pour but le développement d’une méthode de HDX reposant sur l’injection de D_2O après la séparation chromatographique par un té pour en diminuer la consommation. Cette étude a été réalisée lors de ma maîtrise, plus tard convertie en doctorat par un passage accéléré, dont l’objectif était d’identifier des métabolites secondaires du sorgho sucré.

3.1.3 Contributions des auteurs

Le plan expérimental a été conçu par Pedro A. Segura (PAS) et Emmanuel Eysseric (EE). Les analyses par spectrométrie de masse ont été réalisées par Xavier Bellerose pour le triple quadripôle et par EE pour la trappe ionique linéaire et le spectromètre de masse en temps de vol. Le développement des méthodes a été réalisé par EE. L’échantillonnage et l’extraction ont été réalisés par EE. Le traitement des données a été réalisé par EE. La visualisation des données a été réalisée par EE et PAS. La rédaction a été réalisée par EE et PAS. La révision a été effectuée par EE, PAS et Jean-Michel Lavoie. La soumission a été réalisée par PAS. Les réponses aux réviseurs ont été réalisées par PAS et EE.

3.1.4 Abstract

In order to improve the certainty that a specific small organic molecule has been detected in a given sample by high-resolution mass spectrometry, other techniques that give conclusive evidence about the chemical structure of a compound like nuclear magnetic resonance (NMR) or

complementary information on its composition such as hydrogen-deuterium exchange (HDX) are often necessary. This study presents a systematic investigation that aims to improve the applicability of post-column HDX for those purposes. Key parameters like mobile phase flow rates, volume percentage of H_2O in the mobile phase and D_2O addition flow rates were optimized in order to provide an isotopic pattern that allows the accurate determination of the number of exchangeable hydrogen atoms in small organic molecules. A loop injection setup was used to emulate chromatographic conditions in the optimization process, and trimethoprim, a widely-used anti-infective, was used as test compound for the experiments. As expected, results showed that deuteration percentage decreased with a higher mobile phase flow rate and increased with higher D_2O flow rate. The post-column HDX technique was then validated with extracts of samples of river water and plants separated by liquid chromatography in hydrophilic interaction or reversed phase modes. Mass spectra showed a completely visible isotopic pattern that allowed assessing correctly and unambiguously the number of exchangeable hydrogens in the compounds of interest. This study shows that post-column HDX can be used as complementary technique to identify unknown small organic molecules in complex matrices. The current paper proposes an efficient, cost-effective, versatile technique of HDX that is helpful to assign a unique structure to a given high-resolution mass spectrometry signal.

Keywords: hydrogen-deuterium exchange; HDX; small molecule; identification; post-column addition; organic contaminants; liquid chromatography; high-resolution mass spectrometry.

3.1.5 Résumé

Dans l'optique d'améliorer le niveau de certitude relatif à l'identité de petites molécules organiques détectées en spectrométrie de masse à haute résolution, plusieurs techniques additionnelles apportant des informations structurales supplémentaires comme la résonnance magnétique nucléaire (RMN), qui amène des informations structurelles exhaustives, ou l'échange hydrogène-deutérium (HDX), qui est une méthode complémentaire, sont souvent nécessaires. La présente étude propose une investigation systématique afin d'améliorer l'applicabilité de la HDX post-colonne dans un contexte d'élucidation structurale. Des paramètres majeurs comme le débit

des phases mobiles, la composition des phases mobiles en H_2O et le débit d'ajout de D_2O ont été optimisés afin de fournir un patron isotopique permettant une détermination précise du nombre d'hydrogènes échangeables dans une petite molécule organique donnée. Un montage avec une boucle d'injection a été utilisé afin d'imiter des conditions chromatographiques lors du procédé d'optimisation alors que le triméthoprime, un anti-infectieux très commun a été utilisé comme composé cobaye. Les résultats obtenus ont montré que le pourcentage de deutération décroissait avec un débit de phase mobile plus élevé et augmentait avec un débit d'ajout de D_2O plus élevé. La méthode a ensuite été validée en effectuant la HDX sur des extraits des échantillons de rivière et des plantes séparés par chromatographie à interactions hydrophiles (HILIC) et par chromatographie en phase inverse, respectivement. Les patrons isotopiques étaient complètement visibles dans les spectres de masses des composés analysés et fournissaient sans ambiguïté le nombre d'hydrogène échangeables dans les composés étudiés. La présente étude démontre que la HDX post-colonne peut être utilisée comme méthode complémentaire de manière rapide et économique dans un contexte d'élucidation structurale.
Mots clés : Échange hydrogène-deutérium, HDX, petites molécules, identification, dérivation post-colonne, chromatographie liquide, spectrométrie de masse, contaminants organiques.

3.1.6 Introduction

Identification of unknown organic compounds and characterization of transformation products of known organic compounds by mass spectrometry is rapidly expanding in the environmental and natural product fields because of the increasing access to tandem mass spectrometers with accurate mass and high resolution capabilities [1]. However, compound identification based on accurate mass measurements and tandem mass spectra is often difficult due to the possibility of numerous isomers. Therefore, techniques giving conclusive evidence such as nuclear magnetic resonance (NMR), or complementary techniques like chemical derivatization or hydrogen-deuterium exchange (HDX) are often used to correctly identify the compounds of interest [2]. The latter is an established technique in protein analysis [3] and can also be used as a mean to improve the level of confidence in the structural identification of small organic molecules of interest, especially when available amounts rule out NMR. HDX is based on the capacity of labile H atoms present in

organic molecules, generally bound to heteroatoms such as N, O or S, to exchange with D atoms from enriched media (usually D_2O) [4]. The result of such exchange is the increase in the *m/z* ratio by *n* mass units depending in the number *n* of labile hydrogens on the compound.
According to a recent publication [5], the identification of small molecules by mass spectrometry can be classed in five different levels of decreasing confidence: from confirmed structure (level 1) to exact mass (level 5). High resolution-mass spectrometry yields the lowest level of confidence, level 5 data. The level of confidence in the assignment of a unique structure to a given high resolution mass spectrometry signal can be improved to probable structure (level 2) by obtaining additional information on the compounds of interest and deducing the structure by eliminating other candidates. HDX can be used for such purposes.

So far, four different approaches have been used to apply HDX for characterization of small organic molecules: offline [6], in-source [7], online [8] and post-column HDX [9]. Each of these techniques has clear advantages as well as disadvantages that may limit their routine application in some laboratories. In offline HDX, purified compounds previously dissolved in deuterated solvents are directly infused into the mass spectrometer. This technique was recently used by Bourgin *et al.* [6] to characterize ozonation by-products of estrone-sulfate, a potential endocrine disruptor. While this approach is the easiest to implement, its major drawback is that compounds have to be previously purified before offline HDX, since no separation is performed before introduction of the samples into the mass spectrometer. In the in-source technique, a pump delivers D_2O to a second atmospheric pressure ionization source, creating a D_2O-rich atmosphere inside the ionization source in order to achieve HDX before the analytes are introduced in the mass spectrometer. This approach was applied by Wolff and Laures [7] to the analysis of antiulceratives and anthelmintics and can be used to determine the number of exchangeable H atoms of an analyte after visual inspection of its isotopic pattern. However, the application of this technique to the identification of unknowns is difficult since it is limited by low exchange levels, in some cases <50%, which makes difficult the determination of the correct number of labile H atoms in a compound. Additionally, a dual-spray source may not be available for many spectrometers and is usually used for internal mass calibration in high resolution mass spectrometers. High deuteration

levels are obtained in the online technique in which a deuterium-enriched mobile phase is used to induce HDX during the liquid chromatography separation process. Online HDX has been successfully applied for the identification of metabolites, pharmaceutical compounds as well as other small molecules [2,8]. As an example, online HDX was shown helpful in reducing the number of potential isomers during the characterization of sulfadiazine metabolites present in pig manure [10]. A significant reduction of deuterated solvents consumption can be achieved using online HDX with hydrophilic interaction liquid chromatography (HILIC), as the composition of the mobile phase in that technique is generally 50 to 95% organic[11]. HDX in the HILIC mode was applied recently to the separation of 4-aminomethylpyridine from its degradation products [12]. The major drawbacks of this approach are the high cost of deuterated solvents and shifts in retention time observed for deuterated analytes, in some cases, up to 21% longer compared to non-deuterated solvents [8]. The post-column HDX technique introduced by Tolonen *et al.* [9] is an interesting cost savings alternative to online HDX since D_2O is added to the chromatographic column effluent using a syringe pump before entering the ionization source. However, a detailed study on the key parameters affecting the performance and applicability of post-column HDX has not yet been done. The objective of the present work is to study the effect of mobile phase composition, mobile phase flow rate and the type of mixing device on the deuteration percentage in the post-column HDX technique. Such study will improve the application of this technique to assist on identification of unknown small organic molecules separated by liquid chromatography.

3.1.7 Experimental

3.1.7.1 Reagents

Water, acetonitrile and 0.1% formic acid in acetonitrile were all LC-MS Optima grade and were obtained from Fisher Scientific. D_2O (99.9%) was obtained from Cambridge Isotope Laboratories, while trimethoprim (purity ≥98%) was purchased from Sigma Aldrich as were caffeine (99%) and theophylline (99%).

3.1.7.2 Study of post-column HDX parameters using loop injections

In a loop injection setup, the mobile phase is used to transport the analyte of interest from an injection loop placed on a divert valve to the ionization source of the mass spectrometer. Loop injections were used to reproduce chromatographic conditions in real time and to study three key parameters that have a major impact on the post-column HDX: water volume percent in the mobile phase (0, 5, 10, 15, 20, 25, 30 % v/v), mobile phase flow rate (150, 200, 250 and 300 μL min^{-1}) as well as D_2O addition flow rate (30, 50, 70 μL min^{-1}). Composition of the mobile phase was tested to measure the impact of the presence of exchangeable H atoms in H_2O. The organic solvents used in the mobile phase had to be void of exchangeable hydrogens since otherwise they could compete with exchangeable H atoms of the target molecules, thus significantly reducing the deuteration percentage. Acetonitrile (ACN) is a widely used aprotic solvent and one of the most common organic solvents in liquid chromatography-mass spectrometry (LC-MS). Therefore it was used with or without formic acid (FA) 0.1% v/v or ammonium formate as additive in this work. Trimethoprim was selected as a test compound because of its 4 exchangeable amine hydrogens present in its molecular structure and therefore its isotopic pattern in cases of incomplete deuteration is larger and can provide more information. Furthermore, this polar compound is a frequently detected environmental contaminant [13]. In the loop injection experiments, trimethoprim was introduced in a 20 μL stainless steel loop through an injector port placed in the mass spectrometer's divert valve using a 500 μL Hamilton syringe. D_2O was added to the mobile phase with a syringe mounted on a syringe pump using a mixing tee connector. Targeted parameters and their effects were measured at least three times to evaluate signal variation. The effects of different mixing devices such as a tee connector (IDEX, part number P-727, swept volume 0.57 μL) and a HPLC mixer (Waters, Acquity BSM zirconia mixer, internal volume 50 μL) were compared to a mixing tee (IDEX, part number U-466, swept volume 2.2 μL) on the deuteration percentage of trimethoprim.

3.1.7.3 Collection and preparation of samples

3.1.7.3.1 River water samples

Water was sampled from the Magog River (Sherbrooke, Quebec, Canada) on May 14th, 2015 in amber HDPE bottles and conserved in an ice cooler until arrival to the laboratory, where it as immediately stored at -20°C. Upon extraction, samples were thawed at room temperature and extracted using a previously published method [14]. In summary, 250 mL of the water was acidified with a solution of orthophosphoric acid to a pH of 2.8 and extracted by solid-phase extraction (SPE) using Phenomenex Strata-X polymeric reversed phase cartridges (bed mass 200 mg, volume 6 mL, particle size 33 µm). Cartridges were conditioned sequentially with 5 mL of methanol and 5 mL of H_2O, and then loaded with the river water samples at a flow rate of ≈ 8 mL min^{-1} using a SPE manifold connected to a vacuum diaphragm pump. Cartridges were eluted with 2 × 3 mL of a 1:1 acetonitrile: methanol mixture. Samples were then reconstituted in 10 mL acetonitrile and spiked with 4 µg L^{-1} of the trimethoprim, caffeine, and theophylline standards. Samples were then injected analysis using a HILIC-ESI(+)-QqTOFMS method.

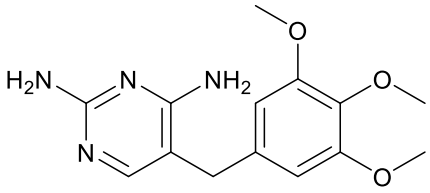


Figure 5. Molecular structure of trimethoprim showing the presence of 4 exchangeable hydrogen atoms. In ESI+, a mass shift of up to 5 mass units can be observed after HDX of trimethoprim because of the formation of a D+ adduct [M(-4H+4D)+D]+.

3.1.7.3.2 Plant samples

Sorghum bicolor, commonly referred as sweet sorghum, was first ground and sieved to particle sizes between 40 and 60 mesh (250-425 µm). It was then put in a steam explosion reactor. A soxhlet extraction was then used on the remaining biomass with toluene and ethanol as solvents in a 1:1 ratio. The solution was then dried with a nitrogen flow apparatus and reconstituted in methanol. The methanol solution was then centrifuged at 3000 *g* for 5 minutes and filtered on a 0.45 µm PVDF filter before being injected for analysis using a RPLC- ESI(+)-QqTOFMS method.

3.1.8 Instruments and methods

Post-column HDX was tested with different instruments in order to evaluate the variability between electrospray ionization (ESI) sources. In all cases, ionization was performed in positive mode. Study of the effect of the HDX parameters on the deuteration percentage was done in triple quadrupole mass spectrometer (QqQMS) and a linear ion trap mass spectrometer (LITMS). Analysis of river samples and plants extracts were done with UHPLC coupled to a quadrupole-time-of-flight mass spectrometer (QqTOFMS).

3.1.8.1.1 ESI(+)-QqQMS

The liquid chromatography-tandem mass spectrometry system used for this work composed was an Acquity Ultra Performance LC from Waters coupled to a Waters Quattro Premier XE triple quadrupole mass spectrometer. Source parameters were as follows: capillary voltage was 3.2 kV, cone voltage was 35 V, extractor voltage was 5 V, source temperature was 120 °C, desolvation temperature was 450 °C, desolvation gas flow was 700 L h^{-1}, cone gas flow was 50 L h^{-1}. The QqQMS parameters were the following: mass range was *m/z* 200 to 400; scan duration was 1 s. Data analysis was performed with Waters' MassLynx V4.1 SCN805.

3.1.8.1.2 ESI(+)-LITMS

HDX parameters were optimized using an Accela liquid chromatograph from Thermo Scientific coupled to a LTQ XL linear ion trap mass spectrometer (LITMS) also from Thermo Scientific and equipped with an electrospray ionization (ESI) source. Ionization was performed in the positive mode, capillary temperature was 275 ºC, the sheath gas flow was 35 and the auxiliary gas flow was 20. Spray voltage was 3 kV, the source current was 100 µA, capillary voltage was 25 V and the tube lens voltage was 57 V. Mass range was *m/z* 290 to 300, scanning was in the zoom mode. The data analysis software was Thermo's Xcalibur (version 2.2) for both the LC and MS systems.

3.1.8.1.3 HILIC-ESI(+)-QqTOFMS

The liquid chromatography-high resolution mass spectrometry system was a Shimadzu Nexera LC-30AD for the pump module, a SIL-30AC for the autosampler module, a CTO-30A for the column oven module and a SPD-M20A Prominence diode array detector. The mass spectrometer was Maxis quadrupole-time-of-flight (QqTOF) from Bruker. Data analysis was performed with Bruker's Compass for otof Series 1.7 patch 2 and was used in conjunction with Bruker's DataAnalysis Version 4.3 (Build 110.102.1532). Analysis of water samples was done in the HILIC mode using an XBridge Amide column (100 × 2.1 mm, 3.5 μm) from Waters. Solvent A was 10 mM ammonium formate in H_2O with 0.05% FA and solvent B was ACN. The chromatographic gradient was the following (% of A): 0 min, 5%; 4 min, 5%; 12 min, 40%; 16 min, 40%, 16.01 min, 5%; 26 min, 5%. Mobile phase flow rate was 300 μL min^{-1}, injection volume was 1 μL and column temperature was 30 °C.

Source parameters were as follows: capillary voltage was 2200 V, end plate offset voltage was 500 V, nebulizer pressure was 4 bar, dry heater was 200 ºC, dry gas flow rate was 10.0 L min^{-1}. QqTOFMS parameters were the following: scan range was from *m/z* 50 to 1200, funnel RF was at 250 Vpp, transfer time was 35 μs and the pre-pulse storage time was 5 μs. The TOF calibrant used was sodium formate.

3.1.8.1.4 RPLC-ESI(+)-QqTOFMS

The LC and MS systems and software used for the analysis of water samples were also employed for the analysis of sweet sorghum extracts. The column used for these experiments was an Acquity UPLC HSS T3 column (50mm×2.1 mm, 1.8 μm). Phase A was H_2O with 0.1% FA and phase B was ACN with 0.1% FA. Chromatographic method for tricin deuterium exchange was as follows: at initial time, 10% of B; at 2.50 minutes, 10% of B; at 20 minutes, 50% of B; at 21 minutes, 100% of B; at 27 minutes, 100% of B; at 27.10 minutes, 10% of B; at 31.10 minutes, 10% of B. Mobile phase total flow rate was maintained at 400 μL min^{-1}. Column temperature was 30°C. Method run time was 31.10 minutes. MS parameters were identical than those used for HILIC chromatography except: capillary voltage was 2000 V, ion cooler RF was 30-200 Vpp and transfer time was 30-60 μs.

A split valve setup was necessary for the HDX of the plant extracts. The water percentage in the mobile phase being of roughly 40% at 400 μL min^{-1} at the time the target compounds were eluting was too high for an effective exchange to take place. The split valve setup was placed before the mixing tee as most of the flow was sent to an UV-DAD and a fraction was sent to the QqTOFMS for the HDX. This allowed to more than double the deuteration percentage of tricin, an O-methylated flavonoid on which HDX was performed.

The deuteration percentage was determined using McCloskey's method [4] in which the contribution of naturally occurring stable isotopes is subtracted from the relative intensities of all peaks of the isotopic pattern of the compound in order to provide corrected relative intensities. Those latter values are then used to calculate molar deuteration percentages, which then are employed to determine the total deuteration percentage in the compound relative to the maximum possible value. A detailed example of this calculation for trimethoprim is shown in the Supplementary material. A theoretical spectrum can be generated with the calculated deuteration percentage using a binomial distribution where the number of possible configurations is the number of coefficients in a row in the Pascal's triangle. A correction with the naturally occurring isotope is then added. An example is given is the Supplementary material (**Figure 27**).

3.1.9 Results and Discussion

3.1.9.1 Study of the post-column HDX parameters using loop injections on the deuteration percentage

Since high deuteration percentages are necessary to correctly identify the total number of exchangeable hydrogens in a given compound, the effect of three key parameters for post-column HDX (H_2O volume percentage in the mobile phase, mobile phase flow rate and D_2O addition flow rate) was studied using loop-injections and mixing tee setup with an ESI-QqQMS system.

First, the effect of the H_2O volume percentage in the mobile phase on the deuteration percentage was measured using 0.1% v/v FA in H_2O (as solvent A) and two different solvents: ACN without additive or 0.1% v/v FA in ACN as solvent B. **Figure 6** shows that the deuteration percentage of

trimethoprim decreases from 77.7 ± 0.6 to 46.1 ± 0.1 and from 60.4 ± 0.6 to 50.4 ± 0.6 when using ACN as compared to a mixture of 0.1% FA in ACN, respectively, when the volume percentage of H_2O in the mobile phase increases from 0 to 30%. As expected, this is caused by an increasing competition between hydrogens of H_2O, FA and trimethoprim for the D atoms of D_2O. The presence of FA in ACN had a significant impact on the deuteration percentage at 0% of water in the mobile phase. This was explained as the result of the competition between FA and the target compound for available D atoms during the optimization. At higher percentages of H_2O in the mobile phase, the effect of FA is much less important since H_2O is present at increasingly higher concentrations than FA.

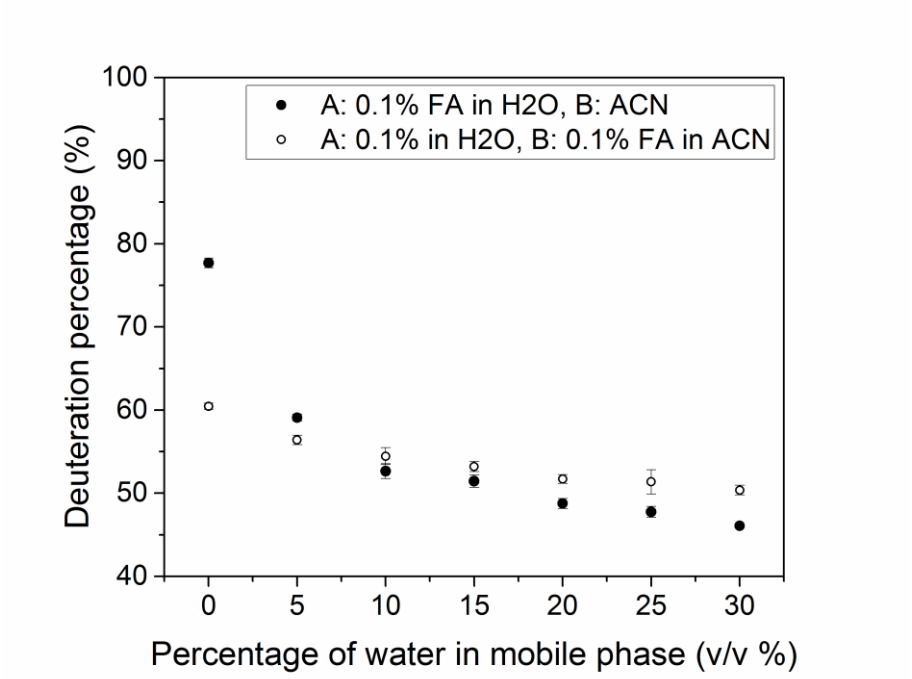


Figure 6. Effect of H_2O volume percent in the mobile phase on the deuteration percentage of trimethoprim using the post-column HDX technique in a QqQMS. Mobile phase flow rate was 300 µL min^{-1}, D2O addition flow rate was 50 µL min^{-1} and the mixing device was a mixing tee. Error bars represent ± 1 standard deviation.

As presented in **Figure 6**, HDX is never complete but remains fairly stable when using 5 to 30% v/v of H_2O. This could be explained by the weak mixing due to diffusion in the capillary tubing, which limits the dilution of the sample volume injected in the loop with mobile phase. Therefore, H_2O content in the mobile phase higher than 5% has reduced effect on the deuteration percentage since the mobile phase mainly "pushes" the sample through the capillary and has little interaction with it before the mixing tee. From the results obtained in this experiment it was observed that in a low resolution mass analyzer such as a QqQMS deuteration percentages ≥45% are enough to identify the correct number of exchangeable H atoms in trimethoprim, 4 in total. Application of the McCloskey method to the isotopic pattern of trimethoprim after post-column HDX showed that the corrected relative intensity of the peak *m/z* 296 was > 1%. That peak corresponds to [M(-

4H + 4D) + D]$^{+}$), a trimethoprim molecule in which 4 hydrogen atoms were replaced by 4 D atoms plus a deuterium adduct formed during ESI+ (**Figure 28**, Supplementary material).

Deuteration percentages ≥45% were obtained with H_2O volume percentages in the mobile phase ≤ 30%, therefore H_2O volume percentage experiments indicate that post-column HDX may not yield sufficiently high deuteration percentages to identify correctly the total amount of exchangeable hydrogen atoms in highly polar compounds separated by reversed phase liquid chromatography (RPLC). Polar organic compounds are weakly retained by the stationary phase in RPLC and are eluted at mobile phase compositions that are usually between 90 to 95% aqueous. However, the post-column technique could be used with less polar and more retained compounds in RPLC or with polar compounds separated by hydrophilic interaction liquid chromatography (HILIC), since maximum volume percentage of H_2O in that mode of chromatography is usually around 40 to 50%.

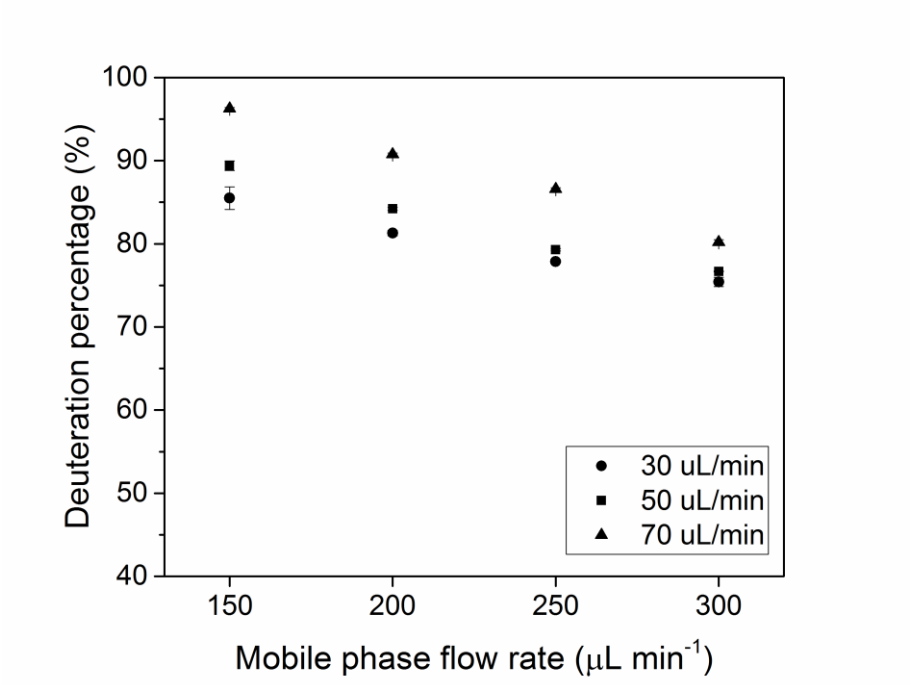


Figure 7. Effect of the mobile phase flow rate and of the D_2O addition flow rate on the deuteration rate of trimethoprim using the post-column HDX technique in a QqQMS. Mobile phase

composition was 100 % ACN and the mixing device was a mixing tee. Error bars represent ± 1 standard deviation. The three series in the insert indicate the D_2O addition flow rate.

The effects of various mobile phase flow rates and post-column D_2O addition flow rates on the deuteration percentage of trimethoprim were then tested with the same previous loop-injection setup but using 100% ACN as mobile phase. Deuteration percentage was estimated to be inversely proportional to the mobile phase flow rate since high flows should reduce the contact time between analytes and D_2O. Such behavior would reduce the possibility for the exchange reaction to take place while increasing the competing HDX reactions that cause back-exchange with water. Furthermore, deuteration percentage should increase with D_2O addition flow rate since a higher amount of D atoms are available for exchange with the target compound. These hypotheses were validated by the results illustrated in **Figure 7**. It can be seen that deuteration percentages (up to 96.2 ± 0.1) are achieved at a low mobile phase flow rate (150 μL min^{-1}) when using at a D_2O addition flow rate of 70 μL min^{-1} while the deuteration percentage decreased to 75.4 ± 0.5 at the highest mobile phase flow rate tested (300 μL min^{-1}) and at a low D_2O addition rate (30 μL min^{-1}). These results show that the present post-column HDX technique would perform better with capillary microbore columns (0.15 to 0.8 mm ID) than with narrow bore columns (1 to 2 mm ID), since the former are generally used at flow rates of 2 to 20 μL min^{-1} while the latter are used generally at flow rates of up to 300 μL min^{-1} [15]. Nonetheless, as discussed earlier, the performance of the present technique is sufficient for identification purposes in common HILIC-MS methods which usually employ narrow bore columns and flow rates between 200 and 300 μL min^{-1}.

3.1.9.2 Effect of mixing devices on the deuteration percentage

A mixing tee, an HPLC solvent mixer, and a regular tee in two different configurations (**Figure 29**, Supplementary material) were employed to study the effect of the type of post-column mixing device on the deuteration percentage, since an optimal mixing between the target compound and D_2O is necessary to attain high deuteration percentages. Results showed that the best deuteration percentage was achieved using the mixing tee (52.4 ± 0.3), while with the regular tee in 90° and

180° configurations deuteration percentages were not significantly different (49 ± 1 and 48 ± 1, respectively). The lowest deuteration percentage (39.2 ± 0.4) was obtained with the HPLC mixer, which is used to improve solvent mixing. However, when comparing the internal volume of this type of mixer (50 μL) with the internal volume of the optimal mixing device according to this experiment (the mixing tee of 2.2 μL), it becomes clear that a good blending of the target compound and D_2O is not the only factor affecting deuteration percentage. A high internal volume has a negative effect on the deuteration percentage since it increases mixing of D_2O with the mobile phase, thus decreasing the effectiveness of HDX with the targeted compound. Therefore, a mixing tee with a low internal volume and a 10 μm porosity frit to aid mixing, such as the one used in the work, is ideal to reduce mobile phase-D_2O blending and maximize deuteration percentage of target compounds.

An additional experiment was performed on a LITMS with the same loop injections setup except for the sample loop volume, which was 5 μL (**Figure 8**). As it can be observed, deuteration percentages are significantly lower (≈ 20 %) than those observed with the QqQMS, which could be explained by the differences in source design, a known factor affecting HDX [4]. Collisions between deuterated species and ion source surface containing layers of adsorbed water and organic non-deuterated material can produce back-exchange, *i.e.* exchange of deuterium incorporated in the analyte of interest with protons from other compounds [4]. Source design also affects back-exchange caused by inadequate desolvation conditions in the ion source [16]. Nonetheless, LITMS results confirmed that a systematic improvement in deuteration percentages was obtained with a mixing tee compared to regular tee connector.

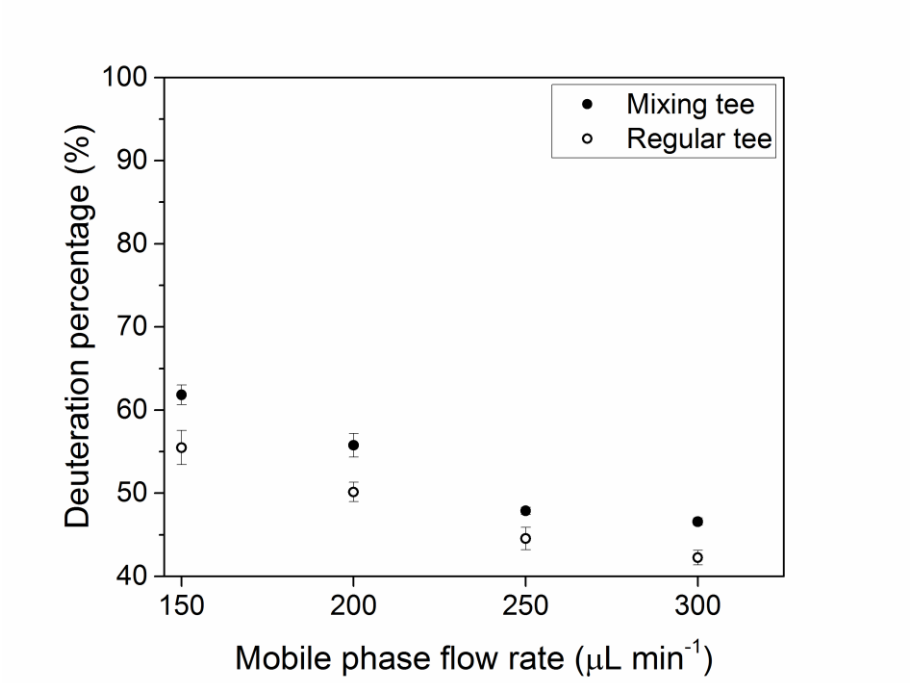


Figure 8. Effect of mixing device type on the deuteration percentage of trimethoprim using the post-column HDX technique in a LITMS. Mobile phase composition was 0.1% FA v/v in ACN and D2O addition flow rate was 70 μL min^{-1}. Error bars represent ± 1 standard deviation.

3.1.9.3 Post-column HDX with river water and plants extracts

Post-column HDX experiments were performed in a QqTOFMS with two types of samples: SPE extracts of water samples collected in a local river and spiked with the target compounds and extracts of sweet sorghum obtained after steam explosion and soxhlet extraction. For river water samples separated by HILIC, deuteration percentages after post-column HDX of the $[M+H]^+$ ions was 51 ± 1 for trimethoprim (**Figure 9**) , 56 ± 1 for theophylline (**Figure 30**, Supplementary material) and 81 ± 4 for caffeine (**Figure 31**, Supplementary material). These results allowed unambiguous identification of the number of exchangeable hydrogens in each molecule in the presence of a complex matrix.

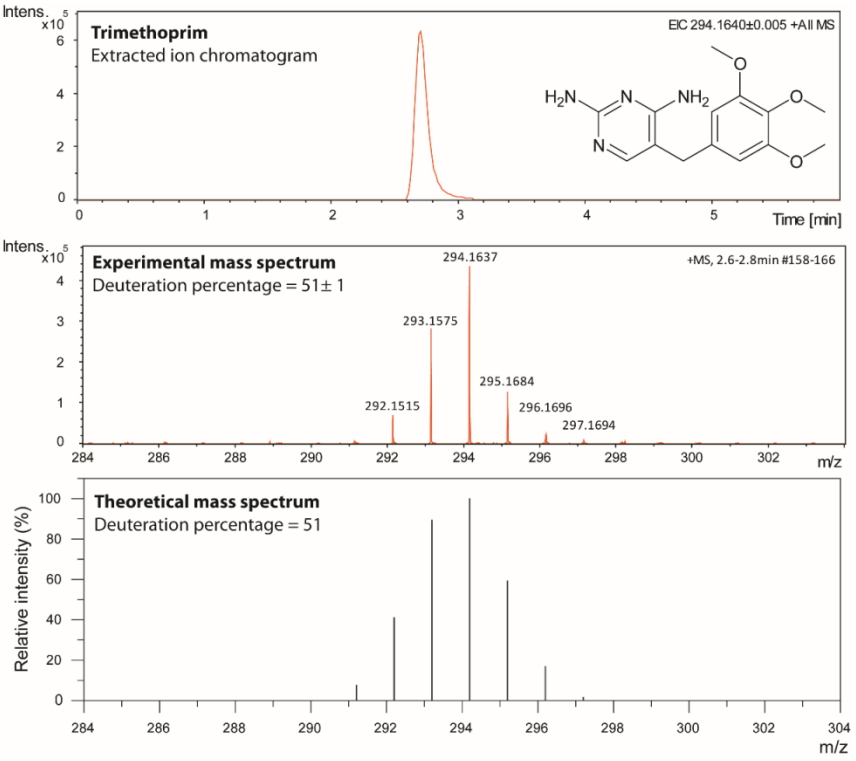


Figure 9. Extracted ion chromatogram and experimental and theoretical mass spectra of trimethoprim after post-column HDX of a SPE extract of spiked river water.

For plant extracts separated by RPLC, post-column HDX was used to further assess the identity of *m/z* 331.0810 (neutral formula $C_{17}H_{14}O_7$, Δm=0.2 mmu) a compound that was suspected to be the O-methylated flavone tricin or the mycotoxin aflatoxin G_2 (**Figure 10**). A split valve was used to reduce the column flow rate before post-column HDX which helped obtain an acceptable level of deuterium exchange (40 %).

The mass spectra confirmed that the compound in question had 3 exchangeable hydrogens. This information, along with tandem mass spectrometry spectra, allowed increasing the confidence level in the identification of tricin from exact mass (level 5) to probable structure (level 2). Ion *m/z* 335 in the experimental spectrum indicates the presence of 4 exchangeable hydrogens; therefore aflatoxin G2 was rejected as a possible structure. In this case, the experimental and theoretical spectra were very similar as can be seen in **Figure 32** (Supplementary material).

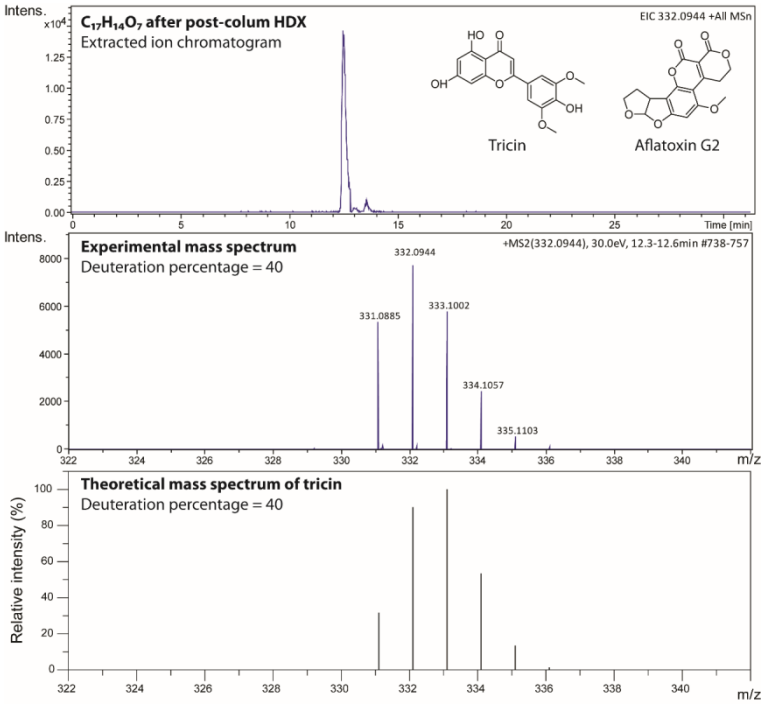


Figure 10. Extracted ion chromatogram and experimental and theoretical mass spectra after post-column HDX of a compound of formula $C_{17}H_{14}O_7$ observed in a sweet sorghum extract. The split ratio was about 1:10 and the D_2O addition flow rate was 30 µL min^{-1}. The experimental mass spectra for compound $C_{17}H_{14}O_7$ was obtained in the MS^2 mode.

3.1.10 Conclusion

The effects of mobile phase composition, mobile phase flow rate and the type of mixing device on the deuteration percentage achieved by post-column HDX were studied in order to use this technique to identify accurately the number of exchangeable hydrogens in small organic molecules. Results showed that the studied parameters can be controlled to obtain acceptable deuteration percentages that allow the accurate identification of the number of exchangeable hydrogens in small organic molecules. In the proposed post-column HDX technique, only about 100 µL of D_2O were used per injection. Compared to the online HDX technique, this represents significant savings in operating costs. The incomplete exchange in post-column HDX is a minor inconvenience as the contribution of naturally occurring stable isotopes, ^{13}C and ^{15}N can be

corrected using the method reported by McCloskey *et al*[4]. The proposed technique can be helpful to improve the level of confidence in the assignment of a unique structure to a given compound present in complex matrix and detected by high-resolution mass spectrometry.

3.1.11 Supplementary material

Supplementary material is available with the article through the journal Web site.

3.1.12 Acknowledgment

Financial support was provided by Mitacs, CÉROM Centre de Recherche sur le Grain, the Faculty of Sciences of Université de Sherbrooke and Natural Sciences and Engineering Research Council of Canada (NSERC). We would also like to thank René Gagnon and Philippe Venne for their technical assistance and Thermo Scientific for their help with the LITMS.

3.1.13 References

(1) Segura, P. A.; Gagnon, C.; Yargeau, V. *Int. J. Env. Poll. Remed.* **2012**, *1*, 85.

(2) Liu, Z. F.; Chan, K. K.; Wang, J. J. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 2581.

(3) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158.

(4) McCloskey, J. A. *Methods Enzymol.* **1990**, *193*, 329.

(5) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 2097.

(6) Bourgin, M.; Gervais, G.; Bichon, E.; Antignac, J.-P.; Monteau, F.; Leroy, G.; Barritaud, L.; Chachignon, M.; Ingrand, V.; Roche, P. *Water Res.* **2013**, *47*, 3791.

(7) Wolff, J. C.; Laures, A. M. F. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3769.

(8) Olsen, M. A.; Cummings, P. G.; Kennedy-Gabb, S.; Wagner, B. M.; Nicol, G. R.; Munson, B. *Anal. Chem.* **2000**, *72*, 5070.

(9) Tolonen, A.; Turpeinen, M.; Uusitalo, J.; Pelkonen, O. *European Journal of Pharmaceutical Sciences* **2005**, *25*, 155.

(10) Pfeifer, T.; Tuerk, J.; Fuchs, R. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1687.

(11) Nguyen, H. P.; Schug, K. A. *J. Sep. Sci.* **2008**, *31*, 1465.

(12) Li, D.; Yu, T.; Zhang, Y.; Yang, M.; Li, Z.; Liu, M.; Qi, R. *Appl. Environ. Microbiol.* **2010**, *76*, 3444.

(13) Segura, P. A.; Takada, H.; Correa, J. A.; El Saadi, K.; Koike, T.; Onwona-Agyeman, S.; Ofosu-Anim, J.; Sabi, E. B.; Wasonga, O. V.; Mghalu, J. M.; dos Santos Junior, A. M.; Newman, B.; Weerts, S.; Yargeau, V. *Environ. Int.* **2015**, *80*, 89.

(14) Segura, P. A.; MacLeod, S. L.; Lemoine, P.; Sauvé, S.; Gagnon, C. *Chemosphere* **2011**, *84*, 1085.

(15) Abian, J.; Oosterkamp, A.; Gelpi, E. *J. Mass Spectrom.* **1999**, *34*, 244.

(16) Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1993**, *115*, 6317.

CHAPITRE 4. DÉVELOPPEMENT D'UNE MÉTHODE DE DÉTERMINATION DE FORMULES MOLÉCULAIRES PAR CONFORMITÉ SPECTRALE

4.1 Avant-propos

Ce chapitre a été publié dans le journal « Analytical Chemistry » sous les références Eysseric et *al*. 10.1021/acs.analchem.7b01761

« Application of spectral accuracy to improve the identification of organic compounds in environmental analysis »

4.1.1 Auteurs et affiliation

Emmanuel Eysseric[1], Killian Barry[1], Francis Beaudry[2], Magali Houde[3], 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, QC J1K 2R1

[2] Groupe de Recherche en Pharmacologie Animale du Québec (GREPAQ), Department of Veterinary Biomedicine, Université de Montréal, Saint-Hyacinthe, QC J2S 2M2

[3] Environment and Climate Change Canada, Montreal, QC H2Y 2E7

4.1.2 Présentation de l'article

Bien que la détermination d'une formule moléculaire univoque soit la première étape dans l'identification d'un composé inconnu et le premier niveau de confiance selon le modèle de Schymanski,[38] cela peut se montrer particulièrement difficile avec les outils de génération de formule disponibles. Les algorithmes utilisent la masse précise calibrée de l'ion inconnu dans un intervalle de tolérance de masse et dans certains cas son patron isotopique. Or, pour les algorithmes qui utilisent seulement la masse de l'ion, obtenir une formule moléculaire sans ambiguïté se complique considérablement pour les composés ayant des masses supérieures à 300 Da. L'utilisation du patron isotopique en conjonction avec la masse permet de bien meilleurs résultats

mais, lorsque plus d'éléments potentiels sont considérés dans la formule chimique en particulier les éléments monoisotopiques comme le Fluor, le Phosphore et l'Iode, un très haut niveau d'incertitude dans la formule moléculaire demeure.

Les algorithmes qui utilisent le patron isotopique pour déterminer les formules procèdent à une centroïdation des pics MS, c'est-à-dire la conversion d'un pic continu, ou profil, en un bâton à l'apex. Si la centroïdation permet un calcul plus rapide de la formule, il en découle une perte d'information qui vient diminuer la qualité du résultat. La conformité spectrale est un outil qui utilise les informations isotopiques en mode profil pour la génération de formules. Cet article explore son potentiel et les principaux paramètres affectant sa performance dans des conditions environnementales avec une matrice d'eau de surface reconstituée et à des concentrations à l'état de trace.

4.1.3 Contributions des auteurs

Le plan expérimental a été conçu par Emmanuel Eysseric (EE), Pedro A. Segura (PAS) et Killian Barry (KB). Les analyses par spectrométrie de masse ont été réalisées par KB sur le triple quadripôle et par EE sur le spectromètre de masse en temps de vol (QTOF) et le spectromètre de masse quadripôle-orbitrap appartenant à Francis Beaudry (FB). Le développement des méthodes a été réalisé par EE et KB. L'échantillonnage et l'extraction ont été réalisés par EE et KB. Le traitement des données a été réalisé par EE pour les données de QTOF et de quadripôle-orbitrap et par KB pour les données de triple-quadripôle. La visualisation des données a été réalisée par EE, KB et PAS. La rédaction a été réalisée par EE et PAS. La révision a été effectuée par EE, KB, PAS, Magali Houde et Christian Gagnon. La soumission a été réalisée par PAS. Les réponses aux réviseurs ont été réalisées par PAS et EE.

4.2 Abstract

Correct identification of a chemical substance in environmental samples based only on accurate mass measurements can be difficult especially for molecules > 300 Da. Here is presented the application of spectral accuracy, a tool for the comparison of isotope patterns toward molecular formula generation, as a complementary technique to assist in the identification process of organic micropollutants and their transformation products in surface water. A set of nine common contaminants (five antibiotics, an herbicide, a beta-blocker, an antidepressant and an antineoplastic) frequently found in surface water were spiked in methanol and surface water extracts at two different concentrations (80 and 300 μg L^{-1}). They were then injected into three different mass analyzers (triple quadrupole, quadrupole-time-of-fight and quadrupole-orbitrap) to study the impact of matrix composition, analyte concentration and mass resolution on the correct identification of molecular formulas using spectral accuracy. High spectral accuracy and ranking of the correct molecular formula were in many cases compound-specific due principally to conditions affecting signal intensity such as matrix effects and concentration. However, in general, results showed that higher concentrations and higher resolutions favoured ranking the correct formula in the top 10. Using spectral accuracy and mass accuracy it was possible to reduce the number of possible molecular formulas for organic compounds of relative high molecular mass (e.g. between 400 and 900 Da) to less than 10 and in some cases, it was possible to unambiguously assign one specific molecular formula to an experimental isotopic pattern. This study confirmed that spectral accuracy can be used as a complementary diagnostic technique to improve confidence levels for the identification of organic contaminants under environmental conditions.

Keywords: mass spectrometry, river water, identification, pharmaceuticals, pesticides, metabolites, transformation products, mass resolution.

4.3 Introduction

The identification of organic micropollutants such as pesticides, pharmaceuticals, personal care products, plastic additives and their metabolites is a real challenge as a large number and variety of compounds are present the environment [1]. One of the first steps to identify a compound using mass spectrometry (MS), is to determine the molecular formula from its mass spectrum. Recently, Schymanski*, et al.* [2] proposed to improve the communication of identification confidence using MS based on a five-level approach. According to the authors, accurate mass represents the lowest confidence (level 5), followed by unequivocal molecular formula (level 4), tentative candidate (level 3, based for example on tandem mass spectrometry or other experimental data), probable structure (level 2, which could be reached using library spectrum match or other diagnostic evidence) and finally confirmed structure (level 1, which requires a reference standard). Determination of an unequivocal chemical formula (level 4) with mass accuracy < 5 ppm for a unknown peak in the mass spectrum is challenging for compounds with molecular masses > 300 Da [3] and can often lead to incorrect conclusions. Using tandem mass spectrometry (MS/MS) databases such as mzCloud or Mass Bank may help to reach confidence level 2 by searching library spectrum matches, however it is difficult to perform such confirmation for less known organic micropollutants or their transformation products that might be absent from those databases. For that reason, complementary techniques have been recently developed to obtain more information on the composition and structure of unknowns such as post-column hydrogen-deuterium exchange [4] and comparison of data acquired in the positive and negative ionization mode[5,6]. The former could be used as diagnostic evidence in order to obtain more information about the presence of specific functional groups (e.g. exchangeable hydrogens present in alcohol and amine groups) on a molecule while the latter can be employed to confirm the presence of certain compounds. Dual ionization has been used for metabolomic profiling in the past to broaden the range of detection of MS methods. Additionally, for some environmental contaminants such as pharmaceutical, herbicides and fluorinated compounds, electrospray in the negative mode is the preferred ionization polarity [7]. Obtaining additional structural information from the mass spectrum for a precursor ion before carrying on MS/MS experiments is of interest since it could save time and resources. It could also help in the identification process of unknown organic micropollutants

and/or their transformation products, often present in samples at low concentrations, since it is difficult or even impossible to obtain meaningful MS/MS spectra for peaks of low abundance.

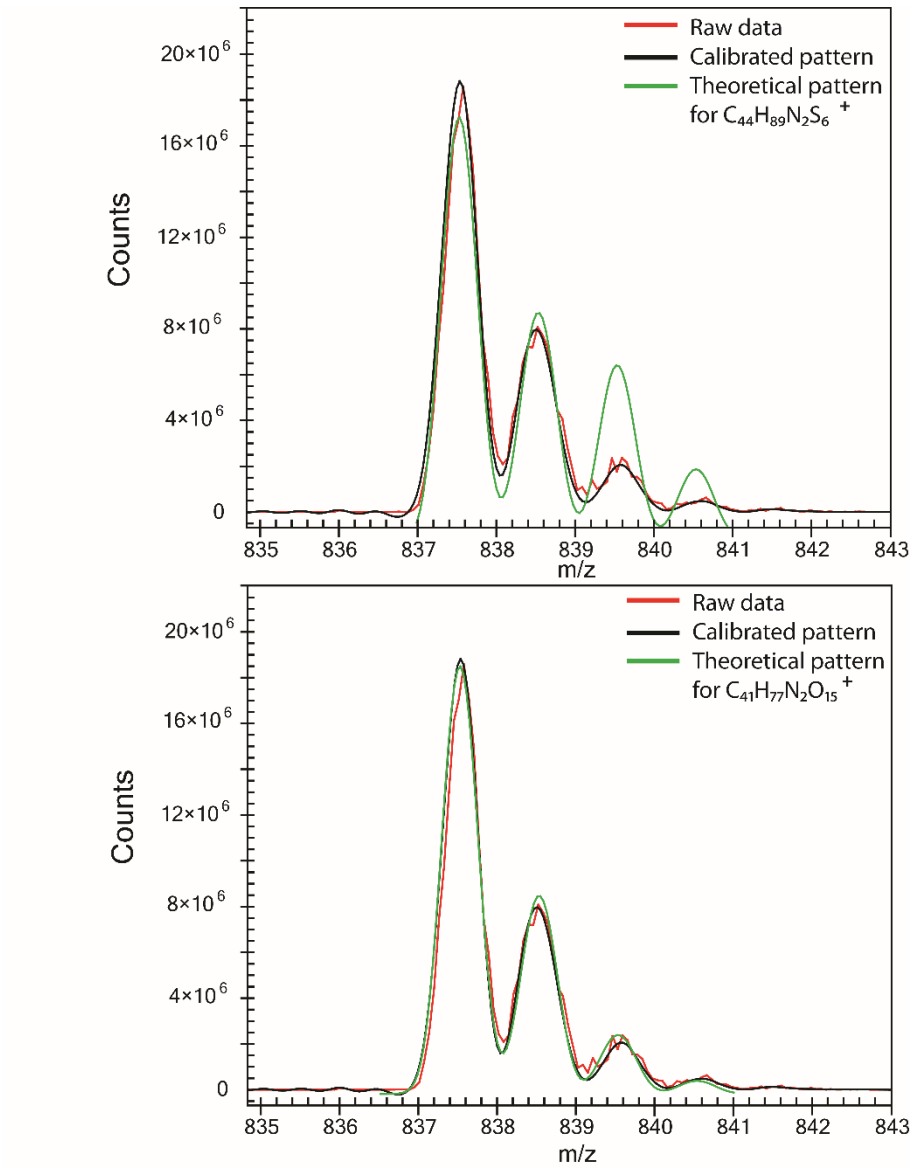


Figure 11. Determination of spectral accuracy in the MassWorks software using a known compound spiked in MeOH and analyzed by LC-ESI(+)-QqQMS. The raw data is first calibrated, i.e. mathematically transformed, and then compared to the theoretical isotopic pattern of molecular formula candidates. The top figure shows that the isotopic pattern corresponding to an ion of formula $C_{44}H_{89}N_2S_6^+$ (Δm=-3.4 mDa), is not a good match for the calibrated isotopic pattern, therefore the spectral accuracy is 75.00% and was ranked 71st out of 71 possible formulas. The

bottom figure shows a better match between the calibrated and the theoretical isotopic pattern of ion $C_{41}H_{77}N_2O_{15}^+$ (Δm=-0.7 mDa) and the spectral accuracy (96.63%), was ranked 6th out of 71 for molecular formulas within the software constraints ($C_{1\text{-}66}$, $H_{0\text{-}109}$, $N_{0\text{-}20}$, $O_{0\text{-}25}$, $S_{0\text{-}13}$, mass tolerance=5 mDa, electron state: even, double bond equivalents range: -0.5 to 50). That ion corresponds to the protonated molecule of roxithromycin, the compound spiked in the sample.

Previous studies have explored different approaches to improve the identification of small organic compounds using MS. The "Seven Golden Rules" established by Kind and Fiehn [8] provide a way to limit the sheer number of possible formulas and are now widely used in the identification process of compounds. In these rules, isotope pattern is the major component for the formula determination, along with other rules such as hydrogen to carbon ratios and probable elements. Recent studies based exclusively on accurate mass to assess molecular formula are rare because MS/MS is normally used to provide additional structural information; however alternatives have been explored. García-Reyes*, et al.* [9] previously developed a workflow for detecting and identifying pesticides and their degradation products using liquid chromatography-time-of-flight MS. The proposed method is efficient for compounds containing S, Cl or Br, which are a common occurrence in pesticides, because of their very easily distinguishable isotopic patterns. However, if an unknown compound does not contain such elements, the number of possible molecular formulas for a given accurate mass within an acceptable tolerance window cannot be reduced significantly, which impairs the compound identification process. For example, 187 possible molecular formulas, within 5 mDa of mass error and having between 0 to 50 atoms of C, H, N, O, P and F with up to 20 double bond equivalents, were found for an hypothetical ion of *m/z* 400.1234. Another approach proposed by Little et al. [10,11], used accurate mass to perform searches on databases such as the Chemical Abstracts Service or ChemSpider to identify unknown compounds. The authors applied orthogonal filters based on the number of literature references to prioritize the list of potential candidates. Though this approach can be highly successful to identify compounds that are commercially available, it can be less suitable for the identification of transformation products of environmental contaminants that may not be integrated in databases.

An interesting alternative technique to improve the confidence on the identification level of a given unknown is spectral accuracy for MS as introduced by Wang and Gu [12]. Spectral accuracy is a metric that describes the similarity between a calibrated experimental profile data of an ion, obtained through a mathematical transformation of its experimental profile data and the theoretical (calculated) isotopic pattern corresponding to a given molecular formula. Thus, high spectral accuracy (e.g. > 98%) indicates that the experimental isotopic pattern fits well to the isotopic pattern of a specific molecular formulae within 2% spectral error [12] (**Figure 11**). The main advantage of spectral accuracy over mass accuracy is that in the latter error is measured at a single point, while in the former error is measured throughout the whole isotopic distribution. Therefore, spectral accuracy uses all the information embedded in the experimental mass spectrum to rank possible molecular formulas accordingly to their level of likeness to the theoretical mass spectrum. Spectral accuracy was successfully applied to the identification of high mass (639 to 1664 Da) natural products in a previous study using a linear ion trap-orbitrap mass spectrometer [13]. Based on both mass and spectral accuracy, it was possible in some cases to eliminate >99% of formula candidates and the correct formula was usually ranked among the top five candidates. However, data were obtained using millimolar concentrations in pure solvents, experimental conditions that do not correspond to environmental analysis. Moreover, previous works on spectral accuracy showed that ion abundance was a major factor impacting quality of results [14].

The main objective of the present work was therefore to determine if spectral accuracy could be used as a complementary technique for the identification of organic contaminants in environmental sample analysis. To answer this question, a set of frequently detected organic contaminants were spiked in methanol and surface water extracts at two different concentrations (80 and 300 $\mu g\ L^{-1}$). Three different types of mass analyzers (i.e., triple quadrupole, quadrupole-time-of-flight and quadrupole-orbital trap) were used to study the impacts of three important factors in environmental analysis (matrix composition, analyte concentration and mass resolution) on spectral accuracy and formula ranking of the selected organic micropollutants (**Figure 33**, Supporting Information).

4.4 Experimental

4.4.1 Reagents and standards

Water, acetonitrile (ACN), methanol (MeOH) and 0.1% formic acid (FA) in ACN were all LC-MS Optima grade and were obtained from Fisher Scientific (Canada). Analytical standards of atrazine (acronym: ATZ, purity: 98.1%), fluoxetine hydrochloride (FLX, 99.95%), josamycin (JOS, ≥ 98%), metoprolol tartrate (MTP, ≥ 98%), ofloxacin (OFL, 99.8%) roxithromycin (ROX, ≥90%) and sulfamethoxazole (SMX, 99.9%) were purchased from Sigma Aldrich (Canada). Trimethoprim (TRI, > 98%) and methotrexate (MTX, ≥98%) were purchased from Santa Cruz Biotechnology (USA).

4.4.2 Collection and preparation of samples

Mixture solutions of the aforementioned compounds were prepared in MeOH each at high (300 μg L^{-1}) and low concentration (80 μg L^{-1}). River surface water samples (250 mL) were collected from the Magog River (Sherbrooke, Quebec) on September 13th, 2016 in amber high density polyethylene bottles and conserved in an ice cooler until arrival to the laboratory, where they were immediately stored at -20 °C. Upon extraction, samples were thawed at room temperature and buffered to pH 7 with phosphoric acid monobasic and phosphoric acid dibasic. Water samples were then extracted using a previously reported method [4]. Briefly, samples were loaded in polymeric Strata-X solid-phase extraction cartridges (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 10 mL with MeOH spiked at the same concentrations as the previous solutions (80 and 300 μg L^{-1}). Therefore, considering a preconcentration factor of 25 (initial volume=250 mL, final volume=10 mL), the high and low concentrations used for each compound are equivalent to 3 and 12 μg L^{-1} respectively, which are in the high range of environmental concentrations of many organic micropollutants detected in surface waters around the globe[15,16]. For all experiments, samples were injected three times in order to evaluate reproducibility of the results.

Another set of analysis at lower concentrations (10, 20, 30, 40 and 50 μg L^{-1}) was added in order to assess the lowe concentration limits of spectral accuracy on the QqTOFMS and the Qq-

OrbitrapMS. Considering the preconcentration factor those concentrations were equivalent to 0.38, 0.75, 1.1, 1.5 and 1.9 μg L^{-1}. For these experiments, the same sample was injected two times.

4.4.3 Instruments and methods

4.4.3.1 Liquid chromatography-triple quadrupole mass spectrometry (LC-QqQMS)

The liquid chromatography-triple quadrupole mass spectrometry instrument (LC-QqQMS) used for this work involved an Acquity Ultra Performance LC coupled to a Quattro Premier XE triple quadrupole mass spectrometer both manufactured by Waters (USA). The LC column was a Waters Acquity UPLC HSS T3 2.1 × 50 mm, 1.8 μm. Two different mobile phases were used: mobile phase A was composed of water with 0.1% (v/v) formic acid while mobile phase B was acetonitrile with 0.1% (v/v) formic acid. Chromatographic separation was obtained using the following elution gradient: at initial time, 5% B; at 5 min, 100% B; at 7 min, 100% B; at 7.01 min, 5% B; at 10 min, 5% B. Run time was 10 min. Mobile phase flow rate was 500 μL min^{-1} throughout the run. Injection volume was 10 μL. The QqQMS system was first mass calibrated via MassLynx with a solution of sodium formate before the acquisition. Additional mass calibration was performed through MassWorks by infusing an external calibrant before the acquisition (sodium formate) or by LC-QqQMS analysis of internal calibrants (MTP, FLX, OFL and JOS) mixed with the samples. A detailed description of experiments carried out to evaluate the mass accuracy stability of the QqQMS system is presented in section SI-1 of the Supporting Information.

4.4.3.2 Liquid chromatography-quadrupole-time-of-flight mass spectrometry (LC-QqTOF-MS)

The instrument used in these experiments involved a LC system manufactured by Shimadzu (Japan) and composed of a Nexera LC-30AD pump module, a SIL-30AC autosampler and a CTO-30A as column oven module. This LC system was coupled to a high-resolution mass spectrometer, the Maxis quadrupole-time-of-fight mass spectrometer (QqTOFMS) made by Bruker (USA). LC conditions were identical as those used in the LC-QqQMS setup. Injection volume was 1 μL. The QqTOFMS was calibrated with a sodium formate solution in HPC mode after waiting 30 min for

the system to stabilize. The mass drift was monitored and all analyses were done within 4 hours of the calibration. No lock mass solution was used. In these conditions, full width at half-maximum mass resolution (R_{FWHM}) at *m/z* 455 was about 25,000.

4.4.3.3 Liquid chromatography-quadrupole-orbitrap mass spectrometry (LC-QqOrbitrap MS)

The instrument used in these experiments was composed of a Dionex Ultimate 3000 Rapid Separation System ultra-high performance liquid chromatograph coupled to a Q-Exactive hybrid quadrupole-orbital ion trap mass spectrometer (QqOrbitrapMS) both manufactured by Thermo (USA). LC conditions were the same as for the LC-QqQMS setup. Injection volume was 2 µL. As for the QqOrbitrapMS parameters, the ion source was ESI in positive mode, spray voltage was at 4.2 kV, capillary temperature was at 300 °C, sheath gas was 50 arbitrary units and auxiliary gas was 25 arbitrary units. R_{FWHM} was set to 70,000 or 140,000 at *m/z* 400. Mass range was *m/z* 200-1000. The instrument was calibrated using the calibrant solution recommended by the manufacturer, a solution containing *n*-butylamine, caffeine, the tripeptide MRFA and Ultramark 1621, a mixture of fluorinated phosphazines.

4.4.4 Software parameters

Raw profile data of the mass spectrum of the test analytes spiked in MeOH and river water samples were extracted from acquisition files and processed in MassWorks (Cerno Bioscience, USA). This software uses two algorithms for the determination of spectral accuracy and formula ranking: calibrated line shape isotope profile search (CLIPS) and self-calibrated line shape isotope profile search (sCLIPS). The former is used for low resolution data and the latter for high-resolution data. Parameters for both algorithms are shown on **Table 3**. As the analyses were performed with electrospray ionization in the positive ionisation mode, the observed charges were mainly +1 ($[M+H]^+$). Doubly charged species were only observed for ROX, albeit it was not the major ion. Mass tolerance was highly dependent on the type of calibration used for the CLIPS algorithm. For the internal calibration with CLIPS, it was observed that a lower mass tolerance could be used than for external calibration. Allowed elements in the formulas generated were those commonly found

in organic contaminants such as pharmaceuticals, pesticides and plastic additives. The number range for each element was set automatically according to the Seven Golden Rules [8]. In some cases, certain parameters had to be individually tuned. Such changes included a higher mass tolerance window if the mass calibration was off. Cl, Br and Si atoms were also withdrawn to reduce computing time when the spectra visually did not suggest the presence of such atoms. Finally, in a few cases, chemically aberrant formulas, i.e. containing an unrealistic number of H, F or P atoms, were eliminated from the list given by MassWorks. Determination of the error of spectral accuracy is detailed in the Supporting Information (Section SI-2).

Table 3. Parameters for MassWorks algorithms

Algorithm	Charge	Mass tolerance	Electron state	Double bond equivalents (DBE) range	Elements allowed
CLIPS	1	1-15 ppm	Even	-0.5 to 30	C, H, N, O, F, P, Si, S, Cl, Br, I
sCLIPS	1	4-6 ppm	Even	-0.5 to 30	C, H, N, O, F, P, Si, S, Cl, Br, I

4.5 Results and discussion

4.5.1 Impact of matrix on spectral accuracy and formula ranking on QqTOFMS data

Mass accuracy for the 80 and 300 μg L^{-1} samples analyzed in the QqTOFMS can be seen in **Table 4**. As expected, values for the QqTOFMS were generally lower or equal than 2 ppm. It was hypothesized that environmental matrices such as river water would have no impact on formula ranking and spectral accuracy for the QqTOFMS since the risk of peak overlaps between analytes and matrix interferences would be much reduced in high resolution mass analyzers. However, results showed that the river water extract (matrix) did have an impact, albeit minor, on rankings (**Table 5**) and spectral accuracies (**Table 6**) measured with the QqTOFMS especially for the compounds with molecular mass > 350 Da. For example, with samples spiked at a concentration of 80 μg L^{-1}, JOS (neutral nominal mass = 827 Da) was ranked 25 ± 24 in the MeOH solution and 34 ± 33 in the matrix. The impact of the matrix was however less pronounced on OFL (361 Da)

and FLX (309 Da) which were always among the top 5 possible formulas in both MeOH and the matrix. Interestingly, OFL had a slightly worse ranking in the pure solvent than in the river extract i.e., 3 ± 2 in MeOH and ranked 2nd or 1st in the matrix. This result was interpreted as a consequence of the presence of matrix effects causing signal enhancement that will be explained later.

Spectral accuracies were generally not affected by the matrix and differences between MeOH and river extracts differed only by approximatively 3 percentage points. The slight decrease in spectral accuracy for SMX could be explained by the presence of matrix interferences near the peaks of the isotope pattern which were included in the spectral accuracy calculation. These interferences can be corrected with the interference rejection parameters which are discussed in the Supporting Information (9.2.3). For OFL spiked at 80 μg L^{-1}, the same trend in the ranking was observed in the spectral accuracy: higher spectral accuracy in the matrix extract (96.3 ± 0.1 %) than in MeOH solution (93.2 ± 0.1 %). Such improvement in ranking and spectral accuracy for OFL could be explained by signal enhancement caused by co-eluting sample components, which has been reported previously [17,18] and resulted in an improved signal-to-noise ratio. The peak area for OFL was indeed about 3 to 20 times higher in the matrix compared to the MeOH solution (**Table 17**, Supporting Information). Matrix effects were quantified by calculating the ratio of peak areas for a test compound in the matrix and MeOH. Thus, for the results presented in **Table 17**, area ratios > 1 indicate signal enhancement and area ratios < 1 signal suppression. Signal enhancement can occur because matrix components that co-elute with target compounds can improve access to the droplet surface during the electrospray ionization process and thus increase ionization efficiency [17]. The opposite effect, known as signal suppression, was also observed during the experiments. Signal suppression can be significant (up to 90%) in some cases in other environmental waters such as hospital and wastewater influents [19]. While these matrix effects can be hardly predicted, they could be reduced by modification in the sample preparation and the chromatographic separation.

4.5.2 Impact of analyte concentration on spectral accuracy and formula ranking on QqTOFMS data

Results in **Table 5** show that concentration (80 vs 300 μg L^{-1} spiked in the river water matrix) was an important factor in deciding the correct formula rank for the larger compounds (> 350 Da). For smaller compounds (< 350 Da), the lower number of possible outcomes for a given accurate mass requires less spectral accuracy for correct formula identification. This was especially true for ATZ and SMX which have very recognizable features with their Cl and S atoms, respectively. Spectral accuracy of MTX (454 Da) (**Table 6**) was the most affected by concentration, from 90.7% ± 0.9 % in the 80 μg L^{-1} matrix solution to 97.9% ± 0.5 % in the 300 μg L^{-1} matrix solution. For the other compounds, differences were < 4 percentage points. This can be explained by a higher signal for the M+2 and M+3 peaks which allows for a more thorough comparison of the calibrated experimental and theoretical peaks in MassWorks. When peak intensities were < 800 counts, major discrepancies are observed between the calibrated experimental and theoretical peaks. These differences result in much lower overall spectral accuracy and less confidence in formula determination. Since the signals for the other compounds were relatively higher than those of MTX in the 80 μg L^{-1} matrix solution, the increase in concentration had a less noticeable effect. Finally, for SMX and OFL, standard deviations > 5 percentage points in spectral accuracies were observed. SMX standard deviation was indeed almost 4 times larger in the 300 μg L^{-1} solution compared to the 80 μg L^{-1} solution. Such high variation is due to an outlier value. When the outlier replicate was removed, the variation in spectral accuracy was within 2 percentage points.

4.5.3 Impact of mass analyzer resolution on spectral accuracy and identification

4.5.3.1 Triple quadruple mass spectrometer (QqQMS)

Measurements with QqQMS were done using both internal and external mass calibrations. Better results were obtained using internal mass calibration and are discussed here. Results obtained using external calibration are discussed in the Supporting Information (section 9.2.1).

Post-acquisition CLIPs internal calibration results in terms of mass accuracy can be seen in **Table 4**. For all compounds, average mass accuracy was ≤ 5 ppm. As expected, better mass accuracy was obtained in both QqTOFS and QqOrbitrapMS, but it is remarkable that a QqQMS can attain such low errors in measuring accurate masses. However, concentration had a major impact on both ranking and spectral accuracy on the QqQMS data. In fact, it was possible to obtain meaningful data only for the samples spiked at 300 μg L^{-1} (**Table 5**, **Table 6**). In general, the matrix slightly increased spectral accuracy but rank was not significantly affected. As explained previously, the higher spectral accuracy in the river matrix could be explained by signal enhancement during the ESI process, in which the co-eluting compounds improve ionization efficiency. As it can be seen in **Table 18** (Supporting Information) for QqQMS data, matrix/MeOH area ratios were between 1.2 and 2.8, indicating signal enhancement due to the matrix, were observed for all compounds except for ROX (0.9). However, for TRI and MTX, lower rankings despite the matrix induced signal enhancement were observed. These results illustrate the two opposite effects of the matrix on spectral accuracy and formula ranking : 1) ions of co-eluting matrix compounds with *overlapping m/z* values with the compound of interest may affect intensities of the isotopic pattern and thus lower spectral accuracy and ranking; 2) ions of co-eluting matrix compounds of *different m/z* from the compound of interest may increase spectral accuracy and formula ranking by increasing the signal-to-noise ratio of all the peaks of the isotopic pattern.

As shown in **Table 5**, obtained rankings were not good enough to allow acceptable certainty in formula determination in most cases. Therefore, these results suggest that the use of a QqQMS for the measurement of accurate masses and spectral accuracy of small organic molecules at low concentrations in complex matrices seems to be very challenging. Nevertheless, it could be very helpful for other routine applications that use simpler matrices and compounds at higher concentrations, e.g. monitoring and confirmation of organic synthesis products or impurity and degradation identification in pharmaceutical products.

4.5.3.2 Quadrupole-orbitrap mass spectrometer (QqOrbitrapMS)

Mass accuracy and ranking results at 80 and 300 μg L^{-1} in MeOH and the matrix solution obtained with QqOrbitrapMS at R_{FWHM} = 70K and 140K are shown in **Table 4** and **Table 5**, respectively.

As expected, mass accuracy for the QqOrbitrapMS were generally lower or equal than 2 ppm. Concerning ranking, both low mass compounds (< 350 Da) and high mass compounds (between 350 and 837 Da) were not affected by resolution considering that rankings were consistently in the top ten for both resolutions. Generally, lower spectral accuracy was observed for the low mass compounds compared to the high mass compounds, but it did not appear to be affected by resolution at lower concentration. For MTP, TRI and OFL in the matrix samples spiked at the higher concentration (300 $\mu g\ L^{-1}$), the largest drops in spectral accuracy caused by an increase in resolution were between 2.2 to 2.7 percentage points, which were statistically significant at the 95% confidence level according to the t-test. Such a reduction is only about half of that reported by a previous study using a linear ion trap-orbitrap mass spectrometer [12], where the spectral accuracy of polar organic compounds (masses between 639 to 1664 Da) was higher (~97%) at R_{FWHM}= 7.5 K than that obtained at R_{FWHM}= 100K (<90%). According to the authors of that study, high resolution hinders high spectral accuracy in the orbitrap mass spectrometer. The authors explained those results as the consequence of a phenomenon called "isotope beating". That phenomenon results from destructive interference of signals having close *m/z* values, e.g. isotopic peaks of multiply charged ions or closely located isotopes under a given isotope cluster such as M+3, and produce errors in the measurements of isotopic abundances[20]. Such effects have been observed with polymers in ion cyclotron resonance (ICR) mass spectrometers [20] and may also be present in the QqOrbitrapMS, which is also an ion trap mass analyzer using Fourier transform signal conversion. Interestingly, such negative correlation between resolution and spectral accuracy was not observed in the present study for compounds with the highest molecular masses (ROX, JOS). It is likely that such issues have been either reduced or corrected in the newer orbitrap models, especially at the high masses (*m/z* > 800). It is possible that enhanced Fourier transform for orbitrap mass spectrometry, introduced in 2014 brought some amelioration in peak shape with the addition of apodization and triple zero-filling. These notably helped reducing spectral leakage and side lobes around peaks [20]. It has yet to be confirmed that those factors were key in reducing spectral accuracy with higher resolutions, but it is out of the scope of this article. However, it is not clear how this improvement affects spectral accuracy as a function of resolution. Another possibility, proposed by Xu*, et al.* [14], is the effect of the reduced-profile mode for the recording of mass spectra in orbitraps. In that mode, the noise is subtracted from the acquired mass

spectra to reduce raw file size. This signal processing can lead to underestimation of the abundance of M+*n* peaks, thus reducing spectral accuracy.

Table 4. Mass accuracy in ppm for the test compounds in three different mass analyzers employed.

Compounds	QqQMS		QqTOFMS (R_{FHWM}=25 K)				QqOrbitrapMS (R_{FHWM}=70 K)				QqOrbitrapMS (R_{FHWM}=140 K)			
	MeOH	Matrix	MeOH		Matrix		MeOH		Matrix		MeOH		Matrix	
	300 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}
ATZ (215)	2 ± 1	5 ± 3	1.5 ± 0.3	0.4 ± 0.3	1.6 ± 0.8	2 ± 1	2 ± 2	3 ± 2	3 ± 2	3 ± 1	1 ± 1	2 ± 1	1.2 ± 0.9	3.3 ± 0.3
SMX (253)	2 ± 2	1.2 ± 0.7	1.0 ± 0.5	0.4 ± 0.3	1.135	0.7 ± 0.4	0.3 ± 0.2	1.0 ± 0.2	2.3 ± 0.2	0.5 ± 0.3	0.3 ± 0.2	0.5 ± 0.2	1.8 ± 0.2	0.4 ± 0.4
MTP (267)	NA	NA	0.2 ± 0.2	0.1 ± 0.1	1.1 ± 0.6	0.9 ± 0.2	0.6 ± 0.2	1.8 ± 0.9	1.0 ± 0.4	0.7 ± 0.4	0.3 ± 0.2	1.1 ± 0.2	1.0436	0.9 ± 0.2
TRI (290)	2 ± 2	4 ± 3	0.3 ± 0.1	0.6 ± 0.2	0.7 ± 0.4	1 ± 1	0.8 ± 0.4	2.1 ± 0.7	0.6 ± 0.2	1.1 ± 0.5	0.3 ± 0.2	1.6 ± 0.3	0.6 ± 0.2	1.1 ± 0.2
FLX (309)	NA	NA	6.0 ± 0.9	3 ± 2	2 ± 2	2 ± 2	1.6 ± 0.2	2.6 ± 0.5	2 ± 2	3 ± 1	0.73	1.4	0.7 ± 0.9	1.5 ± 0.4
OFL (361)	NA	NA	1.5 ± 0.8	1.9 ± 0.8	0.70 ± 0.03	0.2 ± 0.2	1.2 ± 0.3	0.7 ± 0.7	1.4 ± 0.2	1.1 ± 0.2	1.4 ± 0.2	0.67	1.1 ± 0.3	0.9 ± 0.2
MTX (454)	3 ± 2	4 ± 3	1.0 ± 0.9	1.6 ± 0.3	2 ± 1	0.2 ± 0.2	0.8 ± 0.1	0.5 ± 0.1	1.4	0.5 ± 0.2	0.9 ± 0.2	0.2 ± 0.2	1.3 ± 0.3	0.5 ± 0.2
JOS (827)	NA	NA	1.2 ± 0.3	1.1 ± 0.7	1 ± 1	0.6 ± 0.3	2.3 ± 0.9	3.9 ± 0.8	1.8 ± 0.3	2.7 ± 0.5	1.2 ± 0.5	2.2 ± 0.2	0.9 ± 0.7	2.3 ± 0.6
ROX (837)	0.7 ± 0.2	0.8 ± 0.1	0.5 ± 0.2	0.7 ± 0.2	1.1 ± 0.7	1.0 ± 0.1	1.2 ± 0.6	4.3 ± 0.2	0.7 ± 0.5	2.1 ± 0.2	1.0 ± 0.4	2.2 ± 0.7	0.9 ± 0.2	2.2 ± 0.3

Table 5. Formula ranking results for the test compounds in three different mass analyzers employed.

Compounds	QqQMS		QqTOFMS (R_{FHWM}=25 K)				QqOrbitrapMS (R_{FHWM}=70 K)				QqOrbitrapMS (R_{FHWM}=140 K)			
	MeOH	Matrix	MeOH		Matrix		MeOH		Matrix		MeOH		Matrix	
	300 μg L^{-1}	300 μg L^{-1}	80 μg L^{-1}	300 μg L^{-1}	80 μg L^{-1}	300 μg L^{-1}	80 μg L^{-1}	300 μg L^{-1}	80 μg L^{-1}	300 μg L^{-1}	80 μg L^{-1}	300 μg L^{-1}	80 μg L^{-1}	300 μg L^{-1}
ATZ (215)	3 ± 2	3.3 ± 0.6	1	1	1	1	1	1	1	1	1	1	1	1
SMX (253)	127 ± 103	107 ± 66	1	1	1	1	1	1	1	1	1	1	1	1
MTP (267)	NA	NA	1	1	1	1	1	1	1	1	1	1	1	1
TRI (290)	9 ± 8	47 ± 69	1	1	1	1	1	1	1	1	1	1	1	1
FLX (309)	NA	NA	1 ± 1	1	1	1	1	1	1	1	1	1	1	1
OFL (361)	NA	NA	3 ± 2	2 ± 1	1 ± 1	1	1 ±1	1	1 ± 1	1	1	1	1	1
MTX (454)	218 ± 110	1177 ± 858	5 ± 3	1	8 ± 3	2 ± 1	7 ± 1	5 ± 2	9 ± 1	2 ± 1	6 ± 2	8 ± 1	7 ± 2	8 ± 1
JOS (827)	NA	NA	25 ± 24	3 ± 1	34 ± 33	11 ± 12	1	1 ± 1	2 ± 1	2 ± 1	1	1	2 ± 2	2 ± 1
ROX (837)	57 ± 15	48 ± 17	2 ± 2	2 ± 1	4 ± 2	2 ± 1	1	1	1	1	1 ± 1	1	1	1

Table 6. Spectral Accuracy results for the test compounds in three different mass analyzers employed.

Compounds	QqQMS		QqTOFMS (R_{FHWM}=25 K)				QqOrbitrapMS (R_{FHWM}=70 K)				QqOrbitrapMS (R_{FHWM}=140 K)			
	MeOH	Matrix	MeOH		Matrix		MeOH		Matrix		MeOH		Matrix	
	300 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}	80 µg L^{-1}	300 µg L^{-1}
ATZ (215)	90 ± 2	91 ± 1	98.5 ± 0.5	98.9 ± 0.1	98.96 ± 0.04	98.8 ± 0.7	97.7 ± 0.2	97.27 ± 0.04	97.5 ± 0.4	97.6 ± 0.2	97.08 ± 0.05	95.8 ± 0.2	97.3 ± 0.2	96.5 ± 0.2
SMX (253)	86 ± 10	94 ± 4	97.0 ± 0.5	98.5 ± 0.7	97.5 ± 0.5	91 ± 6	85.7 ± 0.5	95.8 ± 0.2	98.2 ± 0.4	98.7 ± 0.1	88 ± 1	95.8 ± 0.3	98 ± 1	98.8 ± 0.1
MTP (267)	NA	NA	98.45 ± 0.02	96 ± 1	98.5 ± 0.2	97 ± 5	98.8 ± 0.5	98.2 ± 0.04	99.55 ± 0.08	98.2 ± 0.6	98.3 ± 0.2	96.7 ± 0.8	98.88 ± 0.03	96.0 ± 0.3
TRI (290)	97 ± 1	97 ± 1	97.98 ± 0.09	96 ± 1	98.6 ± 0.2	99.3 ± 0.2	97.70 ± 0.08	98.0 ± 0.2	98.5 ± 0.3	97.3 ± 0.8	98.1 ± 0.1	95.0 ± 0.8	98.2 ± 0.4	95 ± 1
FLX (309)	NA	NA	97.0 ± 0.4	97.4 ± 0.2	99.1 ± 0.1	97 ± 3	95.2 ± 0.4	98.6 ± 0.1	99.60 ± 0.06	99.3 ± 0.2	98 ± 1	98.90 ± 0.05	99.5 ± 0.1	99.13 ± 0.06
OFL (361)	NA	NA	93.17 ± 0.08	99.32 ± 0.07	98.7 ± 0.2	96 ± 6	96.6 ± 0.4	98.07 ± 0.02	98.4 ± 0.5	98.0 ± 0.3	93 ± 1	97.42 ± 0.07	99.03 ± 0.02	95.4 ± 0.4
MTX (454)	82 ± 9	86 ± 5	89 ± 6	92.2 ± 0.2	93.3 ± 0.2	97.9 ± 0.5	97.4 ± 0.6	97.82 ± 0.07	96.8 ± 0.1	98.52 ± 0.06	96.7 ± 0.2	96.501 ± 0.009	97.1 ± 0.2	97.2 ± 0.1
JOS (827)	NA	NA	92.19 ± 0.06	99.2 ± 0.1	96.8 ± 0.3	97 ± 2	99.5 ± 0.1	99.5 ± 0.1	99.0 ± 0.2	99.2 ± 0.1	99.29 ± 0.07	99.53 ± 0.02	99.18 ± 0.09	99.5 ± 0.1
ROX (837)	93 ± 3	95 ± 2	93.3 ± 0.1	99.2 ± 0.1	98.1 ± 0.1	98 ± 1	99.46 ± 0.06	99.46 ± 0.02	99.3 ± 0.1	99.22 ± 0.08	99.4 ± 0.2	99.3 ± 0.1	99.4 ± 0.1	99.65 ± 0.04

Although QqTOFMS are known to measure isotopic patterns very precisely [8] and accurately [13] compared to other mass analyzers, results showed that better spectral accuracies for the tested compounds were obtained with the QqOrbitrapMS at R_{FWHM}=70 K and 140 K than with the QqTOFMS at R_{FWHM}=25 K. As it can be seen in **Table 6**, in the data acquired with the QqTOFMS employing the matrix spiked at 300 μg L^{-1}, only 3 out of 9 test compounds had average spectral accuracy ≥ 98 %, while in the QqOrbitrapMS data (R_{FWHM}=70 K), 7 out of 9 compounds had average spectral accuracy ≥ 98 %. Formula rankings were also better with the QqOrbitrapMS at both resolutions compared to the QqTOFMS.

It would be interesting to perform experiments at ultra-high resolution with a ICRMS to see what full separation of the isotope fine structure would imply for spectral accuracy and formula ranking. Nevertheless, the sCLIPS algorithm does already factors in unresolved isotope fine structures in the calculation of mass accuracy since it takes into account instrument resolution to calibrate the mass spectrum signal.

An important observation from the QqOrbitrapMS results shown in **Table 5** and **Table 6** is that low spectral accuracy (e.g. < 98 %) does not necessarily mean a low ranking as some compounds consistently get low spectral accuracies, in our case SMX data clearly illustrates that. It seldom or never had > 98% spectral accuracy but it was always ranked 1st nevertheless. Since there are less possible formulas for compounds with low molecular masses for a given accurate mass compared to compounds with high molecular masses, the former requires less spectral accuracy for correct formula identification than the latter.

4.5.4 Lower concentration limits for the measurement of spectral accuracy in environmental samples

Solutions of lower concentrations, ranging from 50 to 10 μg/L, equivalent to environmental concentrations of 0.38 to 1.9 μg L^{-1} considering solid-phase extraction preconcentration factor, were analyzed with the intent to assess a minimum working concentration for spectral accuracy.

Rankings and spectral accuracies can be seen in **Figure 12** and **Figure 13** respectively. In the QqTOFMS, the larger compounds such as JOS in particular and ROX were heavily affected by the drop of signal intensity. JOS notably had very low signal intensity; thus M+2 peaks and onward were undistinguishable. Very high variability in the rankings was also observed in the cases of JOS (4 to 157) and, to a lesser degree, ROX (5 to 61). ROX did however improved significantly in ranking at 40 and 50 μg L^{-1}. ROX signal intensity was also much superior to that of JOS (16 500 counts at apex at 50 μg L^{-1} compared to 2300 in the same conditions respectively); M+2 and M+3 were well-defined. Rankings were never lower than 4 for ATZ, SMX, MTP, TRI, FLX and OLF in all instances. MTX ranking varied from 3 to 12.

Those lower concentrations mixtures were also analyzed with the QqOrbitrapMS at both 70K and 140K resolutions. Ranking and spectral accuracy results can be seen in **Figure 12** and **Figure 13**, respectively. Data showed excellent rankings (within top 5) for all compounds at all concentrations in both resolutions. Except for MTX, which ranked consistently over 5. OFL had higher ranking at the lower concentrations for both resolutions. JOS was also affected at 140K resolution for the 10 and 20 μg L^{-1}. Spectral accuracies were low for SMX, between 5.6 and 38.7% for 70K and between 3.0 and 37.7% for 140K. This was caused by an impurity of high intensity at *m/z* 256, which although well separated from the M+2 peak, lowered the match between calibrated and theoretical isotopic patterns of SMX. This impurity was not observed in the QqTOFMS data, thus spectral accuracy for SMX on that instrument was superior (54.2 to 85.1%) than in the QqOrbitrapMS.

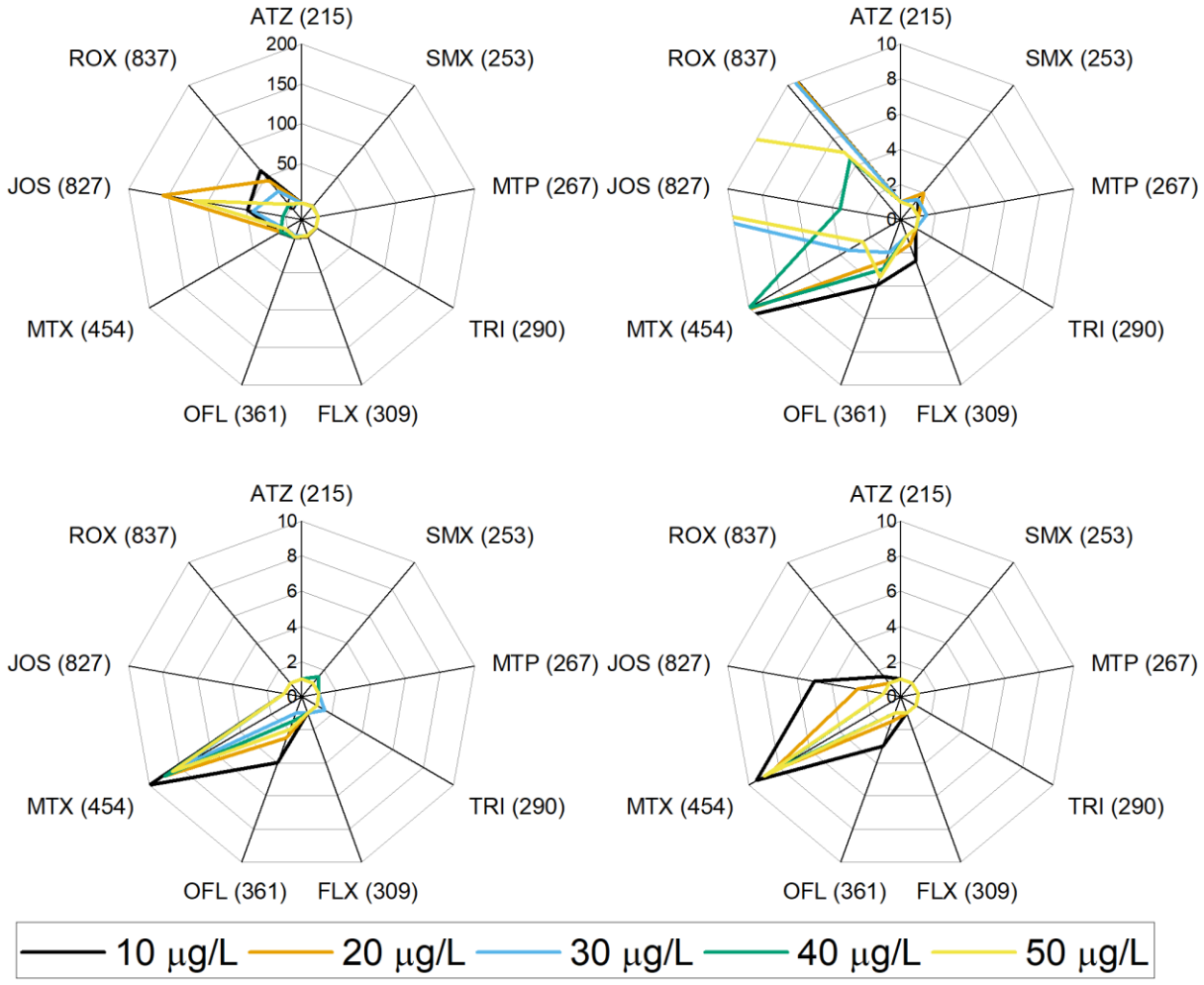


Figure 12. Radar plots representing mean formula ranking for the target micropollutants in spiked matrix samples. Top left: QqTOFMS (R_{FHWM}=25 K), top right: QqTOFMS results zoomed in, bottom left: QqOrbitrapMS (R_{FHWM}=70 K) and bottom right : QqOrbitrapMS (R_{FHWM}=140 K).

As discussed in the previous sections, adequate measurement of spectral accuracy depends on many factors such as nature of the compound, matrix composition and even the chromatographic and sample preparation methods employed. Therefore, the lower concentration limits for the measurement of spectral accuracy can be improved by using more selective solid-phase extraction or by improving the chromatographic separation of analytes from co-eluting matrix compounds having overlapping *m/z* values with the compound of interest. In summary, as long as the isotopic

pattern is free of interferences and significantly higher than the noise, measurement of spectral accuracy will yield a dependable value that could be use to improve the level of confidence in the assignment of a molecular formula to an accurate mass.

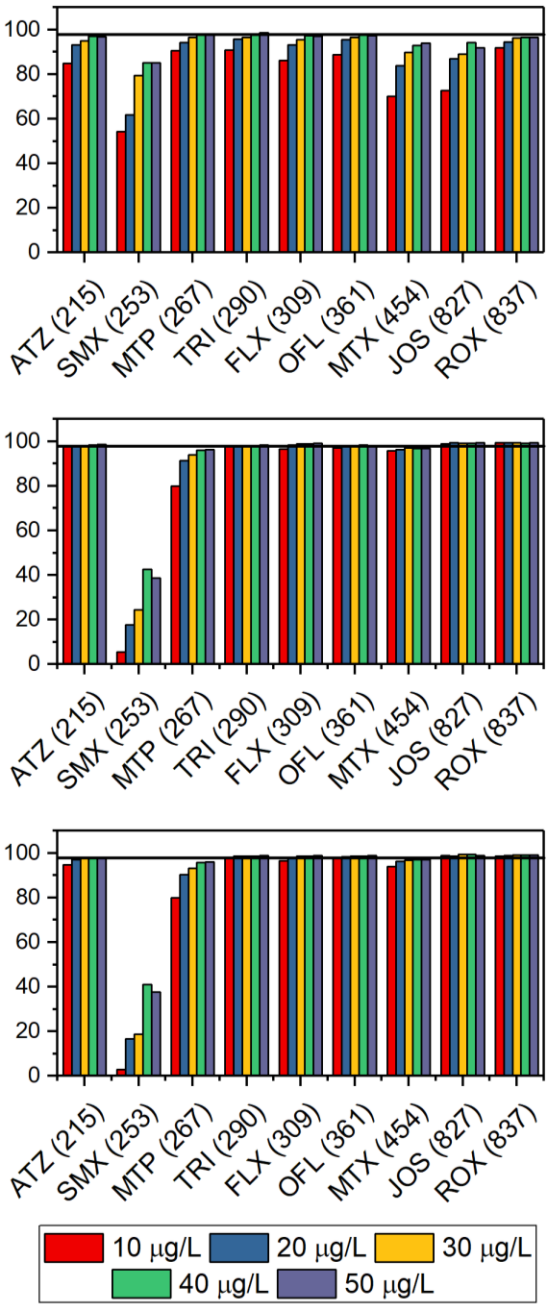


Figure 13. Bar plots representing mean spectral accuracy for the target micropollutants in spiked matrix samples. Top: QqTOFMS (R_{FHWM}=25 K), middle: QqOrbitrapMS (R_{FHWM}=70 K) and

bottom : QqOrbitrapMS (R_{FHWM}=140 K). Straight line indicates the threshold of high spectral accuracy (98%). The same sample was injected twice.

4.5.5 Rules and limitations

Comparison of spectral accuracy to other techniques used for formula determination showed that the algorithms used by MassWorks obtained better ranking of the correct formula of larger compounds (>350 Da) and it was more robust when using lower intensity signals (Supporting Information, section S-4). However, this spectral accuracy determination has also a few drawbacks. Computing time necessary for determination of ranking and spectral accuracy of the high mass compounds such as JOS (827 Da) and ROX (837 Da) was the biggest downside found in this technique. Depending on the computer performance, the time to generate thousands of formula candidates could go up to 40 minutes if limitations were not made. Atoms like F and P being monoisotopic means they can be fitted in most formulas while a compound with high M+1 relative abundance like Si compensated for the lack of M+1 distribution from the inclusion of F and P. This results in a high number of generated formulas with low C and H and high Si and P, which are not realistic. In cases where a broad range of atoms are allowed in the formulas, pre- and post-research rules need to be defined to optimize formula generation using MassWorks. First, one must observe the mass spectra and visually assess the presence or absence of Cl and Br in the compounds. These two atoms are very distinct with unmistakeable isotope patterns and they could be withdrawn from the allowed atom list if abundant M+2, M+4 or higher isotopes are not observed. Such procedure will divide by two the total number of generated formulas and thus save computing time. Additionally, the maximum number of C for a formula using the empirical parameters (seven golden rules) needs to be used with utmost caution since it may exclude the correct formula from consideration. It was observed in one case that the empirical parameters on MassWorks underestimated the maximum number of C on a formula and the correct formula did not appear in the list. Such event was observed with a background contaminant, the plasticizer diisooctyl phthalate ($C_{24}H_{38}O_4$), where the maximum number of C was 23 based on the empirical rules implemented while the compound has 24 C. Setting a more reasonable minimum number of C also helps to reduce the number of generated formulas and saves computing time. A corresponding number of H can be added for the minimum and maximum values. A maximum H/C ratio of 2 was found to help reduce the number of formulae. P tends to be inserted in all

formulas and can be monitored by a O/P ratio of minimum 3 as it mainly occurs in the form of organophosphates with high O per P; double bond equivalents (DBE) then also should be monitored as a P-O double bond implies a higher DBE. Na or K adducts were not observed for the selected compounds in this work, but adding them to the allowed atoms might help to uncover the accurate formula in case positive alkaline adducts are formed. Adjusting pre-search parameters is not necessary with smaller molecules as there are only few possibilities for a compound with a narrow mass error window. Finally, lack of automation is also an issue as each compound mass spectrum needs to be evaluated individually.

4.6 Conclusion

This study showed that spectral accuracy is a powerful tool for the determination of chemical formulas from accurate mass data. Spectral accuracy allowed to reduce the number of likely molecular formulas for organic micropollutants of relative high molecular mass (e.g. between 400 and 900 Da) to less than 10, and in some cases, it assigned unambiguously one specific molecular formula to an experimental isotopic pattern. Experiments showed that the major parameter affecting spectral accuracy and correct formula ranking for a set of common organic micropollutants is signal intensity. Thus, conditions increasing signal intensity, such as signal enhancement by the matrix and higher compound concentration, favoured higher spectral accuracy and ranking of the correct molecular formula. A significant improvement of both ranking and spectral accuracy was also obtained with higher mass resolution of the mass analyzer. Contrary to a previous study[13], a moderate (≈7 percentage points) decrease in spectral accuracy with higher resolution in the orbitrap mass spectrometer was not observed.

Results also showed that high spectral accuracy (e.g. > 98 %) and identification of the correct molecular formula were not necessarily correlated for low molecular mass compounds (< 350 Da). It was however more prevalent for high molecular mass compounds (>350 Da). Using MassWorks software it was possible to acquire accurate mass data with less than 5 ppm mass accuracy in a QqQMS. While the low resolution of the QqQMS impairs accurate mass and spectral accuracy determination in complex matrices such as river water, application of spectral accuracy to routine analyses is of interest for laboratories without access to high resolution MS technology.

Experiments demonstrated that for some compounds, high spectral accuracies and rankings can be obtained at concentrations as low as 10 μg L^{-1} and in general, if the isotopic pattern of the compound is free of major interferences and the signal is above the noise of the instrument, it is possible to measure spectral accuracy correctly.

Finally, this study confirmed that spectral accuracy could be used as a complementary technique to eliminate formula candidates corresponding to an observed accurate mass during identification workflows of organic micropollutants based on liquid chromatography-high resolution MS. Thus, spectral accuracy is a powerful tool to elevate level 5 data (accurate mass) to level 4 (unequivocal molecular formula) according to the identification confidence levels proposed by Schymanski et al. [2]. In this way, identification of unknowns present in environmental samples can be a more efficient process.

4.7 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.
(section 9.2). Molecular structures of the test compounds used in this study; 9.2.1. External calibration and stability of the QqQMS system; 9.2.2. Error in spectral accuracy determination; 9.2.3. Interference rejection; **Table 18**. Determination of matrix effects in data acquired the three mass spectrometers; 9.2.4. Software comparison.

4.8 Acknowledgements

This study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery grant awarded to P.A. Segura. We would like to thank Yongdong Wang and Don Kuehl of Cerno Biosciences for his help with MassWorks software. Funding sources did not have any involvement in the design, experiments, data interpretation, writing, revision or submission of this study. We would like to thanks Thermo Fisher Scientific for providing a generous access to a QqOrbitrapMS.

4.9 References

(1) Boxall, A.; Rudd, M.; Brooks, B.; Caldwell, D.; Choi, K.; Hickmann, S.; Innes, E.; Ostapyk, K.; Staveley, J.; Verslycke, T.; Ankley, G.; Beazley, K.; Belanger, S.; Berninger, J.; Carriquiriborde, P.; Coors, A.; Deleo, P.; Dyer, S.; Ericson, J.; Gagné, F., et al. *Environ. Health Perspect.* **2012**, *120*, 1221–1229.

(2) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 2097-2098.

(3) Gross, J. H. *Mass Spectrometry: A Textbook*, 2nd ed.; Springer: Berlin, Germany, 2011, p 716.

(4) Eysseric, E.; Bellerose, X.; Lavoie, J.-M.; Segura, P. A. *Can. J. Chem.* **2016**, *94*, 781-787.

(5) Yuan, M.; Breitkopf, S. B.; Yang, X.; Asara, J. M. *Nature protocols* **2012**, *7*, 872.

(6) Zhang, Y.; Ren, Y.; Jiao, J.; Li, D.; Zhang, Y. *Anal. Chem.* **2011**, *83*, 3297-3304.

(7) Cortés-Francisco, N.; Flores, C.; Moyano, E.; Caixach, J. *Anal. Bioanal. Chem.* **2011**, *400*, 3595-3606.

(8) Kind, T.; Fiehn, O. *BMC bioinformatics* **2007**, *8*, 1.

(9) García-Reyes, J. F.; Hernando, M. D.; Molina-Díaz, A.; Fernández-Alba, A. R. *TrAC Trends in Analytical Chemistry* **2007**, *26*, 828-841.

(10) Little, J. L.; Cleven, C. D.; Brown, S. D. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 348-359.

(11) Little, J. L.; Williams, A. J.; Pshenichnov, A.; Tkachenko, V. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 179-185.

(12) Wang, Y.; Gu, M. *Anal. Chem.* **2010**, *82*, 7055-7062.

(13) Erve, J. C.; Gu, M.; Wang, Y.; DeMaio, W.; Talaat, R. E. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2058-2069.

(14) Xu, Y.; Heilier, J.-F.; Madalinski, G.; Genin, E.; Ezan, E.; Tabet, J.-C.; Junot, C. *Anal. Chem.* **2010**, *82*, 5490-5501.

(15) Hughes, S. R.; Kay, P.; Brown, L. E. *Environ. Sci. Technol.* **2013**, *47*, 661-677.

(16) Segura, P. A.; Takada, H.; Correa, J. A.; El Saadi, K.; Koike, T.; Onwona-Agyeman, S.; Ofosu-Anim, J.; Sabi, E. B.; Wasonga, O. V.; Mghalu, J. M.; dos Santos Junior, A. M.; Newman, B.; Weerts, S.; Yargeau, V. *Environ. Int.* **2015**, *80*, 89-97.

(17) Taylor, P. J. *Clin. Biochem.* **2005**, *38*, 328-334.

(18) Segura, P. A.; Gagnon, C.; Sauvé, S. *Anal. Chim. Acta* **2007**, *604*, 147-157.

(19) Gros, M.; Rodríguez-Mozaz, S.; Barceló, D. *J. Chromatogr. A* **2013**, *1292*, 173-188.

(20) Lange, O.; Damoc, E.; Wieghaus, A.; Makarov, A. *Int. J. Mass Spectrom.* **2014**, *369*, 16-22.

CHAPITRE 5. DÉPISTAGE NON CIBLÉ DE CONTAMINANTS ORGANIQUES PAR ANALYSE COMBINATOIRE DE SPECTRES DE MASSE EN TANDEM

5.1 Avant propos

Ce chapitre a été publié dans le journal « Talanta » sous les références Eysseric et al. 10.1016/j.talanta.2021.122293

« 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 »

5.1.1 Auteurs et affiliation

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.

3Environment and Climate Change Canada, Montréal, Canada

5.1.2 Présentation de l'article

Le dépistage non ciblé se base fortement sur l'utilisation de banques de spectres de masse en tandem à haute résolution ($HRMS^2$) pour identifier des composés inconnus. Or, ces banques de données sont souvent limitées dans la quantité de composés à intérêt environnemental qui les composent. Cela est d'autant plus vrai pour les produits de transformation qui sont pour la grande majorité peu ou pas connus. De plus, il y a très peu de métadonnées disponibles pour les additifs de produits de consommation et les intermédiaires de réaction. Il y a donc un goulot d'étranglement très important dans le dépistage non ciblé causé par la taille des banques de données $HRMS^2$.

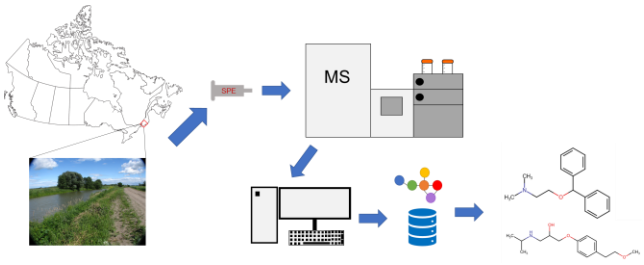
Pour remédier à cette situation, il existe des outils computationnels, ou *in silico*, qui permettent de générer des spectres $HRMS^2$ à partir des informations contenues dans les grands recueils de

composés comme PubChem ou ChemSpider. Les identifications ont en revanche un plus haut risque d'être erronées que celles réalisées par des banques de données empiriques. Dans cet article, trois outils d'identification de composés par correspondance de spectres $HRMS^2$, dont deux *in silico* et une plateforme de banques de spectres $HRMS^2$ empiriques, sont utilisés dans le cadre d'un dépistage non ciblé dans des échantillons d'eau de surface de la rivière Yamaska en amont et en aval de la station de traitement des eaux usées de la ville de Granby.

5.1.3 Contributions des auteurs

Le plan expérimental a été conçu par Emmanuel Eysseric (EE) et Pedro A. Segura (PAS). Les analyses par spectrométrie de masse ont été réalisées par EE sur le spectromètre de masse quadripôle-orbitrap appartenant à Francis Beaudry (FB). Le développement des méthodes a été réalisé par EE. L'échantillonnage et l'extraction ont été réalisés par EE. Le traitement des données a été réalisé par EE. La visualisation des données a été réalisée par EE. La rédaction a été réalisée par EE et PAS. La révision a été effectuée par EE, PAS, FB et Christian Gagnon. La soumission a été réalisée par PAS. Les réponses aux réviseurs ont été réalisées par PAS et EE.

Graphical abstract



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

5.2 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.

5.3 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, 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 14**. 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 14**. 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.

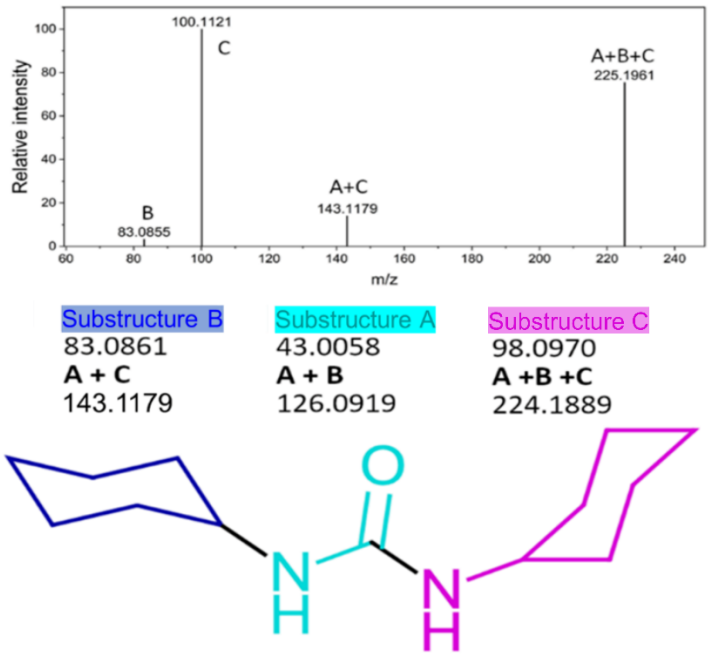


Figure 14. Four ions are found in the MS^2 spectrum of dicyclohexylurea on mzCloud. All four ions can be explained by substructures in the figure with the addition/subtraction of one or two hydrogen atoms. These complementary substructures were generated by systematically disconnecting the breakable 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.

5.4 Experimental section

5.4.1 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.

5.4.2 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 39**). 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 40** 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.

5.4.3 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.

5.4.4 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 (**Figure 41**).

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 (**Figure 42**).

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 (**Figure 43**).

5.4.5 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 **Figure 41**; details about how the field and instrumental blanks were used to look for contaminants with GNPS are given in **Figure 42**; details about how the background subtraction were applied with patRoon and MetFrag are given in **Figure 43**.

5.4.6 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.

5.5 Results and discussion

5.5.1 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 7** for SPS and in **Figure 15** for GNPS and MetFrag. In **Table 7**, 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 7. 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 15a**, 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 15b**, 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 16a**): 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 8**). A more detailed account of the matched product ions can be found in **Table 21**. 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 44** (Supporting Information).

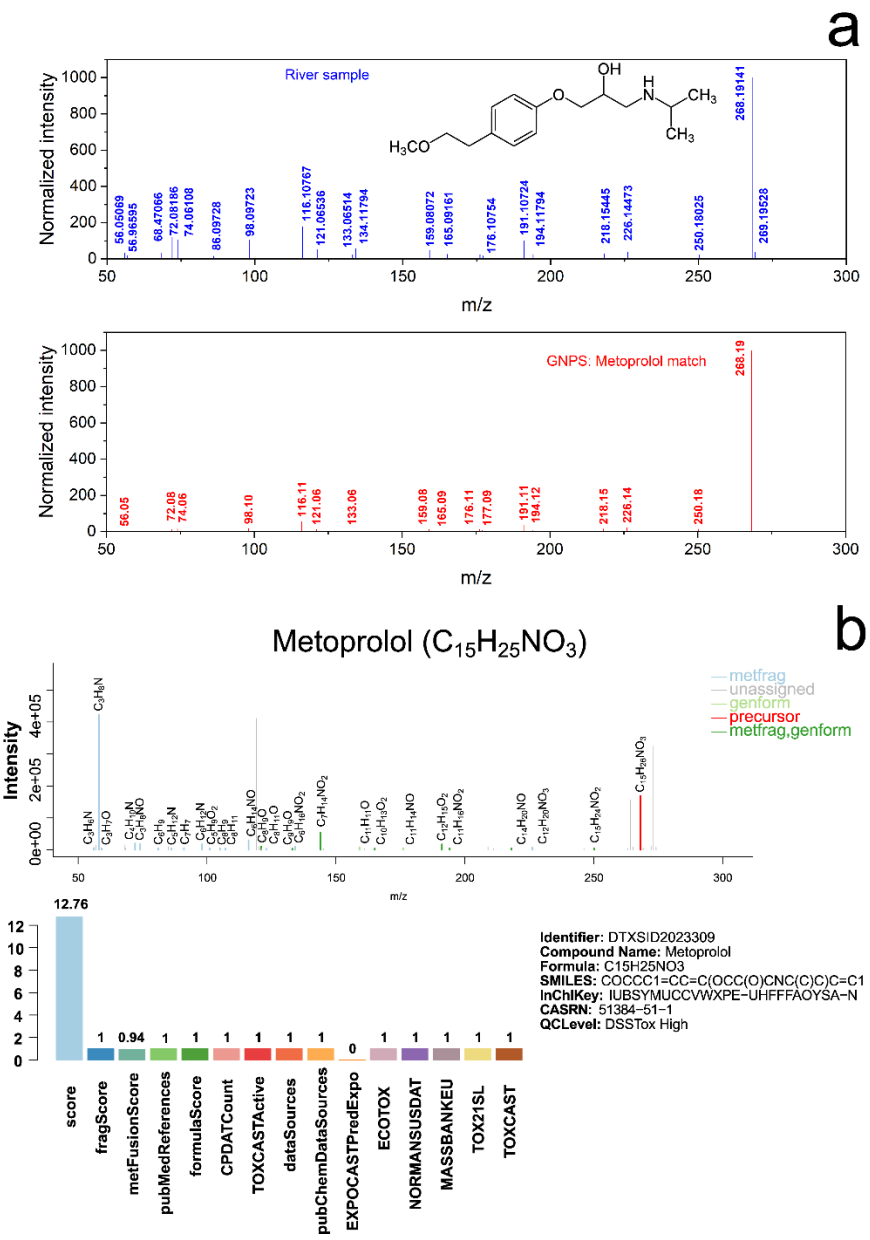


Figure 15. Example of probable structure match for metoprolol a) from GNPS and b) from MetFrag. FragScore is the MetFrag score; metFusion score combines MetFrag score with MassBank when applicable i.e., when MassBank has an entry for the compound; pubMedReference is the number of times a compound is 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.

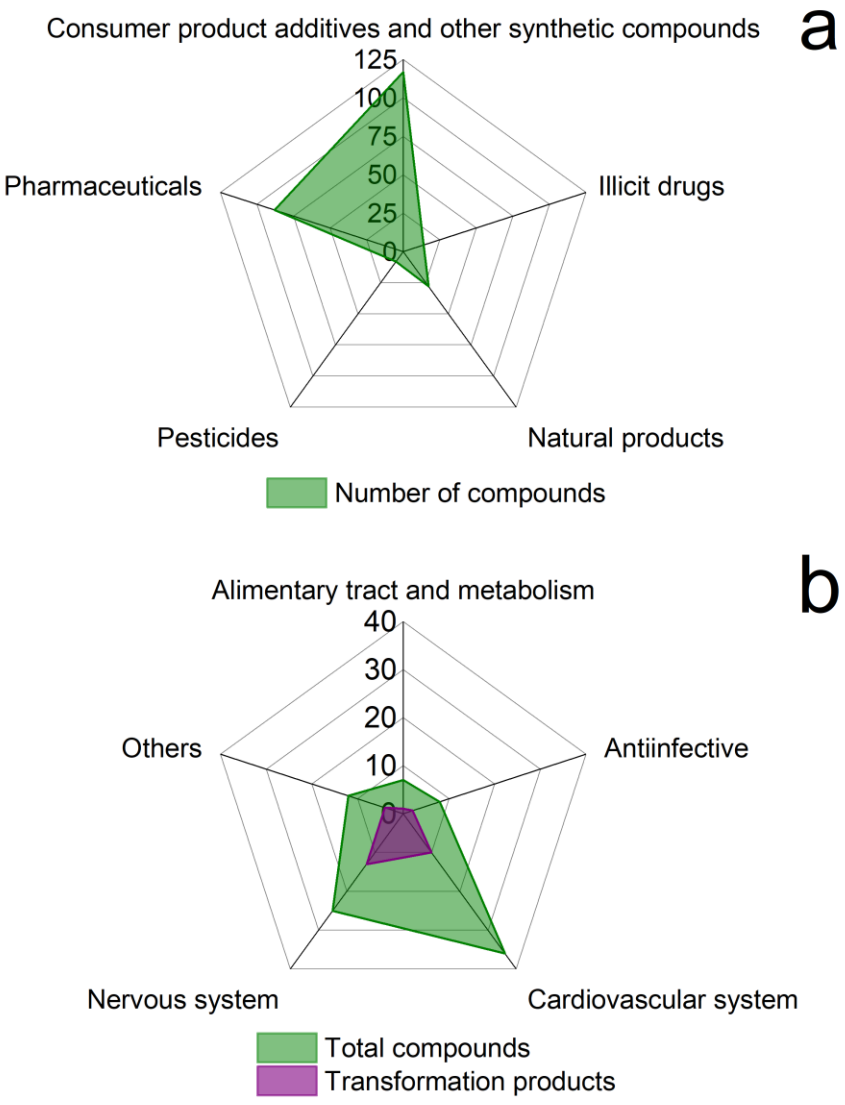


Figure 16. a: Generic classes of the compounds identified as probable or confirmed structures by the multi-tool method. **b:** Anatomical Therapeutic Chemical (ATC) classes of pharmaceutical

compounds identified as probable or confirmed structures. All the identified chemicals are found in the "IdentifiedCompounds.xslx" file (Supporting Information).

In **Table 8**, 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 16b**) 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 45**, Supporting Information).

Table 8. 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 17**. 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.

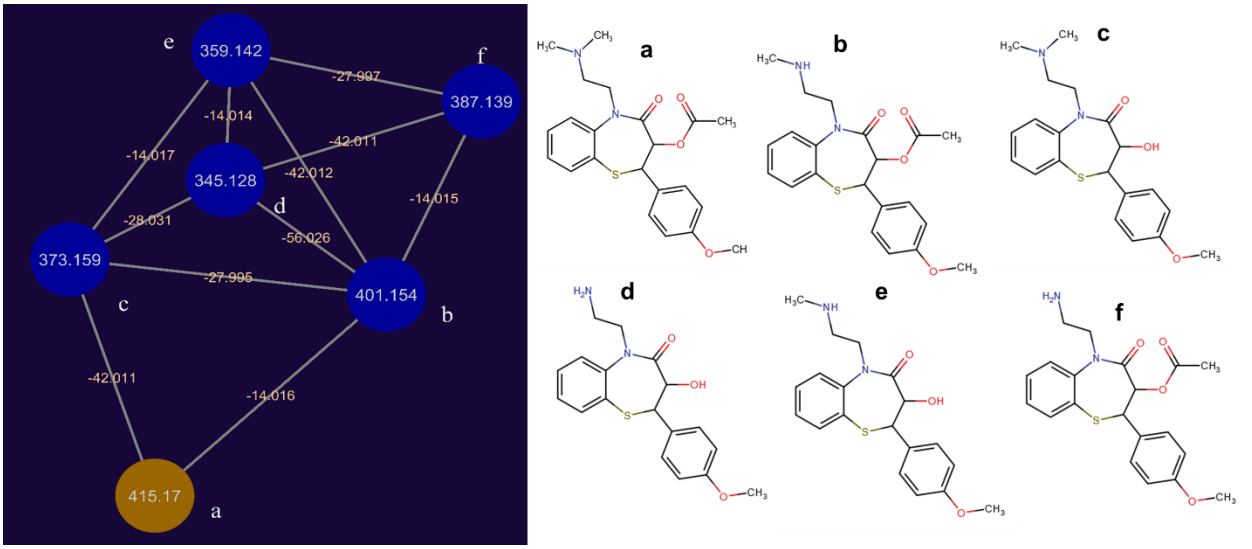


Figure 17. Molecular network of the calcium channel blocker diltiazem (a) and its transformation products (b to f) with proposed structures to the left. Each detected precursor is a node linked in the network with precursors that have similar MS^2 spectra. In the figure, **a** is diltiazem, **b** is 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 46** (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 8**. For the other congeners' confirmation, please see **Figure 47** (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 48** (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.

5.6 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.

5.7 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.

5.8 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.

CHAPITRE 6. ÉTUDE EN LABORATOIRE DU DEVENIR DE QUATRE CONTAMINANTS ORGANIQUES PRIORITAIRES ET CONFIRMATION DE LEUR PRÉSENCE DANS LES EAUX DE SURFACE

6.1 Avant propos

Ce chapitre a été publié dans le journal avec comité de lecture « Chemosphere » sous la référence : Eysseric *et al.* (2022). DOI : 10.1016/j.chemosphere.2021.133408

« Uncovering transformation products of four organic contaminants of concern by photodegradation experiments and analysis of real samples from a local river »

6.1.1 Auteurs et affiliation

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

[1]Department of Chemistry, Université de Sherbrooke, Sherbrooke, Canada
[2]Environment and Climate Change Canada, Montreal, Canada,

* Tel: 1-(819) 821-7922. E-mail: pa.segura@usherbrooke.ca

6.1.2 Présentation de l'article

Il y a plusieurs centaines de milliers de composés et mélanges de composés enregistrés pour la production dans le monde. Ces composés, lorsque relâchés dans l'environnement sont soumis à des conditions menant à leur potentielle dégradation dont la photolyse qui joue un rôle particulièrement important pour les composés pharmaceutiques. Or, leur devenir environnemental est très peu connu et encore moins l'identité de leurs potentiels produits de transformation. Pour connaître le comportement de dégradation de composés dans l'environnement, on les expose à des conditions imitant celles que l'on peut y trouver en laboratoire. Les produits de transformation sont ensuite identifiés à la suite des expositions. Cependant, peu d'études viennent confirmer la

présence environnementale des produits de transformation générés en laboratoire; sans cette validation, il y a moins de certitude sur leur occurrence dans l'environnement et donc sur la pertinence de les incorporer à des listes de suspects pour les études de dépistages. Dans cet article, les produits de transformation de quatre contaminants préoccupants selon leur occurrence, leur toxicité et ou leur persistance sont identifiés après une exposition à des conditions imitant la dégradation du soleil. La présence environnementale de ces composés est ensuite confirmée dans un dépistage des suspects.

6.1.3 Contribution des auteurs

Le plan expérimental a été conçu par Emmanuel Eysseric (EE), Christian Gagnon (CG) et Pedro A. Segura (PAS). Les analyses par spectrométrie de masse ont été réalisées par EE sur le spectromètre de masse quadripôle-orbitrap. Le développement des méthodes a été réalisé par EE. L'exposition des composés et leur extraction ont été réalisées par EE. L'échantillonnage et l'extraction pour le dépistage des suspects ont été réalisés par EE. Le traitement des données a été réalisé par EE. La visualisation des données a été réalisée par EE. La rédaction a été réalisée par EE, CG et PAS. La révision a été effectuée par EE, PAS et CG. La soumission a été réalisée par PAS. Les réponses aux réviseurs ont été réalisées par CG, PAS et EE.

6.2 Abstract

In this study, photodegradation experiments simulating the exposure conditions of sunlight on the commonly detected in surface and wastewater contaminants atorvastatin (ATV), bezafibrate (BEZ), oxybenzone (OXZ), and tris(2-butoxyethyl)phosphate (TBEP) were conducted as the fate of these compounds and their transformation products (TPs) was followed. Then a nontargeted analysis was carried out on an urban river to confirm the environmental occurrence of the TPs after which the ECOSAR software was used to generate predicted effect levels of toxicity of the detected TPs on aquatic organisms. Five TPs of ATV were tentatively identified including two stable ones at the end of the experiment: ATV_TP557a and ATV_TP575, that were the product of hydroxylation. Complete degradation of OXZ was observed in the experiment with no significant TP identified. BEZ remained stable and largely undegraded at the end of the exposure. Five TPs

of TBEP were found including four that were stable at the end of the experiment: TBEP_TP413, TBEP_TP415, TBEP_TP429, and TBEP_TP343. In the nontargeted analysis, ATV_TP557b, a positional isomer of ATV_TP557a, ATV_TP575 and the 5 TPs of TBEP were tentatively identified. The predicted concentration for effect levels were lower for ATV_TP557b compared to ATV indicating the TP is potentially more toxic than the parent compound. All the TPs of TBEP showed lower predicted toxicity toward aquatic organisms than their parent compound. These results highlight the importance of conducting complete workflows from laboratory experiments, followed by nontargeted analysis to confirm environmental occurrence to end with predicted toxicity to better communicate concern of the newfound TPs to monitoring programs.

Keywords: pharmaceuticals and personal care products; flame retardants; transformation products; ecotoxicity prediction; nontarget analysis.

6.3 Introduction

There are tens of thousands of chemicals that are commercially available and that may thus be released in the environment which includes surface waters (Hollender et al., 2019). Among them, pharmaceutically active compounds are a significant cause of concern because of their toxicity on aquatic organisms (Daughton and Ternes, 1999; Jin et al., 2012). In surface waters, organic contaminants are exposed to multiple degradation pathways and can undergo various biotic and abiotic reactions such as photolysis, hydrolysis, and dealkylation (Fatta-Kassinos et al., 2011). Photolysis through sunlight exposure notably plays an important role in the degradation of pharmaceutical compounds to form transformation products (TPs) with new structures, properties, toxicity and fate (West and Rowland, 2012; Lin et al., 2013). These TPs can be numerous and, in some instances, more toxic than the original compounds (Wang et al., 2018b). There is a major lack in knowledge on the fate of most organic contaminants as most TPs and their structure are largely unknown. As such, monitoring programs and prospective targeted analysis of surface waters may be looking at compounds that have been transformed into new TPs and miss potentially concerning contamination. For nontargeted screening, prioritizing and then identifying new TPs never reported before is a daunting and highly time-intensive task that is not ready to be deployed routinely despite some successful applications (Bletsou et al., 2015; Hollender et al., 2017; Zahn

et al., 2019; Eysseric et al., 2021; Tian et al., 2021). This lack of information results in increased uncertainty in risk assessment and toxicity prediction models.

There are multiple studies dedicated to assessing the degradation kinetics of organic contaminants and identify their TPs with advanced oxidation processes (Kong et al., 2018; Lecours et al., 2018; Henning et al., 2019; Konstas et al., 2019). There have also been laboratory studies simulating real environmental conditions in which pharmaceuticals are exposed to artificial light in surface waters (West and Rowland, 2012; Poirier-Larabie et al., 2016; Wang et al., 2018a). This approach allows controlling the parameters of the exposure experiments such as working concentrations, time of exposure and matrix composition. The fate of industrial compounds in laboratory-controlled environmental conditions have also been successfully investigated (Zahn et al., 2019). While identifying new TPs is a valuable information, only few studies will seek to "validate" these TPs by looking for them in real samples (Zahn et al., 2019). Undoubtedly, there is a need to not only expand the knowledge on the fate and TPs of concerning contaminants using controlled experiments, but also to confirm the existence of the TPs in real environmental samples.

As nontargeted screening continues to expand, so does the need to obtain high levels of confidence (Schymanski et al., 2014) on the structures of priority contaminants generated through transformation experiments and confirmation of their occurrence in environmental samples. This systematic bottom-up approach was successfully applied for several industrial contaminants (Zahn et al., 2019). Among the diverse contaminants of emerging concern present in the aquatic environment, atorvastatin (ATV), bezafibrate (BEZ), oxybenzone (OXZ) and tris(2-butoxyethyl)phosphate (TBEP) are of interest because of high toxicity and ubiquity. Also, for these compounds, knowledge on the fate in surface water and the identity and occurrence of TPs in real environmental samples is still lacking (Yin et al., 2017; Henning et al., 2019; Zahn et al., 2019).
ATV is a widely consumed statin used in the treatment of cardiovascular events that is considered to have high stability in wastewater and sewers in its free acid or lactone form (Hermann et al., 2005; Lin et al., 2021). ATV has been detected in river and drinking waters in the past (Vanderford and Snyder, 2006) and has shown significant toxicity to various aquatic organisms such as duckweed (Brain et al., 2006), rainbow trout, and zebrafish in both free acid and lactone forms (Ellesat et al., 2010; Ellesat et al., 2012). The fate and TPs of ATV during a photolysis experiment

have been investigated in the past in river water conditions (Wang et al., 2018a). There was, however, a discrepancy between the stable TPs found in the study and the observed metabolites in wastewaters and surface water.

Bezafibrate (BEZ) is a fibrate widely used and prescribed in the treatment of cardiovascular events. It is highly stable in wastewater (Lin et al., 2021) and has shown toxicity toward both freshwater and saltwater organisms (Duarte et al., 2019). There are multiple studies on the removal of BEZ by advanced oxidation techniques (Dantas et al., 2007; Sui et al., 2016; Gallardo-Altamirano et al., 2021). However, there is not as much information on the TPs of BEZ generated in surface waters.

Oxybenzone (OXZ) is a UV filter used in sunscreen. It has been linked to coral reef bleaching and has been found in various species of fish (DiNardo and Downs, 2018; Schneider and Lim, 2019). Finally, TBEP is a polymer additive used as a plasticizer and a flame retardant. TBEP is a ubiquitous high production volume chemical that has been detected in surface water and wastewater [154], drinking water (Lee et al., 2016), human breast milk (Kim et al., 2014) and urine (Ingle et al., 2020). Although the photocatalytic degradation of TBEP has been investigated (Konstas et al., 2019), its fate in surface water is still largely unknown to our knowledge.

The objective of this paper was to apply a systematic bottom-up approach based on controlled laboratory experiments under environmental conditions to study the photolysis of ATV, BEZ, OXZ, and TBEP, confirm the environmental occurrence of their identified TPs and then compare the predicted toxicity of parent compounds and their TPs.

6.4 Material and methods

6.4.1 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). Atorvastatin (ATV) calcium (certified reference material), bezafibrate (BEZ) (>98.5%), oxybenzone (OXZ) (certified reference material), and tris(2-bytoxyethyl)phosphate (TBEP) (>95%) were all obtained from Millipore Sigma (St-Louis, MO, USA).

6.4.2 Photolysis experiments

The physical properties of ATV, BEZ, OXZ, and TBEP are shown in **Table 22**. For the photolysis experiments, dechlorinated tap water was spiked with 5 mg L^{-1} Pahokee peat humic substances purchased from the International Humic Substances Society (IHSS) (Denver, Colorado) to mimic river surface water conditions. An air bubbling system (Optima Pump 1000 cm^3 min^{-1}, 4 psi) was used to keep natural aerobic conditions over the exposure period which was 24 days for ATV and BEZ and 25 days for OXZ and TBEP. The selected compounds were spiked with 200 ng mL^{-1} for ATV, 50 ng mL^{-1} for BEZ and TBEP, and 500 ng mL^{-1} for OXZ in 2 L-glass Erlenmeyer flasks. The concentrations of the compounds were adjusted based on signal intensity with the mass spectrometer. Compounds with a lower concentration to signal ratio like OXZ and ATV were spiked in higher concentration to get a higher intensity to facilitate the structural elucidation of the TPs. Large volume samples (2 liters) with a single solution per compound. A blank solution with the same humic substances and exposure parameters was also prepared for quality control. Each batch of analysis, an instrumental blank was also prepared. These spiked water samples were exposed to artificial sunlight using an Exo Terra Solar Glo 125 W lamp at a distance of 50 cm (**Figure 49**). More information on the lamp and parameters of exposure is given in the Supplementary material.

6.4.2.1 Extraction

Water samples were extracted in duplicates at the beginning of the experiment and after 1, 2, 3, 7, 9, 16 and 24 days for ATV and BEZ and after 1, 2, 3, 7, 9, 16 and 25 days for OXZ and TBEP. Prior to the extractions, the solution was homogenised with a Teflon coated magnetic stir bar as to ensure the concentrations stayed the same throughout the experiment. The extraction was performed with Oasis HLB or Oasis MAX solid phase extraction cartridges made by Waters (Milford, MA). BEZ and ATV were tested with both HLB and MAX cartridges considering their acid moieties. The MAX had the highest yield of extraction in both instances and were selected for the experiment. For OXZ and TBEP, the HLB had the highest yield of extraction. Briefly, for ATV and BEZ, a first extraction was done for the neutral compounds with MeOH. The acidic compounds were extracted with 5% formic acid (FA) in acetonitrile (ACN) subsequently. For

OXZ and TBEP, a single extraction was performed with 2% FA in ACN. Details about the extraction parameters are shown in **Table 23** (Supplementary material).

6.4.3 Collection and preparation of samples for the nontargeted screening

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. More information about the collection and preparation of samples is available in the supplementary information.

6.4.4 Instruments and methods

6.4.4.1 Identification of phototolysis products

Reversed-phase ultra high-performance liquid chromatography (UHPLC) coupled by electrospray ionization in the positive mode to a quadrupole time-of-flight mass spectrometer (QqTOFMS) was used for the identification of the photolysis products. The QqTOFMS was mass calibrated with a sodium formate solution in high precision calibration mode after waiting 30 min for the system to stabilize. Mass drift was monitored, and all analyses were conducted within 4 hours of the calibration. Internal mass calibration was not employed. Full width at half-maximum mass resolution (R_{FWHM}) at *m/z* 337 was about 24 000. The parameters and detailed information pertaining to the UHPLC system are available in the Supplementary Information.

6.4.4.2 Nontargeted screening of surface waters

Mass spectrometry coupled with reverse phase ultra high-performance liquid chromatography (UHPLC) was used for the nontargeted screening of surface waters. For the mass spectrometry, the ion source was a pneumatic assisted heated electrospray ion source at its parameters were the following: positive mode, capillary temperature was 275 ◦C; sheath gas was 45; auxiliary gas was 10; and spray voltage was 3800 V. Data dependent acquisition (DDA) was used for detection with one MS^1 survey scan (*m/z* 100–1000) acquired at R_{FWHM} = 35 000 and precursor ions meeting user-defined criteria for monoisotopic precursor intensity (dynamic acquisition of MS^2 based on the 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). Fragment ions were detected in the Orbitrap at R_{FWHM} = 17 500. Instrument calibration was performed prior to all analyses and mass accuracy was notably below 1 ppm using Thermo Pierce calibration solution and the automated instrument protocol. The calibration mixture was composed of caffeine, n-butylamine, the tetrapeptide MRFA, and Ultramark 1621, a mixture of fluorinated phosphazenes, in an ACN-MeOH-acetic acid solution. The parameters and detailed information pertaining to the UHPLC system and parameters are available in the supplementary information.

6.4.5 Sofware parameters and data management

Files were converted into mzML with MSConvert from ProteoWizard (Kessner et al., 2008). Data treatment was effectuated with R (version 4.04) in the RStudio interface (version 1.4.1106). The packages used were part of the Bioconductor software (Gentleman et al., 2004; Huber et al., 2015). The XCMS package (version 3.13) was used for peak integration and peak grouping (Smith et al., 2006; Tautenhahn et al., 2008; Benton et al., 2010). The CAMERA package (version 3.13) was used for statistical tests and isotope grouping (Kuhl et al., 2012). Volcano plots were plotted with Origin 2021.

Compounds that were extracted with the same SPE cartridges - ATV/ BEZ pair and the OXZ/TBEP one - were compared in pairwise analysis. Replicates of a same compound from a specific day of exposure were pooled in a group while the replicates from the other compound with the same day of exposure formed the other group. Peaks were then integrated and merged if their *m/z* were within ±10 ppm and their retention times differed by less than 20 s to avoid multiple features for a single compound. The peak list was submitted to a Welch t-test where the most relevant compounds in terms of p value and fold were selected. This allowed prioritizing transformation products of the compounds instead of humic acid transformation products.

The molecular formulas of all TPs were verified by measuring their spectral accuracy using MassWorks software (version 4.0) (Wang and Gu, 2010) from Cerno Bioscience (Las Vegas, NV). Spectral accuracy evaluates the similarity between experimental and calculated isotopic patterns of candidate molecular formulas for an ion as a percentage (Eysseric et al., 2017a). Therefore, high spectral accuracy (>98%) indicates a high degree of correspondence between a compound and its

experimental isotopic pattern, thus suggesting that the correct molecular formula for a given *m/z* was assigned. This tool has been used to confirm molecular formulas in complex matrices (Zhou et al., 2011; Ochiai et al., 2012; Eysseric et al., 2017b; Eysseric et al., 2021). To generate the molecular formulas, the following were used: mass tolerance was set to 5 ppm**,** empirical rules were used for the ratio of atoms in a formula according to the molecular weight; the atoms in the molecular formula had to be plausible in regard with formula of the parent compounds (C, H, N, O, P, S, Cl).

Proposed molecular structures of TPs were not confirmed with standards and they were assigned according to the changers observed in their tandem mass (MS^2) spectra compared to the parent compound. Thus, structures of all observed TPs were considered as probable structures (level 2b) according to the identification confidence scale proposed by Schymanski et al. (2014).

The software Ecological Structure Activity Relationships Predictive Model (ECOSAR, version 2.0) from the US EPA was used to generate toxicity values for the TPs that were identified in environmental samples. Briefly, the software predicts effect levels on specific exposure duration for the predicted effect levels with endpoints (LC_{50}, EC_{50} or Chronic value) for the effect level for a given compound. The predicted effects are based on "classes" that are moiety each carrying a different predicted effect level. Values are given for the following organisms: daphnid, fish in fresh and salt water, green algae in fresh and salt water, *Lemna gibba*, and mysid in fresh and salt water. In all cases, the most conservative values (i.e., the lowest ones) were taken for each compound. ECOSAR was selected because of its accuracy compared to other predictive models according to a benchmarking study (Melnikov et al., 2016), ease of use and free access.

6.5 Results and discussion

6.5.1 Photodegradation experiments

6.5.1.1 Atorvastatin

Photodegradation experiments for ATV show complete degradation by day 7 (**Figure 18a**). The most statistically significant features at the 24th day of exposure in terms of *p*-value and fold change

were prioritized as potential stable transformation products and can be seen in the plots of ATV for the acid (**Figure 19a**) and neutral (**Figure 19b**) fractions.

ATV_TP557a (**Figure 18b**), that began to supersede ATV in peak area by day 3, appears to be the result of a hydroxylation of ATV lactone, a form that ATV partly transforms into when in aqueous solution (Hermann et al., 2005; Lee et al., 2009). It should also be noted that at day 0, ATV_TP557a had a non-zero peak area which means that it was already formed prior to the photolysis experiment. The site of the hydroxylation is located in the phenyl moiety as can be seen in the MS^2 spectrum of the compound (**Figure 20a**). The location of the hydroxylation site is consistent to what was observed in photodegradation settings (Wang et al., 2018a), but different from hydroxylated human metabolites of ATV where it is found in the phenyl amide (Hermann et al., 2005). Furthermore, spectral accuracy was used to confirm the molecular formula of ATV_TP557a ($C_{33}H_{33}FN_2O_5$) that corresponds to the proposed structure (**Table 9**). ATV_TP557a had the most abundant signal of all ATV TPs at the 24^{th} day of light exposure. The mass spectra in **Figure 20a-c** do not rule out the possibility that the hydroxylation is located on the fluorophenyl group. Hydroxylation of monosubstituted fluorobenzenes mediated by P450 enzymes has been reported (Koerts et al., 1997) and such reaction has actually more favorable kinetics compared to benzene hydroxylation (Burka et al., 1983). However, to the authors knowledge, fluorophenol TPs or metabolites have not been reported for ATV (Hermann et al., 2005; Park et al., 2008; Lee et al., 2009; Wang et al., 2018a)**.**

ATV_TP575 is potentially a hydroxylated TP of ATV. As with ATV_TP557a, the site of hydroxylation is located on the phenyl moiety (**Figure 20b**) and not on the phenyl amide which is also the case for human metabolites of ATV (Hermann et al., 2005). The molecular formula of the compound was unequivocally confirmed with spectral accuracy (**Table 9**). ATV_TP575 had the second most abundant signal of all of ATV TPs and shown to be persistent over the course of the experiment.

ATV_TP555, whose molecular formula ($C_{33}H_{31}FN_2O_5$) was also confirmed by spectral accuracy (94%) (**Table 9**), shows a net loss of H_2O compared to ATV_TP573 ($C_{33}H_{33}FN_2O_6$), the product of another hydroxylation on the phenyl ring of ATV_TP557a (**Figure 18b**) as seen in the MS^2

spectra of these two compounds (**Figure 50a**). ATV_TP555 could be the result of the oxidation of the alcohol moiety on the lactone or a rearrangement following the loss of H_2O by ATV_TP573. The peak areas of both ATV_TP555 and ATV_TP573 increased after a sharp decrease of peak area by ATV_TP557a which could suggest that they are both TPs of the latter.

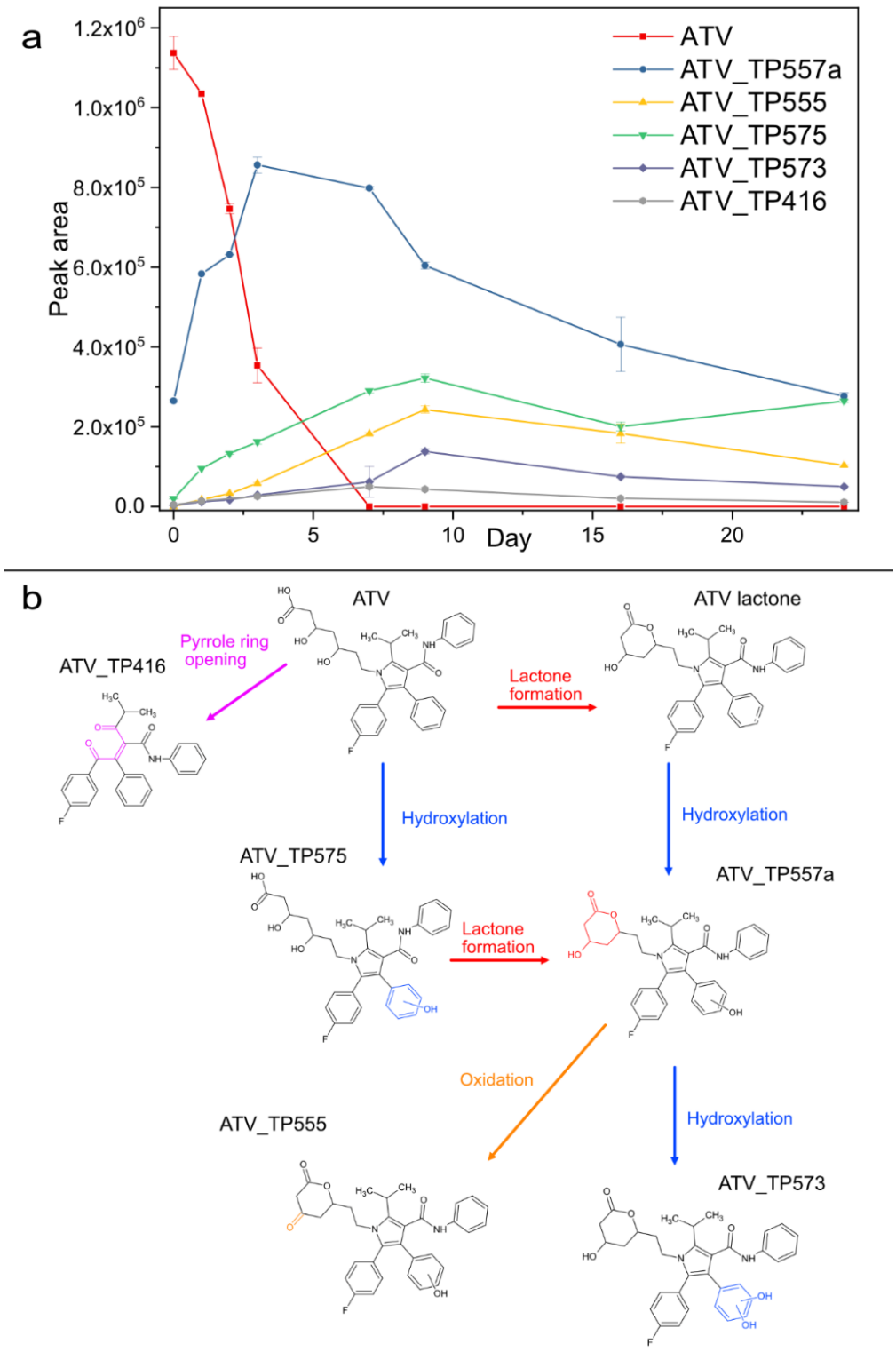


Figure 18. Photodegradation kinetics (a) and proposed degradation pathways and structures (b) for atorvastatin (ATV) and its transformation products.

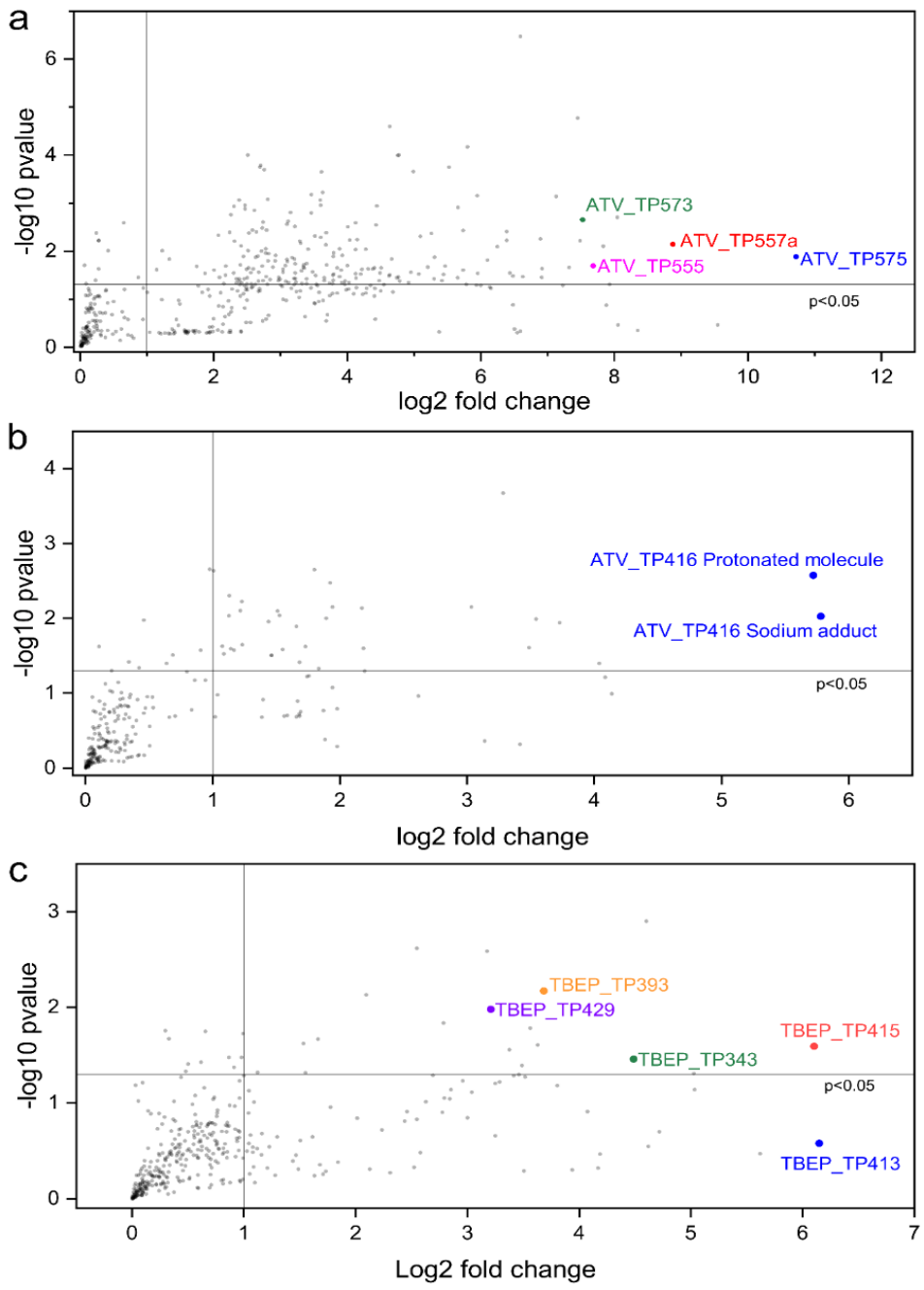


Figure 19. Plots of -log10 p-value vs log2 fold change for (a) ATV with the anionic exchange cartridge after the 24th day of exposure, (b) ATV with the neutral cartridge after the 24th day of exposure and (c) TBEP after the 25th day of exposure. The main transformation products are labeled. These transformation products had the highest fold changes while being statistically significant. The ions with lower fold changes but also lower p-values were, in both instances, either

the M+1 and M+2 isotopes or sodium adducts or protonated molecules and thus referring to the same transformation products.

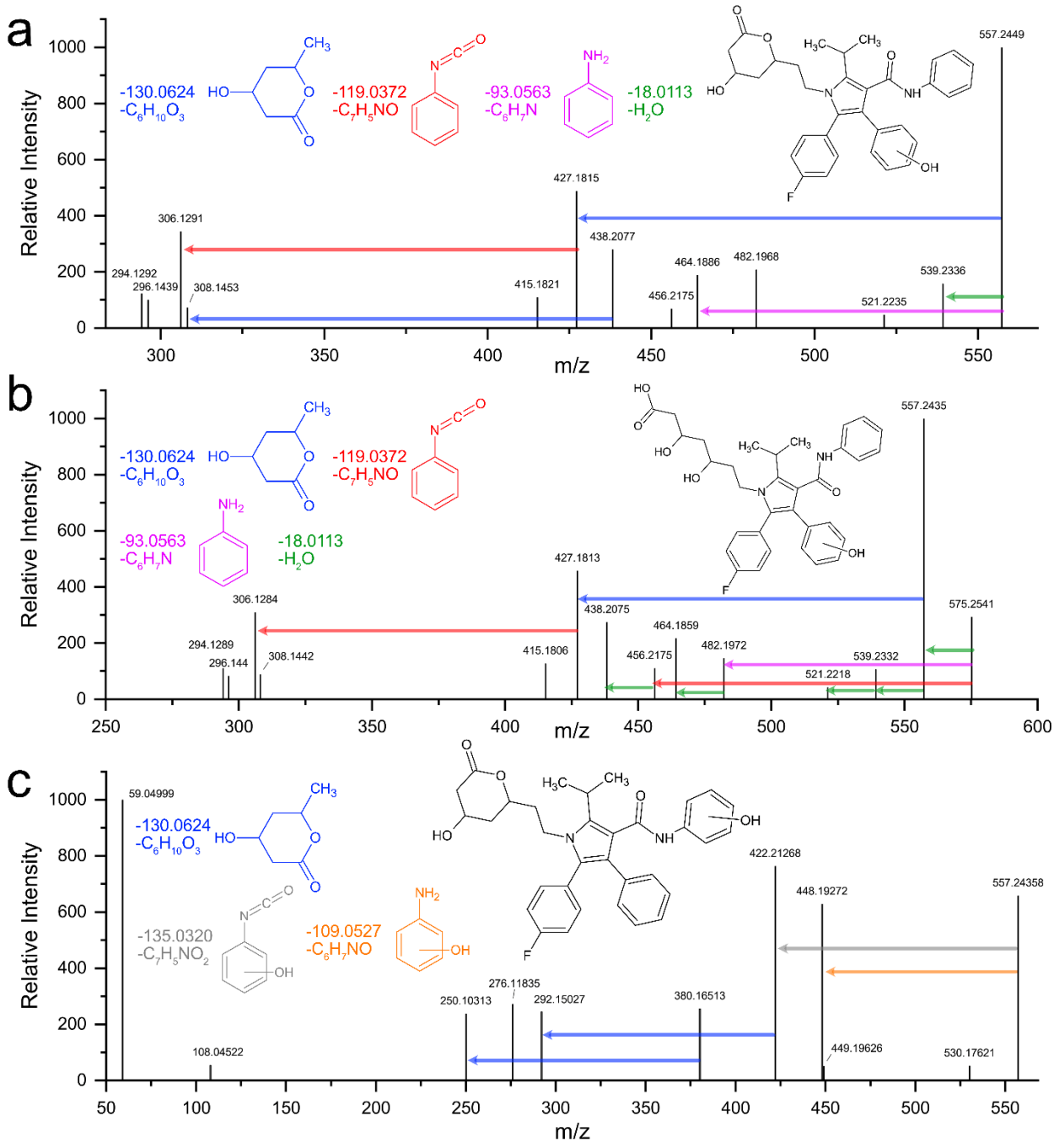


Figure 20. MS^2 spectra of ATV_TP557a **(a)**, ATV_TP575 **(b)**, and ATV_TP557b **(c)**. ATV_TP557b was observed only in the nontarget screening analysis. In **(a)** and **(b)**, the hydroxylation does not occur on the phenylamide moiety as the mass shift corresponding to the loss of an aniline (fragment in purple, 93 Da) and a phenylamide (fragment in red, 119 Da) are

observed. In **(c)**, the mass shift corresponding to hydroxy-aniline and hydroxy-phenylamide are observed indicating the site of hydroxylation is located in the latter moiety. However, the exact location on the ring remains uncertain.

Table 9. Formula ranks and spectral accuracies of the studied transformation products in the photodegradation and river samples.

Compound	Molecular formula	Photodegradation samples			River samples		
		Mass accuracy (mDa)	Spectral accuracy (%)	Formula rank	Mass accuracy (mDa)	Spectral accuracy (%)	Formula rank
ATV_TP557a	$C_{33}H_{33}FN_2O_5$	-0.100	99.2	1	NA	NA	NA
ATV_TP555	$C_{33}H_{31}FN_2O_5$	0.023	94.0	1	NA	NA	NA
ATV_TP575	$C_{33}H_{35}FN_2O_6$	0.009	99.6	1	-0.292	83.7	1
ATV_TP573	$C_{33}H_{33}FN_2O_6$	0.359	99.0	1	NA	NA	NA
ATV_TP416	$C_{26}H_{22}FNO_3$	-1.100	96.7	1	NA	NA	NA
ATV_TP557b	$C_{33}H_{33}FN_2O_5$	NA	NA	NA	-0.350	94.3	1
TBEP_TP299	$C_{12}H_{27}O_6P$	-0.900	99.4	1	0.300	67.9	1
TBEP_TP343	$C_{14}H_{31}O_7P$	-0.500	98.6	1	-0.186	99.3	1
TBEP_TP371	$C_{15}H_{31}O_8P$	0.569	93.6	1	0.267	42.8	1
TBEP_TP413	$C_{18}H_{37}O_8P$	-0.526	76.9	1	-0.152	84.1	1
TBEP_TP415	$C_{18}H_{39}O_8P$	-0.276	99.2	1	-1.056	73.0	1

TBEP_TP 429	$C_{18}H_{37}O_9P$	-0.340	81.7	1	-0.875	78.2	1

ND: not detected.

The proposed structure of ATV_TP573 is consistent to what was proposed in a previous study on ATV degradation (Wang et al., 2018a) and the MS^2 spectrum (**Figure 50b**). What appears to be a case of pyrrole ring opening was observed with ATV_TP416 (**Figure 18b**). This is supported by the fact that ATV_TP416 was detected in the neutral fraction of the SPE cartridge and its MS^2 spectrum does not exhibit the loss of the lactone unlike all the other TPs (**Figure 50c**). An explanation of the MS^2 spectra of ATV_TP555, ATV_TP573 and ATV_TP416 is shown in the supplementary information. At the end, ATV_TP557a and ATV_TP575 had the most abundant signals and the trends of their peak area showed that they were the most stable TPs (**Figure 18a**) with the highest fold change at day 24 (**Figure 19a**).

6.5.1.2 Bezafibrate and oxybenzone

Bezafibrate (BEZ) proved to be highly resistant to light exposure as the average peak area at the 24^{th} day was 56.5% of its value at initial time (**Figure 51a**). As can also be seen in the section of the log10 p-value vs log2 fold change plot corresponding to BEZ (**Figure 52a**), the parent compound remained by far the most significant feature. This would tend to confirm the previous studies on the stability of BEZ in surface waters and wastewaters (Lin et al., 2021).

For oxybenzone (OXZ), a reduction of over 96% in peak area was observed by day 1 and over 99% at the 25th day compared to the initial time (**Figure 51b**). However, no statistically significant TP in terms of *p*-value ($p<0.05$) and fold change (fold change >3) in the -log10 p-value vs log2 fold change plot (**Figure 52b**) was observed. The lack of transformation product for OXZ could be attributed to the cartridges and the ionization mode used in the experiments.

6.5.1.3 Tris(2-bytoxyethyl) phosphate

Tris(2-bytoxyethyl) phosphate (TBEP) lost 66% of its peak area in the first day of exposure after which it slowly degraded and ended with a 94% loss in peak area by the 25th day (**Figure 21a**). The most significant compounds at the 25th day of exposure were TBEP_TP393, TBEP_TP415, TBEP_TP343, TBEP_TP413 and TBEP_TP429 (**Figure 19c**, **Figure 21a**).

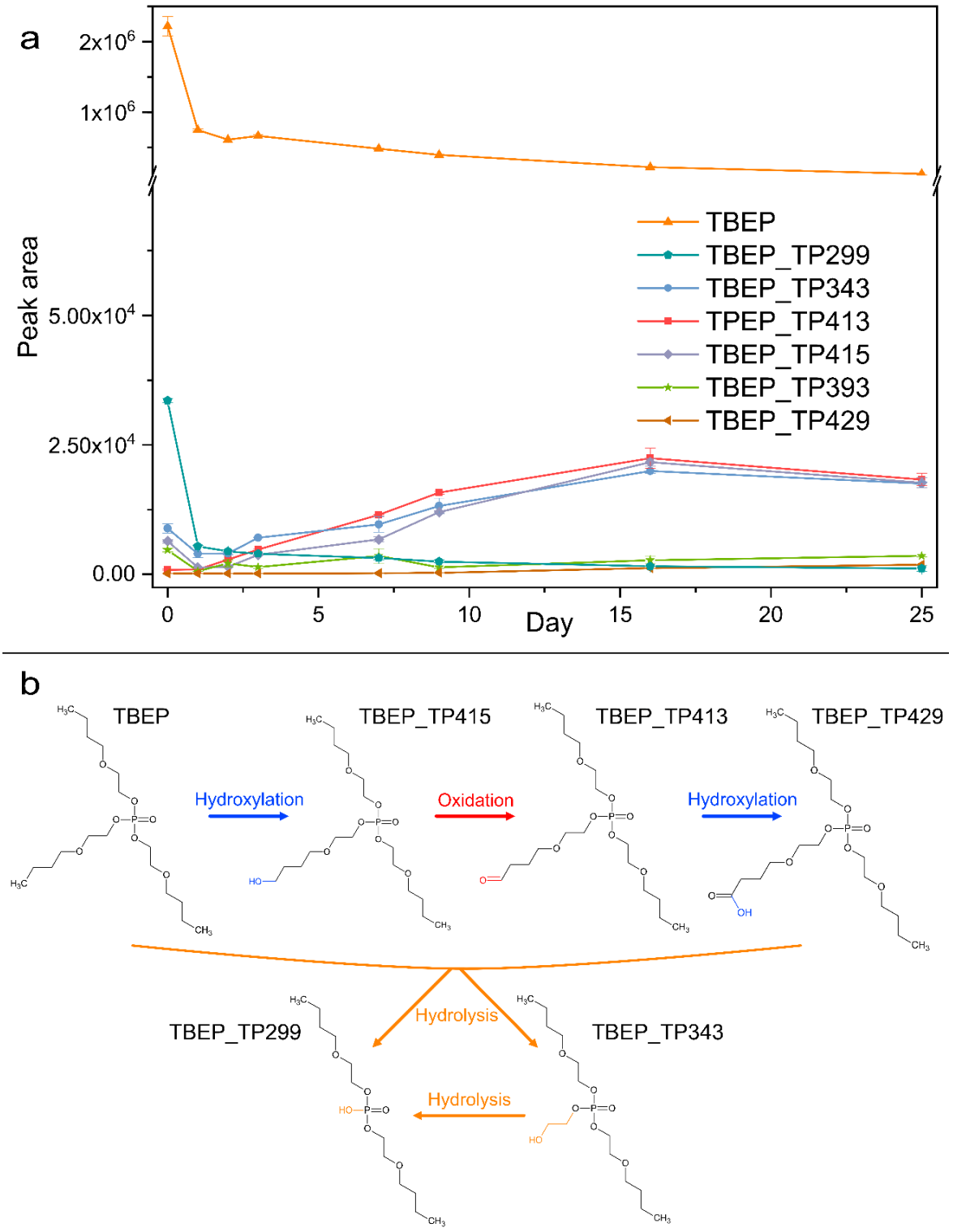


Figure 21. Photodegradation kinetics (a) and proposed degradation pathways and structures (b) for Tris(2-bytoxyethyl) phosphate (TBEP) and its transformation products.

The transformation products TBEP_TP413, TBEP_TP415, and TBEP_TP429 are named after the *m/z* of their protonated molecules for nomenclature consistency with the other transformation products but were more intense in their Na^+ adduct form with *m/z* of 435.2113, 437.2272, 451.2064, respectively, as can be seen in **Table 9**.

TBEP_TP415 results from the hydroxylation of a methyl group on the end of the aliphatic chain of TBEP. Then, TBEP_TP413 is the result of the oxidation of the newly formed alcohol into an aldehyde. Thereupon, the aldehyde of TBEP_TP413 is hydroxylated resulting in the acid TP TBEP_TP429 (**Figure 21b**). The molecular formula of these TBEP transformation products was confirmed by measuring their spectral accuracy (**Table 9**). Two additional TPs resulting from the hydrolysis of observed: TBEP_TP343 and TBEP_TP299. TBEP_TP343 shows a loss of 56 Da compared to TBEP corresponding to the net loss of one of the outer n-butyl moieties (**Figure 21b**) while the net loss of one butoxyethyl moiety was observed in the case of TBEP_TP299. The molecular formula of both TBEP_TP343 ($C_{14}H_{31}O_7P$) and TBEP_TP299 ($C_{12}H_{27}O_6P$) was also confirmed with spectral accuracy (**Table 9**). The degradation kinetics of TBEP_TP415, TBEP_TP413 and TBEP_TP343 show that these three transformation products had the most abundant ions and were the most stable at the 25th day of exposure. TBEP_TP429 constantly rose in peak area despite much lower values as can be seen in the relative photodegradation kinetics of TBEP (**Figure 53**). TBEP_TP299 had similar photodegradation kinetics to TBEP as the peak area decreased at an even faster rate than TBEP. This could suggest that it was an impurity of the latter, but not the result of an in-source fragmentation since the observed retention times of TBEP and TBEP_299 were different. TBEP_TP393 had *roller coaster-like* degradation kinetics as can be seen in the relative photodegradation kinetics of TBEP (**Figure 53**) which means it could be an impurity or an intermediate that is continuously produced and transformed into other TPs.

The structures that were observed for the TPs of TBEP share some similitudes with a previous study on the photocatalytic degradation of TBEP in which TPs with the same molecular formula as TBEP_TP415, TBEP_TP413, TBEP_TP343, TBEP_TP299 were tentatively identified (Konstas et al., 2019). Hydroxylation and oxidation were the most observed reactions and formed stable TPs in the cases of ATV and TBEP which were to be expected considering the oxidizing environment. Assuming these oxidizing conditions with light exposure are also found on the

surface of the water in rivers, the stable TPs could also be found in real samples of surface waters. As such, these compounds were then targeted as suspects in the following nontargeted screening study.

6.5.2 Nontargeted screening

The results of the nontargeted screening pertaining to compounds other than those related to this article are currently in press. Features corresponding to ATV_TP557a and ATV_TP575 were detected while no signal of ATV itself was observed. Upon further examination, however, the compound detected in the river samples and ATV_TP557a were different. This new compound, which was named ATV_TP557b, is a positional isomer of ATV_TP557a, observed during the photodegradation experiment. Indeed, the hydroxylation site is located on the phenyl amide moiety rather than the phenyl as can be seen in the MS^2 spectrum (**Figure 20c**). ATV_TP557b has a level of confidence of 3 (Schymanski et al., 2014) considering the exact location of the hydroxylation on the phenyl amide is uncertain. Ortho and para hydroxy-ATV, in both free acid lactone forms, have been recorded as human metabolites of ATV (Hermann et al., 2005) but never in surface waters to our knowledge. Furthermore, the molecular formula of ATV_TP557b ($C_{33}H_{33}FN_2O_5$) was unequivocally confirmed based on spectral accuracy (94.3%). What allowed to differentiate both compounds and tentatively identify ATV_TP557b was that the acquisition was set to the DDA mode, which automatically generates product ion scans in a nontargeted fashion. Regarding the presence of ATV_TP557b and the absence of ATV_TP557a, it could be explained by the fact that, because ATV is metabolized at an over 50% rate in humans (Hermann et al., 2005), the pathway in which ATV is hydroxylated on the phenyl moiety might be missing. This could explain the discrepancy observed between controlled photolysis experiments and natural photolysis. The presence of ATV_TP575 appeared to be detected, but it was unfortunately not abundant enough to be selected in the quadrupole during the DDA experiment and thus lacks an MS^2 product ion scan spectrum. However, the molecular formula ($C_{33}H_{35}FN_2O_6$) of this TP was unequivocally confirmed with spectral accuracy (83.7%) which gives the tentative TP a level of confidence of 4 (**Table 9**). As with ATV_TP557b, it is probable that the hydroxylation on ATV_TP575 occurred on the phenyl amide rather than the phenyl one.

Regarding BEZ, it was not detected in the nontargeted screening experiments while OXZ was observed. The presence of OXZ was confirmed with a reference standard.

TBEP as well as all 5 TPs reported in this paper were detected in the nontargeted screening. All compounds were abundant enough to be selected in the quadrupole during the DDA experiment which allowed to identify them tentatively at the level of confidence 2b for the TPs while TBEP was confirmed with a reference standard. The MS^2 spectrum indicates that the hydroxylation site on TBEP_TP415 is located on the n-butyl moiety of the aliphatic chain as are the oxidation for TBEP_TP413 and that an additional hydroxylation occurs to form a carboxylic acid function in TBEP_TP429. The MS^2 spectra also suggests that TBEP_TP299 and TBEP_TP343 are hydrolysis reaction products as can be seen by the net loss of an n-butoxyethyl and an n-butyl, respectively (**Figure 22**).

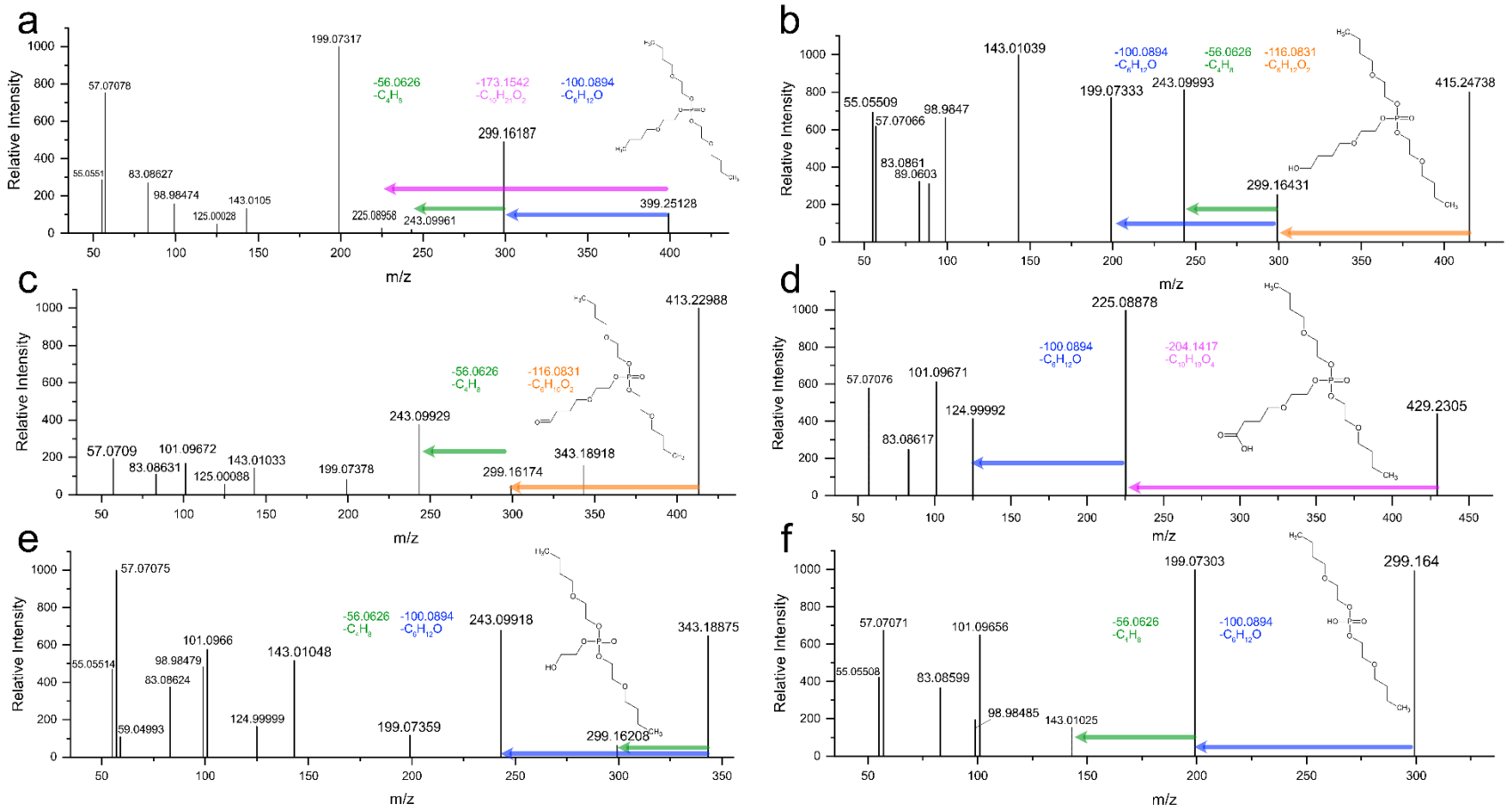


Figure 22. MS^2 spectra for **(a)** TBEP, **(b)** TBEP_TP415, **(c)** TBEP_TP413, **(d)** TBEP_TP429, **(e)** TBEP_TP343, and **(f)** TBEP_TP299.

The number of transformation products that were tentatively identified and then detected highlights the dynamic fate of ATV and TBEP in surface waters and the usefulness of studies in controlled settings like this one. Performing a data-dependent nontargeted analysis with a high-resolution mass spectrometer also proved to be highly fruitful as we would not have been able to detect ATV_TP557b with a targeted analysis since the selected transition would have been different. Since the analysis of the TPs in the photolysis experiments and in the nontargeted screening was performed with two different instruments in different matrices, a semi-quantitative comparison of the peak areas is impossible.

6.5.3 Predicted effect levels with ECOSAR

Given the presence of ATV and TBEP TPs in surface waters, the publicly available QSAR program ECOSAR by the US EPA was used to generate their predicted effect levels (**Table 10**). For ATV_TP557b, the structures given as input corresponded to hydroxylation in *ortho* and *para* of the phenyl amide moiety based on known ATV metabolites in the human body (Hermann et al., 2005; Ellesat et al., 2010). In all organisms, the *ortho* hydroxylated ATV_TP57b showed higher predicted toxicity than ATV (or lower predicted effect levels) as shown in **Table 10**. The *para* hydroxylated ATV_TP557b also showed higher toxicity than ATV on fish and green algae in fresh and salt water, but in a lesser extent than the *ortho* one. In the case of ATV_TP575, we supposed that the hydroxylation sites were the same as in ATV_TP557b. The *ortho*-hydroxylated structure had a higher predicted toxicity than the *para* one, but both free acid structures had lower predicted toxicity vales than their lactone counterparts. These predictions are in line with previous empirical assays on rainbow trouts showing that ATV exhibits higher toxicity in lactone form of than free acid (Ellesat et al., 2010). None of the TPs of TBEP showed higher predicted toxicity than the parent compound which suggests an empirical toxicity assay of these compounds might not be needed. These predicted values need to be confirmed with empirical assays using relevant species for a clearer picture of their potential toxicity level.

Table 10. Predicted effect levels in terms of chronic values and LC50 on different organisms in fresh and salt water for ATV, TBEP and their respective transformation products

Organism	Duration	End Point	Predicted effect level of toxicity (mg L^{-1})										
			ATV	*o*-OH-ATV lactone (ATV_TP557b)	*p*-OH-ATV lactone (ATV_TP557b)	*o*-OH-ATV (ATV_TP575)	*p*-OH-ATV (ATV_TP575)	TBEP	TBEP_TP415	TBEP_TP413	TBEP_TP429	TBEP_TP343	TBEP_TP299
Daphnid	NA	ChV	15×10^{-3}	0.30×10^{-3}	7.4×10^{-3}	20×10^{-3}	49×10^{-3}	6.7×10^{-3}	25×10^{-3}	26×10^{-3}	210×10^{-3}	39×10^{-3}	21
Fish	NA	ChV	15×10^{-3}	0.60×10^{-3}	9.3×10^{-3}	23×10^{-3}	75×10^{-3}	14×10^{-3}	23×10^{-3}	23×10^{-3}	220×10^{-3}	23×10^{-3}	40
Fish (SW)	96h	LC_{50}	1.2	55×10^{-3}	190×10^{-3}	310×10^{-3}	1.1	79×10^{-3}	220×10^{-3}	220×10^{-3}	1.93	290×10^{-3}	49
Green Algae	NA	ChV	190×10^{-3}	17×10^{-3}	37×10^{-3}	110×10^{-3}	300×10^{-3}	3.36	23§	8.39	180§	47	39
Green Algae (SW)	96h	LC_{50}	NA	0.41×10^{-3}	3.4×10^{-3}	1.5×10^{-3}	13×10^{-3}	NA	NA	NA	NA	NA	NA
Lemna gibba	7d	EC_{50}	NA	31×10^{-3}	110×10^{-3}	170×10^{-3}	620×10^{-3}	NA	NA	NA	NA	NA	NA
Mysid	96h	LC_{50}	NA	73×10^{-3}	250×10^{-3}	NA	NA	12	200§	210§	1400§	640	470§
Mysid (SW)	NA	ChV	0.68×10^{-3}	2.0×10^{-3}	0.86×10^{-3}	1.0×10^{-3}	4.4×10^{-3}	0.15×10^{-3}	0.50×10^{-3}	0.50×10^{-3}	4.3×10^{-3}	0.72×10^{-4}	44

SW: salt water; ChV: chronic value; NA: not available. Ortho and para hydroxyATV lactone are the two possible structures of ATV_TP557b. Ortho and para hydroxyATV are two potential structures of ATV_TP575. §: Indicates instances where the chemical may not be soluble enough to measure this predicted effect.

6.6 Conclusion

This study used a vertical bottom-up approach starting from compound selection, then to a laboratory-controlled photolysis experiment to study the fate and identify the TPs of the selected compounds, followed by a nontargeted analysis of river samples to look for the aforementioned TPs to end with a simulation to generate predicted effect levels for the TPs that were detected in the nontargeted screening. The obtained results demonstrate that laboratory degradation studies are crucial for uncovering the structures of new transformation products of contaminants of concern in controlled settings. Tentative identification of 5 stable TPs of TBEP was possible because the controlled settings of the experiment allowed us to use statistical tools to prioritize the most significant features. This prioritization would not have been possible in a nontargeted analysis and these compounds would thus have stayed unresolved. However, confirmation of the environmental occurrence of TPs by nontargeted analysis is paramount. Indeed, the compound ATV_TP557b would not have been detected nor tentatively identified had we performed a targeted analysis of ATV_TP557a. This also exemplifies the importance of performing data-dependent or independent acquisitions to gather MS^2 information for higher confidence in the identification. Finally, the ECOSAR software indicated that there might be a cause for concern in the case of ATV_TP557b whose predicted effect levels are lower than the parent compounds while the concern seems lower for the TPs of TBEP. As data-analysis algorithms and software like molecular networking continue to develop, more "top-down" approaches might be seen in environmental studies, but the complexity of identifying unknown transformation products in environmental samples is currently too high for current methods to be relied upon. The TPs discovered in bottom-up approaches can be searched in retrospective analysis of nontargeted "top-down" workflows for data-mining purposes. The bottom-up approach used in this study is valuable for monitoring and risk assessment programs to get a better idea of the different forms of contaminants in the water and how concerning it is.

6.6.1 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. 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 Lou Paris from OBV

Yamaska and Clémence Soulié for assistance with sample collection. Also, we thank Audrey Roy-Lachepelle and Sylvie Poirier-Larabie from Environment and Climate Change Canada for their help with the photodegradation experimental setup.

6.7 Credit author statement

Emmanuel Eysseric: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing

Christian Gagnon: Funding Acquisition, Resources, Resources, Supervision, Writing – Review & Editing

Pedro A. Segura: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Writing – Review & Editing, Project administration

6.8 References

Benton, H.P., Want, E.J., Ebbels, T.M., 2010. Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. Bioinformatics 26, 2488-2489.

Bletsou, A.A., Jeon, J., Hollender, J., Archontaki, E., Thomaidis, N.S., 2015. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. TrAC, Trends Anal. Chem. 66, 32-44.

Brain, R.A., Reitsma, T.S., Lissemore, L.I., Bestari, K., Sibley, P.K., Solomon, K.R., 2006. Herbicidal effects of statin pharmaceuticals in Lemna gibba. Environ. Sci. Technol. 40, 5116-5123.

Burka, L.T., Plucinski, T.M., Macdonald, T.L., 1983. Mechanisms of hydroxylation by cytochrome P-450: metabolism of monohalobenzenes by phenobarbital-induced microsomes. P Natl Acad Sci USA 80, 6680-6684.

Dantas, R.F., Canterino, M., Marotta, R., Sans, C., Esplugas, S., Andreozzi, R., 2007. Bezafibrate removal by means of ozonation: Primary intermediates, kinetics, and toxicity assessment. Water Res. 41, 2525-2532.

Daughton, C.G., Ternes, T.A., 1999. Pharmaceuticals and personal care products in the environment: agents of subtle change? Environ. Health Perspect. 107, 907-938.

DiNardo, J.C., Downs, C.A., 2018. Dermatological and environmental toxicological impact of the sunscreen ingredient oxybenzone/benzophenone-3. J. Cosmet. Dermatol. 17, 15-19.

Duarte, B., Prata, D., Matos, A.R., Cabrita, M.T., Caçador, I., Marques, J.C., Cabral, H.N., Reis-Santos, P., Fonseca, V.F., 2019. Ecotoxicity of the lipid-lowering drug bezafibrate on the bioenergetics and lipid metabolism of the diatom Phaeodactylum tricornutum. Sci. Total Environ. 650, 2085-2094.

Ellesat, K.S., Holth, T.F., Wojewodzic, M.W., Hylland, K., 2012. Atorvastatin up-regulate toxicologically relevant genes in rainbow trout gills. Ecotoxicol. 21, 1841-1856.

Ellesat, K.S., Tollefsen, K.-E., Åsberg, A., Thomas, K.V., Hylland, K., 2010. Cytotoxicity of atorvastatin and simvastatin on primary rainbow trout (Oncorhynchus mykiss) hepatocytes. Toxicol. in vitro 24, 1610-1618.

Eysseric, E., Barry, K., Beaudry, F., Houde, M., Gagnon, C., Segura, P.A., 2017a. Application of spectral accuracy to improve the identification of organic compounds in environmental analysis. Anal. Chem. 89, 9805–9813.

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.

Eysseric, E., Ghislain, T., Duret, X., Lalonde, O., Segura, P.A., Lavoie, J.-M., 2017b. Effect of steam treatments on the availability of various families of secondary metabolites extracted from green sweet sorghum. Ind. Crops Prod. 104, 120-128.

Fatta-Kassinos, D., Vasquez, M., Kümmerer, K., 2011. Transformation products of pharmaceuticals in surface waters and wastewater formed during photolysis and advanced oxidation processes–degradation, elucidation of byproducts and assessment of their biological potency. Chemosphere 85, 693-709.

Gallardo-Altamirano, M., Maza-Márquez, P., Pérez, S., Rodelas, B., Pozo, C., Osorio, F., 2021. Fate of pharmaceutically active compounds in a pilot-scale A2O integrated fixed-film activated sludge (IFAS) process treating municipal wastewater. J. Environ. Chem. Eng. 9, 105398.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, 1-16.

Henning, N., Falås, P., Castronovo, S., Jewell, K.S., Bester, K., Ternes, T.A., Wick, A., 2019. Biological transformation of fexofenadine and sitagliptin by carrier-attached biomass and suspended sludge from a hybrid moving bed biofilm reactor. Water Res. 167, 115034.

Hermann, M., Christensen, H., Reubsaet, J., 2005. Determination of atorvastatin and metabolites in human plasma with solid-phase extraction followed by LC–tandem MS. Anal. Bioanal. Chem. 382, 1242-1249.

Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? Environ. Sci. Technol. 51, 11505-11512.

Hollender, J., Van Bavel, B., Dulio, V., Farmen, E., Furtmann, K., Koschorreck, J., Kunkel, U., Krauss, M., Munthe, J., Schlabach, M., 2019. High resolution mass spectrometry-based non-target screening can support regulatory environmental monitoring and chemicals management. Environ. Sci. Eur. 31, 1-11.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 12, 115-121.

Ingle, M.E., Mínguez-Alarcón, L., Carignan, C.C., Butt, C.M., Stapleton, H.M., Williams, P.L., Ford, J.B., Hauser, R., Meeker, J.D., 2020. The association of urinary phosphorous-containing flame retardant metabolites and self-reported personal care and household product use among couples seeking fertility treatment. J. Expo. Sci. Env. Epid. 30, 107-116.

Jin, Z.P., Luo, K., Zhang, S., Zheng, Q., Yang, H., 2012. Bioaccumulation and catabolism of prometryne in green algae. Chemosphere 87, 278-284.

Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., 2008. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24, 2534-2536.
Kim, J.-W., Isobe, T., Muto, M., Tue, N.M., Katsura, K., Malarvannan, G., Sudaryanto, A., Chang, K.-H., Prudente, M., Viet, P.H., 2014. Organophosphorus flame retardants (PFRs) in human breast milk from several Asian countries. Chemosphere 116, 91-97.
Koerts, J., Velraeds, M.M., Soffers, A.E., Vervoort, J., Rietjens, I.M., 1997. Influence of substituents in fluorobenzene derivatives on the cytochrome P450-catalyzed hydroxylation at the adjacent ortho aromatic carbon center. Chem. Res. Toxicol. 10, 279-288.
Kong, X., Wu, Z., Ren, Z., Guo, K., Hou, S., Hua, Z., Li, X., Fang, J., 2018. Degradation of lipid regulators by the UV/chlorine process: Radical mechanisms, chlorine oxide radical (ClO•)-mediated transformation pathways and toxicity changes. Water Res. 137, 242-250.
Konstas, P.S., Hela, D., Giannakas, A., Triantafyllos, A., Konstantinou, I., 2019. Photocatalytic degradation of organophosphate flame retardant TBEP: kinetics and identification of transformation products by orbitrap mass spectrometry. Int. J. Environ. Anal. Chem. 99, 297-309.
Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T.R., Neumann, S., 2012. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal. Chem. 84, 283-289.
Lecours, M.-A., Eysseric, E., Yargeau, V., Lessard, J., Brisard, G.M., Segura, P.A., 2018. Electrochemistry-High Resolution Mass Spectrometry to Study Oxidation Products of Trimethoprim. Environments 5, 1-18.
Lee, H.-B., Peart, T.E., Svoboda, M.L., Backus, S., 2009. Occurrence and fate of rosuvastatin, rosuvastatin lactone, and atorvastatin in Canadian sewage and surface water samples. Chemosphere 77, 1285-1291.
Lee, S., Jeong, W., Kannan, K., Moon, H.-B., 2016. Occurrence and exposure assessment of organophosphate flame retardants (OPFRs) through the consumption of drinking water in Korea. Water Res. 103, 182-188.
Lin, A.Y.-C., Wang, X.-H., Lee, W.-N., 2013. Phototransformation determines the fate of 5-fluorouracil and cyclophosphamide in natural surface waters. Environ. Sci. Technol. 47, 4104-4112.
Lin, W., Huang, Z., Gao, S., Luo, Z., An, W., Li, P., Ping, S., Ren, Y., 2021. Evaluating the stability of prescription drugs in municipal wastewater and sewers based on wastewater-based epidemiology. Sci. Total Environ. 754, 142414.
Melnikov, F., Kostal, J., Voutchkova-Kostal, A., Zimmerman, J.B., Anastas, P.T., 2016. Assessment of predictive models for estimating the acute aquatic toxicity of organic chemicals. Green Chem. 18, 4432-4445.
Ochiai, N., Sasamoto, K., MacNamara, K., 2012. Characterization of sulfur compounds in whisky by full evaporation dynamic headspace and selectable one-dimensional/two-dimensional retention time locked gas chromatography–mass spectrometry with simultaneous element-specific detection. J. Chromatogr. A 1270, 296-304.
Park, J.-E., Kim, K.-B., Bae, S., Moon, B.-S., Liu, K.-H., Shin, J.-G., 2008. Contribution of cytochrome P450 3A4 and 3A5 to the metabolism of atorvastatin. Xenobiotica 38, 1240-1251.
Poirier-Larabie, S., Segura, P.A., Gagnon, C., 2016. Degradation of the pharmaceuticals diclofenac and sulfamethoxazole and their transformation products under controlled environmental conditions. Sci. Total Environ. 557, 257-267.

Schneider, S.L., Lim, H.W., 2019. Review of environmental effects of oxybenzone and other sunscreen active ingredients. J. Am. Acad. Dermatol. 80, 266-271.
Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. 48, 2097-2098.
Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem. 78, 779-787.
Sui, Q., Yan, P., Cao, X., Lu, S., Zhao, W., Chen, M., 2016. Biodegradation of bezafibrate by the activated sludge under aerobic condition: Effect of initial concentration, temperature and pH. Emerging Contaminants 2, 173-177.
Tautenhahn, R., Böttcher, C., Neumann, S., 2008. Highly sensitive feature detection for high resolution LC/MS. BMC bioinformatics 9, 1-16.
Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., 2021. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371, 185-189.
Vanderford, B.J., Snyder, S.A., 2006. Analysis of pharmaceuticals in water by isotope dilution liquid chromatography/tandem mass spectrometry. Environ. Sci. Technol. 40, 7312-7320.
Wang, M., Li, J., Shi, H., Miao, D., Yang, Y., Qian, L., Gao, S., 2018a. Photolysis of atorvastatin in aquatic environment: influencing factors, products, and pathways. Chemosphere 212, 467-475.
Wang, W.-L., Wu, Q.-Y., Huang, N., Xu, Z.-B., Lee, M.-Y., Hu, H.-Y., 2018b. Potential risks from UV/H2O2 oxidation and UV photocatalysis: a review of toxic, assimilable, and sensory-unpleasant transformation products. Water Res. 141, 109-125.
Wang, Y., Gu, M., 2010. The concept of spectral accuracy for MS. Anal. Chem. 82, 7055-7062.
West, C.E., Rowland, S.J., 2012. Aqueous phototransformation of diazepam and related human metabolites under simulated sunlight. Environ. Sci. Technol. 46, 4749-4756.
Yin, L., Wang, B., Yuan, H., Deng, S., Huang, J., Wang, Y., Yu, G., 2017. Pay special attention to the transformation products of PPCPs in environment. Emerging Contaminants 3, 69-75.
Zahn, D., Mucha, P., Zilles, V., Touffet, A., Gallard, H., Knepper, T., Frömel, T., 2019. Identification of potentially mobile and persistent transformation products of REACH-registered chemicals and their occurrence in surface waters. Water Res. 150, 86-96.
Zhou, W., Zhang, Y., Xu, H., Gu, M., 2011. Determination of elemental composition of volatile organic compounds from Chinese rose oil by spectral accuracy and mass accuracy. Rapid Commun. Mass Spectrom. 25, 3097-3102.

CHAPITRE 7. ÉVALUATION DES OUTILS DE RÉSEAUX MOLÉCULAIRES ET AGGLOMÉRATION DE SPECTRES DE MASSE EN TANDEM DANS UNE APPROCHE NON CIBLÉE DESCENDANTE

7.1 Avant propos

Ce chapitre a été publié dans le journal avec comité de lecture « Science of the total environment » sous la référence : Eysseric *et al.* (2022). DOI : 10.1016/j.scitotenv.2022.153540

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

7.1.1 Auteurs et affiliation

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

[1]Department of Chemistry, Université de Sherbrooke, Sherbrooke, Canada
[2]Environment and Climate Change Canada, Montreal, Canada,

* Tel: 1-(819) 821-7922. E-mail: pa.segura@usherbrooke.ca

7.1.2 Présentation de l'article

Il y a plusieurs centaines de milliers de composés organiques et mélanges de composés organiques qui sont enregistrés pour la production dans le monde et qui peuvent donc être potentiellement relâchés dans l'environnement. Les additifs de produits de consommation représentent une importante partie de ce nombre. Or, les structures individuelles de plusieurs de ces composés sont mal définies car il s'agit de mélanges complexes d'homologues. Le fait que leur structure

individuelle même soit inconnue dans certains cas cause un important problème pour leur identification dans le cadre de dépistage non ciblé. À ce problème s'ajoute celui de l'identification de produits de transformation de contaminants organiques en général et d'additifs de produits de consommation; les études de dégradation de composés en conditions contrôlées en laboratoires ne peuvent pas réalistement identifier les produits de transformation de tous les contaminants aquatiques potentiellement préoccupants. Il existe des algorithmes permettant de regrouper des composés selon la similarité de leur spectre MS^2, mais ils sont principalement utilisés en analyse de protéines et de métabolites humains. Cet article consiste un dépistage non ciblé « top down » ou descendant utilisant ces outils de regroupement de spectres MS2 similaires en conjonction avec des outils de correspondance de spectre $HRMS^2$ *in silico* pour identifier des congénères d'additifs de produits de consommation et des produits de transformation en amont et en aval des stations de traitement d'eaux usées de trois villes le long de la rivière Yamaska.

7.1.3 Contribution des auteurs

Le plan expérimental a été conçu par Emmanuel Eysseric (EE), Christian Gagnon (CG) et Pedro A. Segura (PAS). Les analyses par spectrométrie de masse ont été réalisées par EE sur le spectromètre de masse quadripôle-orbitrap. Le développement des méthodes a été réalisé par EE. La recherche des outils informatiques a été réalisée par EE. L'échantillonnage et l'extraction pour le dépistage des suspects ont été réalisés par EE. Le traitement des données a été réalisé par EE. La visualisation des données a été réalisée par EE. La rédaction a été réalisée par EE, CG et PAS. La révision a été effectuée par EE, PAS et CG. La soumission a été réalisée par PAS. Les réponses aux réviseurs ont été réalisées par CG, PAS et EE.

7.2 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 oxides 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.

7.3 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 sometime 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, or 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 can not 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 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 CPA 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. The use of *in-silico* spectra matching algorithms such as MetFrag (Ruttkies et al., 2016) and SPS (Sweeney, 2014) allows for much larger 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 CPA 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 CPA with a high level of confidence. To do so, a NTS of water samples from a local river impacted with industrial, urban, and agricultural contamination sources was carried out. Molecular networking from GNPS and CluMSID were used to identify congeners and transformation products from CCMs.

7.4 Material and Methods

7.4.1 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.

7.4.2 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 54** (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 CPA, 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 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 55** (Supplementary Material).

7.4.3 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 55** (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 of 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 Material (NTA_SRT_wPlot-and-ScoreTable.xlsx).

7.4.4 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 2×10^5 AGC target). The frequency of acquisition was at 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 55** (Supplementary Material).

7.4.5 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 55** 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 [140]. 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.

7.4.6 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 need 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 were part of the score calculation performed by the patRoon tool. The criteria for identification can be seen in **Figure 55**.

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.

7.5 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 11**), 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 Material (IdentifiedCompounds.xlsx).

The chemicals were classified into five superclasses: consumer product additives (CPAs), illicit drugs, natural products, pesticides, and pharmaceuticals, based on the metadata on their PubChem and US-EPA Comptox Chemistry Dashboard profiles. CPA 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 24**, Supplementary Material).

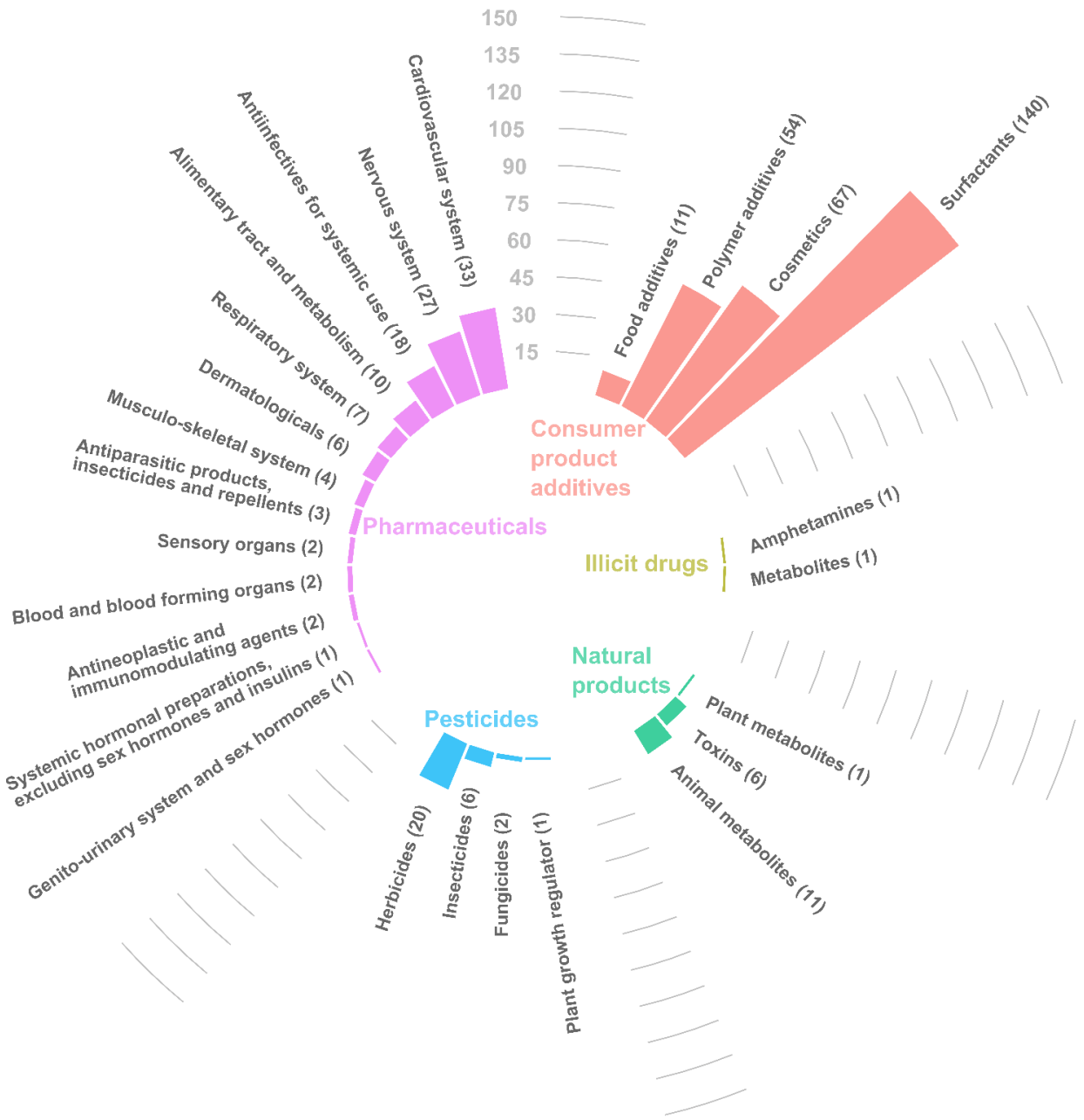


Figure 23. Total compounds tentatively identified and confirmed by superclass and class in all sampling points

7.5.1 Consumer Product Additives

Consumer product additives (CPAs) were the most numerous of the superclasses with 273 tentative identifications and 27 confirmed structures (**Figure 23**). CPA contamination was generalized to

all points ranging from 186 compounds tentatively identified or confirmed in Downstream Cowansville to 244 in downstream Farnham (**Figure 24**).

Table 11. 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 12. 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 56**
Polyethylene glycols	24	CPA (cosmetics)	**Figure 57**
Betaines and alkylamidopropyl dimethylamines	11	CPA (cosmetics)	**Figure 3**
Diethanolamines and diethanolamides	6	CPA (cosmetics)	**Figure 58**
Polyoxyethylene alkyl ethers and esters	64	CPA (surfactants)	**Figure 59**, **Figure 60** and **Table 25**
Octylphenol ethoxylates	16	CPA (surfactants)	**Figure 61**
Alkylphenols ethoxylates acids	21	CPA (surfactants)	**Figure 62**
Metolachlor transformation products	3	Pesticides (herbicides)	**Figure 64**
Beta-blockers and transformation products	4	Pharmaceuticals (cardiovascular system)	**Figure 66** and **Figure 67**
Irbesartan and transformation products	5	Pharmaceuticals (cardiovascular system)	**Figure 67** and **Figure 69**
Diltiazem and transformation products	4	Pharmaceuticals (cardiovascular system)	**Figure 67** and **Figure 70**

Gliclazide and transformation products	2	Pharmaceuticals (Alimentary tract and metabolism)	**Figure 71**

CPA: Consumer product additive.

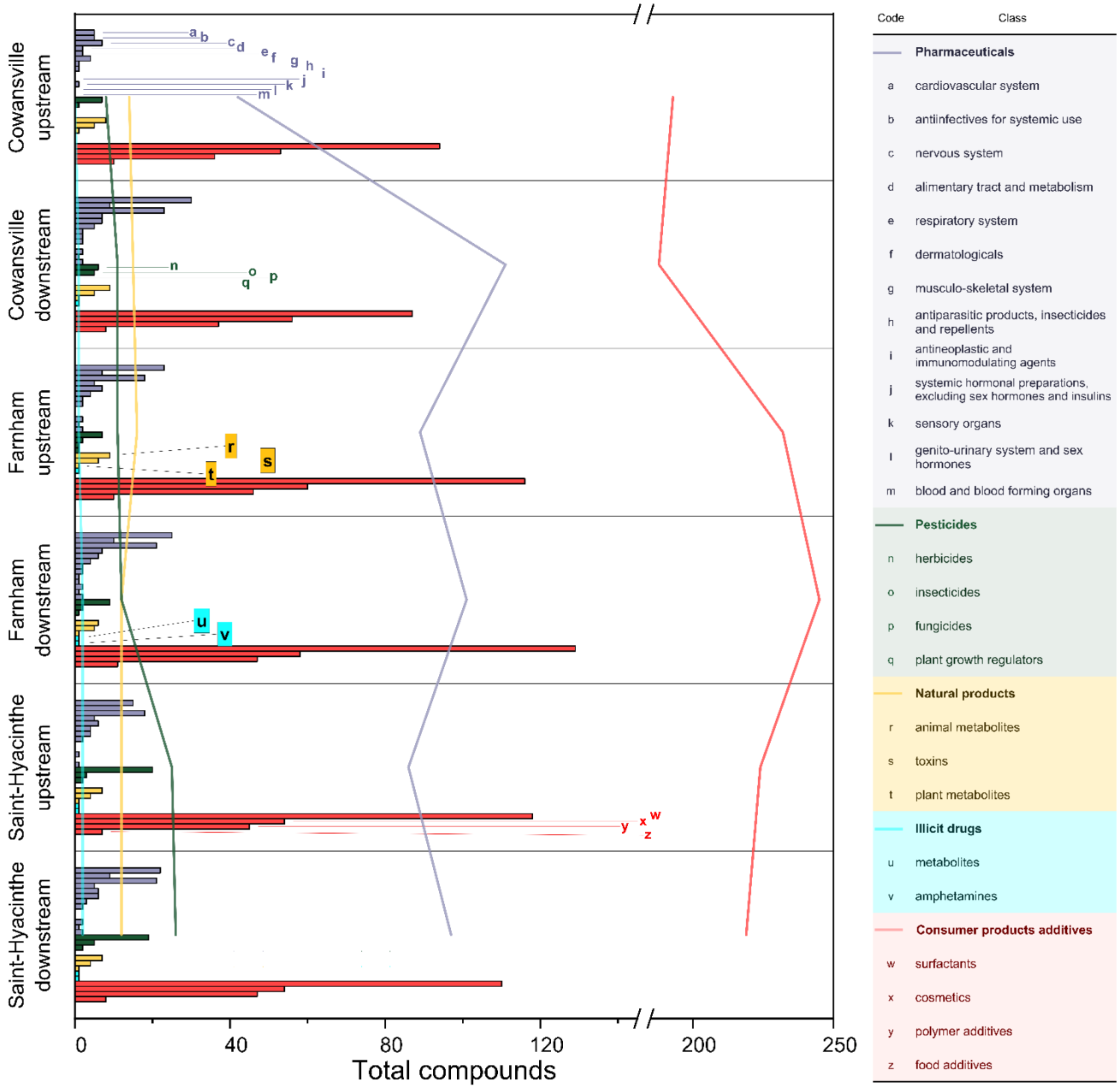


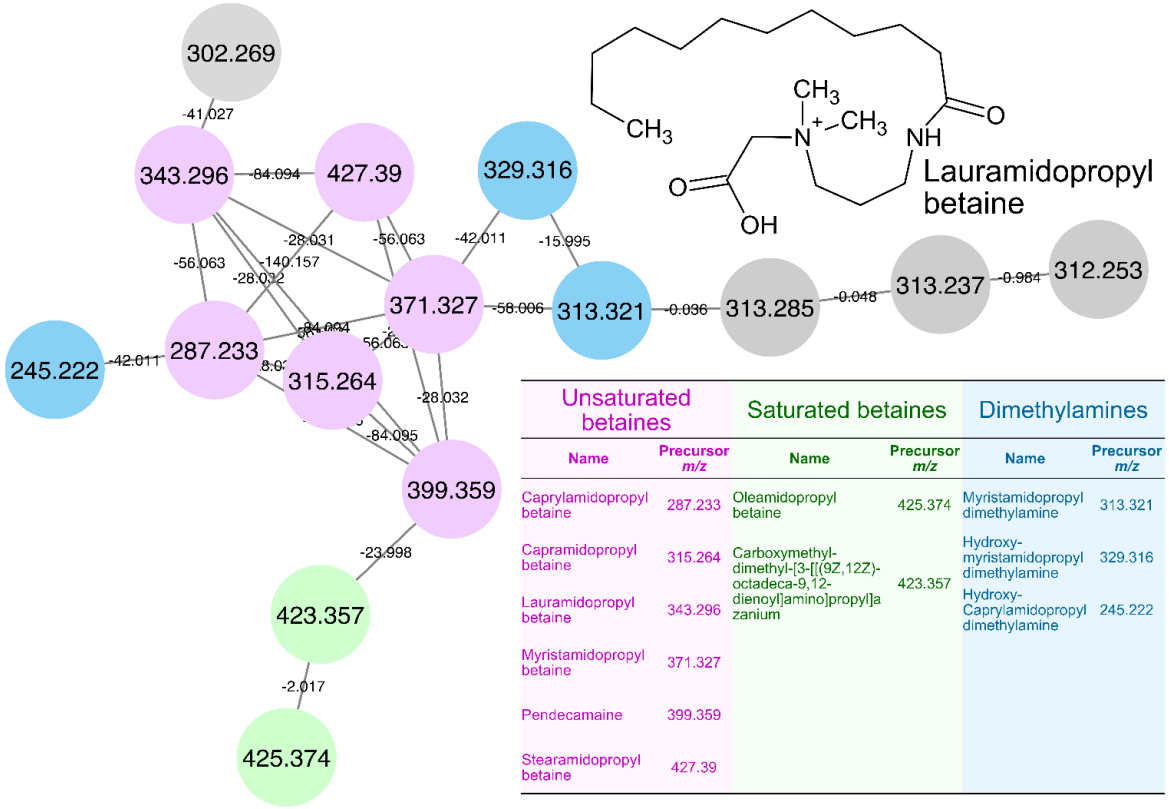
Figure 24. Frequency of detection of confirmed and tentatively identified compounds per station by superclass and class

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 56**, 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 57**, Supplementary Material). PEGs have a very wide variety of uses in cosmetics like 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 25**). The compound lauramidopropyl betaine was confirmed with a reference standard (**Table 11**); 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 would mean 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 58**, 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 25. 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.

The 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 24**). These surfactants are part of multiple very large families of PEG based congeners. A molecular network (**Figure 59**, **Figure 60**, Supplementary Material) and a cluster (**Table 25**, 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 two molecular networks (**Figure 61**, Supplementary Material) and confirmed with reference standards. Additionally, 27 carboxylic acid TPs of OPEOs and closely related nonylphenol ethoxylates (NPEO) in several molecular networks (**Figure 62**, Supplementary Material) were tentatively identified. OPEOs and NPEOs have been known to biodegrade into carboxylic acids TPs under aerobic conditions [155]. 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 63**, 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.

7.5.2 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 64**, Supplementary Material). The number of pesticides identified went up sharply at the stations of upstream and downstream Saint-Hyacinthe (**Figure 24**). 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 54**, Supplementary Material).

The amphetamine MDMA and the main metabolite of cocaine benzoylecgonine, were confirmed. The natural products tentatively identified included 6 toxins (**Figure 23**). 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 24**).

7.5.3 Pharmaceuticals

A total of 116 pharmaceuticals subdivided into 13 classes (**Figure 23**) 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 24**. 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 23**). 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 a hydroxylation followed by an oxidation (**Figure 65**, 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 66**, Supplementary Material) as well as next to metoprolol in the dendrogram of all precursors from downstream Cowansville, where it was detected (**Figure 67**, Supplementary Material). Furthermore, one previously unknown TP of irbesartan was tentatively identified: irbesartan_TP445 is result of a hydroxylation at the end of the aliphatic chain (**Figure 68**, Supplementary Material). Three other TPs that were found in the same network, irbesartan_TP443, the result of an 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 68**, 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 69**). 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 70**, 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 67**, 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 for 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 23**) 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 development of bio resistance as 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 information (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 71**, Supplementary Material). To the authors knowledge, this TP was previously unknown and unreported in the literature.

7.6 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 between these compounds including several unknown ones, as can be seen in **Table 12**. 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.

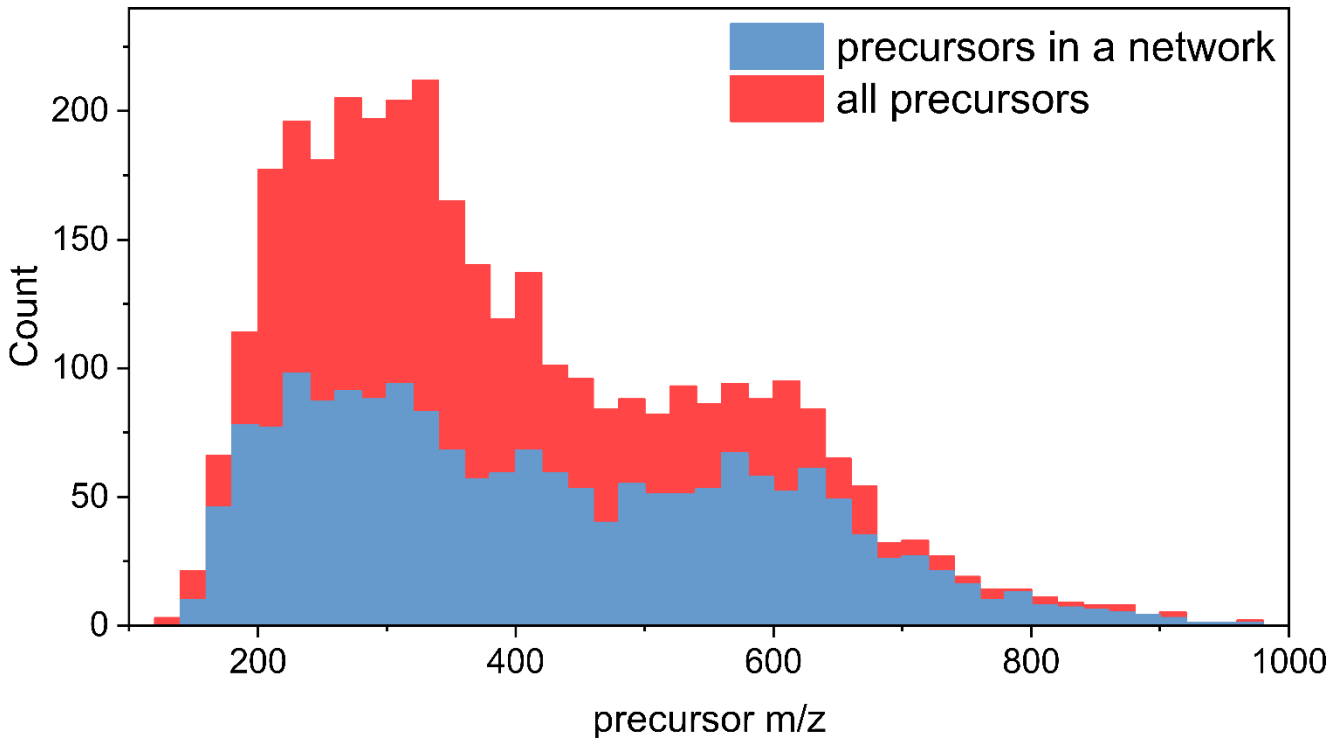


Figure 26. Repartition of the number of precursors in a network per precursor m/z on the total number of precursors across all sampling points. The background signal, noise and blank features were removed and are not presented in this figure.

Networking and clustering tools showed powerful capacities when it came to uncovering the structures of oligomers such as alkylphenol ethoxylates, polyoxyethylenes, polyoxyethylene akyl 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 12**. 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 63**. All the polyoxyethylene congeners can be seen in an Excel File in the Supplementary Material (IdentifiedCompounds.xlsx)

However, 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 72** (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 57**, **Figure 60**, **Table 25**, and **Figure 61** (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 25**), alkyldiethanolamines and alkyldiethanolamides (**Figure 58**, Supplementary Material), and metolachlor (**Figure 64**, 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 inconvenient of both tools was how time-consuming they are and cannot be applied to routine analysis as of now.

7.7 Conclusion

A "top-down" workflow consisting of a non-targeted water analysis of 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.

7.8 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.

7.9 References

Altenburger R, Brack W, Burgess RM, Busch W, Escher BI, Focks A, et al. 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, et al. 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: 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, et al. 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.
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.
Lara-Ortega FJ, Robles-Molina J, Brandt S, Schütz A, Gilbert-López B, Molina-Díaz A, et al. 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.
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, et al. 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, et al. 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, et al. 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.
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, et al. 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, et al. 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, et al. 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, et al. 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, et al. 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, et al. 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, et al. Identification of potentially mobile and persistent transformation products of REACH-registered chemicals and their occurrence in surface waters. Water research 2019; 150: 86-96.

CHAPITRE 8. CONCLUSION GÉNÉRALE

À travers ses trois objectifs, cette thèse s'articule autour du thème de l'identification de produits de transformation et d'homologues de produits de consommation par le développement de méthodes, techniques et outils analytiques en spectrométrie de masse et la génération de produits de transformation en laboratoire. Ainsi, l'échange hydrogène-deutérium post colonne en CHAPITRE 3, en intervenant à la fin du processus d'identification, a permis de fournir une technique de diagnostic très utile identifier un composé inconnu en permettant de passer d'un candidat potentiel à une structure probable selon le schéma des niveaux de confiance illustré en **Figure 2**. Cependant, son manque de flexibilité, l'importante quantité de temps nécessaire pour une identification et la nécessité d'un signal fort le rendent moins intéressant pour l'identification de produits de transformation à faible concentration. En revanche, la conformité spectrale au CHAPITRE 4 quant à elle s'est avérée un indicateur très puissant pour aider à la confirmation de contaminants avant l'achat d'étalons de référence au CHAPITRE 5 et s'est avérée essentielle pour l'identification des produits de transformation générés en laboratoire au CHAPITRE 6. Passer d'un *m/z* précis et exact avec un patron isotopique à une formule moléculaire sans ambiguïté est très utile lorsqu'un spectre MS^2 n'est pas disponible. Les outils de correspondance de spectres de masse en tandem *in silico* et empirique en conjonction avec les algorithmes d'agglomération de spectres de masse en tandem, déployés dans le CHAPITRE 5 et le CHAPITRE 7 ont mis en lumière la présence de produits de transformation et d'homologues de produits de consommation toxiques dans la rivière Yamaska. Ces derniers incluent les acides d'éthoxylates d'alkylphénols et les produits de transformation de l'atorvastatine et du diltiazem. Les méthodes ciblées traditionnelles utilisées dans les programmes de monitoring n'auraient pas pu identifier ces composés préoccupants de même que des analyses non-ciblées utilisant seulement des banques de données empiriques de taille limitée. Les éthoxylates d'alkylphénols et leurs produits de transformation mériteraient d'ailleurs des études plus approfondies pour déterminer leur concentration et leur impact sur le milieu aquatique de la rivière Yamaska.

Par d'un côté l'identification de produits de transformation générés en laboratoire suivie de la confirmation de leur présence environnementale et de l'autre leur identification directe dans les

échantillons, cette thèse propose une synthèse des approches ascendante, *bottom-up*, et descendante, *top-down*. Ces approches ont été consolidées par le développement et l'application de méthodes, techniques et outils analytiques puissants et novateurs pour identifier des contaminants organiques posant problème aux méthodes traditionnelles. En soulignant les forces et faiblesses de ces deux paradigmes analytiques, elle fournit aussi un état des lieux des capacités analytiques de pointe en spectrométrie de masse à haute résolution pour l'identification de contaminants organiques peu ou pas connus et les perspectives à l'horizon dans cette discipline. L'approche ascendante demeure la plus sûre et robuste pour identifier des produits de transformation mais représente un important investissement en temps et en autres ressources; elle devrait continuer à être utilisée pour les contaminants prioritaires, mais ne peut pas réalistement être appliquée à tous les composés préoccupants. Il revient ainsi aux approches descendantes de caractériser la contamination chimique dans l'environnement. Cependant, les approches descendantes nécessitent des spectres de masse en tandem de bonne qualité et donc des intensités de signal suffisamment élevées. Celles-ci sont fonction du rendement d'ionisation et de la concentration des composés. Or, les composés au signal le plus intense ne sont pas forcément les plus toxiques. Les produits de transformation toxiques doivent être identifiés par l'approche ascendante comme le dépistage des suspects nécessite un signal moins important. Le fardeau analytique en analyse non ciblée est imposant mais le déploiement des outils, méthodes et techniques proposés dans cette thèse dans des programmes de surveillance pourrait à terme permettre d'aider à mieux obtenir un portrait plus clair de la pollution chimique.

La rétrospectivité de l'analyse non ciblée permet aussi la possibilité de procéder à du *data mining* sur les anciennes analyses avec les outils nouvellement développés. Cependant, il y a à ce jour une importante redondance dans le traitement des données qui est causée par le dédoublement de nombreuses étapes pour l'utilisation d'outils. Une uniformisation de ces étapes pour leur utilisation est nécessaire pour leur application à grande échelle. Le développement d'instruments plus performants amène certes de nouvelles perspectives mais aussi de nouveaux défis dans la gestion et le traitement des données; la spectrométrie de masse est un domaine révolutionné par de nouvelles technologies plusieurs fois par décennie. Il m'apparaît que l'implémentation de l'analyse non ciblée doit se faire à travers le dépistage des suspects où leur signal est recherché dans échantillons à partir d'une liste étendue. Cette liste inclurait des produits de transformation dont la

présence environnementale a été confirmée. L'identification suivrait le concept des niveaux de confiance alors qu'une constellation d'outils analytiques dont la conformité spectrale seraient employés pour consolider la confiance dans une identification. Les outils d'analyse combinatoire de spectres de masse en tandem ainsi que les réseaux moléculaires permettraient également d'augmenter la confiance dans les dépistages de suspects implémentés dans les programmes de monitoring. Des dépistages non ciblés seraient réalisés ponctuellement pour analyser les tendances de contamination. L'analyse donnée-indépendante, aussi appelée DIA, permet de plus d'obtenir une meilleure sensibilité d'identifier des composés moins sensibles qui auraient pu être manqués par analyse donnée-dépendante. Les résultats de DIA sont plus complexes à analyser mais permettraient de potentiellement identifier un plus grand nombre de composés. Parallèlement, le potentiel de l'analyse non ciblée quantitative doit être exploré; l'identification suivie de la quantification de centaines de composés par analyse rendrait les institutions de monitoring équipées pour caractériser la pollution chimique. Les horizons sont ouverts pour l'identification de produits de transformation et nos connaissances sur leur présence environnementale et leur devenir continuera à croître avec les développements analytiques.

CHAPITRE 9. ANNEXES

9.1 ANNEXE A. Informations supplémentaires du CHAPITRE 4

Post-column hydrogen-deuterium exchange technique to assist in the identification of small organic molecules by mass spectrometry

Emmanuel Eysseric[1], Xavier Bellerose[1], Jean-Michel Lavoie[2], Pedro A. Segura[1, *]

[1] Department of Chemistry, Université de Sherbrooke, 2500 Boulevard de l'Université, Sherbrooke, QC, Canada J1K 2R1

[2] Department of Chemical and Biotechnical Engineering, Université de Sherbrooke

[*] Corresponding author: e-mail: pa.segura@usherbrooke.ca, Tel: 1-819-7922, Fax: 1-819-821-8017

Supplementary material

9.1.1 Determination of deuteration percentage

The deuteration percentage of trimethoprim was calculated using an algorithm previously published[1]. Here we present a detailed account of how this algorithm was employed to determine deuteration percentages in the experiments described in the manuscript.

For a small molecule with a formula of $C_wH_xO_yN_z$ such as trimethoprim ($C_{14}H_{18}O_3N_4$), the main isotopic contributions are ^{13}C (1.1% per carbon atom) and ^{15}N (0.37% per nitrogen atom) whereas ^{17}O, ^{18}O and ^{2}H contributions are negligible.

First, the abundance due to naturally occurring X+1 isotopes ($A_{natural}$) is determined with the equation:

$$A_{natural} = (1.1\% \times w) + (0.37\% \times z)$$

Where w and z indicate the number of carbon and nitrogen atoms, respectively. For trimethoprim, $A_{natural}$ in the M+n peaks is 16.88% of the M+(n-1) peak, therefore this value must be subtracted to obtain the artificial contribution due to HDX. For example, the following data was obtained for trimethoprim using the post-column HDX technique (**Table 13**).

Table 13. Relative intensities of all the peaks of the isotopic pattern of trimethoprim obtained after loop injections using a flow rate of 300 μLmin^{-1} for the mobile phase (A: 30% of 0.1% FA in H_2O, B: 70% of 0.1% FA in ACN) and an addition rate of 50 μLmin^{-1} of D_2O in the QqQMS.

Peak	Ion *m/z*	Relative intensity (%)
M	291	3.59
M+1	292	24.35
M+2	293	76.71
M+3	294	100.00
M+4	295	40.60
M+5	296	8.42
M+6	297	1.13

The corrected relative intensity for the M+1 peak is calculated as follows:

$$Corr.rel.intensity\ (M+1) = [rel.intensity\ (M+1)] - [rel.intensity\ M \times A_{natural}]$$

Therefore, the corrected relative intensity for the M+1 of trimethoprim in **Table 13** is 23.74 %. The same procedure is applied to the following M+*n* peaks to obtain **Table 14**:

Table 14. Corrected relative intensity values for all peaks of the trimethoprim isotopic pattern after post-column HDX.

Peak	Ion *m/z*	Relative intensity (%)	Correction value	Corrected relative intensity (%)
M	291	3.59	0	3.59
M+1	292	24.35	0.61	23.74
M+2	293	76.71	4.11	72.60
M+3	294	100.00	12.95	87.05
M+4	295	40.60	16.88	23.72
M+5	296	8.42	6.85	1.57
M+6	297	1.13	1.42	-0.29

As it can be observed, this calculation shows that the corrected relative intensity for the M+6 peak is negative, indicating that only 5 deuterium atoms were exchanged (4 labile hydrogens plus the proton adduct). Then, the ratio between each corrected intensity and the sum of all corrected

relative intensities is calculated and expressed as a molar percentage. For trimethoprim, the sum all of corrected intensities is 211.98%. Therefore, we obtain the values shown in**Table 15**:

Table 15. Molar percentage for all peaks of the trimethoprim isotopic pattern after post-column HDX.

Peak	Ion *m/z*	Corrected relative intensity (%)	Molar percentage (%)
M	291	0.64	1.69
M+1	292	8.34	11.20
M+2	293	47.27	34.25
M+3	294	91.78	41.07
M+4	295	26.14	11.19
M+5	296	0.93	0.74

Finally, the deuteration percentage is calculated adding the products of the molar percentage (MP) of each peak by their maximum deuterium fraction (DF). Results are shown in**Table 16**.

$$Deuteration\ percentage = \sum_{i=1}^{n} DF_i \times MP_i$$

Table 16. Percentage of deuterium for all peaks of the trimethoprim isotopic pattern after post-column HDX.

Peak	Ion *m/z*	Molar percentage (%)	Maximum deuterium fraction	Percentage of deuterium
M	291	1.69	0	0.00
M+1	292	11.20	0.2	2.24
M+2	293	34.25	0.4	13.70
M+3	294	41.07	0.6	24.64
M+4	295	11.19	0.8	8.95
M+5	296	0.74	1	0.74

The maximum deuterium fraction is calculated according to the total number of deuterium atoms exchanged. For trimethoprim, the total number is five; therefore, there is a difference of 1/5 between the maximum deuterium fractions of peak having 1 mass unit of difference in the isotopic

pattern. The deuteration percentage is calculated by adding the percentage of deuterium of all peaks. For this specific experiment the deuteration percentage was 50.3%.

Generation of theoretical mass spectra

Once the deuteration percentage is calculated, the theoretical mass spectra can be generated using a binomial distribution also known as the Pascal's triangle. The number of coefficients in a row must be equal to the number of possible configurations. Then, the percentage of ^{1}H is the first coefficient and the percentage of ^{2}H (deuterium) is the second one.

For example, trimethoprim has 4 exchangeable hydrogens plus 1 in positive ionization for a total of 5 exchangeable hydrogens, so 6 possible configurations. The corresponding row is the one where n = 5 and there are 6 coefficients: 1, 5, 10, 10, 5 and 1. The two terms of the distribution here are the calculated percentage of ^{1}H, noted ^{H}P and the calculated percentage of ^{2}H, noted ^{D}P. The binomial distribution then looks like this:

1.1. $$(^{H}P + {}^{D}P)^{5} = {}^{H}P^{5} + 5\,{}^{H}P^{4} * {}^{D}P + 10\,{}^{H}P^{3} * {}^{D}P^{2} + 10\,{}^{H}P^{2} * {}^{D}P^{3} + 5\,{}^{H}P * {}^{D}P^{4} + {}^{D}P^{5}$$

Each one of the six values represent the 6 possible configurations in the spectrum, from the 0% deuterated form (291) to the 100% deuterated one (296).

From then, a correction is applied because of the contribution of the naturally occurring isotopes, ^{13}C and ^{15}N principally, on the spectra. The contribution is basically the reverse of what was done previously, approximately 17% of the M intensity is added to the M + 1 signal.

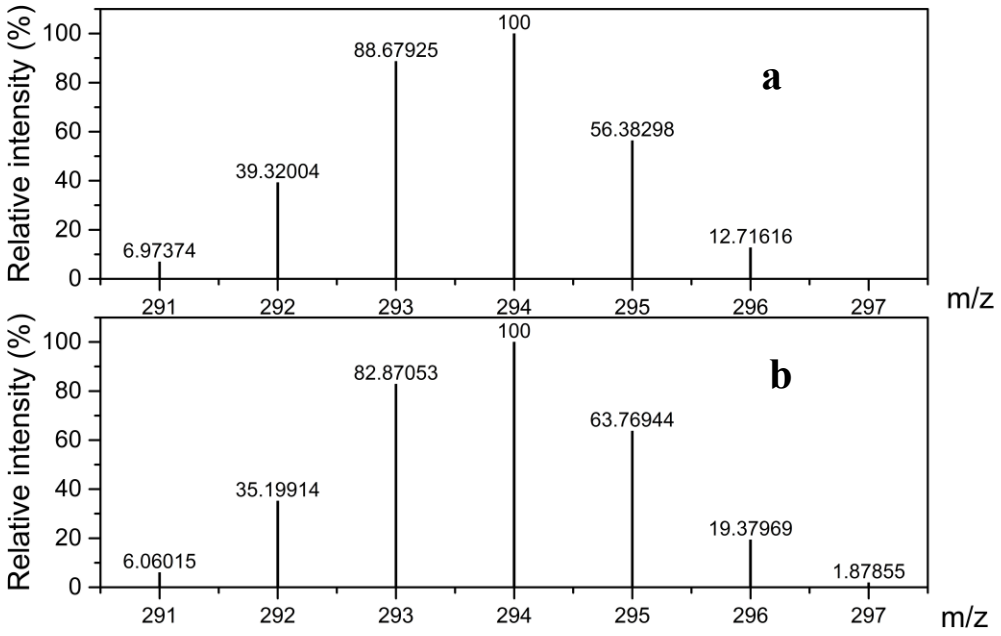


Figure 27. Theoretical mass spectra of trimethoprim without (a) and with (b) the naturally occurring isotope correction. Both spectra were generated using the aforementioned method.

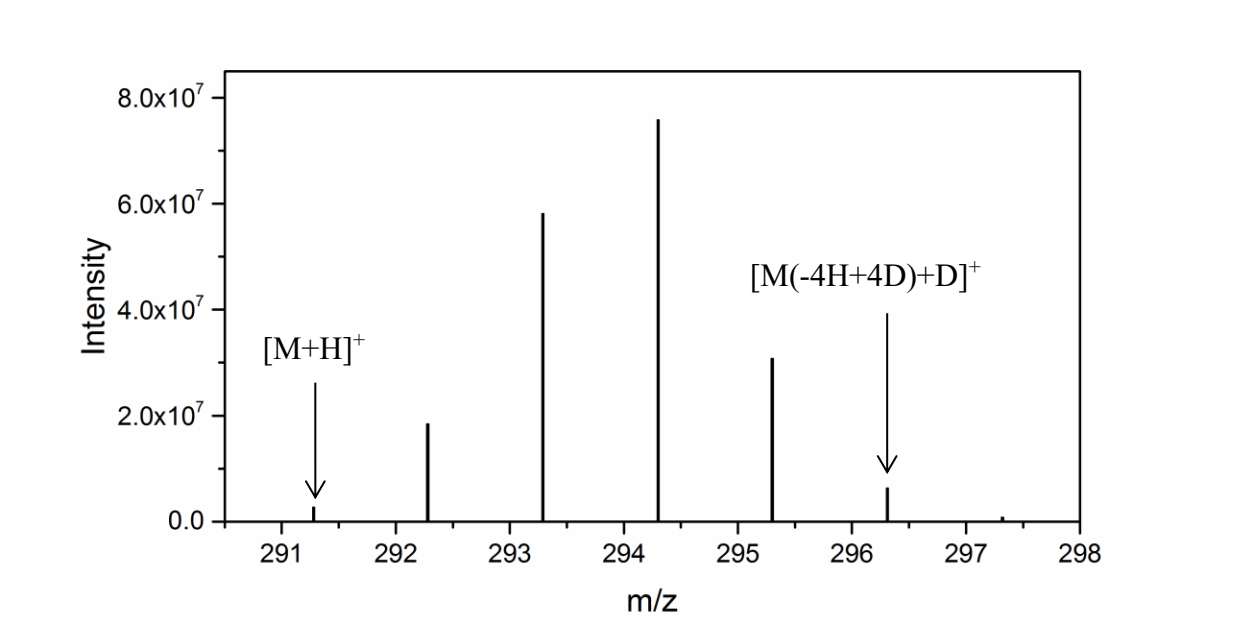


Figure 28. Mass spectrum of trimethoprim after HDX in a QqQMS. Mobile phase flow rate was 300 μL min^{-1} **(A: 30% of 0.1% FA in H_2O, B: 70% of 0.1% FA in ACN)**, D_2O addition flow rate was 50 μL min^{-1} and the mixing device was a mixing tee.

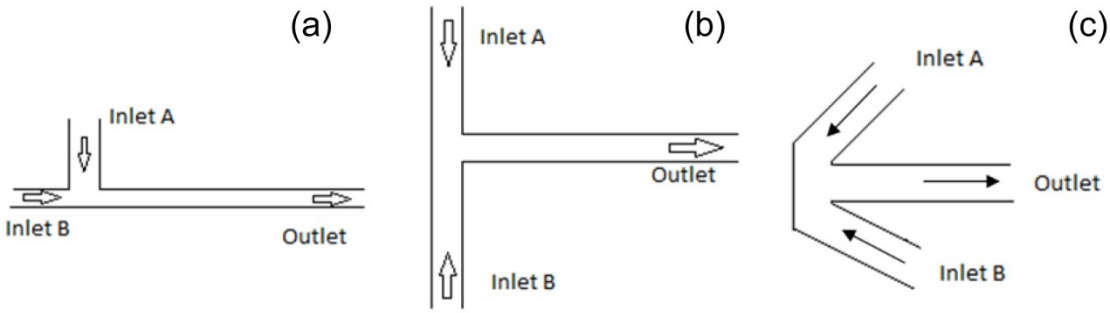


Figure 29. Representation of the flow direction in different tee connectors shapes: tee connector 90° (a), tee connector 180° (b) and mixing tee (c). Inlet A represents D_2O and inlet B the mobile phase.

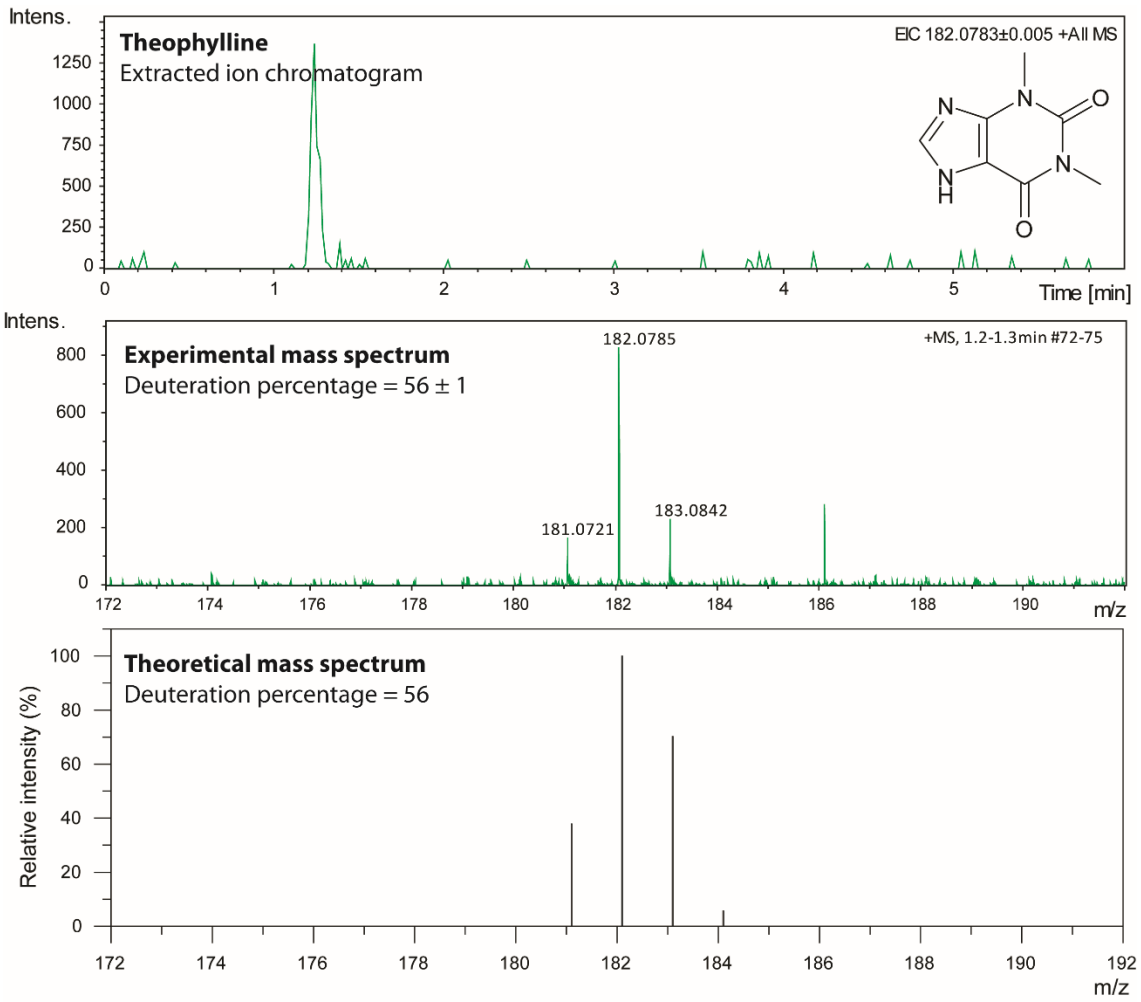


Figure 30. Extracted ion chromatogram and experimental and theoretical mass spectra of theophylline after post-column HDX of a SPE extract of spiked river water.

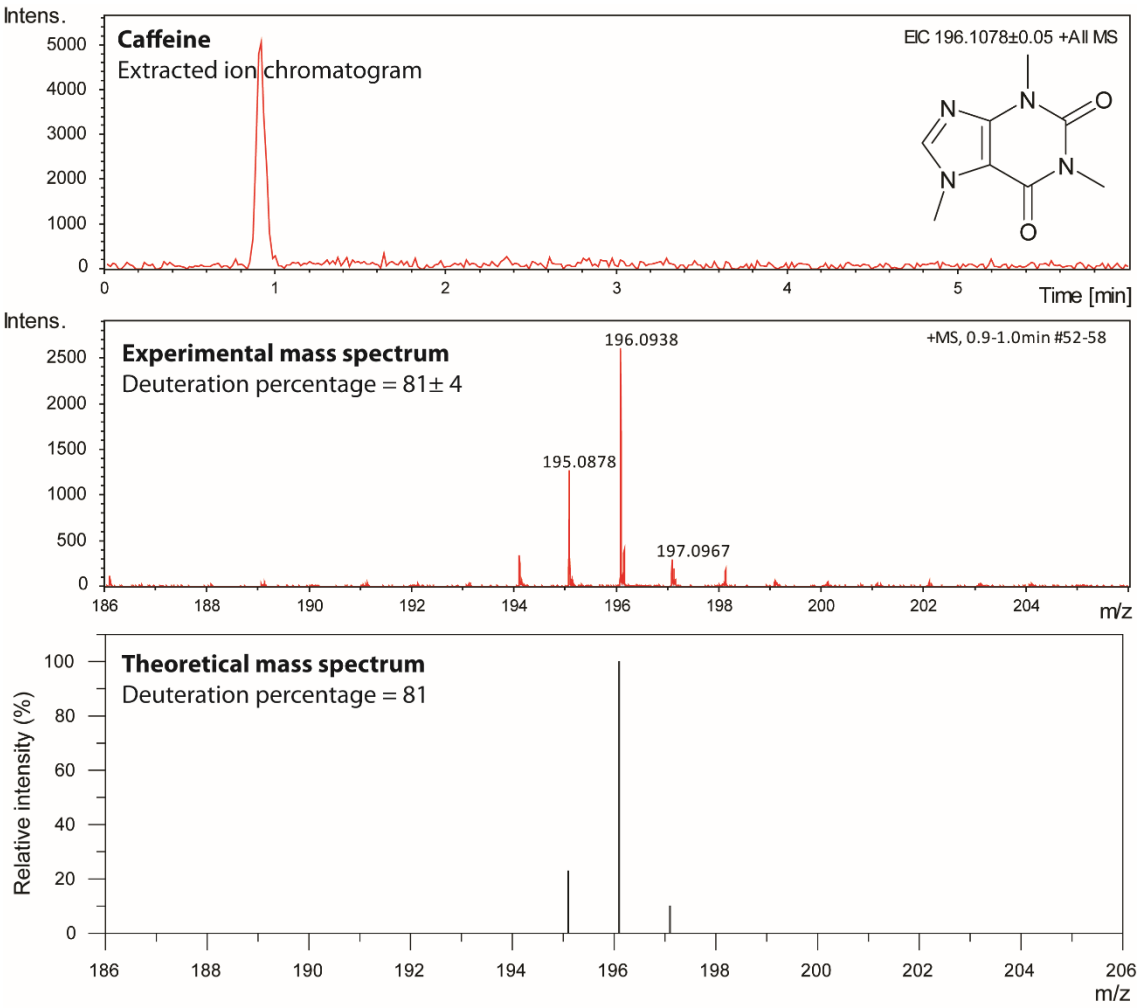


Figure 31. Extracted ion chromatogram and experimental and theoretical mass spectra of caffeine after post-column HDX of a SPE extract of spiked river water.

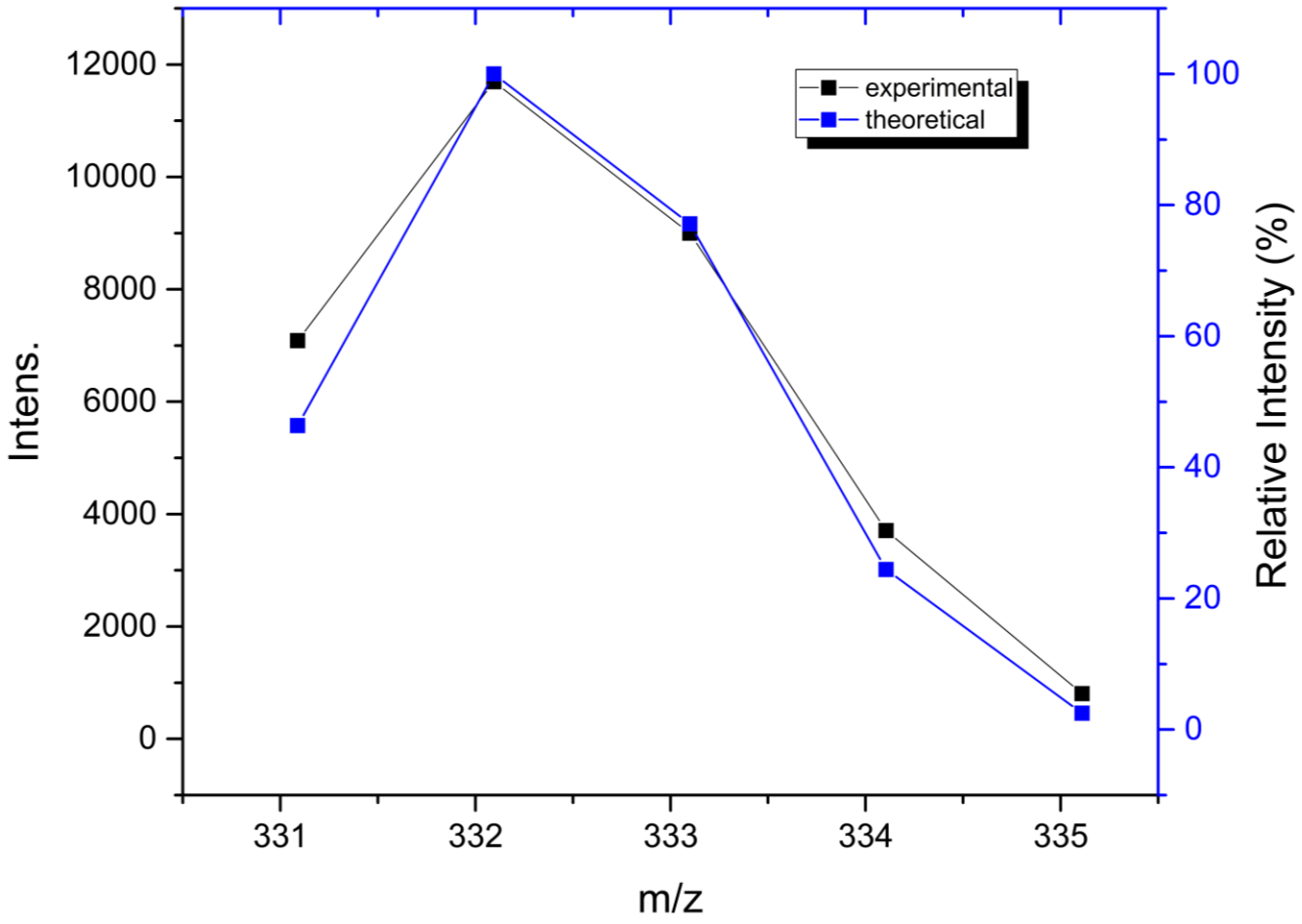


Figure 32. Comparaison of intensities of experimental and theoretical and corrected peaks of tricin at 40% deuteration presented in Fig. 6.

9.1.2 References

(1) McCloskey, J. A. Methods Enzymol. 1990, 193, 329.

9.2 ANNEXE B – Informations supplémentaires du CHAPITRE 4

Application of spectral accuracy to improve the identification of organic compounds in environmental analysis

Emmanuel Eysseric[1], Killian Barry[1], Francis Beaudry[2], Magali Houde[3], Christian Gagnon[3], Pedro A. Segura[1,*]

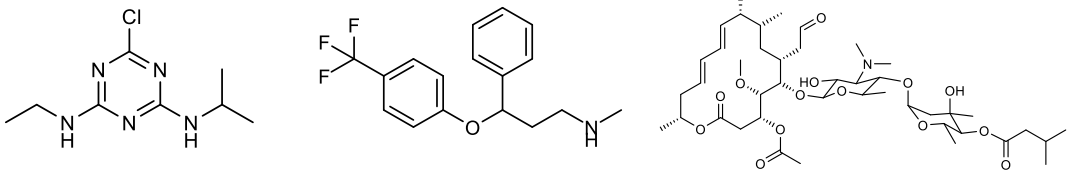
* Tel: 1-(819) 821-7922. Fax: 1-(819) 821-8019. E-mail: pa.segura@usherbrooke.ca

[1] Department of Chemistry, Université de Sherbrooke, Sherbrooke, QC J1K 2R1

[2] Groupe de Recherche en Pharmacologie Animal du Québec (GREPAQ), Department of Veterinary Biomedicine, Université de Montréal, Saint-Hyacinthe, QC J2S 2M2

[3] Environment and Climate Change Canada, Montreal, QC H2Y 2E7

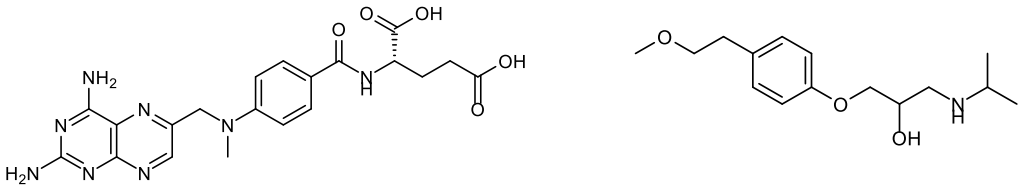
Supporting Information



Atrazine (ATZ)
$C_8H_{14}N_5Cl$ (215 Da)

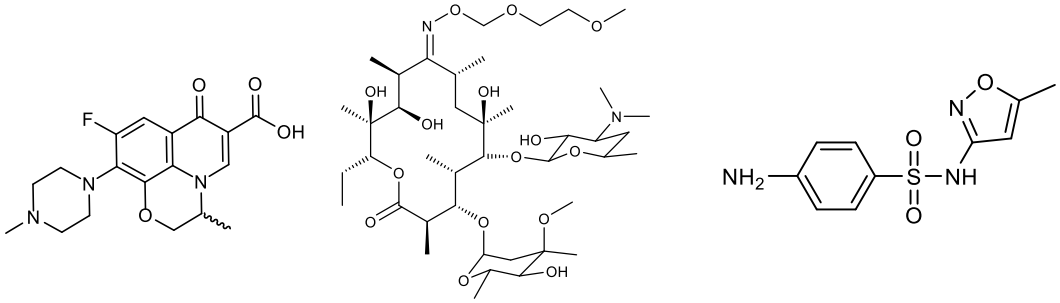
Fluoxetine (FLX)
$C_{17}H_{18}F_3NO$ (309 Da)

Josamycin (JOS)
$C_{42}H_{69}NO_{15}$ (827 Da)



Methotrexate (MTX)
$C_{20}H_{22}N_8O_5$ (454 Da)

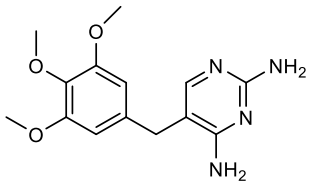
Metoprolol (MET)
$C_{15}H_{25}NO_3$ (267 Da)



Ofloxacin (OFL)
$C_{18}H_{20}N_3O_4F$ (361 Da)

Roxithromycin (ROX)
$C_{41}H_{76}N_2O_{15}$ (837 Da)

Sulfamethoxazole (SMX)
$C_{10}H_{11}N_3O_3S$ (253 Da)



Trimethoprim (TRI)
$C_{14}H_{18}N_4O_3$ (290 Da)

Figure 33. Acronym, neutral nominal mass and molecular structure of the test compounds used in this study.

9.2.1 External calibration and stability of the QqQMS system

A sodium formate solution was used for external calibration and four of the test compounds (MTP, FLX, OFL and JOS) were used as internal calibrants for ATZ, SMX, TRI, MTX and ROX. Figure SI-2 shows the impact of the calibration method. As expected, mass accuracy significantly improved when internal calibration was applied. In the case of ATZ, it improved from 524 ppm (113 mDa) to 6 ppm (1.3 mDa), allowing to get a better ranking (from 3 ± 1 to 1st) even with a relatively low spectral accuracy (in both cases ≤ 91%).

Mass accuracy stability of the QqQMS using external calibration was evaluated by successive injections of a mixture of ATZ, SMX and TRI spiked at 300 μg L^{-1} in MeOH during a 12 h period (Figure SI-3). The results showed that the average mass accuracy was 74 ± 17 mDa for ATZ, 66 ± 30 mDa for SMX and 46 ± 23 mDa for TRI. Those results indicated that the QqQ mass spectrometer was relatively stable overtime when using external calibration, however the average mass accuracy was too high, between 77 and 113 mDa, which negatively affects the ranking of the correct formula. It was not possible to obtain results for ROX using external calibration because its large mass error and large molecular mass lead to a high number of generated formulas. Major improvement of the results as illustrated by ATZ (Figure SI-2) were observed for all compounds using internal calibration.

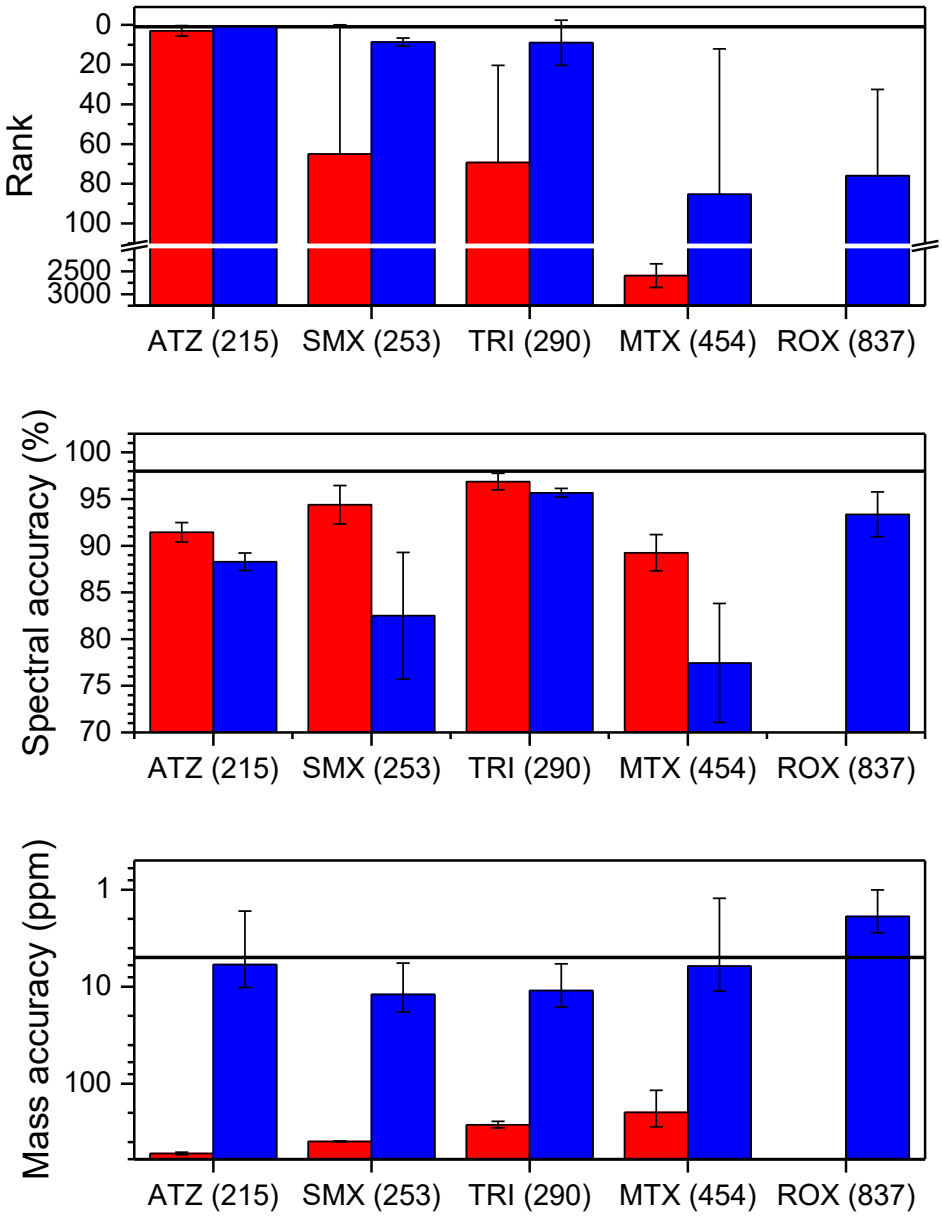


Figure 34. Ranking (top), spectral accuracy (middle) and mass accuracy (bottom) of the test compounds at a concentration of 300 µg L^{-1} in MeOH measured with the QqQMS using external (red) and internal calibration (blue). Straight lines indicate: expected value for ranking, 1st (top); the threshold of high spectral accuracy, 98% (middle) and accepted value for maximum mass accuracy, 5 ppm (bottom).

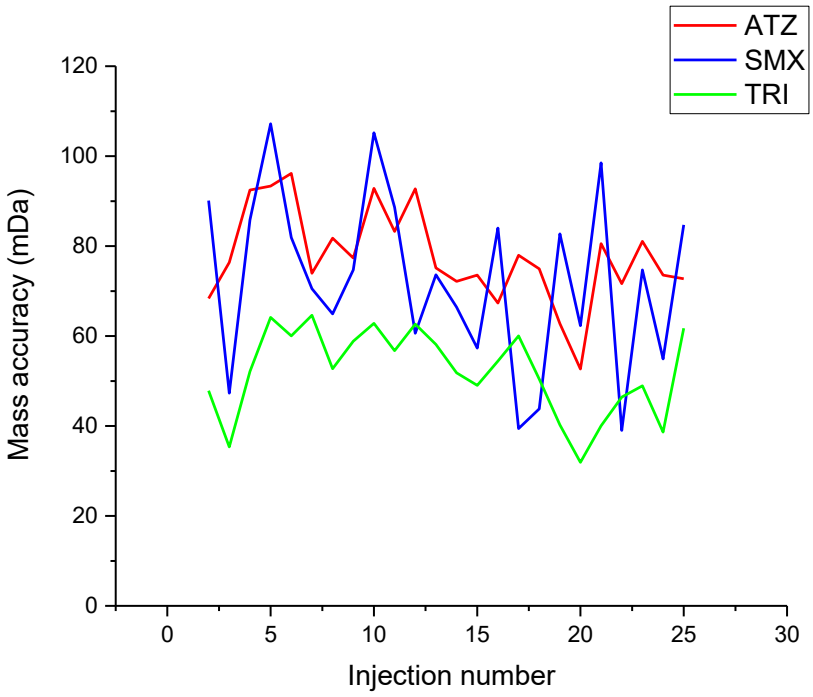


Figure 35. Stability of the mass accuracy obtained with external calibration in the QqQMS instrument over a 12 h period for ATZ, SMX and TRI spiked at 300 µg L^{-1} in MeOH.

A previous study demonstrated that post-acquisition analysis with MassWorks of low resolution data, obtained with gas chromatography-quadrupole mass spectrometry using perfluorotributylamine as external calibrant, can be employed to identify unknown compounds [1]. Therefore, a second experiment of stability of mass accuracy over time was performed by injecting a solution of sodium formate each hour within a period of 12 h in the QqQMS. In those experiments the solution of sodium formate at 0.5 µM was used in chromatographic conditions to study the evolution of mass accuracy in the QqQMS over time. Chromatographic conditions were the following: mobile phase flow rate was 50 µLmin^{-1}, and the mobile phase was a mixture of 2-propanol and water (9:1, v/v). Run time was 1 hour and injection volume was 10 µL. No chromatographic column was used as there was no need for separation. A union was used instead of the column. QqQMS source and ion optics parameters were the same as those used for the analysis of the selected compounds. In total 11 injections were made over a period of 12 h. Sodium formate spectra were then calibrated post-acquisition with MassWorks CLIPS algorithm. The second injection spectrum was used to calibrate the rest of the data.

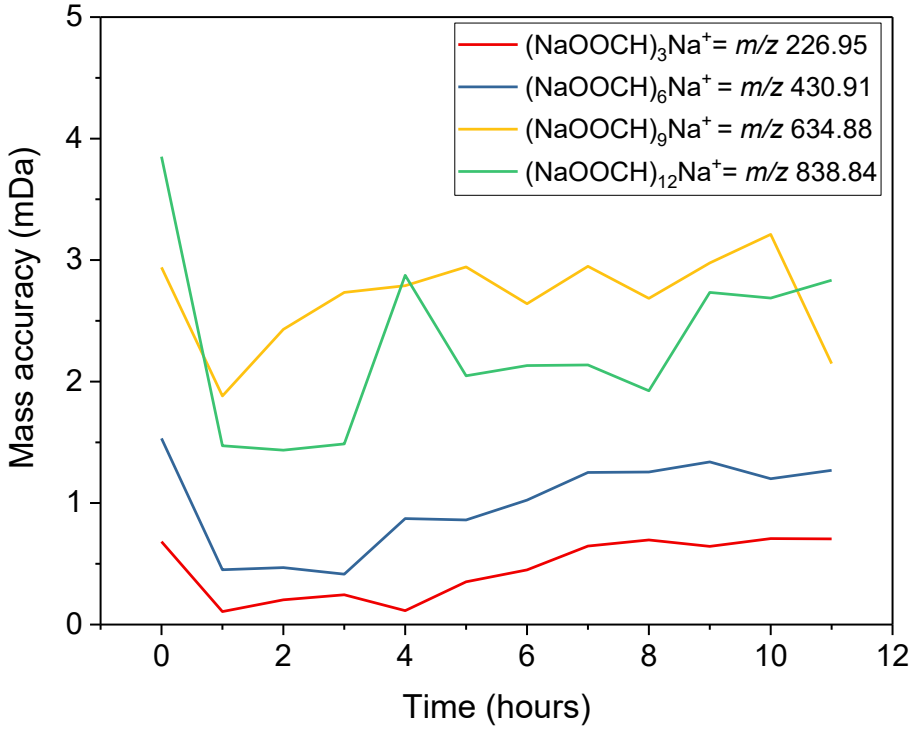


Figure 36. Post-acquisition CLIPS-calibrated stability of the mass accuracy obtained with external calibration in the QqQMS instrument over a 12 h period for sodium formate at 0.5 µM in 2-propanol- water (9:1, v/v). Injection #2 (time = 1 hour) was used to calibrate all other masses.

As can be seen in Figure SI-4, post-acquisition CLIPS-calibrated masses showed little variation (difference between maximum and minimum values was <3 mDa) and had good mass accuracy ($\Delta m \leq 4$ mDa) over the span of 11 hours. This is a large discrepancy compared to the external mass calibration results for ATZ, SMX and TRI discussed previously (Figure SI-3). It could be partly explained by completely different chromatographic conditions: mobile phase composition, flow rate, and separation (or the absence thereof) were all dissimilar. It is also worth noting that all masses were calibrated with their own *m/z* values contrary to the external calibration in which the *m/z* values of the calibrants were close but different to those of the test compounds. Hence, at this point is not clear why a major systematic error in the mass accuracy of the QqQMS data with external calibration could not be corrected by the software. That issue is out of the scope of the present work and will be the topic of future research.

9.2.2 Error in spectral accuracy determination

For a reported spectral of certain value, the spectral error is (100 – spectral accuracy). The square of the spectral error follows a Chi-square distribution. Then the confidence interval (CI) of the spectral error is given by:

$$(100 - spectral\ accuracy) \times \sqrt{\frac{\chi^2_{\left(\frac{\alpha}{2}\right)}}{k-2}} < spectral\ error\ < (100 - spectral\ accuracy) \times \sqrt{\frac{\chi^2_{(1-\frac{\alpha}{2})}}{k-2}}$$

where χ^2 is the Chi-squared critical value (upper- or lower-trail) for a given significance level α and degrees of freedom and *k-2* represents the degrees of freedom that are calculated according to the number of data points in the profile isotopic pattern. A value of 2 is subtracted from the data points because of the baseline and the pure theoretical mass spectrum. For the spectral accuracy of JOS at 50 ng mL^{-1}, we would have spectral accuracy of 96.5686%, and 450 data points across the profile isotopic pattern, therefore the 95% confidence limits of the spectral error are:

$$Upper\ limit = (100 - 96.5686) \times \sqrt{\frac{536.6167}{448}} = 3.63\%$$

$$Lower\ limit = (100 - 96.5686) \times \sqrt{\frac{368.5509}{448}} = 3.21\%$$

9.2.3 Interference rejection

It is expected that ions from coeluting matrix compounds could lower the spectral accuracy and the ranking if they are found in the spectral region of interest, i.e. between the peaks of the relevant isotope pattern. The interference rejection function in MassWorks was designed to correct those potential issues. This function allows the exclusion of a sub-spectral region from the calibrated experimental spectrum if the relative theoretical abundance in that region is less than a defined fractional value. For example, interference rejection with a value of 0.001 means that any spectral region where the relative theoretical abundance is less than 0.1%, relative to the most intense isotope, would be ignored and not factored into the spectral accuracy calculation. The effect of this function was measured and is displayed on Figure SI-5. At 80 $\mu g L^{-1}$ spiked in the matrix, the effect of interference rejection on the ranking was unnoticeable for most compounds except for MTX. However spectral accuracy improved slightly with interference rejection. Interference rejections results for the 300 $\mu g\ L^{-1}$ solution are presented in Figure SI-6 and Table SI-1 and followed a similar trend. Since the matrix can contain a myriad of compounds of low abundance, there are a multitude of peaks that can decrease the similarity between calibrated and theoretical isotopic patterns. Therefore, by rejecting those low abundance peaks that were not part of the isotopic pattern of the compound of interest, spectral accuracy increases. As expected, the effect of interference rejection was more important for larger compounds with more significant M+3 or M+4 peaks such as MTX, JOS and ROX. While the interference rejection is a useful tool to correct spectral accuracy calculations in MassWorks, it must be used with care and only when known interferences of *m/z* value close to the compound of interest co-elute. Otherwise use of this feature can lead to wrong conclusions in the identification process.

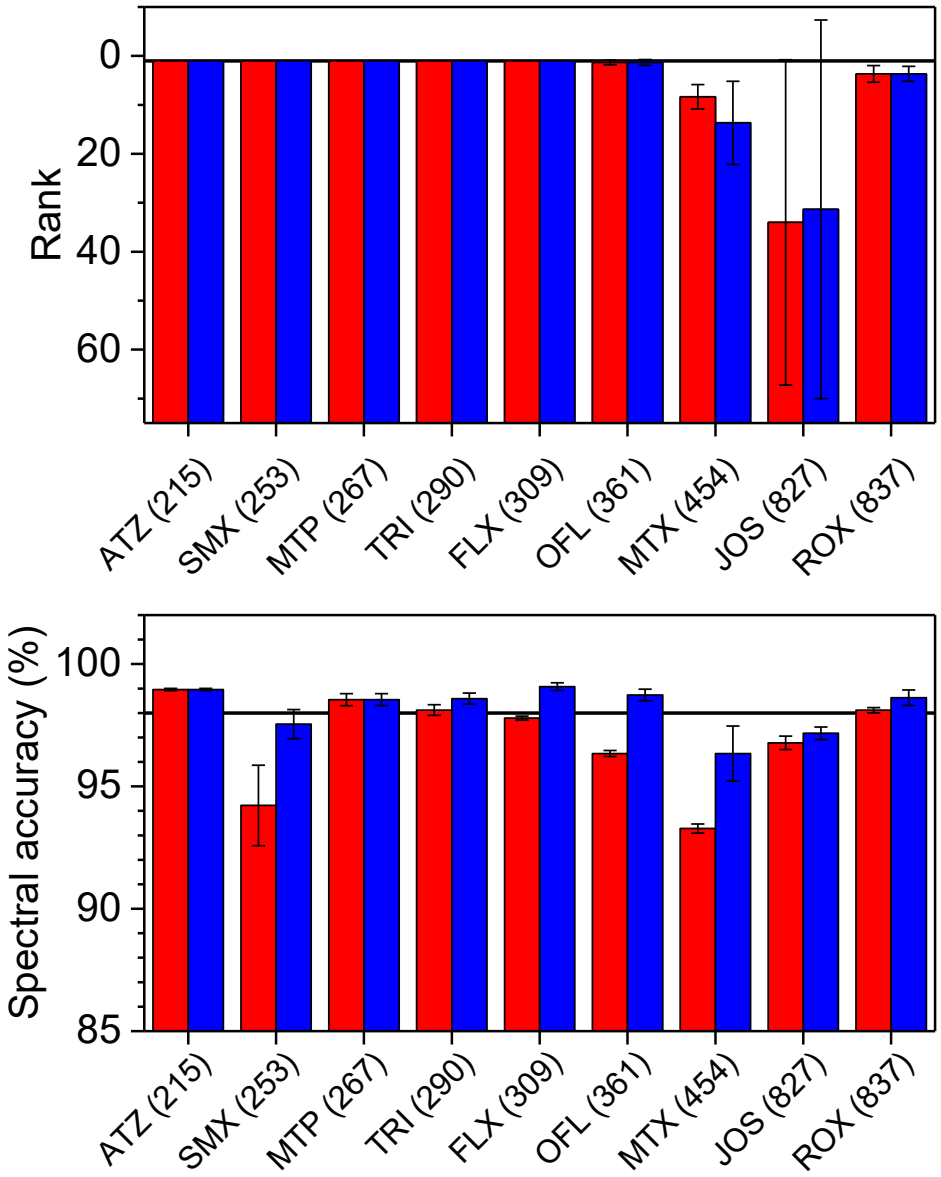


Figure 37. Rank and spectral accuracies measured with a QqTOFMS in the 80 µg L^{-1} matrix solutions with (blue) and without (red) interference rejection. Straight lines indicate: expected value for ranking, 1st (top) and threshold of high spectral accuracy, 98%(bottom).

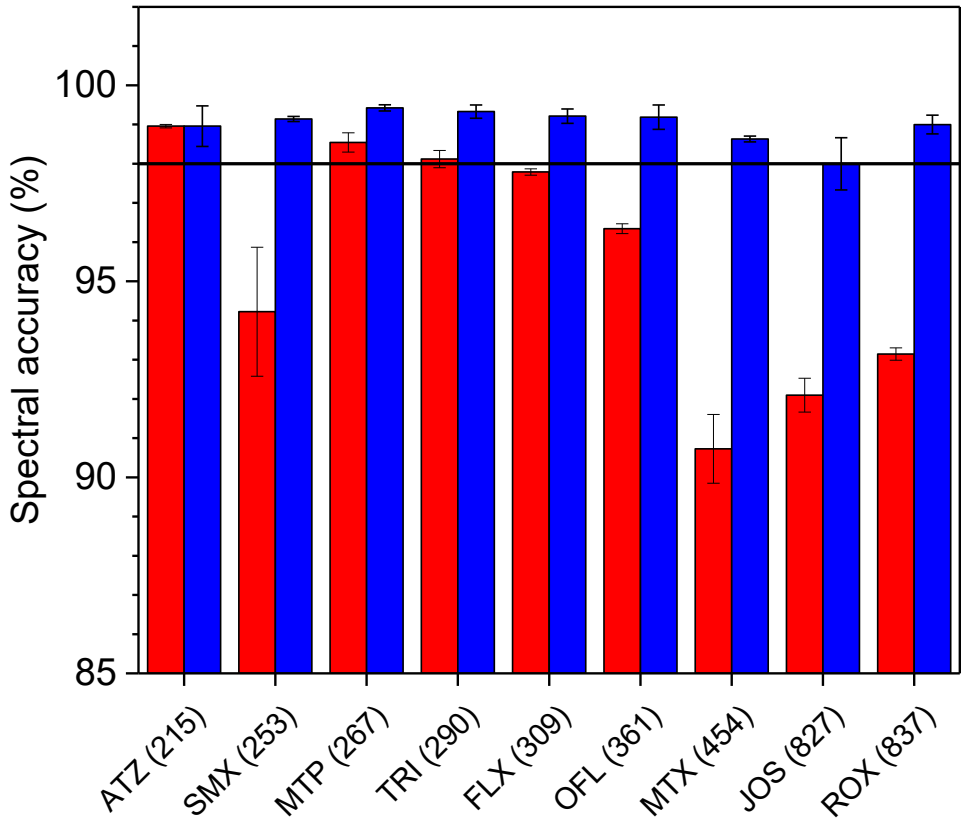


Figure 38. Spectral accuracies measured with a QqTOFMS in the 300 µg L^{-1} matrix solutions with (blue) and without (red) interference rejection. The straight line indicates the threshold of high spectral accuracy, 98%.

Table 17. Impact of interference rejection on correct formula ranking.

Compound	**Matrix spiked at 300 μg L^{-1}**	
	Without interference rejection	With interference rejection
ATZ (216)	1	1
SMX (254)	1	1
MTP (268)	1	1
TRI (291)	1	1
FLX (310)	1	1
OFL (362)	1	1
MTX (455)	2 ± 1	2 ± 1
JOS (827)	11 ± 12	13 ± 14
ROX (837)	2 ± 1	2 ± 1

Table 18. Determination of matrix effects in data acquired with the three mass spectrometers

Compound	Order of elution	Matrix/MeOH areas ratio 300 µg L^{-1} QqQMS	Matrix/ MeOH areas ratio 80 µg L^{-1} QqTOFMS (R_{FHWM}=25 K)	Matrix/ MeOH areas ratio 300 µg L^{-1} QqTOFMS (R_{FHWM}=25 K)	Matrix/ MeOH areas ratio 80 µg L^{-1} QqOrbitrapMS (R_{FHWM}=70 K)	Matrix/ MeOH areas ratio 300 µg L^{-1} QqOrbitrapMS (R_{FHWM}=70 K)	Matrix/ MeOH areas ratio 80 µg L^{-1} QqOrbitrapMS (R_{FHWM}=140 K)	Matrix/ MeOH areas ratio 300 µg L^{-1} QqOrbitrapMS (R_{FHWM}=140 K)
ATZ (215)	8	1.2 ± 0.1	0.832 ± 0.003	0.7 ± 0.1	0.96 ±0.03	1.04 ± 0.01	0.94 ± 0.03	1.04 ± 0.04
SMX (253)	5	2.8 ± 0.4	0.6 ± 0.4	0.49 ± 0.01	1.03 ± 0.02	1.02 ± 0.02	0.95 ± 0.02	0.99 ± 0.02
MTP (267)	4	NA	0.7 ± 0.1	0.78 ± 0.05	0.89 ± 0.01	1.02 ± 0.02	0.81 ± 0.02	1.03 ± 0.01
TRI (290)	2	1.6 ± 0.2	0.62 ± 0.02	0.72 ± 0.04	1.01 ± 0.04	1.09 ± 0.04	0.99 ± 0.03	1.02 ± 0.03
FLX (309)	6*	NA	0.9 ± 0.1	0.72 ± 0.03	0.93 ± 0.02	1.07 ± 0.02	0.87 ± 0.02	1.12 ± 0.05
OFL (361)	3	NA	2.8 ± 0.4	20 ± 2	4.9 ± 0.3	7.0 ± 0.2	6.1 ± 0.2	8.8 ± 0.3
MTX (454)	1	1.5 ± 0.2	3.6 ± 0.3	1.1 ± 0.2	1.04 ± 0.04	1.07 ± 0.02	1.04 ± 0.07	1.07 ± 0.04
JOS (827)	7	NA	0.6 ± 0.2	0.36 ± 0.02	0.83 ± 0.03	1.06 ± 0.03	0.79 ± 0.02	1.01 ± 0.02
ROX (837)	6*	0.90 ± 0.02	0.9 ± 0.1	0.72 ± 0.03	1.03 ± 0.04	1.05 ± 0.02	0.94 ± 0.04	1.07 ± 0.04

* These two compounds co-eluted. NA: Not available.

9.2.4 Software comparison

9.2.4.1 Methods

Molecular Formula Finder of the ChemCalc web application [2] (http://www.chemcalc.org/mf_finder) was used to determine the number of total formulas corresponding to specific accurate masses measured in the QqQMS. To perform a reasonable comparison, the parameters used were as similar as possible as those used for MassWorks. For example, allowed elements and their number were determined by MassWorks based on the seven golden rules and varied depending on the compound. Double bond equivalents (0 to 999) and reference values (2012) were the default values. Mass error (tolerance) was determined experimentally according to mass accuracy and values were between 1 and 4 mDa.

For QqTOFMS data, the built-in tool for Bruker's Data Analysis Smart Formula was used for formula determination. All parameters were equivalent for Smart Formula and MassWorks in order to have a meaningful comparison.

9.2.5 Molecular formula finder and MassWorks using QqQMS data

One of the goals of using spectral accuracy is to reduce the number of potential formulas corresponding to an accurate mass within a given mass accuracy. Using data of the matrix spiked at 300 μg L^{-1}, the results obtained with MassWorks were compared to a tool that generates formulas only from mass accuracy, Molecular Formula Finder. As shown in Table S-3, results indicated that MassWorks allows a significant reduction (up to 96%) in the number of potential molecular formulas and lead to drastically improvement of the rankings. For example, accurate mass in the QqQMS of the protonated ion of TRI for the three replicates was *m/z* 291.1492, 291.1467 and 291.1454. Using those values and parameters indicated previously, Molecular Formula Finder listed 8380 ±45 possible formulas while MassWorks, based on both spectral accuracy and mass accuracy, only returned 336 ± 15.

It should be highlighted that MassWorks gives the ability to perform accurate mass measurements with a system that is not designed for this kind of experiments. Indeed, the QqQMS used is neither a high resolution nor a high-end mass analyzer with accurate mass capabilities. Therefore, the QqQMS instrument used would not be able to perform accurate mass measurements without the MassWorks software. Although the results were greatly improved using MassWorks, the obtained rankings were not good enough to allow formulae determination with a high degree of certitude. It is also important to keep in mind that these measurements were not possible with low concentration tested, 80 μg L^{-1}, spiked in the river extract. Therefore, signal intensity for the QqQMS data was critical. For example, MTX obtained a low ranking (260 ± 120) because signal had a low signal-to-noise ratio.

Table 19. Number of possible formulas and ranking of the test compounds in ChemCalc and MassWorks for the results obtained with the QqQMS using the matrix spiked at 300 μg L^{-1}.

Compound	**Number of possible formulas**		**Rank**	
	ChemCalc	MassWorks	ChemCalc	MassWorks
ATZ (215)	1662 ± 10	93 ± 6	230 ± 140	3 ± 2
SMX (253)	4763 ± 12	554 ± 21	360 ± 100	130 ± 100
TRI (290)	8380 ± 45	336 ± 15	>1000	9 ± 7
MTX (454)	142400 ± 1400	4978 ± 341	>1000	260 ± 120
ROX (837)	13800 ± 200	4485 ± 49	>1000	49 11

9.2.6 Smart Formula and MassWorks using QqTOFMS data

Data acquired with the QqTOFMS using the test compounds spiked in the matrix were processed with SmartFormula, an algorithm developed by Bruker that uses a similar approach to MassWorks to determine the best molecular formula match. While details about the SmartFormula algorithm were not provided, it compares experimental and theoretical isotopic patterns without MS peak shape calibration by calculating a statistical match factor, the Sigma value [3]. Therefore, formulas associated to lower Sigma values are ranked higher since their theoretical isotopic patterns are

more similar to the experimental isotopic pattern. As can be seen in Table S-4, both MassWorks and Smart Formula consistently ranked first the correct formula for test compounds < 350 Da (ATZ, FLX, MTP, SMX and TRI) at 80 μg L^{-1}. MassWorks showed significantly better formula ranking for compounds with molecular mass > 350 Da at 80 and 300 μg L^{-1} except for JOS at 300 μg L^{-1} (11 ± 12) which had a larger standard deviation than the ranking obtained with SmartFormula (11 ± 4). An inaccurate ranking value for a single injection was the cause of the high standard deviation observed for MassWorks. This latter value was due a significant discrepancy between the experimental and theoretical M+1 peaks, an effect lost in the ranking with Smart Formula. Based on these results, the performance of the formula determination algorithm of MassWorks was superior than the algorithm used by SmartFormula, since the former obtained better ranking of the correct formula of larger compounds (>350 Da) and it was more robust when using lower intensity signals.

Table 20. Accurate formula ranking with MassWorks and Smart Formula of the target compounds spiked in the river matrix.

Compound	Spiked at 80 μg L^{-1}		Spiked at 300 μg L^{-1}	
	MassWorks rank	Smart Formula rank	MassWorks rank	Smart Formula rank
ATZ (216)	1	1	1	1
SMX (254)	1	1	1	1
MTP (268)	1	1	1	1
TRI (291)	1	1	1	1
FLX (310)	1	1	1	1
OFL (362)	1 ± 1	4 ± 4	1	2 ± 1
MTX (455)	8 ± 2	38 ± 21	2 ± 1	8 ± 2
JOS (827)	34 ± 33	102 ± 73	11 ± 12	11 ± 4
ROX (837)	4 ± 2	34 ± 19	2 ± 1	19 ± 14

9.2.7 References

(1) Gu, B.; Wang, Y. *Spectroscopy* **2008**.

(2) Patiny, L.; Borel, A. *J. Chem. Inf. Model.* **2013**, *53*, 1223-1228.

(3) Bristow, T.; Constantine, J.; Harrison, M.; Cavoit, F. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1213-1222.

9.3 ANNEXE C – Informations supplémentaires du CHAPITRE 5

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, Montreal, Canada

9.3.1 Experimental section

9.3.1.1 Reagents and standards

Analytical standards of atenolol, atorvastatin, azithromycin, benzoylecgonine, caffeine, carbamazepine, cetirizine, citalopram, cocaine, N,N-diethyl-meta-toluamide, denatonium, O-desmethylvenlafaxine, diltiazem, diphenhydramine, fexofenadine, irbesartan, 3,4-methylenedioxymethamphetamine, methadone, octaethylene glycol, oxazepam, pentaethylene glycol, quetiapine fumarate, telmisartan, temazepam, tris(2-butoxyethyl) phosphate, triton X-100, valsartan and venlafaxine were purchased from Sigma Aldrich (St-Louis, Missouri, USA). All were of certified reference material grade except for Triton X-100 which was laboratory grade.

9.3.1.2 Sampling location

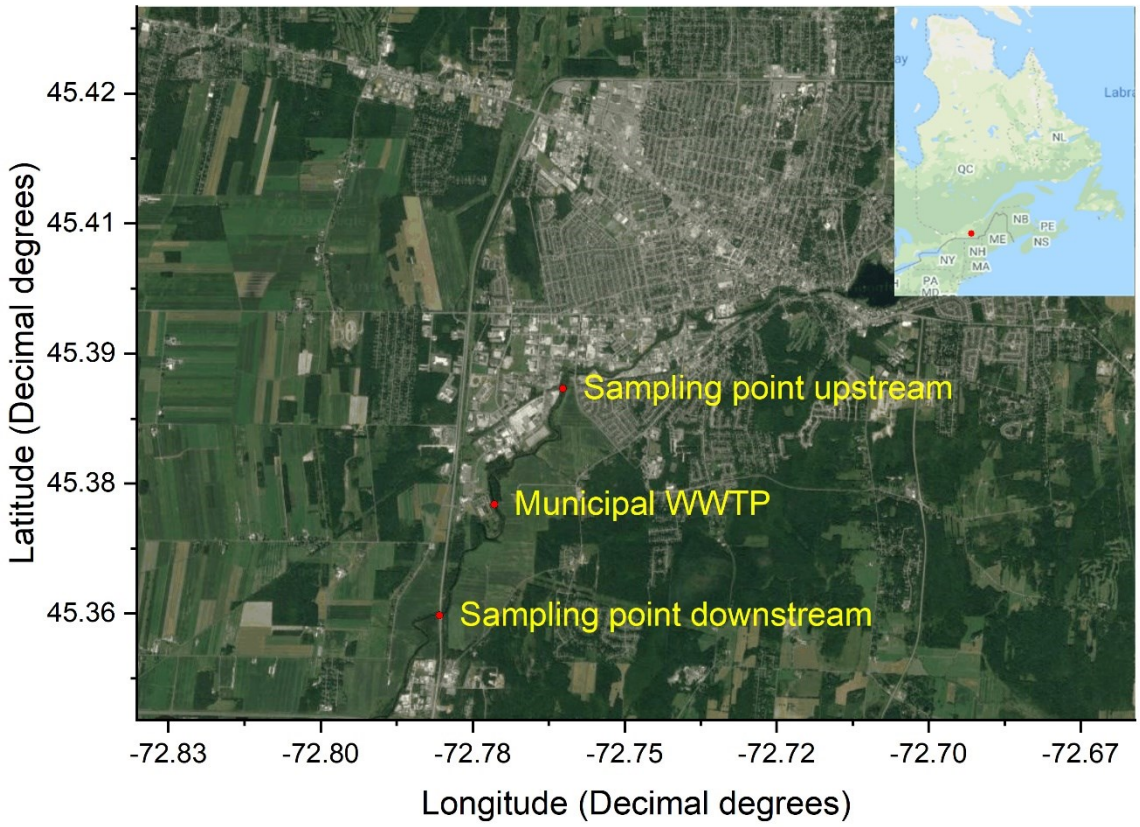


Figure 39. List of the sampling points in the Yamaska River in Granby.

9.3.1.3 Instruments and method

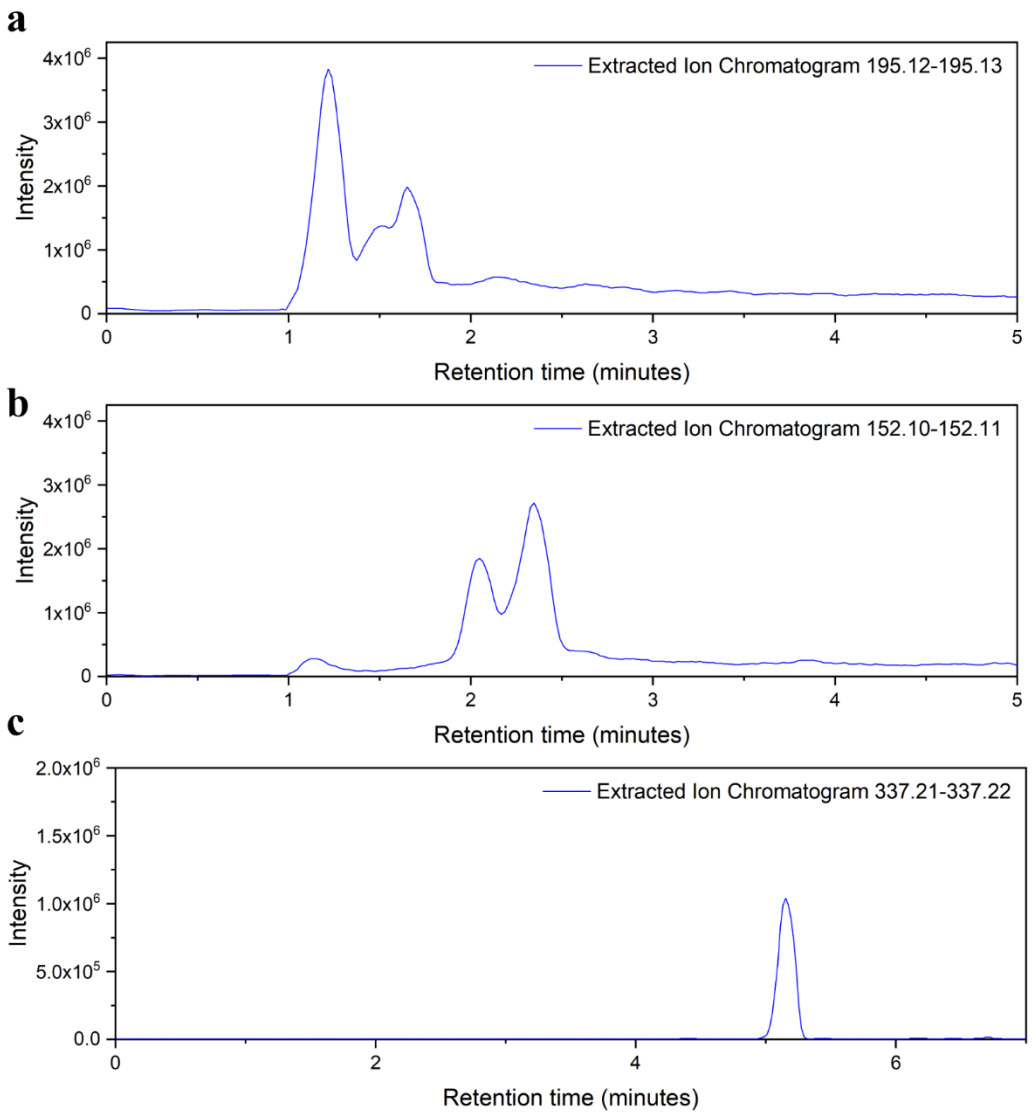


Figure 40. Extracted Ion Chromatograms of tetraethylene glycol (a), norephedrine (b) and acebutolol (c). Peak distortion was observed for tetraethylene glycol and norephedrine. The effect of MeOH as sample reconstitution solvent ceased to matter for acebutolol at about 5 minutes into to the chromatography (approximatively 13% (v/v) of MeOH in the mobile phase).

Injecting the samples in an aqueous solvent would have led to better peak shapes at the beginning of the chromatography. The concentration factor of 400, however, as for the more hydrophobic compounds would have precipitated. Reducing the volume of injection was also considered in this effect to reduce peak decoupling; nevertheless, a higher signal intensity for a wide array of compounds was favored at the expense of the chromatographic integrity in the first minutes.

9.3.1.4 Data conversion and processing for SPS

MSConvertGUI from the ProteoWizard tool Suite [1] was used to convert data files from vendor format (.raw) to ms1 and mgf formats. These formats contain the data for the MS^1 and MS^2 parts of the analysis respectively. The conversion parameters were set as follow: for ms1 conversion, peak picking was vendor msLevel=1-1; threshold peak filter was count 600 most-intense; binary encoding precision was 32-bit, "write index" and "TPP compatibility" boxes were checked but "use zlib compression" and "package in gzip" were left unchecked. For mgf conversion, the similar boxes were checked and unchecked respectively while peak picking was vendor msLevel=2-2 and threshold peak filter was count 200 most intense.

Both ms1 and mgf files were then compressed together into a zip folder and uploaded to an Amazon Web Service S3 folder for the SPS algorithm from MathSpec Inc. (USA) to process. Once processed, the results were saved into two comma-separated values (csv) files. Then a background subtraction from the blank was performed with an Excel add-in. The process was as follow: after being treated by the SPS algorithm, contents from the selected background file and the one the subtraction is to be performed on are pasted on an Excel sheet. The add-in then compares features from both files and subtracts entries that are less than 5 times as abundant in the sample file than in the background one. The subtracted results were then imported with Microsoft Access into a database template. A schematized workflow is presented in Figure SI-2.

Molecules for which there was a match had a hyperlink redirect to their Compound Summary on PubChem. Database matches were shown with a similarity score and number of synonyms (alternative names for the same compound), both of which were used to determine the compound's identity. Matches originating from only one spectrum were not considered as valid. For a chemical to be assigned a probable structure, a minimum of two spectra within 10 seconds of each other with the same precursor ion were required.

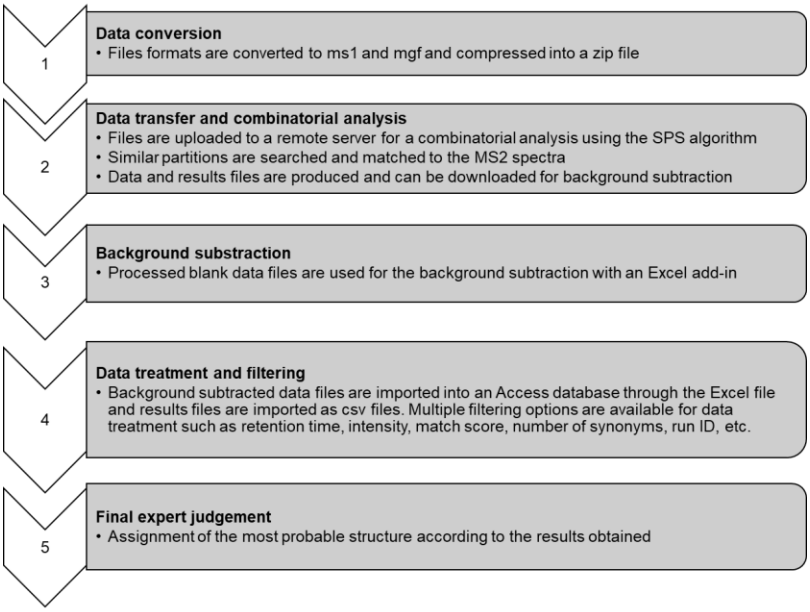


Figure 41. Data processing of DDA acquisition files for combinatorial analysis using the Similar Partition Search algorithm.

9.3.1.5 Data conversion and processing for GNPS

Raw DDA files were converted with MSConvert into mzXML format and uploaded to the GNPS server with FileZilla where the search and networks were conducted. The conversion parameters were as follow: peak picking was vendor msLevel=1-2; binary encoding precision was 32-bit, "write index" and "TPP compatibility" boxes were checked but "use zlib compression" and "package in gzip" were left unchecked.

GNPS parameters were the following: Precursor Ion Mass Tolerance and Fragment Ion Mass Tolerance were both 0.02 Da as suggested by Quinn et al.[2]. Min pairs cosine as minimal value to form an edge between two nodes was 0.7 Cosine; Network top K was 10; Maximum Connected Component Size (Beta) was 0; Minimum Matched Fragment Ions was 5 and Minimum Cluster size was 1. Networking was inspired based on articles by Bouslimani et al. [3, 4]. Library research was also performed after the networking. Parameters for the library identification were as follows: Library Search Min Matched Peaks was 4; Score Threshold was 0.7; analogs were not searched, and Maximum Analog Search Mass Difference was 100.00.

Once the files were processed by GNPS, the files were downloaded and treated with Cytoscape, a non-profit and open-source software platform (available at https://cytoscape.org/download.html) for visualizing complex networks and integrating these with any type of attribute data [5].

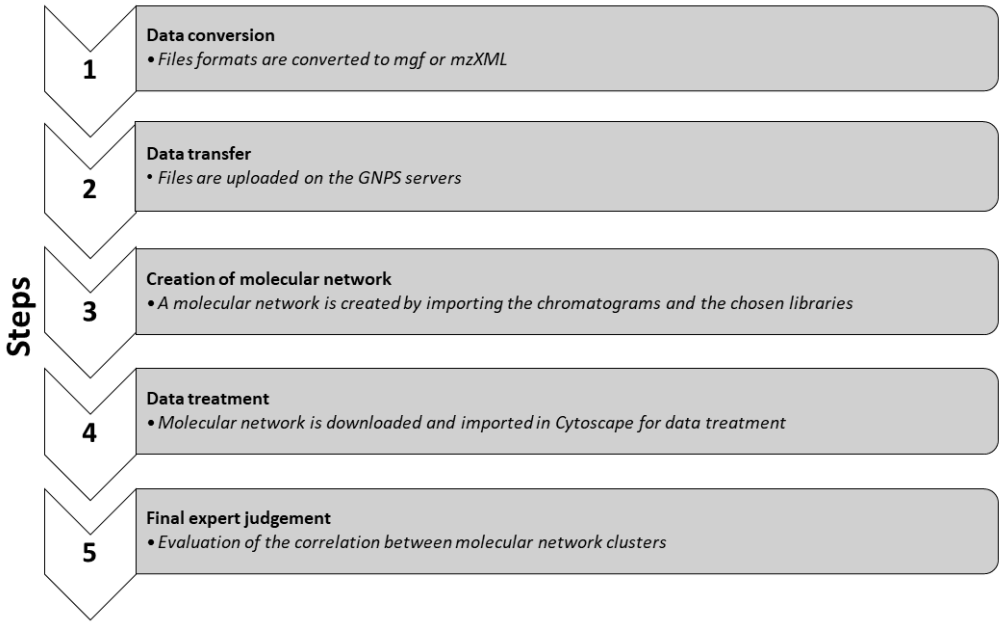


Figure 42. Data processing of DDA acquisition files for analysis using the GNPS molecular networks platform.

9.3.1.6 Data conversion and processing for MetFrag

The following PatRoon script was used for the MetFrag workflow:

```
## Script automatically generated on Mon Nov 11 10:05:01 2019
library(patRoon)

# --------------------------
# initialization
# --------------------------
workPath <- "D:/QE/rfiles/Yamaska/2020-03-12"
setwd(workPath)
getwd()
# Take example data from patRoonData package (triplicate solvent blank + triplicate standard)
Yamaska <- generateAnalysisInfo(paths = "D:/QE/2020/Yamaska/2020-03-03_QE/mzML",
                                groups = c(rep("B", 1), rep("GM", 1), rep("GV", 1)),
                                blanks = "B")
# --------------------------
# features
# --------------------------

# Find all features.

# NOTE: see XCMS manual for many more options
fList <- findFeaturesXCMS3(Yamaska, param = xcms::CentWaveParam(ppm=5, peakwidth = c(4,20), noise
= 50000), verbose = TRUE)

# Group and align features between analysis
fGroups <- groupFeatures(fList, "xcms", rtalign = TRUE, retcorArgs = list(method = "obiwarp"))
```

```
# Basic rule based filtering
fGroups <- filter(fGroups, preAbsMinIntensity = 50000, absMinIntensity = 50000,
                  relMinReplicateAbundance = 1, maxReplicateIntRSD = 0.75,
                  blankThreshold = 5, removeBlanks = TRUE,
                  retentionRange = c(10, 1260), mzRange = c(120, 1000))
tableoffeatures <- as.data.table(fGroups)
write.csv(tableoffeatures, file = 'tableoffeatures.csv')

# -------------------------
# annotation
# -------------------------

# Retrieve MS peak lists
avgPListParams <- getDefAvgPListParams(clusterMzWindow = 0.005)
mslists <- generateMSPeakLists(fGroups, "mzr", maxMSRtWindow = 5, precursorMzWindow = 5,
                               avgFeatParams = avgPListParams, avgFGroupParams = avgPListParams)
avgPListParams <- getDefAvgPListParams(clusterMzWindow = 0.005,
                                       topMost = 10, avgFun = mean, pruneMissingPrecursorMS = TRUE,
                                       method = "hclust", retainPrecursorMSMS = TRUE)
mslists <- generateMSPeakLists(fGroups, "mzr", maxMSRtWindow = 10, precursorMzWindow = 10,
                               avgFeatParams = avgPListParams, avgFGroupParams = avgPListParams)

# Calculate formula candidates
formulas <- generateFormulas(fGroups, "genform", mslists, relMzDev = 5, hetero = TRUE, adduct =
"[M+H]+",
                             elements = "CHNOFPSCl",  MSMode = "both", isolatePrec = TRUE, oc =
TRUE,
                             timeout = 120, topMost = 5, calculateFeatures = FALSE, featThreshold
= 0.75,
                             maxProcAmount = getOption("patRoon.maxProcAmount"),batchSize = 25)

# Find compound structure candidates
compToxCompounds <- generateCompounds(fGroups, mslists, "metfrag", database = "comptox", adduct =
"[M+H]+")
compounds <- addFormulaScoring(compToxCompounds, formulas, TRUE)
table.of.compounds <- as.data.table(compToxCompounds)
write.csv(table.of.compounds, file = "tableOfCompounds.csv")
```

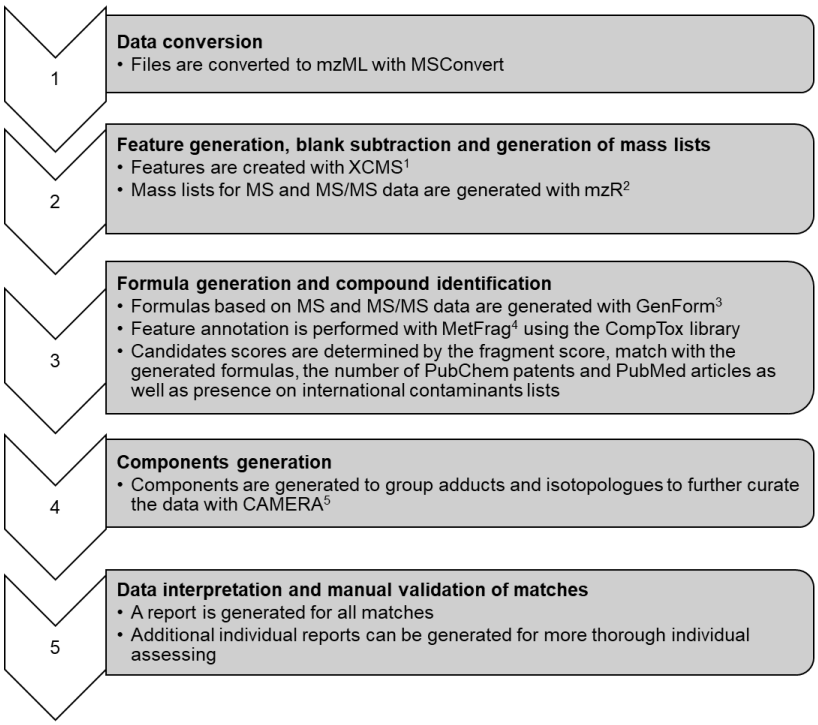


Figure 43. Data processing of DDA acquisition files for analysis using the MetFrag workflow.

9.3.2 Results

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

Table 21. Compounds verified with reference standards in MS^2 mode

Probable structure	Precursor (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)	Product ions (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)
Octylphenol ethoxylate-9	625.3925	625.3937	-1.20	331.1744	331.1743	0.10
				567.3160	567.3138	2.20
				419.2271	419.2267	0.40
				375.1979	375.2012	-3.30
				347.1677	347.1683	-0.60
				303.1439	303.1410	2.90
				287.1479	287.1465	1.40
				271.1530	271.1518	1.20
				243.1199	243.1221	-2.20
				215.0899	215.0916	-1.70
				199.0945	199.0955	-1.00
Pentaethylene glycol	239.1497	239.1489	0.80	73.0655	73.0648	0.70
				87.0446	87.0441	0.50
				89.0603	89.0597	0.60
				107.0707	107.0703	0.40
				133.0860	133.0859	0.10
				151.0964	151.0965	-0.10
				177.1122	177.1121	0.10
				195.1232	195.1227	0.50
Octaethylene glycol	371.2280	371.2276	0.40	87.0447	87.0441	0.60
				89.0603	89.0597	0.60
				91.0762	91.0754	0.80
				107.0706	107.0703	0.30
				131.0703	131.0703	0.00
				133.0860	133.0859	0.10
				177.1123	177.1121	0.20
				221.1386	221.1384	0.20
				239.1487	239.1489	-0.20
				283.1755	283.1751	0.40
				327.2019	327.2013	0.60

Probable structure	Precursor (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)	Product ions (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)
Atenolol	267.1708	267.1704	0.46	56.0506	56.0505	0.08
				72.0818	72.0817	0.10
				74.0610	74.0609	0.08
				86.0973	86.0974	-0.11
				98.0970	98.0970	0.00
				116.1076	116.1074	0.18
				134.1180	134.1179	0.09
				145.0660	145.0650	1.05
				164.0713	164.0704	0.92
				178.0872	178.0863	0.82
				180.1021	180.1019	0.29
				190.0867	190.0864	0.34
				208.0964	208.0971	-0.69
				225.1244	225.1235	0.87
Atorvastatin	559.2628	559.2609	1.91	440.2248	440.2239	0.87
				422.2123	422.2125	-0.19
				466.2036	466.2022	1.34
				441.2275	441.2303	-2.78

Probable structure	Precursor (river sample)	Precursor ion (pure standard)	Mass accuracy (mDa)	Product ions (river sample)	Precursor ion (pure standard)	Mass accuracy (mDa)

	(*m/z*)	(*m/z*)		(*m/z*)	(*m/z*)	
Azithromycin	375.2635	375.2587	4.78	83.0502	83.0492	1.01
				115.0761	115.0769	-0.75
				116.1077	116.1071	0.52
				117.1108	117.1108	0.03
				127.0760	127.0756	0.42
				158.1181	158.1183	-0.23
				186.1504	186.1483	2.03
				398.2918	398.2907	1.07
				416.3031	416.3020	1.09
				434.3128	434.3102	2.69
				573.4142	573.4107	3.50
				591.4237	591.4214	2.34
Benzoylecgonine	290.1398	290.1387	1.16	82.0663	82.0660	0.30
				84.0814	84.0818	-0.42
				84.9606	84.9604	0.21
				93.0706	93.0705	0.11
				100.0762	100.0760	0.21
				105.0342	105.0340	0.16
				119.0500	119.0498	0.24
				124.1127	124.1123	0.44
				150.0920	150.0916	0.43
				168.1024	168.1019	0.43
				267.0590	267.0583	0.64
				272.1284	272.1280	0.37
Carbamazepine	237.1028	237.1016	1.19	192.0813	192.0806	0.70
				194.0970	194.0975	-0.51
				195.1007	195.0987	1.94

Probable structure	Precursor (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)	Product ions (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)
Cetirizine	389.1637	389.1620	1.69	166.0780	166.0804	-2.38
				201.0472	201.0477	-0.50
				202.0512	202.0495	1.71
Citalopram	325.1721	325.1714	0.70	109.0456	109.0446	1.03
				116.0502	116.0503	-0.07
				123.0247	123.0246	0.13
				144.0450	144.0440	0.95
				156.0815	156.0805	0.96
				166.0657	166.0663	-0.60
				184.0763	184.0767	-0.37
				221.0644	221.0651	-0.64
				222.0714	222.0725	-1.07
				234.0720	234.0722	-0.22
				236.0877	236.0890	-1.31
				238.0667	238.0678	-1.09
				242.0963	242.0979	-1.56
				247.0794	247.0792	0.29
				250.1033	250.1022	1.05
				260.0871	260.0895	-2.41
				262.1034	262.1029	0.42
				280.1141	280.1130	1.10
				307.1613	307.1613	-0.01

Probable structure	Precursor (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)	Product ions (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)
Cocaine	304.1555	304.1546	0.88	53.0486	53.0516	-2.96
				69.0710	69.0710	-0.02
				82.0660	82.0660	0.00
				84.9608	84.9604	0.39
				93.0708	93.0707	0.12
				97.0656	97.0652	0.33
				99.0448	99.0444	0.37
				105.0344	105.0342	0.24
				113.0603	113.0605	-0.18
				182.1181	182.1178	0.34
				183.1198	183.1211	-1.33
N,N-Diethyl-meta-toluamide (DEET)	192.1388	192.1383	0.49	119.0498	119.0489	0.82
				120.0532	120.0532	0.00
Denatonium	325.2280	325.2277	0.31	58.0662	58.0662	-0.02
				84.0817	84.0816	0.11
				84.9605	84.9606	-0.11
				86.0974	86.0972	0.18
				91.0551	91.0550	0.18
				92.0585	92.0581	0.42
				216.9016	216.9003	1.30
				233.1656	233.1650	0.58
				234.1687	234.1683	0.41
				234.9114	234.9116	-0.14
				260.8919	260.8904	1.43
				278.9014	278.9007	0.70

Probable structure	Precursor (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)	Product ions (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)
O-Desmethylvenlafaxine	264.1966	264.1962	0.37	107.0499	107.0491	0.83
				133.0651	133.0648	0.33
				145.0652	145.0652	0.09
				201.1279	201.1281	-0.15
				202.1302	202.1303	-0.08
				246.1863	246.1855	0.77
Diltiazem	415.1695	416.1680	1.55	121.0655	121.0659	-0.38
				137.0602	137.0602	-0.06
				150.0381	150.0380	0.03
				178.0326	178.0330	-0.42
				179.0359	179.0360	-0.17
				191.0405	191.0412	-0.63
				223.0901	223.0915	-1.35
				310.0905	310.0910	-0.49
				312.1062	312.1065	-0.27
				355.1494	355.1497	-0.27
				370.1128	370.1128	0.01
				373.1616	373.1591	2.55
Diphenhydramine	256.1703	256.1702	0.06	165.0703	165.0705	-0.27
				167.0860	167.0861	-0.12
				168.0895	168.0891	0.38
Fexofenadine	502.2967	502.2933	3.39	129.0697	129.0690	0.64
				131.0858	131.0850	0.83
				143.0853	143.0866	-1.23
				171.1172	171.1170	0.18
				172.1204	172.1193	1.05

				217.1220	217.1251	-3.12
				233.1175	233.1178	-0.28
				246.1488	246.1477	1.07
				250.1596	250.1633	-3.74
				262.1602	262.1608	-0.61
				280.1695	280.1714	-1.88
				466.2756	466.2746	0.98
				467.2803	467.2776	2.70
				484.2862	484.2853	0.97
				485.2906	485.2877	2.88
Irbesartan	429.2405	429.2417	-1.16	167.1544	167.1550	-0.59
				180.0811	180.0819	-0.79
				190.0655	190.0647	0.84
				192.0825	192.0809	1.59
				195.1499	195.1496	0.29
				196.1526	196.1513	1.37
				206.0839	206.0845	-0.58
				207.0924	207.0924	0.02
				208.0959	208.0957	0.23
				235.0982	235.0991	-0.92
				386.2242	386.2241	0.09
				400.2242	400.2258	-1.60
				401.2357	401.2359	-0.14

Probable structure	Precursor (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)	Product ions (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)
3,4-Methylenedioxy methamphetamine (MDMA)	194.1181	194.1173	0.79	56.9659	56.9658	0.15
				58.0663	58.0660	0.24
				59.0502	59.0499	0.39
				68.9335	68.9333	0.26
				71.9529	71.9524	0.49
				72.0818	72.0814	0.36
				89.0605	89.0601	0.38
				107.9607	107.9602	0.51
				108.9618	108.9613	0.53
				109.9594	109.9589	0.50
				111.9689	111.9679	1.09
				126.9722	126.9717	0.54
				127.9698	127.9691	0.69
				135.0445	135.0439	0.64
				136.0469	136.0473	-0.41
				141.9707	141.9703	0.47
				149.0605	149.0595	1.02
				151.9500	151.9494	0.61
				152.9516	152.9508	0.85
				153.9489	153.9481	0.72
				163.0707	163.0751	-4.44
				169.9606	169.9598	0.81
				170.9616	170.9612	0.37
				170.9743	170.9733	0.96
				171.9594	171.9586	0.76

Methadone	310.2163	310.2159	0.46	57.0347	57.0344	0.31
				59.0503	59.0499	0.46
				91.0553	91.0546	0.69
				100.1131	100.1122	0.88
				105.0343	105.0338	0.45
				105.0709	105.0702	0.67
				117.0702	117.0701	0.06
				119.0860	119.0858	0.14
				129.0705	129.0699	0.58
				131.0866	131.0854	1.21
				135.0810	135.0801	0.92
				143.0858	143.0858	0.00
				145.1021	145.1012	0.89
				147.0811	147.0807	0.41
				159.1175	159.1165	0.98
				187.1129	187.1112	1.66
				195.1174	195.1166	0.78
				219.1177	219.1165	1.25
				223.1127	223.1113	1.39
				247.1491	247.1475	1.60
				258.9387	258.9373	1.37
				264.8772	264.8756	1.56
				265.1595	265.1581	1.37
				266.1636	266.1616	2.01
				286.9336	286.9317	1.95
				304.9426	304.9426	0.00

Probable structure	Precursor (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)	Product ions (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)
Oxazepam	287.0588	287.0575	1.28	84.9606	84.9604	0.20
				104.0503	104.0501	0.21
				128.0266	128.0258	0.78
				166.0057	166.0052	0.44
				231.0690	231.0675	1.51
				241.0531	241.0521	1.05
				242.0370	242.0351	1.92
				242.0573	242.0567	0.56
				269.0482	269.0470	1.25
				270.0515	270.0505	0.95
Quetiapine	384.1763	384.1742	2.05	158.1184	158.1173	1.15
				159.1176	159.1156	2.07
				221.1076	221.1069	0.66
				253.0800	253.0797	0.38
				254.0814	254.0824	-0.94
				279.0953	279.0953	0.03
Tris(2-butoxyethyl) phosphate	399.2506	399.2499	0.67	199.0737	199.0739	-0.23
				243.0989	243.0967	2.16
				299.1634	299.1638	-0.48
Telmisartan	515.2459	515.2453	0.64	193.0660	193.0658	0.20
				211.0759	211.0766	-0.73
				276.1371	276.1382	-1.13
				303.1578	303.1606	-2.84
				305.1769	305.1770	-0.15
				306.1796	306.1757	3.88

				317.1792	317.1799	-0.69
				497.2349	497.2348	0.14
				498.2388	498.2365	2.26

Probable structure	Precursor (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)	Product ions (river sample) (*m/z*)	Precursor ion (pure standard) (*m/z*)	Mass accuracy (mDa)
Temazepam	301.0749	301.0733	1.59	84.9607	84.9602	0.47
				180.0215	180.0208	0.63
				228.0577	228.0569	0.81
				236.9224	236.9210	1.43
				255.0690	255.0677	1.36
				256.0740	256.0714	2.56
				283.0643	283.0626	1.65
				284.0671	284.0662	0.95
Valsartan	436.2356	436.2353	0.35	57.0709	57.0707	0.27
				180.0817	180.0804	1.39
				190.0661	190.0646	1.43
				194.0968	194.0961	0.70
				207.0926	207.0914	1.17
				210.0917	210.0910	0.69
				235.0990	235.0980	1.02
				291.1501	291.1485	1.56
				306.1717	306.1707	0.98
Venlafaxine	278.2133	278.2127	0.59	107.0865	107.0851	1.38
				121.0655	121.0657	-0.13
				147.0812	147.0813	-0.12
				159.0806	159.0808	-0.19
				173.0972	173.0962	1.02

				215.1439	215.1442	-0.30
				260.2016	260.2013	0.22

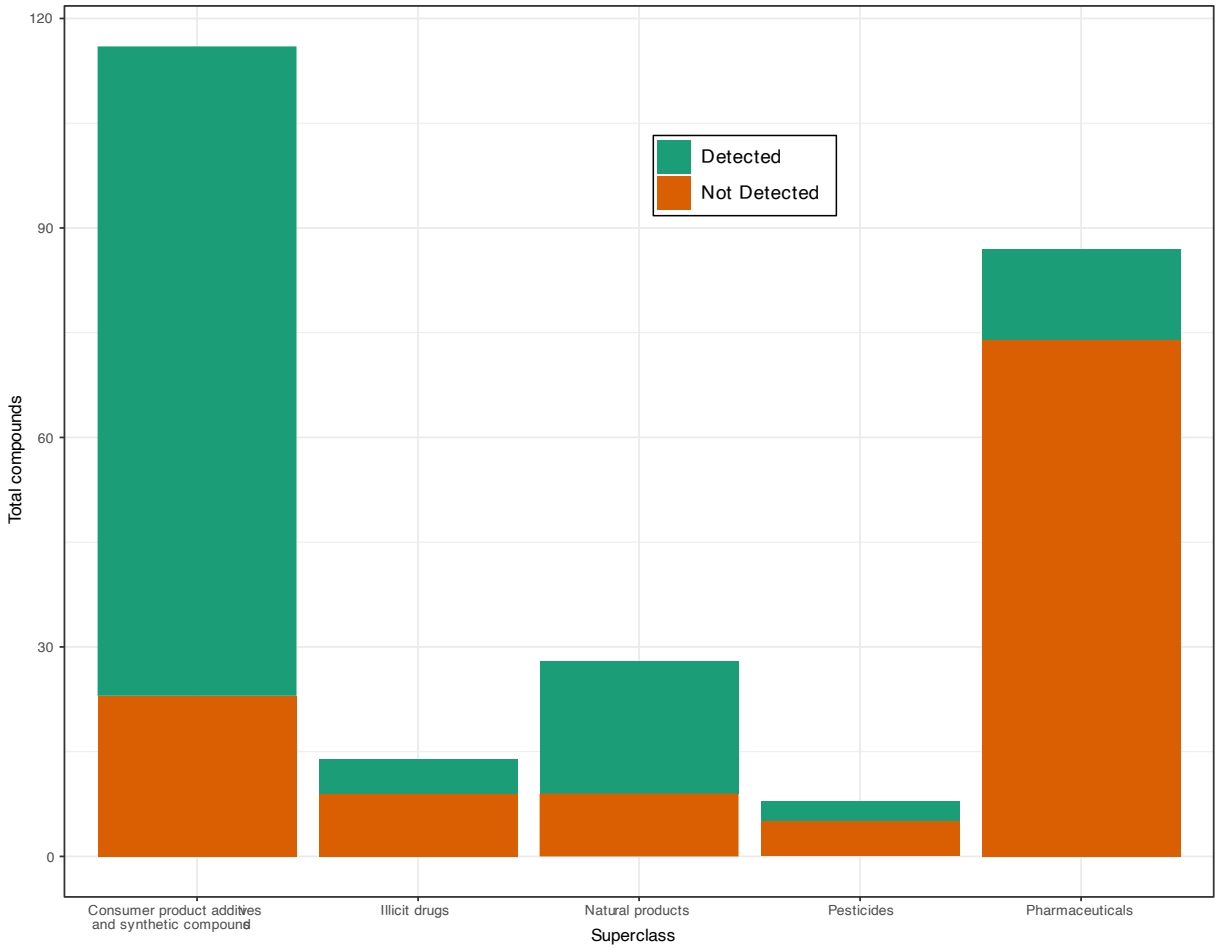
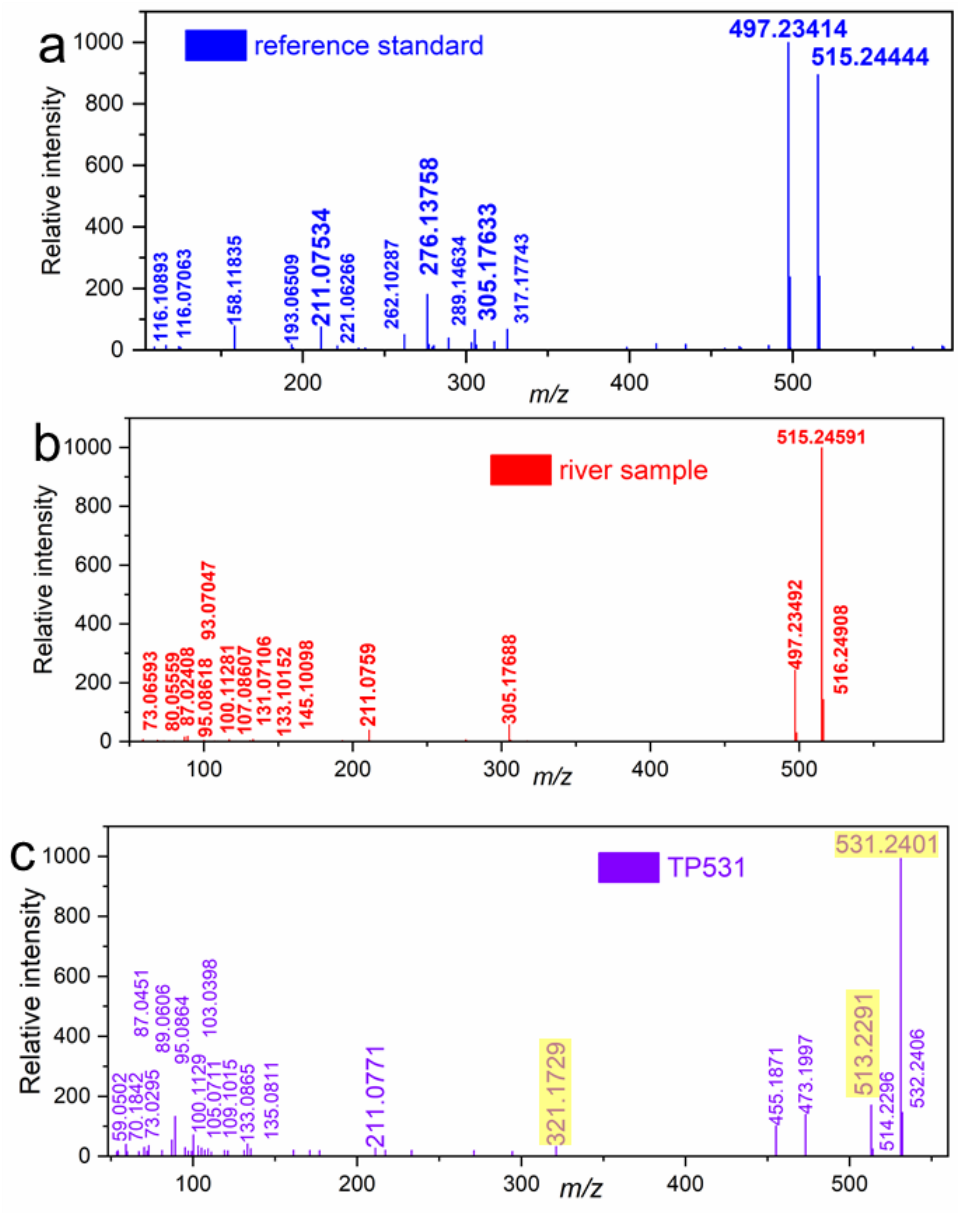


Figure 44. Compounds detected per category upstream Granby's wastewater treatment plant



m/z	Intensity	*m/z*	Intensity	*m/z*	Intensity
158.1184	78	59.0502	7	161.0964	21
211.0753	75	68.9451	5	171.1173	21
262.1029	50	87.0450	16	177.0918	19
276.1376	181	89.0607	18	211.0771	28
289.1463	39	100.1128	5	217.1220	21
303.1616	24	117.0552	7	233.0703	20
305.1763	66	133.0862	8	271.1341	19
317.1774	27	211.0759	40	294.4771	16
325.1710	67	276.1371	6	321.1729	33
416.3025	21	305.1769	58	455.1871	102
497.2341	999	306.1796	5	473.1997	140
497.3970	19	497.2349	243	513.2291	171
498.2371	238	498.2388	30	514.2296	25
515.2444	895	515.2459	999	531.2401	999
516.2478	240	516.2491	143	532.2406	147

Figure 45. Spectra of telmisartan from the reference standard (a), the river sample (b) and of TP531 (c) with the 15 most intense fragment ions in tables. A mass shift of 15.9994 corresponding to a hydroxylation along is highlighted in (c).

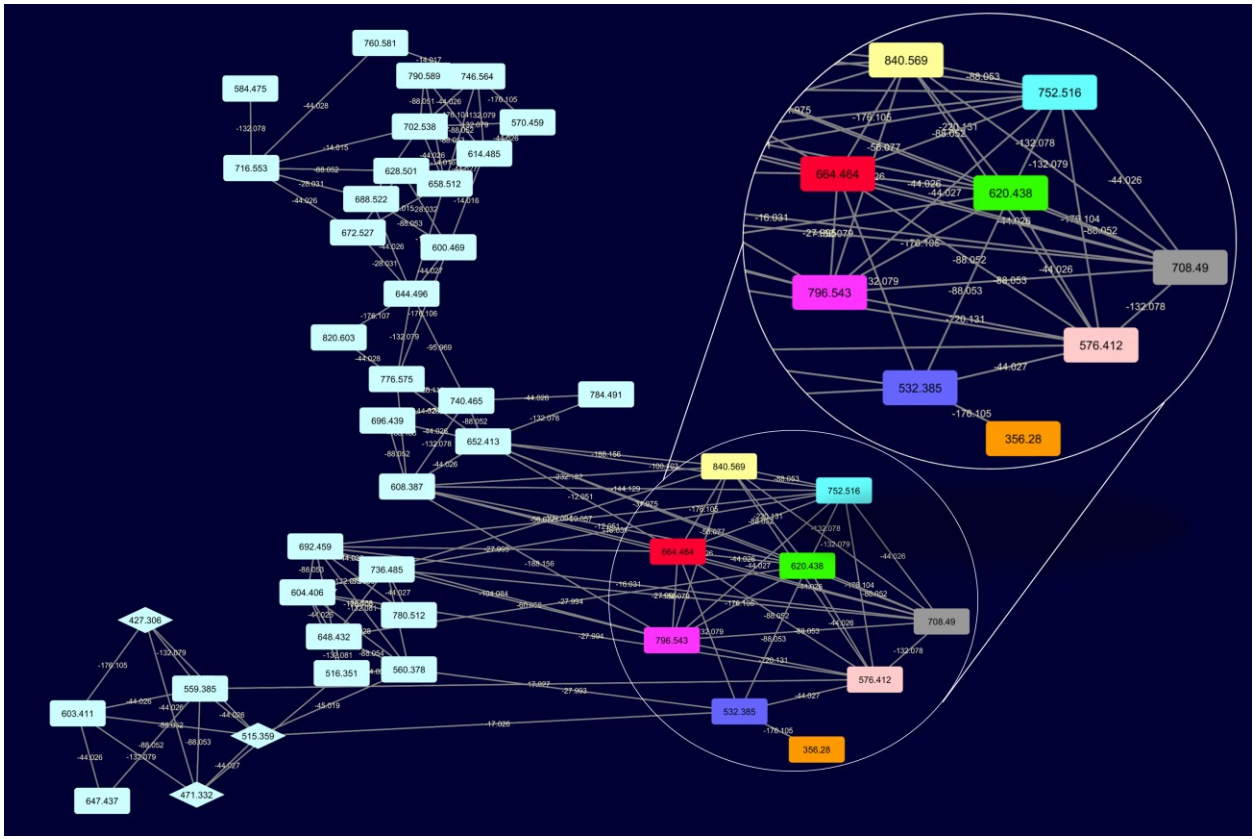
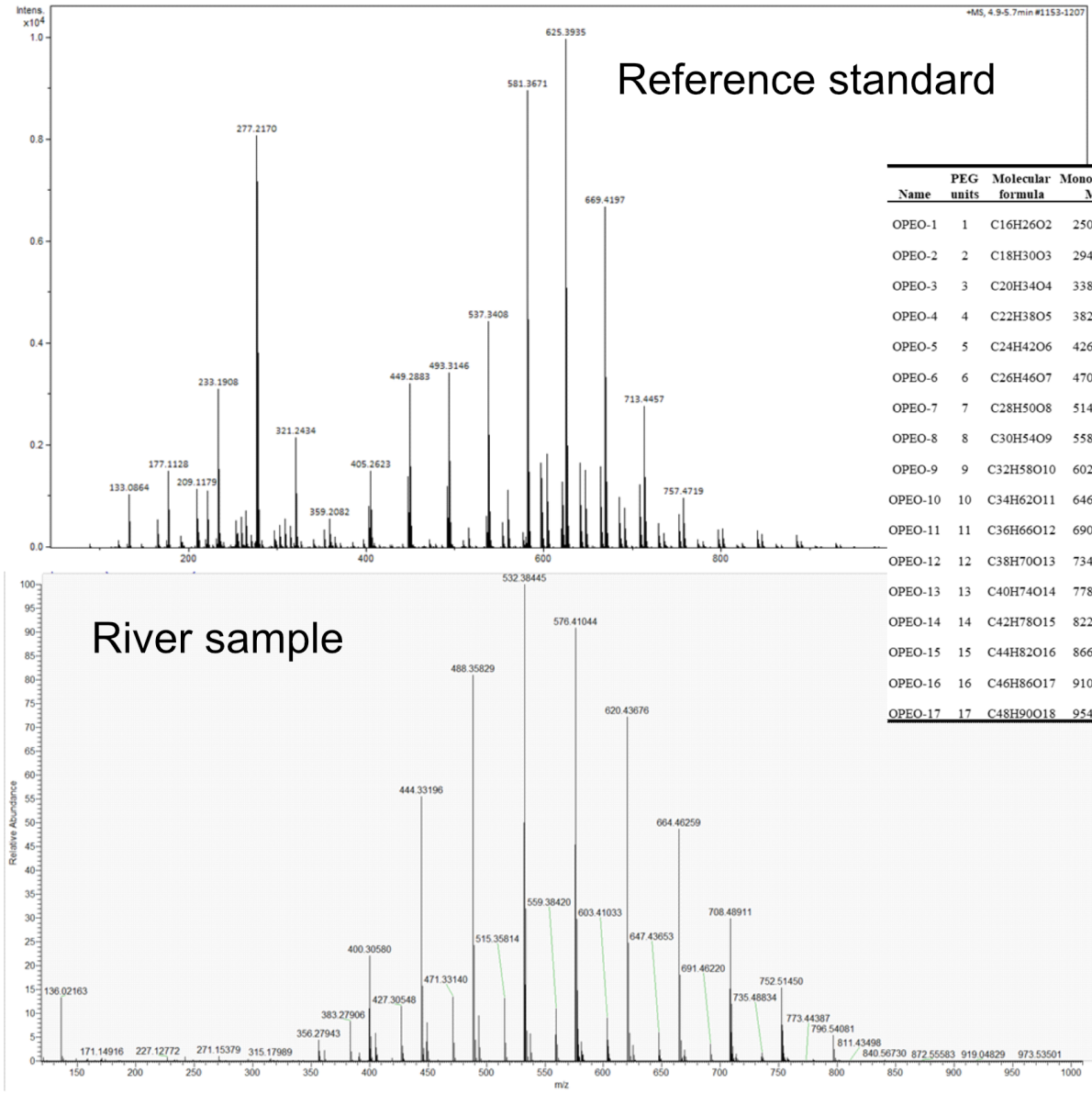
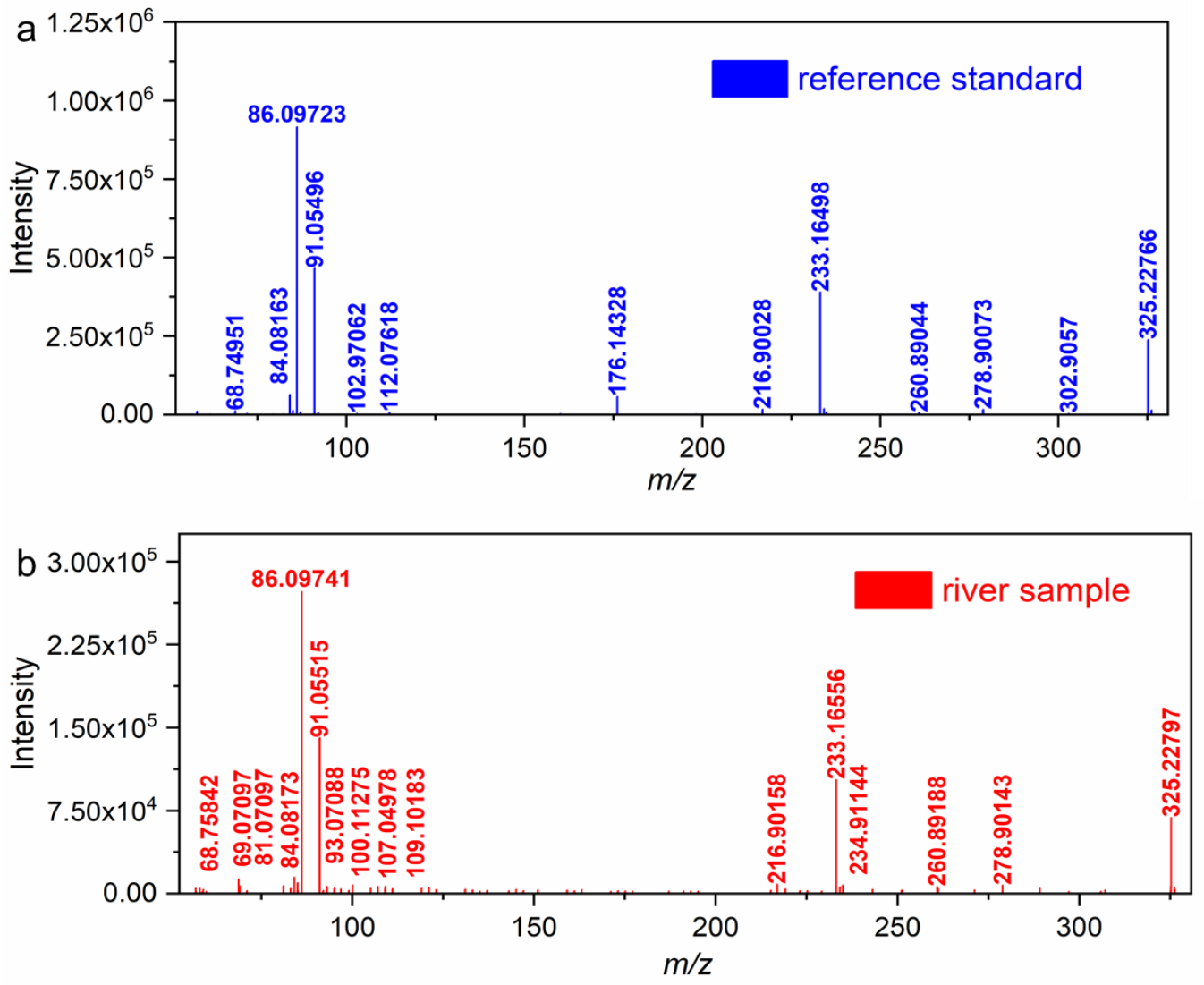


Figure 46. Cluster in GNPS with a zoom on several identified OPEOs. In orange is OPEO-3, in blue OPEO-7, in beige OPEO-8, in green OPEO-9, in red OPEO-10, in grey OPEO-11, in cyan OPEO 12, in purple OPEO 13 and in sand OPEO-14. All had a $[M+NH_4]^+$ adduct.



Name	PEG units	Molecular formula	Monoisotopic MW	M+H	M+Na	M+NH4
OPEO-1	1	C16H26O2	250.1933	251.2006	273.1825	268.2271
OPEO-2	2	C18H30O3	294.2189	295.2262	317.2082	312.2528
OPEO-3	3	C20H34O4	338.2446	339.2519	361.2338	356.2784
OPEO-4	4	C22H38O5	382.2703	383.2776	405.2595	400.3041
OPEO-5	5	C24H42O6	426.2959	427.3032	449.2852	444.3298
OPEO-6	6	C26H46O7	470.3216	471.3289	493.3108	488.3554
OPEO-7	7	C28H50O8	514.3473	515.3546	537.3365	532.3811
OPEO-8	8	C30H54O9	558.3729	559.3802	581.3622	576.4068
OPEO-9	9	C32H58O10	602.3986	603.4059	625.3878	620.4324
OPEO-10	10	C34H62O11	646.4243	647.4316	669.4135	664.4581
OPEO-11	11	C36H66O12	690.4499	691.4572	713.4392	708.4838
OPEO-12	12	C38H70O13	734.4756	735.4829	757.4648	752.5094
OPEO-13	13	C40H74O14	778.5013	779.5085	801.4905	796.5351
OPEO-14	14	C42H78O15	822.5269	823.5342	845.5162	840.5608
OPEO-15	15	C44H82O16	866.5526	867.5599	889.5418	884.5864
OPEO-16	16	C46H86O17	910.5783	911.5855	933.5675	928.6121
OPEO-17	17	C48H90O18	954.6039	955.6112	977.5932	972.6378

Figure 47. MS Spectra of octylphenol ethoxylates from the Triton X-100 reference standard and the river sample. On the right is a table of the different adducts' m/z of each congener.



m/z	Intensity	*m/z*	Intensity
58.0662	10861	68.7584	13008
68.7495	12867	69.0710	7081
84.0816	64248	81.0710	7106
84.9606	13632	84.0817	14751
86.0972	915911	84.9605	9853
87.1006	9557	86.0974	272749
91.0550	466323	91.0551	140755
176.1433	57453	100.1127	8039
216.9003	16989	109.1018	6657
233.1650	391078	216.9016	8396
234.1683	18791	233.1656	103064
234.9116	9161	234.9114	7775
278.9007	16208	278.9014	7708
325.2277	238530	325.1435	7073
326.2315	14290	325.2280	68643

Figure 48. Spectra of denatonium from the reference standard (a) and the river sample (b) with the 15 most intense fragment ions in table.

9.3.3 References

[1] 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. *Nature biotechnology* **2012**, *30*, 918.

[2] 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 in Pharmacological Sciences* **2017**, *38*, 143.

[3] A. Bouslimani, C. Porto, C. M. Rath, M. Wang, Y. Guo, A. Gonzalez, D. Berg-Lyon, G. Ackermann, G. J. M. Christensen, T. Nakatsuji. Molecular cartography of the human skin surface in 3D. *Proceedings of the National Academy of Sciences* **2015**, *112*, E2120.

[4] A. Bouslimani, A. V. Melnik, Z. Xu, A. Amir, R. R. da Silva, M. Wang, N. Bandeira, T. Alexandrov, R. Knight, P. C. Dorrestein. Lifestyle chemistries from phones for individual profiling. *Proceedings of the National Academy of Sciences* **2016**, *113*, E7645.

[5] 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* **2003**, *13*, 2498.

9.4 ANNEXE D – Informations supplémentaires du CHAPITRE 6

Uncovering new transformation products of concerning organic contaminants by photodegradation experiments and analysis of real samples from a local river

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

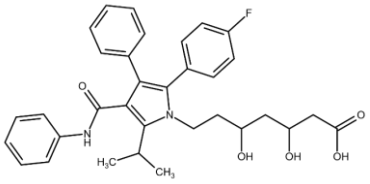
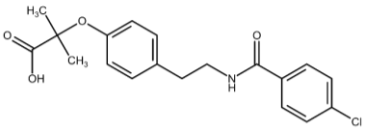
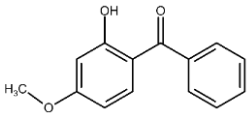
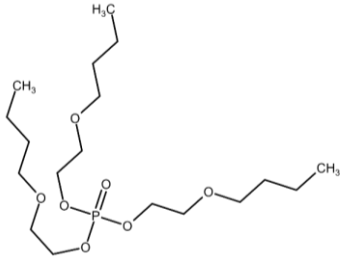
[1]Department of Chemistry, Université de Sherbrooke, Sherbrooke, Canada
[2]Environment and Climate Change Canada, Montreal, Canada
Corresponding author: Pedro A. Segura. pa.segura@usherbrooke.ca

Supplementary Information

9.4.1 Materials and method

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). Triton X-100 (laboratory grade) was purchased from Sigma Aldrich (St-Louis, Missouri). Atorvastatin (ATV), bezafibrate (BEZ), oxybenzone (OXZ) and tris(2-butoxyethyl) phosphate (TBEP) were purchased from Sigma Aldrich. Pahokee peat humic acid was purchased from the International Humic Substances Society (IHSS) (Denver, Colorado).

Table 22. Chemical properties and classification of the selected compounds in the photodegradation study.

Compound	**Acronym**	**Class**	**Molecular formula**	**Monoisotopic mass**	**Structure**	**Log K_{ow}**
Atorvastatin	ATV	Cardiovascular system	$C_{19}H_{20}ClNO_4$	558.2530		6.36
Bezafibrate	BEZ	Cardiovascular system	$C_{33}H_{34}FN_2O_5$	361.1081		3.8
Oxybenzone	OXZ	Cosmetic (sunscreen)	$C_{14}H_{12}O_3$	228.0786		3.79
Tris(2-butoxyethyl) phosphate	TBEP	Polymer additive (flame retardant)	$C_{18}H_{39}O_7P$	398.2433		3.75

9.4.1.1 Photodegradation experiments

9.4.1.1.1 Exposure parameters

At 50 cm, The Exo Terra Solar Glo 125 W lamp emits visual light, UVA, UVB and infrared radiation. Light is emitted at 86,000 lx at the front of the reactors and 12,100 lx which passes through behind them. Knowing that direct sunlight corresponds to 32,000–100,000 lx and a sunny day without direct exposure corresponds to 10,000 to 25,000 lx, one can conclude that exposure of the sample to artificial sunlight corresponds to the bright sun illumination. Exposure to light was timed 12 hours on followed by 12 hours off with a NOMA Outdoor Heavy Duty 24-Setting Timer.

Figure 49. Setup for the photodegradation study.

9.4.1.1.2 Extraction

Table 23. Solid phase extraction (SPE) parameters for the compounds in the photo degradation study

	atorvastatin	*bezafibrate*	*oxybenzone*	*tris(2-butoxyethyl) phosphate*
SPE cartridge type	Oasis MAX	Oasis MAX	Oasis HLB	Oasis HLB
Conditioning	6 mL methanol	6 mL methanol	6 mL methanol	6 mL methanol
Equilibration	6 mL water	6 mL water	6 mL water	6 mL water
Loading	75 mL sample	75 mL sample	75 mL sample	75 mL sample
Washing	6 mL 95:5 water : methanol	6 mL 95:5 water : methanol	6 mL 25 mM ammonium acetate in water	6 mL 25 mM ammonium acetate in water
Elution 1	2 x 3 mL methanol	2 x 3 mL methanol	2 x 3 mL 2% formic acid in water	2 x 3 mL 2% formic acid in water
Elution 2	2 x 3 mL 5% formic acid in water	2 x 3 mL 5% formic acid in water	N/A	N/A

9.4.1.1.3 Instruments and methods

The LC system was manufactured by Shimadzu (Japan) and composed of a Nexera LC-30AD pump module, a SIL-30AC autosampler and a CTO-30A column oven module. This LC system was coupled a Maxis quadrupole-time-of-fight mass spectrometer made by Bruker (Billerica, MA). The LC column was a Waters Acquity UPLC HSS T3 2.1 × 50 mm, 1.8 µm. 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 following elution gradient was used for the chromatographic separation in MS1 experiments: at initial time, 5% B; at 5 min, 100% B; at 7 min, 100% B; at 7.01 min, 5% B; at 10 min, 5% B. Run time was 10 min. Mobile phase flow rate was 500 µL min-1 throughout the run. Injection volume was 2 µL. The following elution gradient was used for the chromatographic separation in the data-dependent acquisition experiments: 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.

9.4.1.2 Nontarget screening

9.4.1.2.1 Collection and preparation of samples

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,

collected 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). Samples of 250 mL were concentrated on Strata-X polymer solid-phase extraction cartridges (200 mg, 6 mL) from Phenomenex (Torrance, CA) and then eluted with 2 × 3 mL of a 1:1 (v/v) 2% formic acid solution of ACN-MeOH. Finally, eluates were evaporated under a nitrogen stream and reconstituted to 625 µL, which resulted in a preconcentration factor of 400.

9.4.1.2.2 Instruments and methods

A Vanquish Flex ultra-high performance liquid chromatography system coupled to a Q-OrbitrapMS model Q Exactive Plus Orbitrap, both manufactured by Thermo Scientific (San Jose, CA) was used for all nontargeted screening analyses. 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.

9.4.2 Results and discussion

9.4.2.1 Photodegradation experiments

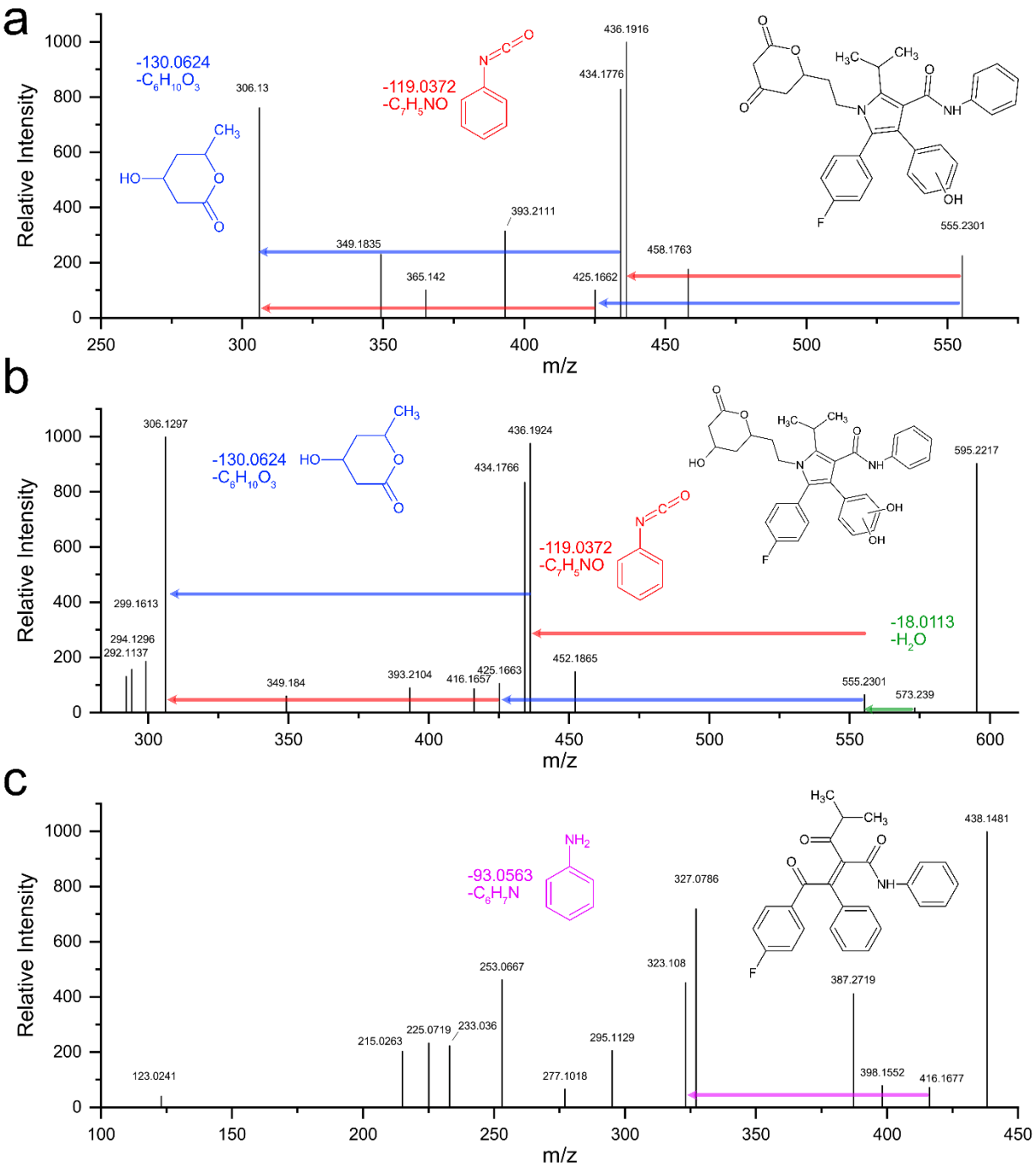
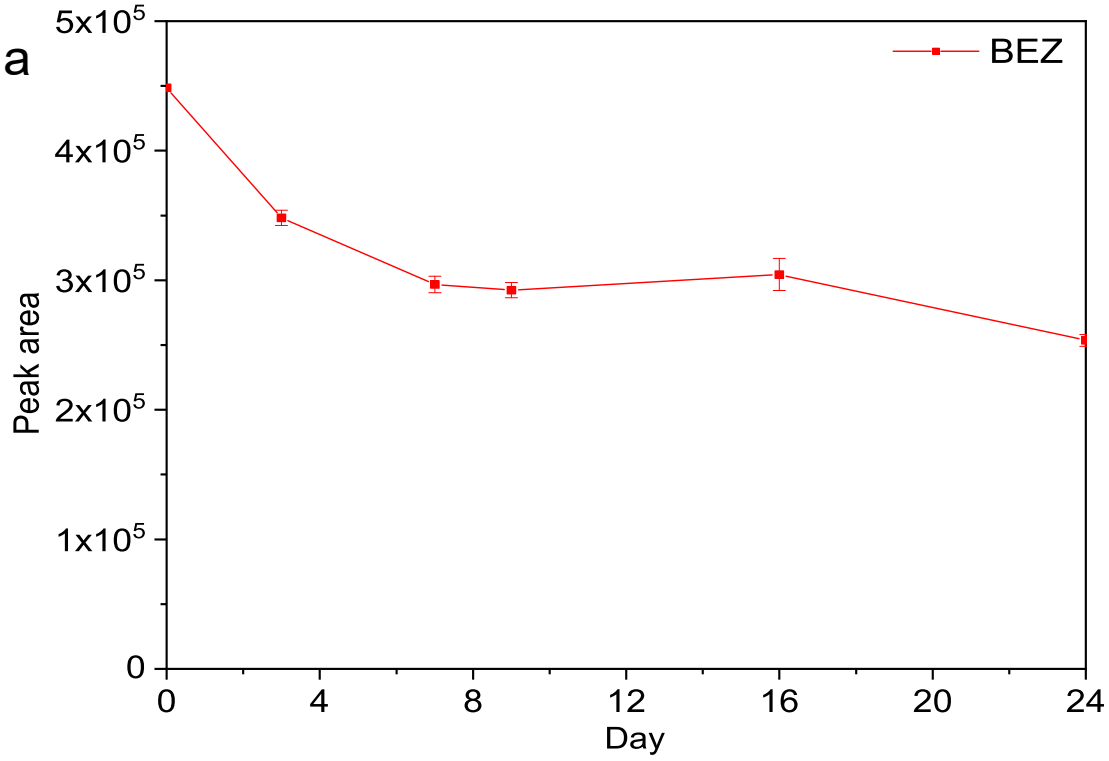


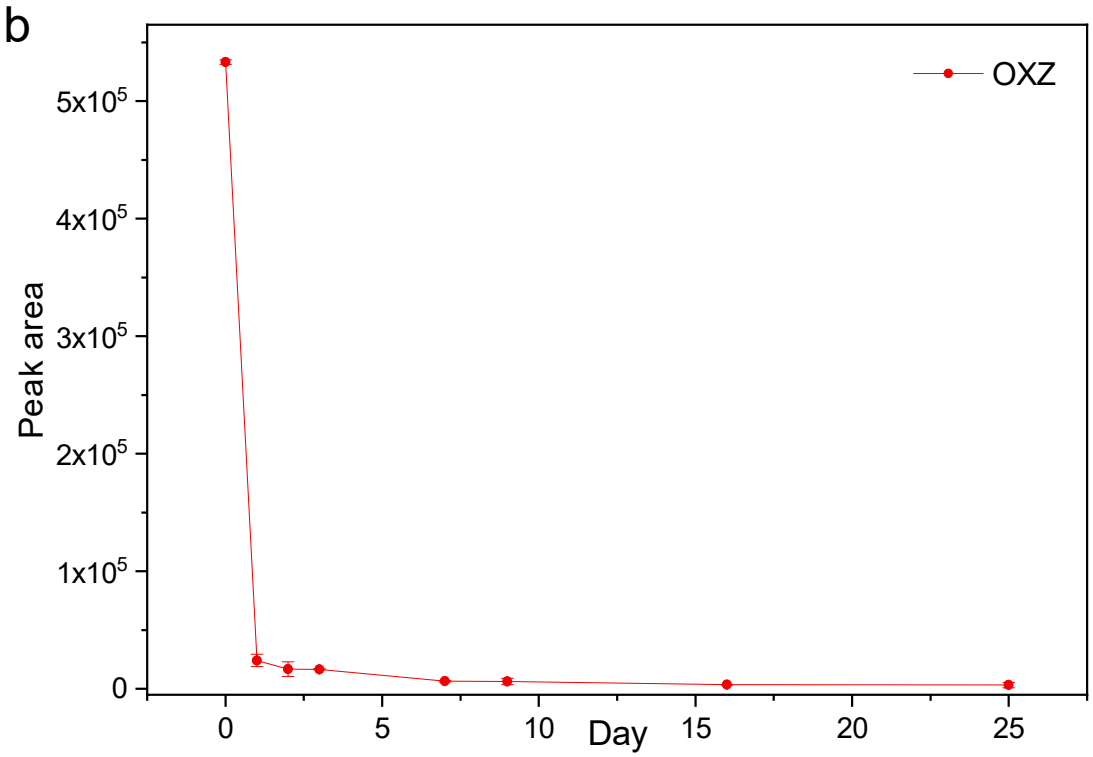
Figure 50. MS^2 spectra of ATV_TP555 (a), ATV_TP573 (b), and ATV_TP416 (c). In a, the loss of the lactone and phenyl amide are observed in blue and red respectively. In b, the sodium adduct ion (m/z

595.2217) is the predominant pseudo molecular ion; a mass shift of 18 corresponding to H2O is observed along with the loss of the lactone and phenyl amide in blue and red respectively. In c, the sodium adduct (m/z 438.1481) is the prevalent pseudo molecular ions while the loss of the aniline seen. There is no loss of lactone seen which would be consistent with the pyrrole ring opening.

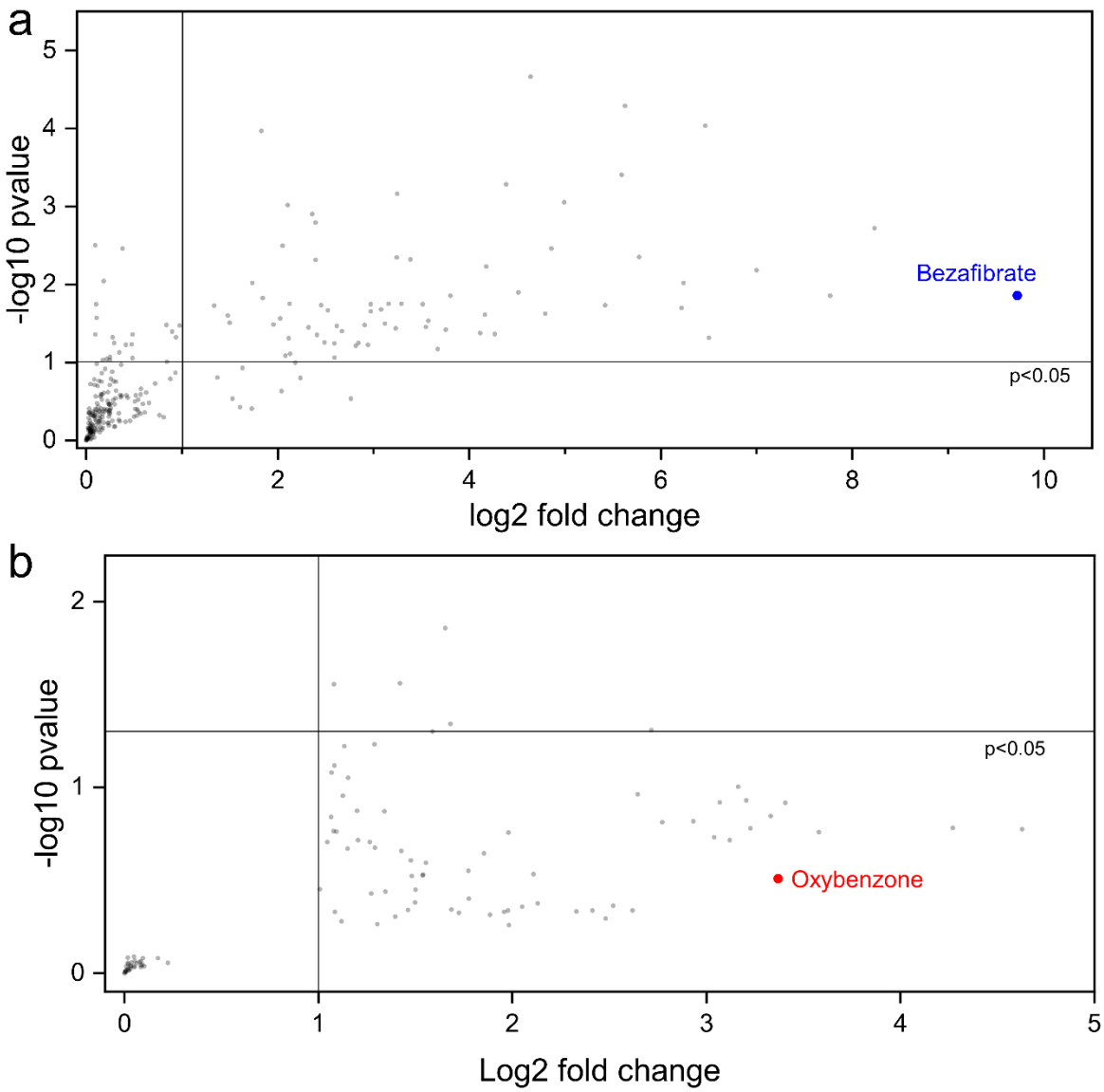
9.4.2.1.1 Explanation of the mass spectra:

For ATV_TP555, the loss of the lactone is observed as well as the loss of the phenyl amide. As with ATV_TP557a, the hydroxylation site could be on the phenyl or the phenyl fluorine. ATV_TP573 also shows the presence of the lactone and the phenyl amide. It is once again likely that the hydroxylation site is located on the phenyl. Finally, for ATV_TP416, the loss of the phenyl amine is observed. The absence of the loss of the lactone group as well as the fact that the compound was found in the neutral fraction of the solid phase extraction cartridge points toward a pyrrole ring opening as was observed by Wang et al. in 2018.


a
BEZ
Peak area
5x10^5
4x10^5
3x10^5
2x10^5
1x10^5
0
0
4
8
12
16
20
24
Day


b
OXZ
Peak area
5x10^5
4x10^5
3x10^5
2x10^5
1x10^5
0
0
5
10
15
20
25
Day

120 **Figure 51.** Photodegradation kinetics of bezafibrate (a) and oxybenzone (b)



121

122 **Figure 52.** Plots of -log10 p-value vs log2 fold change for bezafibrate (a) at the 24th day of exposure and
123 for oxybenzone (b) at the 25th day of exposure. No statistically significant transformation products in
124 terms of p-value and fold change were observed for both compounds.

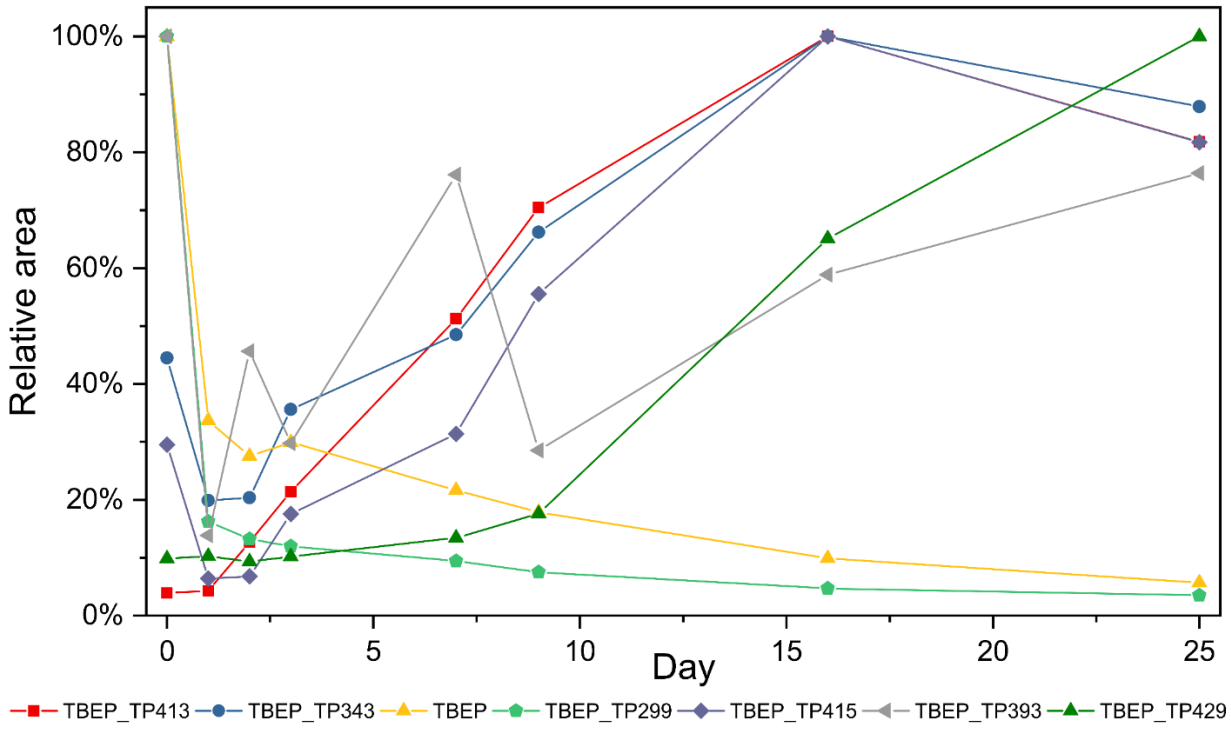


Figure 53. Relative photodegradation kinetics of tris(2-butoxyethyl) phosphate (TBEP) along with its transformation product.

9.5 ANNEXE E – Informations complémentaires du CHAPITRE 7

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]

[1]Department of Chemistry, Université de Sherbrooke, Sherbrooke, Canada
[2]Environment and Climate Change Canada, Montreal, Canada
Corresponding author: Pedro A. Segura. pa.segura@usherbrooke.ca

Keywords: Molecular networking, CluMSID, consumer products additives, pharmaceuticals, high-resolution mass spectrometry

Supplementary Material

9.5.1 Materials and methods

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). Triton X-100 (laboratory grade) was purchased from Sigma Aldrich (St-Louis, Missouri, USA).

9.5.1.1 Nontarget screening

9.5.1.1.1 Collection and preparation of samples

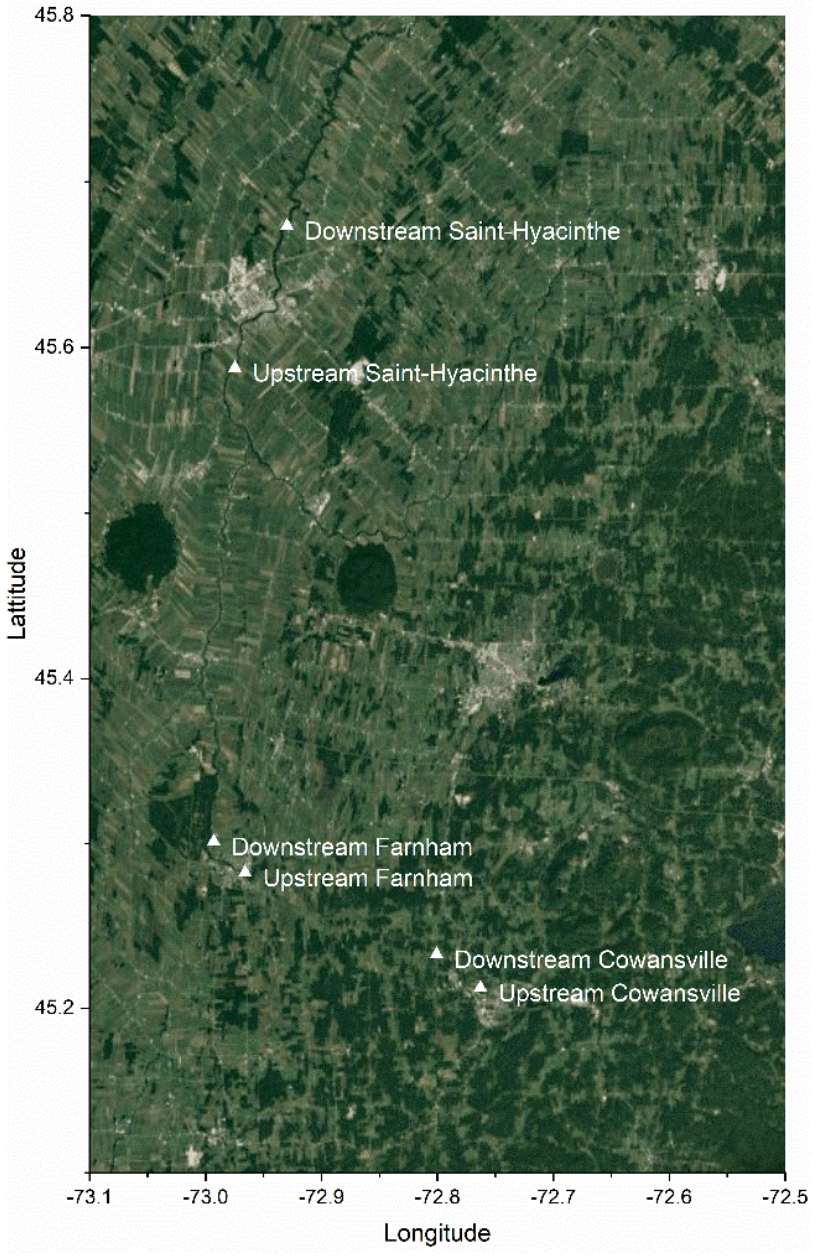


Figure 54. Map of the sampling points

9.5.1.2 Software parameters

The files in vendor format (Thermo RAW) were converted into mzML, ms1, and mgf with MSConvertGUI. MetFrag used mzML files, SPS used ms1 and mgf files and GNPS and CluMSID used mgf files.

MetFrag was used as part of the patRoon R package. XCMS was used for pick picking with mass tolerance of 5 ppm, peak width of 4 to 20 seconds and noise value of 50 000. The range of *m/z* was 120-1000. MzR was used for the creation of the MS^2 mass lists. The database used with MetFrag was CompTox and the adduct was $[M+H]^+$.

For Similar Partition Searching (SPS), ms1 files were used for intensities and mgf files were used for MS^2 matching. Both files were zipped and exported into an Amazon S3 server where they were treated. Then, results were imported from the server and treated with Microsoft Access.

For GNPS, mgf files were uploaded into the GNPS platform. Molecular networks were as follow: tolerance was 0.02 Da, min cosine value to form an edge between two nodes was 0.7, minimum matched fragment ions was 5. Library search parameters were as follow: minimum matched peaks was 4, score threshold was 0.7. Analogs were not searched.

Sampling Composite field blank, freezing upon arrival
Filtration 0.45 µm mixed cellulose ester membranes
Extraction Strata-X polymer solid phase extraction cartridges

UHPLC separation HSS T3 column 2.1 x 50mm, 1.8 µm; H_2O 0.1% FA and ACN:MeOH 3:2 0.1% FA
MS analysis ESI positive mode, DDA top 10, AGC target 2e5, auto exclusion, survey scan 35 000 resolution, MS^2 product ion scan 17 500 resolution, stepped normalized energy 25,35,45 units

SPS

Data conversion ms1 level 1-1; mgf level 2-2, 50 most intense product ions; software used: MsconverGUI; files compressed and sent to zip file

Data transfer and combinatorial analysis zip file uploaded to Amazon S3 server with Cloudberry Explorer for Amazon S3 software; similar partitions searched and matched to MS^2 spectra; data and result files downloaded

Blank subtraction The processed blank data files are used to subtract the features in the blank with less than 5 times the peak area with an excel add in

Data treatment and filtering The blank filtered data are imported to an Access database where they are treated, compounds with an absolute score over 70 are filtered and examined manually

GNPS

Data conversion mgf level 2-2, 50 most intense product ions; software used: MsconverGUI

Data transfer, molecular library search and networks settings The mgf files are uploaded to the GNPS servers with WinSCP; Precursor Ion Mass tolerance 0.02 Da, Fragment Ion Mass Tolerance 0.02 Da, Library Search Min Matched Peaks = 4, Score Threshold = 0.7, min pair cos = 0.7, Minimum Matched Fragment Ions = 4, Maximum Connected Component Size = 100, Network TopK = 10, groups are created with the blank and each sampling point in a separate group

Data treatment The network containing all nodes, edges and library matches and scores information is downloaded and treated on Cytoscape; the list of IDs is treated on Excel

Annotation The level of confidence is assigned based on the quality of the MS^2 data and contextual information, the IDs are manually added on a table

CluMSID

Data conversion mgf level 2-2, 50 most intense product ions; software used: MsconverGUI

Data and feature list import The mgf files are individually imported into R with the previously made annotations

Generation of the distance matrix The distance is created from the peak list

Visualization of the distance matrix Precursor networks, OPTICS plots, OPTICS tables, dendrograms and heatmaps are created on vector type files (PDF)

Data treatment The data is treated on R, Excel and the graphs

Annotation New levels of confidence are assigned based on the new cluster information available

R based NTS with patRoon interface

Data conversion mzML level 1-2; software used: MsconverGUI

Peak picking Groups are created for blanks and sampling points, XCMS settings for feature generation are 5 ppm tolerance, peakwidth = 4 to 20 seconds, noise = 50 000, rt alignment with Obiwarp method, blank threshold = 5

Creation of mass list Mass lists are created with mzR. maxMSRtWindow = 5, topmost = 20

Combinatorial library search and molecular formula generation Library search performed with MetFrag. Database is the MetFrag compTox DB, adducts are $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, fragRelMzDev = 5 ppm.
Molecular formulas generated with GenForm, elements are CHNOFPSClBrI.
The score is based on the number of matched product ions, the correct molecular formula, and the presence of the compounds on several international suspects lists (NORMANSUSDAT, REACH...)

Data treatment The data is treated on R or Excel. Only IDs with a minimum of 4 matched fragment ions are selected.

Figure 55. Steps of the different methods and modules in this study

9.5.2 Results and discussion

Table 24. Physical and chemical properties of the sampling points

Sampling point	**Temperature**	**Dissolved oxygen (%)**	**Conductivity (μScm^{-1})**	**pH**	**Weather**
Cowansville Upstream	22.8	93.7	117.5	7.70	Sunny
Cowansville Downstream	23.0	77,9	185,6	7,55	Sunny
Farnham Upstream	26.8	95,6	241,0	7,96	Sunny
Farnham Downstream	26.7	124,8	238,4	9,08	Sunny
Saint-Hyacinthe Upstream	27.9	117,3	275,0	8,74	Sunny
Saint-Hyacinthe Downstream	27.2	85,0	341,5	8,34	Sunny

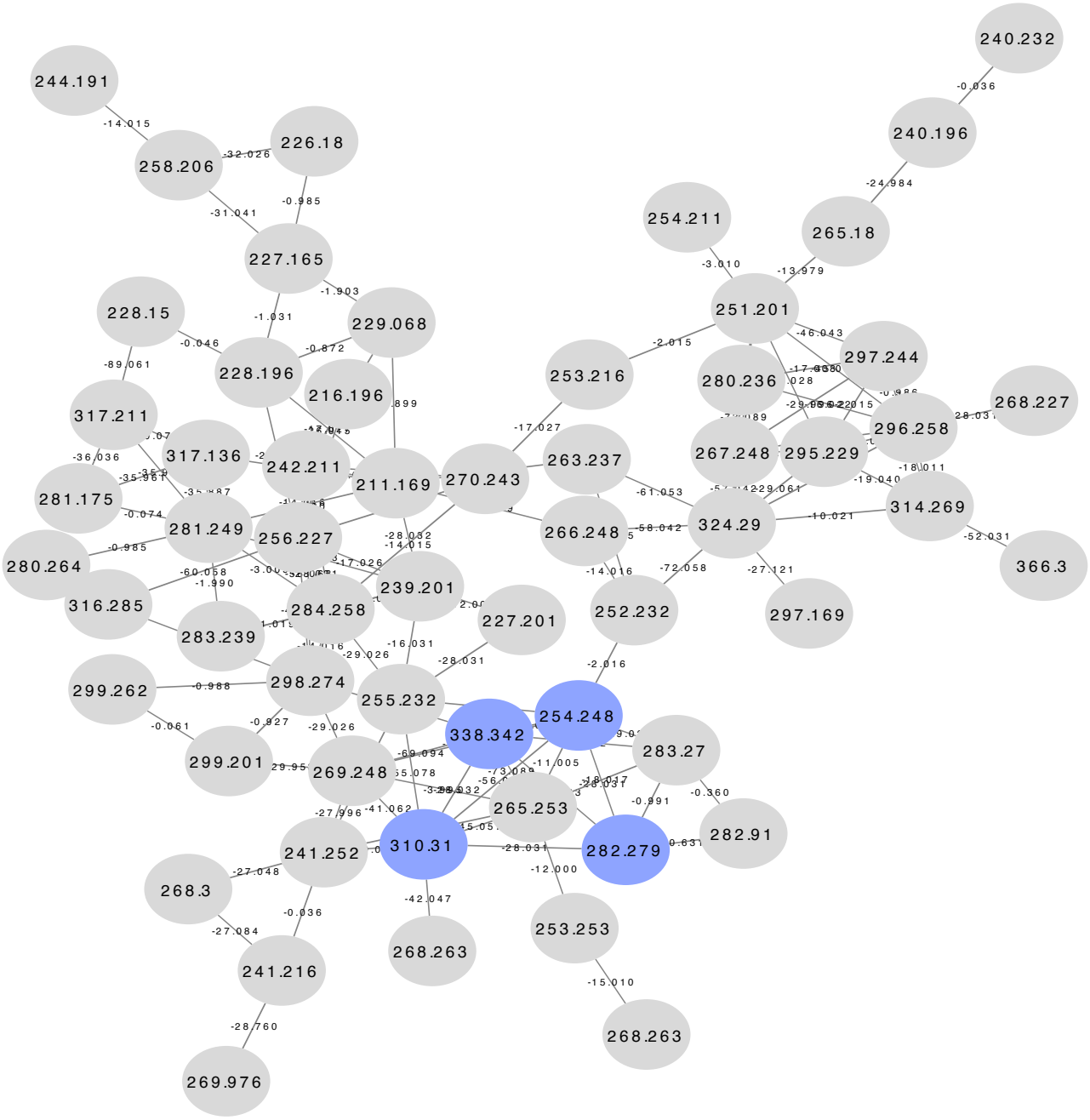
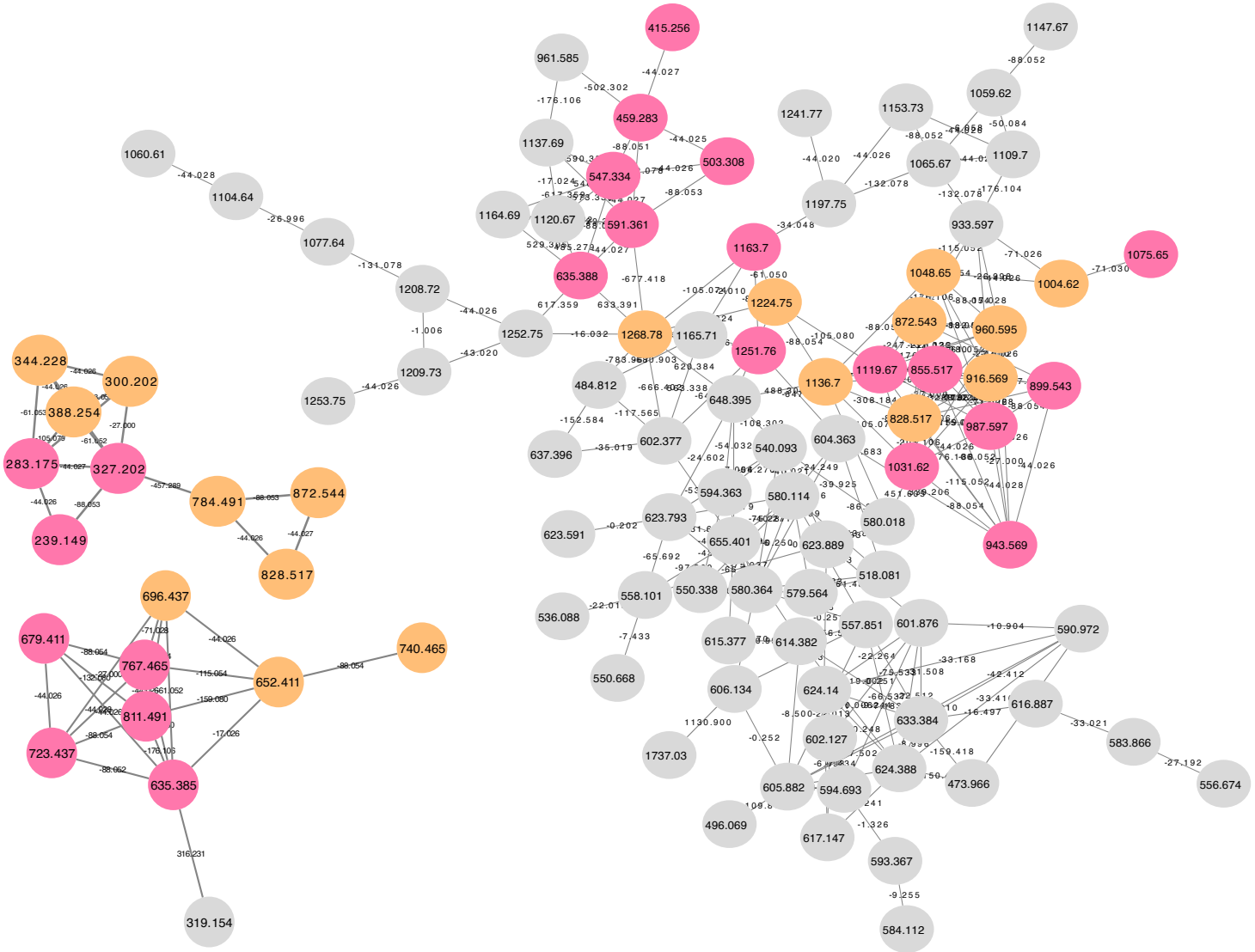


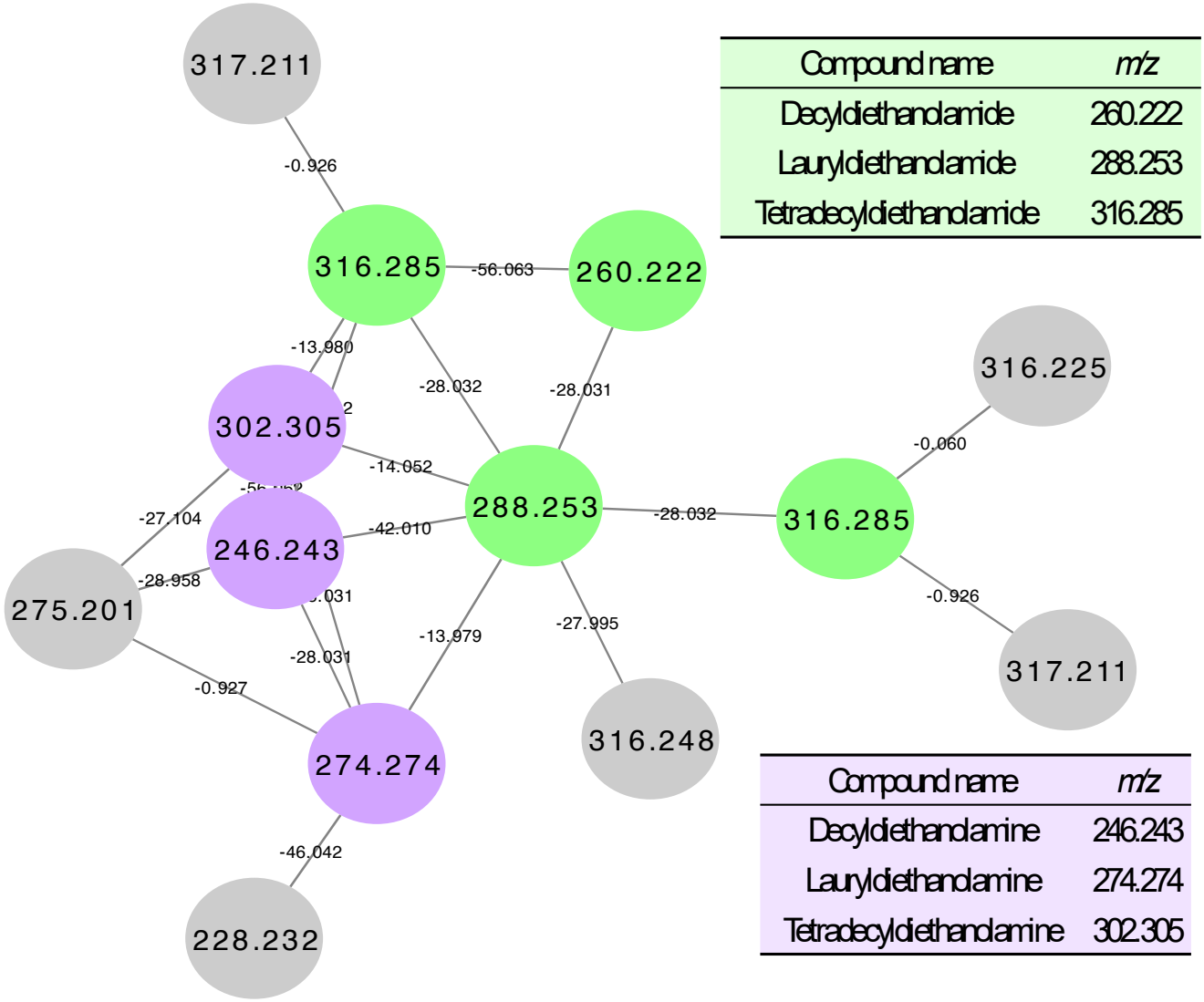
Figure 56. Molecular network of fatty amides lubricants 9-hexadecenamide (m/z 254.248), oleamide (m/z 282.279), eicosenamide (m/z 310.31), and erucamide (m/z 338.342) in blue. In grey are unannotated precursors.



[M+H]⁺ adduct		[M+NH₄]⁺ adduct	
Compound	*m/z*	Compound	*m/z*
PEG-05	239.149	PEG-06	300.202
PEG-06	283.175	PEG-07	344.228
PEG-07	327.202	PEG-08	388.254
PEG-09	415.256	PEG-14	652.411
PEG-10	459.283	PEG-15	696.437
PEG-11	503.308	PEG-16	740.465
PEG-12	547.334	PEG-17	784.491
PEG-13	591.361	PEG-18	828.517
PEG-14	635.388	PEG-19	872.544
PEG-15	679.411	PEG-20	916.569
PEG-16	723.437	PEG-21	960.595
PEG-17	767.465	PEG-22	1004.62
PEG-18	811.491	PEG-23	1048.65
PEG-19	855.517	PEG-25	1136.7
PEG-20	899.543	PEG-27	1224.75
PEG-21	943.597	PEG-28	1268.78
PEG-22	987.597		
PEG-23	1031.62		
PEG-24	1075.65		
PEG-25	1119.67		
PEG-26	1163.7		
PEG-28	1251.76		

1

Figure 57. Molecular network of the polyethylene glycol (PEG) congeners with the hydrogen and ammonium adducts. In grey are unannotated precursors.



Compound name	*m/z*
Decyldiethanolamide	260.222
Lauryldiethanolamide	288.253
Tetradecyldiethanolamide	316.285

Compound name	*m/z*
Decyldiethanolamine	246.243
Lauryldiethanolamine	274.274
Tetradecyldiethanolamine	302.305

Figure 58. Molecular network of cosmetics diethanolamines (purple) and diethanolamides (green). In grey are the unannotated precursors that are probably isobars of other precursors in the network that were selected to the quadrupole at the same time.

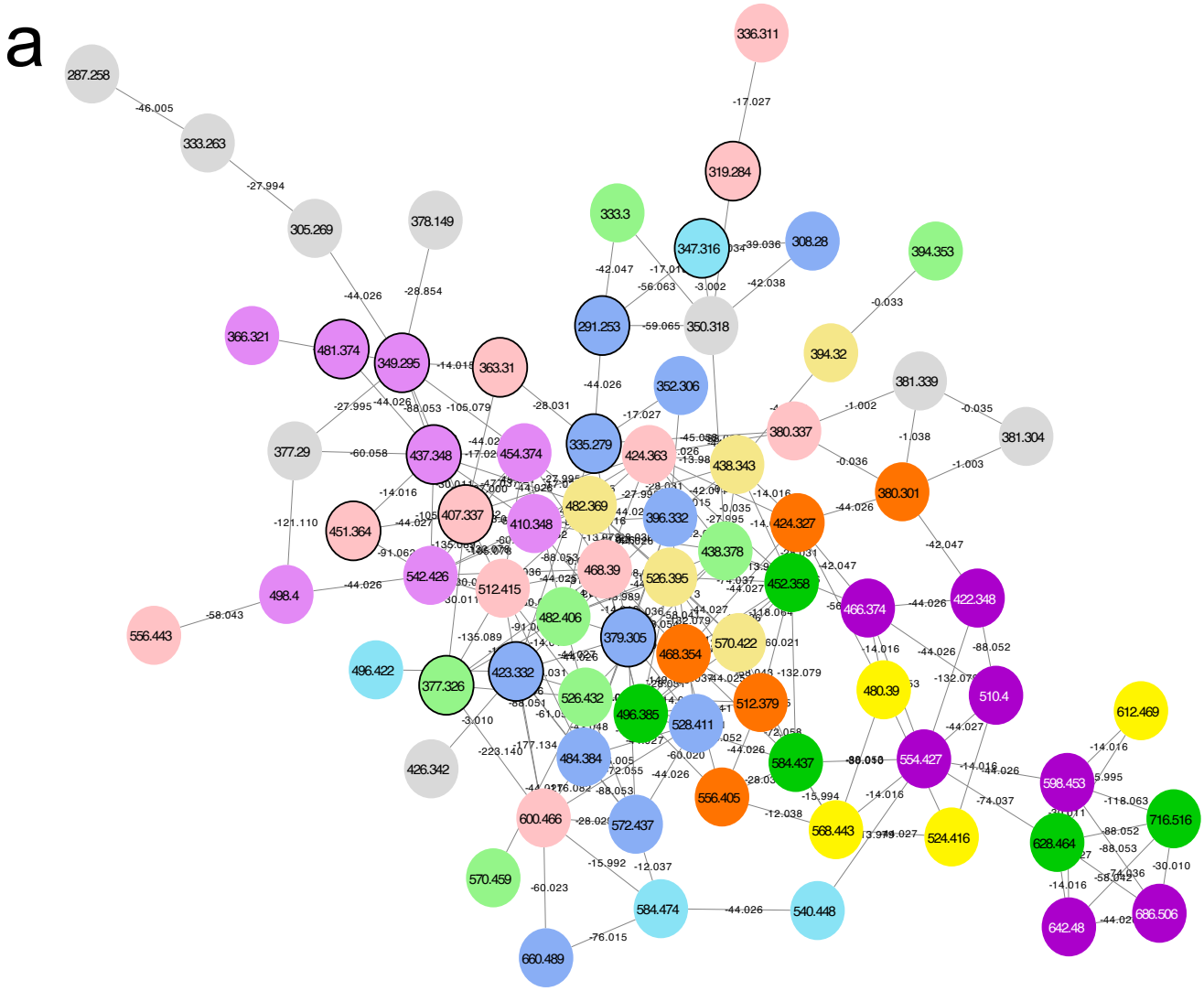
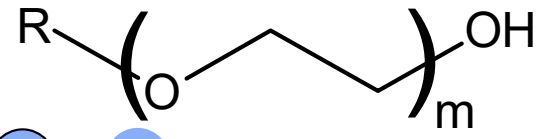


Figure 59. Molecular network generated by GNPS of polyoxyethylene alkyl ethers and esters. The color coding for the subcategories is indicated in **Figure 60**.

b



$R=C_{10}H_{21}$

	$[M+H]^+$	$[M+NH_4]^+$
Triethylene glycol decyl ether	291.253	308.28
Tetraethylene glycol decyl ether	335.279	352.306
Pentaethylene glycol decyl ether	379.305	396.332
Hexaethylene glycol decyl ether	423.332	
Heptaethylene glycol decyl ether		484.384
Octaethylene glycol decyl ether		528.411
Nonaethylene glycol decyl ether		572.437
Undecaethylene glycol decyl ether		660.489

$R=C_{11}H_{23}$

	$[M+H]^+$	$[M+NH_4]^+$
Tetraethylene glycol undecyl ether	349.295	366.321
Pentaethylene glycol undecyl ether		410.348
Hexaethylene glycol undecyl ether	437.348	454.374
Heptaethylene glycol undecyl ether	481.374	498.4
Octaethylene glycol undecyl ether		542.426

$R=C_{12}H_{25}$

	$[M+H]^+$	$[M+NH_4]^+$
Triethylene glycol monododecyl ether	319.284	336.311
Tetraethylene glycol monododecyl ether	363.31	380.337
Pentaethylene glycol monododecyl ether	407.337	424.363
Hexaethylene glycol monododecyl ether	451.365	468.39
Heptaethylene glycol monododecyl ether		512.415
Octaethyleneglycol monododecyl ether		556.443
Nonaethyleneglycol monododecyl ether		600.466

$R=C_{13}H_{27}$

	$[M+H]^+$	$[M+NH_4]^+$
Triethylene glycol monotridecyl ether	333.3	
Tetraethylene glycol monotridecyl ether	377.326	394.353
Pentaethylene glycol monotridecyl ether		438.378
Hexaethylene glycol monotridecyl ether		482.406
Heptaethylene glycol monotridecyl ether		526.432
Octaethylene glycol monotridecyl ether		570.458

$R=C_{14}H_{29}$

	$[M+H]^+$	$[M+NH_4]^+$
Triethylene glycol monotetradecyl ether	347.316	
Hexaethylene glycol monotetradecyl ether		496.422
Heptaethylene glycol monotetradecyl ether		540.448
Octaethylene glycol monotetradecyl ether		584.474

$R=C_{11}H_{21}O$

	$[M+NH_4]^+$
Polyoxyethylene (4) monoundecanoate	380.301
Polyoxyethylene (5) monoundecanoate	424.327
Polyoxyethylene (6) monoundecanoate	468.354
Polyoxyethylene (7) monoundecanoate	512.379
Polyoxyethylene (8) monoundecanoate	556.405

$R=C_{12}H_{23}O$

	$[M+NH_4]^+$
Polyoxyethylene (4) monolaurate	394.32
Polyoxyethylene (5) monolaurate	438.343
Polyoxyethylene (6) monolaurate	482.369
Polyoxyethylene (7) monolaurate	526.395
Polyoxyethylene (8) monolaurate	570.422

$R=C_{13}H_{25}O$

	$[M+NH_4]^+$
Polyoxyethylene (5) monotridecanoate	452.358
Polyoxyethylene (6) monotridecanoate	496.385
Polyoxyethylene (8) monotridecanoate	584.437
Polyoxyethylene (9) monotridecanoate	628.464
Polyoxyethylene (11) monotridecanoate	716.516

$R=C_{14}H_{27}O$

	$[M+NH_4]^+$
Polyoxyethylene (4) monotetradecanoate	422.348
Polyoxyethylene (5) monotetradecanoate	466.374
Polyoxyethylene (6) monotetradecanoate	510.4
Polyoxyethylene (7) monotetradecanoate	554.427
Polyoxyethylene (8) monotetradecanoate	598.453
Polyoxyethylene (9) monotetradecanoate	642.48
Polyoxyethylene (10) monotetradecanoate	686.506

$R=C_{15}H_{29}O$

	$[M+NH_4]^+$
Polyoxyethylene (5) monopentadecanoate	480.39
Polyoxyethylene (6) monopentadecanoate	524.416
Polyoxyethylene (7) monopentadecanoate	568.443
Polyoxyethylene (8) monopentadecanoate	612.469

Figure 60. Base structure of polyoxyethylene alkyl ethers (left) and esters (right) and their color coding and coordinates in the network in Figure S-5. The oxyethylene unit is repeated m times with the prefix in the case of ethers and the number in parenthesis in the case of esters.

Table 25. Clusters of 64 polyoxyethylene alkyl ethers and esters congeners generated with CluMSID

Feature coordinate	Precursor ID	Cluster ID	Feature coordinate	Precursor ID	Cluster ID
M584.44T877.44	Polyoxyethylene (8) monotridecanoate	21	M658.51T848.96	Decaethylene glycol monotridecyl ether	21
M672.49T866.45	Polyoxyethylene (10) monotridecanoate	21	M598.49T962.01	Octaethylene glycol pentadecyl ether	21
M628.46T871.76	Polyoxyethylene (9) monotridecanoate	21	M660.49T692.88	Undecaethylene glycol decyl ether	21
M556.41T776.88	Polyoxyethylene (8) monoundecanoate	21	M554.46T966.55	Heptaethylene glycol pentadecyl ether	21
M600.43T771.43	Polyoxyethylene (9) monoundecanoate	21	M686.51T901.09	Polyoxyethylene (10) monotetradecanoate	21
M512.38T782.89	Polyoxyethylene (7) monoundecanoate	21	M528.41T696.84	Octaethylene glycol decyl ether	21
M644.46T766.94	Polyoxyethylene (10) monoundecanoate	21	M598.45T912.01	Polyoxyethylene (8) monotetradecanoate	21
M656.5T968.35	Polyoxyethylene (9) monopentadecanoate	21	M498.4T736.72	Heptaethylene glycol undecyl ether	21
M700.52T963.21	Polyoxyethylene (10) monopentadecanoate	21	M440.36T700.17	Hexaethylene glycol decyl ether	21
M716.52T861.28	Polyoxyethylene (11) monotridecanoate	21	M424.33T796.11	Polyoxyethylene (5) monoundecanoate	21
M612.47T972.94	Polyoxyethylene (7) monopentadecanoate	21	M396.33T701.37	Pentaethylene glycol decyl ether	21
M468.35T788.84	Polyoxyethylene (6) monoundecanoate	21	M484.38T698.85	Heptaethylene glycol decyl ether	21
M468.39T799.54	Hexaethylene glycol monododecyl ether	21	M540.45T912.82	Heptaethylene glycol tetradecyl ether	21
M568.44T978.48	Polyoxyethylene (7) monopentadecanoate	21	M496.42T916.85	Hexaethylene glycol tetradecyl ether	21
M760.54T856.1	Polyoxyethylene (12) monotridecanoate	21	M480.39T989.22	Polyoxyethylene (5) monopentadecanoate	21
M642.48T907.66	Polyoxyethylene (9) monotetradecanoate	21	M572.44T695.1	Nonaethylene glycol decyl ether	21
M688.48T764.01	Polyoxyethylene (11) monoundecanoate	21	M554.43T916.8	Polyoxyethylene (7) monotetradecanoate	21
M526.39T822.85	Polyoxyethylene (7) monolaurate	21	M452.36T891.22	Polyoxyethylene (5) monotridecanoate	21
M482.37T830.3	Polyoxyethylene (6) monolaurate	21	M614.48T852.85	Nonaethylene glycol monotridecyl ether	21
M454.37T738.53	Hexaethylene glycol undecyl ether	21	M704.52T692.07	Dodecaethylene glycol decyl ether	21
M524.42T983.67	Polyoxyethylene (6) monopentadecanoate	21	M510.4T922.98	Polyoxyethylene (6) monotetradecanoate	21
M732.51T759.33	Polyoxyethylene (12) monoundecanoate	21	M363.31T815.01	Tetraethylene glycol monododecyl ether (H+)	21
M804.57T850.93	Polyoxyethylene (13) monotridecanoate	21	M380.34T814.81	Tetraethylene glycol monododecyl ether (NH4+)	21
M349.29T746.73	Tetraethylene glycol undecyl ether (H+)	21	M335.28T703.6	Tetraethylene glycol decyl ether (H+)	21
M616.46T693.98	Decaethylene glycol decyl ether	21	M408.33T901.68	Polyoxyethylene (4) monotridecanoate	21
M644.49T796.31	Decaethyleneglycol monododecyl ether	21	M352.31T703.5	Tetraethylene glycol decyl ether (NH4+)	21
M512.41T806	Heptaethylene glycol monododecyl ether	21	M380.3T802.98	Polyoxyethylene (4) monoundecanoate	21
M556.44T803.08	Octaethyleneglycol monododecyl ether	21	M422.35T931.35	Polyoxyethylene (4) monotetradecanoate	21
M716.55T897.71	Undecaethylene glycol tetradecyl ether	21	M510.44T969.82	Hexaethylene glycol pentadecyl ether	21
M628.5T905.12	Nonaethylene glycol tetradecyl ether	21	M436.36T995.76	Polyoxyethylene (5) monopentadecanoate	21
M584.47T909.36	Octaethylene glycol tetradecyl ether	21	M424.36T811.83	Pentaethylene glycol monododecyl ether	21
M600.46T801.86	Nonaethyleneglycol monododecyl ether	21	M496.39T889.76	Polyoxyethylene (6) monotridecanoate	21
M672.53T900.49	Decaethylene glycol tetradecyl ether	21	M380.3T772.74	Polyoxyethylene (4) monoundecanoate	21

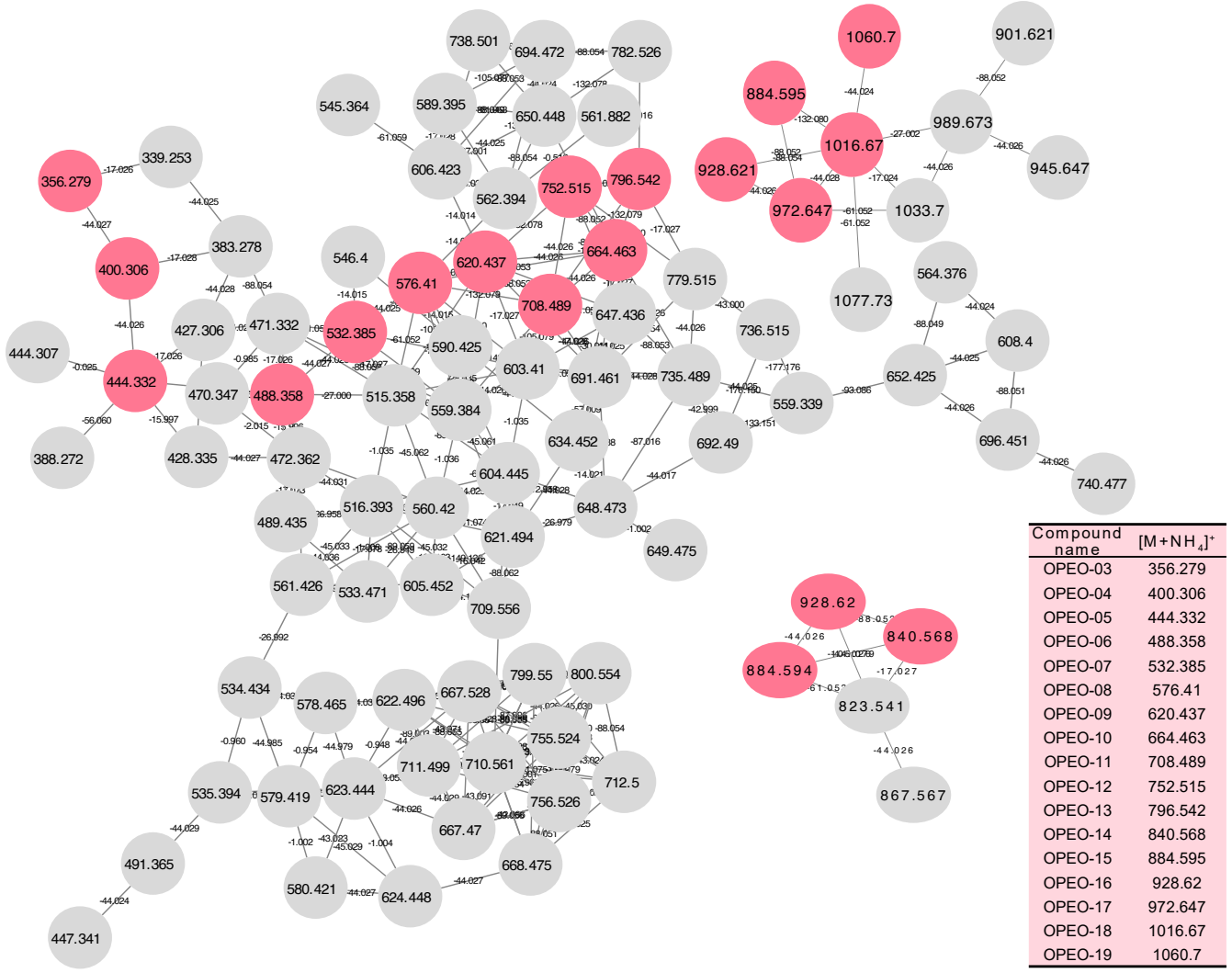
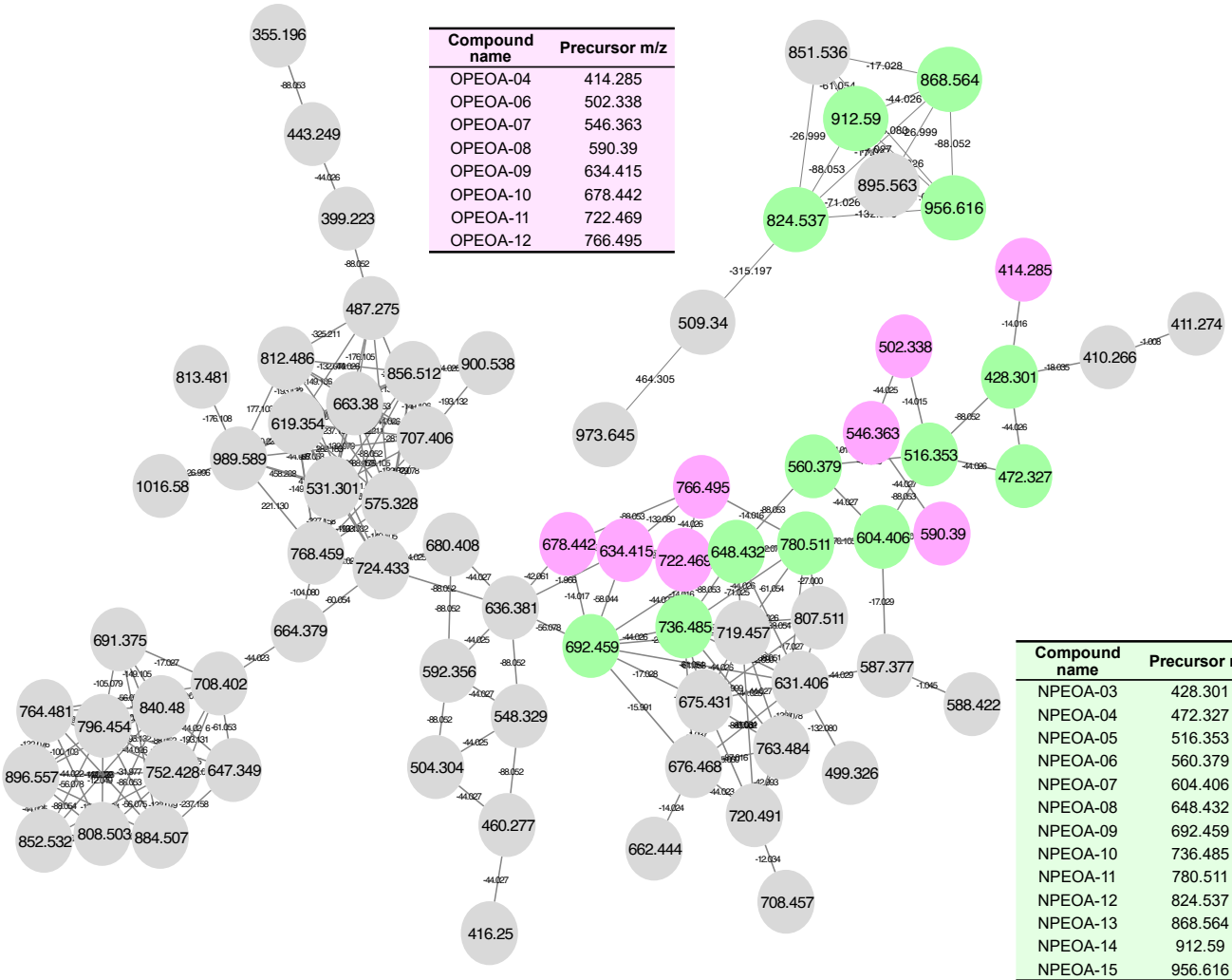


Figure 61. Molecular network of nonionic surfactants octylphenol ethoxylates (OPEOs) with the number of ethylene oxide units given by the number following it. In grey are the unannotated precursors.



Compound name	Precursor m/z
OPEOA-04	414.285
OPEOA-06	502.338
OPEOA-07	546.363
OPEOA-08	590.39
OPEOA-09	634.415
OPEOA-10	678.442
OPEOA-11	722.469
OPEOA-12	766.495

Compound name	Precursor m/z
NPEOA-03	428.301
NPEOA-04	472.327
NPEOA-05	516.353
NPEOA-06	560.379
NPEOA-07	604.406
NPEOA-08	648.432
NPEOA-09	692.459
NPEOA-10	736.485
NPEOA-11	780.511
NPEOA-12	824.537
NPEOA-13	868.564
NPEOA-14	912.59
NPEOA-15	956.616

Figure 62. Molecular network of octylphenol ethoxylates acids (OPEOA) in purple and nonylphenol ethoxylates acids (NPEOA) in light green. Unannotated nodes are in grey. While 27 congeners of OPEOs and NPEOs were tentatively identified, the number of precursors in this network amounts to 21. The 6 other TPs did not figure in the molecular network but still exhibited product ions characteristic to APEOAs MS2 fragmentation. The complete list of all compounds tentatively identified can be seen in the Excel file IdentifiedCompounds.xlsx (Supplementary Material).

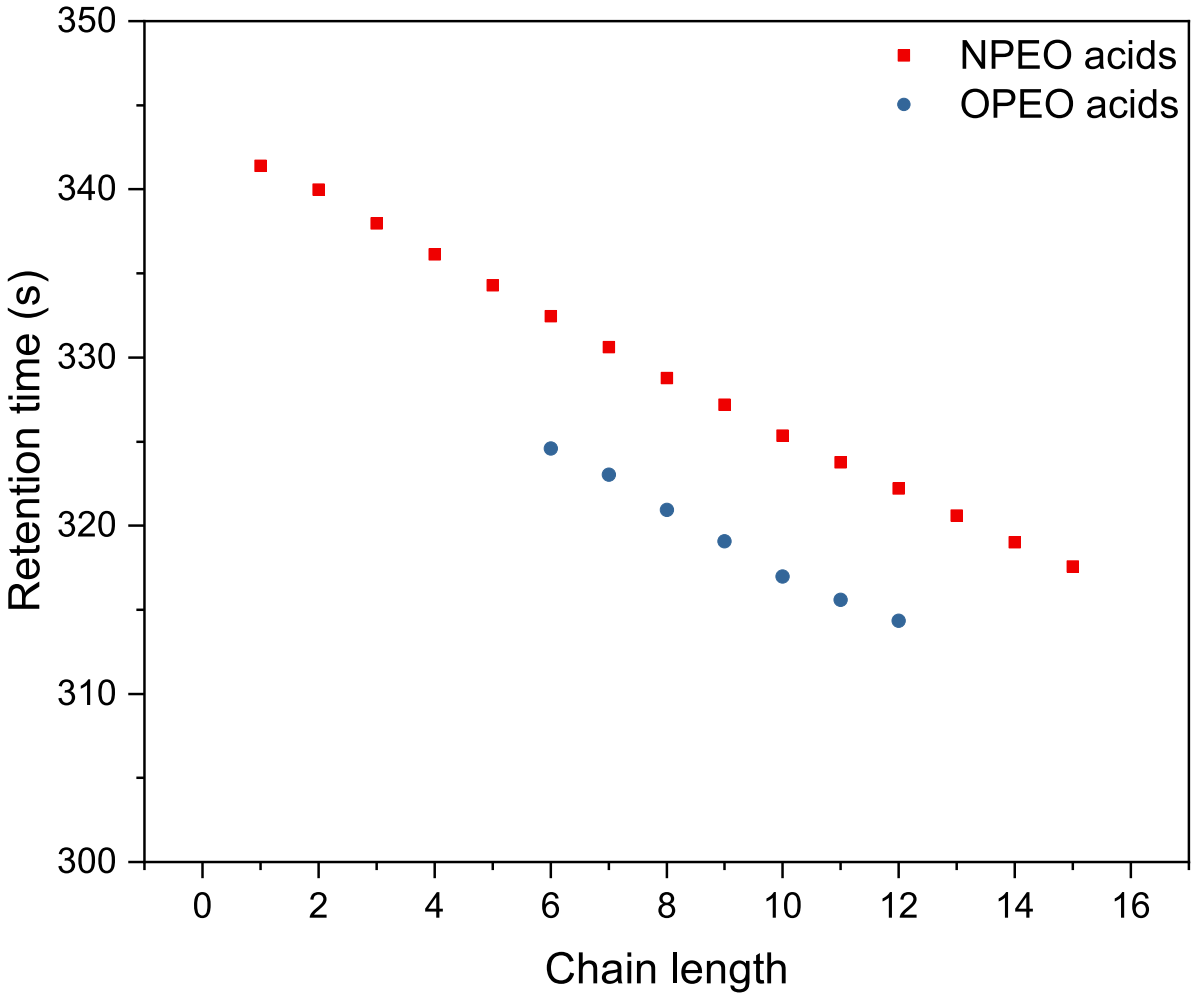
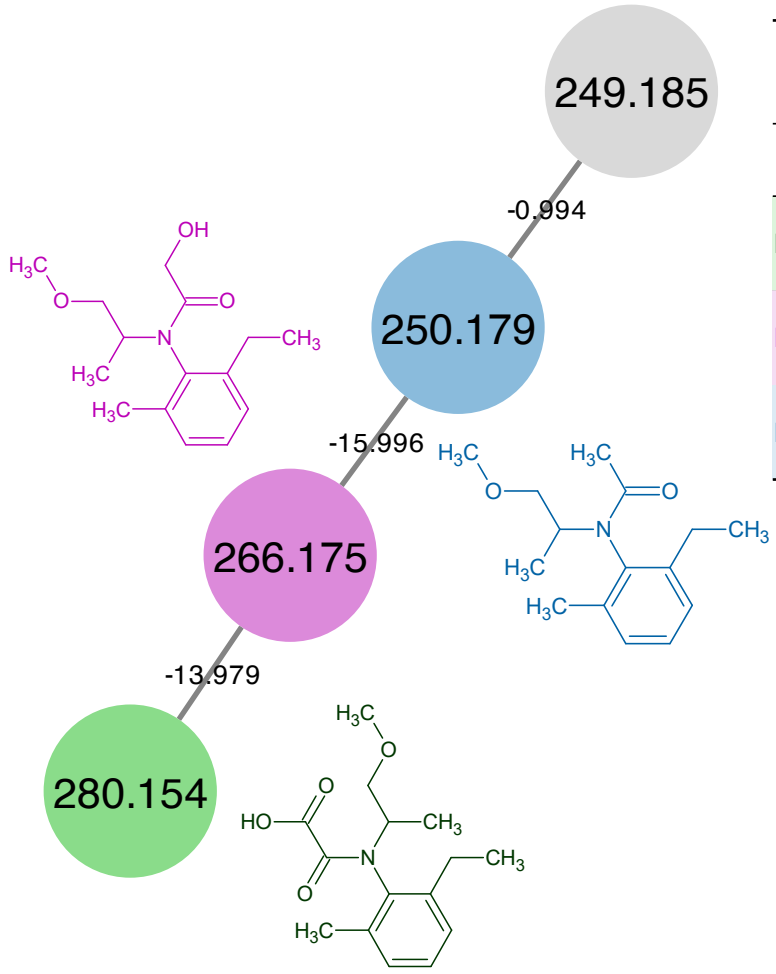


Figure 63. Retention time of the nonylphenol ethoxylates (NPEO) and octylphenol ethoxylates (OPEO) acids as a function of ethylene oxide chain length.



Metolachlor transformation products	
Name	**Precursor *m/z***
Metolachlor oxalic acid	280.154
Metolachlor_TP266	266.175
Metolachlor_TP250	250.179

Figure 64. Molecular network of metolachlor oxalic acid in light green, metolachlor TP_266 and a proposed structure in purple, and metolachlor_TP250 in blue a proposed structure. The unannotated node in grey is an isobaric contamination of metolachlor_TP250.

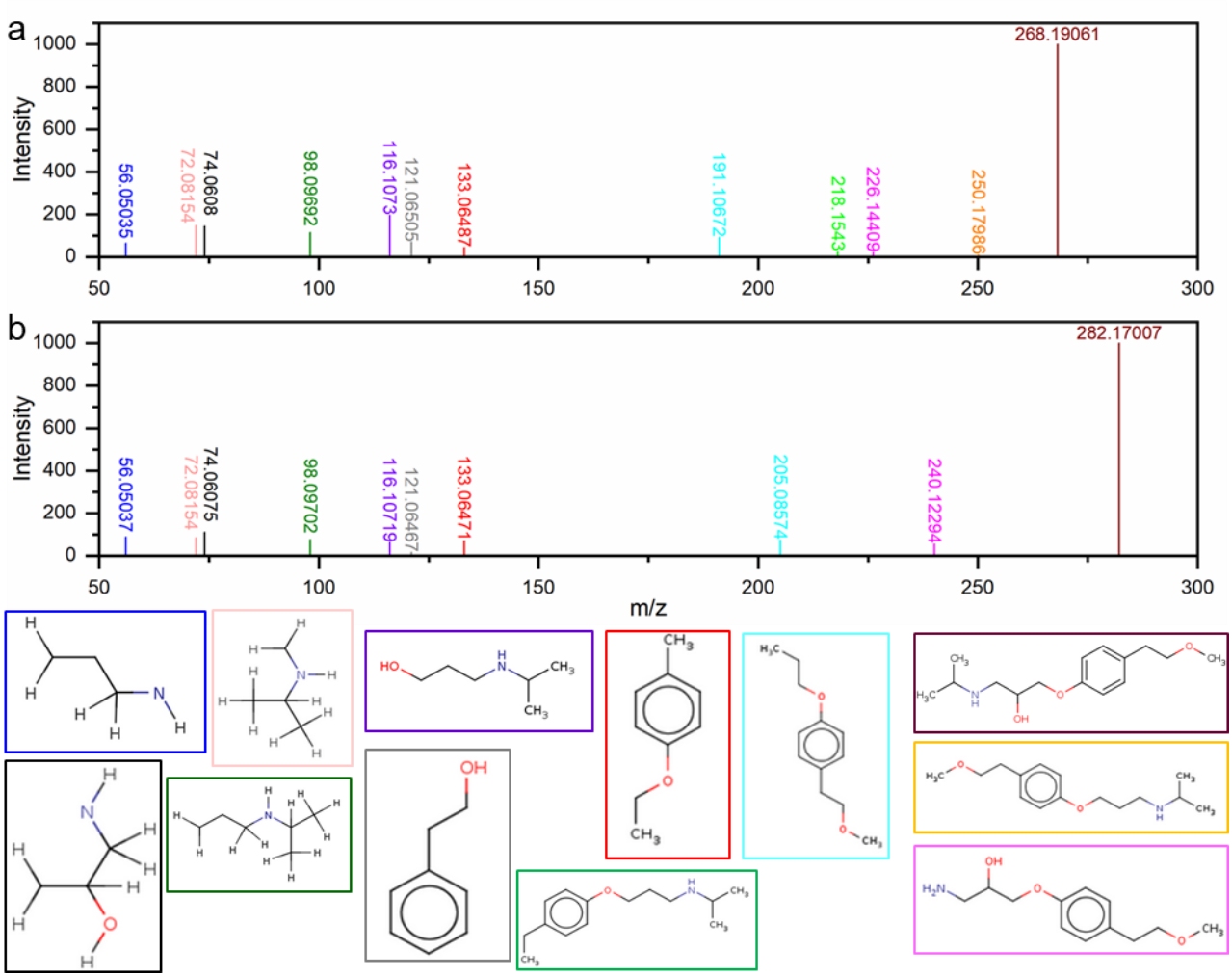


Figure 65. MS^2 spectra of metoprolol (a) and metoprolol_TP282 (b) and the color-coded structures of metoprolol fragments from Metlin. A shift of 13.979 Da, which corresponds to adding an oxygen atom but removing 2 hydrogen atoms, is seen for the cyan (m/z 191 to m/z 205) and pink (m/z 226 to m/z 240) product ions and the brown precursor ions. Since the mass shift is not observed for the red product ions (m/z 133) nor any of the smaller other ones, the hydroxylation site can only be in one place, at the end of the chain.

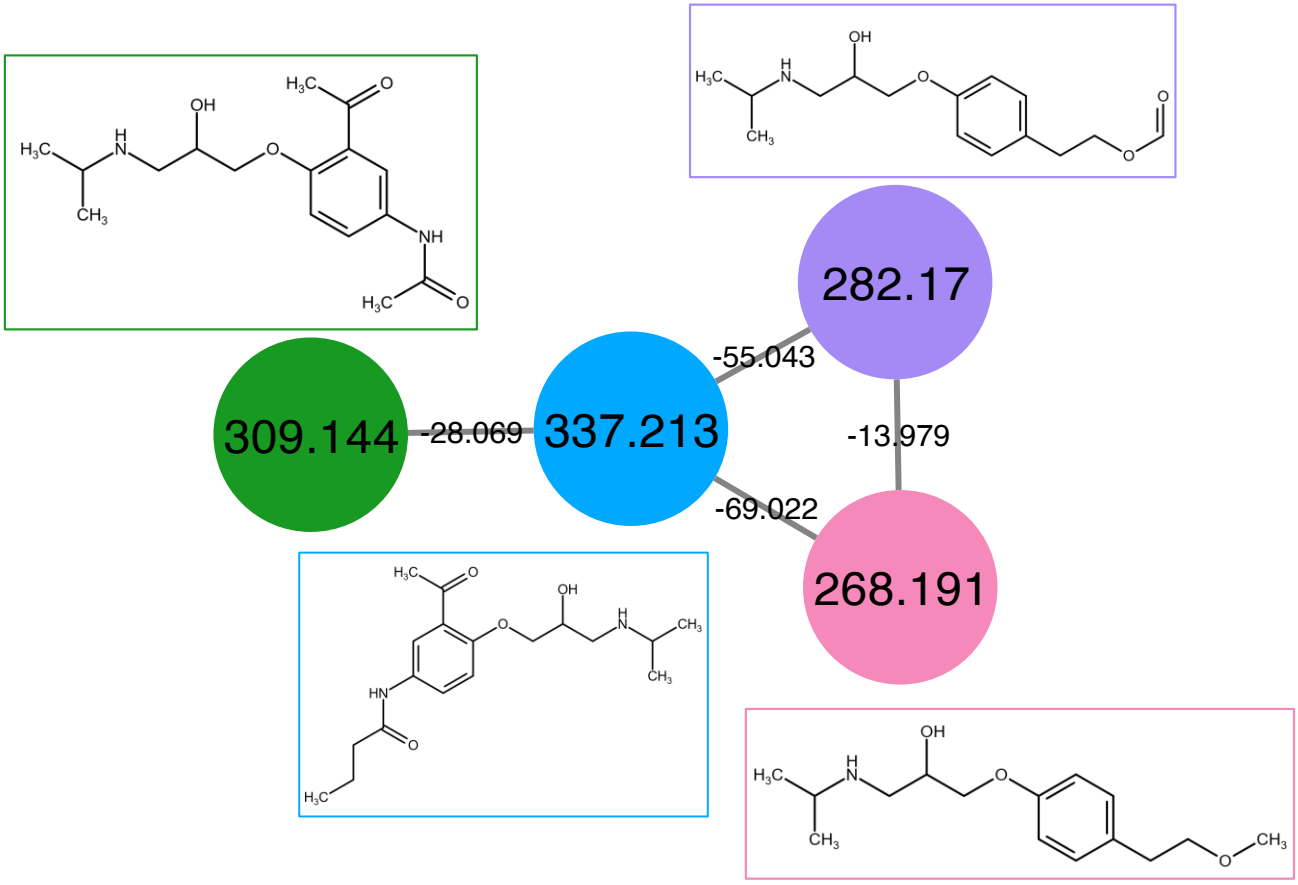


Figure 66. Molecular network generated with GNPS of the beta-blockers diacetalol (green; m/z 309.144), acebutolol (blue; m/z 337.213), metoprolol (orchid; m/z 268.191), and the proposed structure of metoprolol_TP282 (lavender; m/z 282.17)

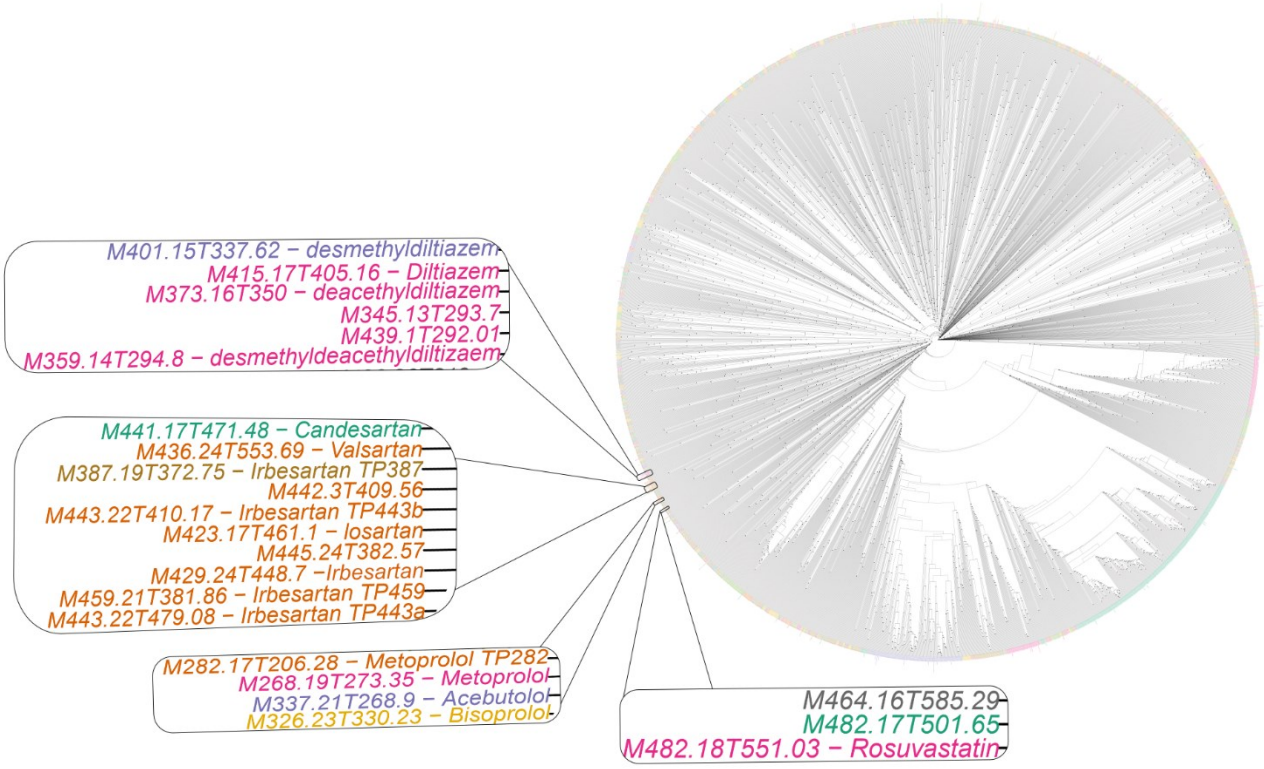


Figure 67. Location of several pharmaceutical compounds and transformation products on the dendrogram generated with CluMSID. Compounds that share similar spectra are neighbours.

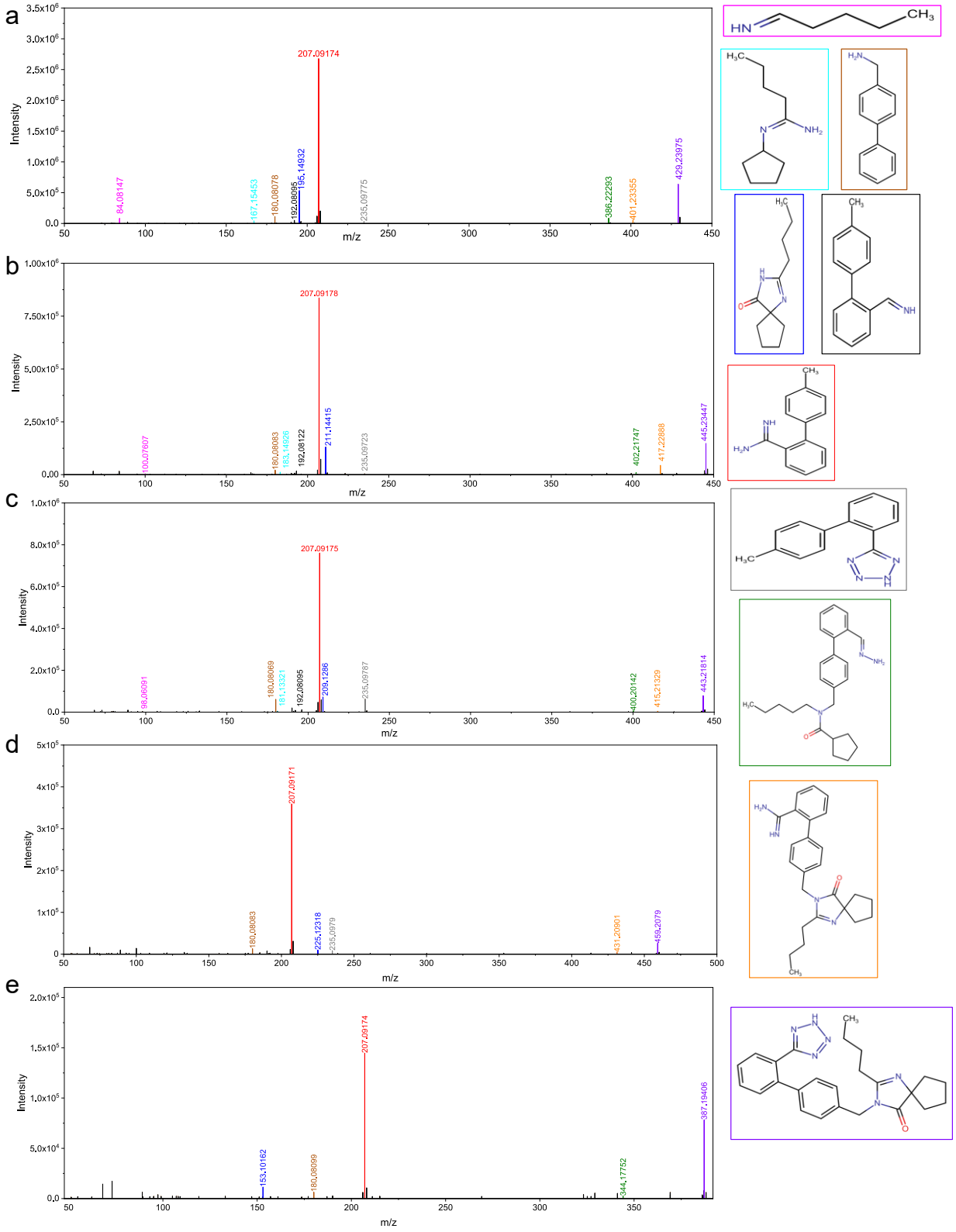

a
Intensity
m/z
207.09174
84.08147
167.15453
180.08078
192.08095
195.14932
235.09775
386.22293
401.23355
429.23975
b
207.09178
100.07607
180.08083
183.14926
192.08122
211.14415
235.09723
402.21747
417.22688
445.23447
c
207.09175
98.06091
180.08069
181.13321
192.08095
209.1286
235.09787
400.20142
415.21329
443.21814
d
207.09171
180.08083
225.12318
235.0979
431.20901
459.2079
e
207.09174
153.10162
180.08099
344.17752
387.19406

Figure 68. MS^2 spectra of Irbesartan (a) with color corresponding fragment ions from Metlin. The observed mass shifts located on the pink, cyan, blue, green, and orange for irbesartan_TP445 (b) indicates toward a hydroxylation on the n-butyl moiety. Irbesartan_TP443 (c) is the oxidized TP of irbesartan_TP445. Irbesartan_TP459 (d) would be the result of another hydroxylation on irbesartan_TP443 to form an acid. Irbesartan_TP387 (e) looks to be the result from the loss of a C_3H_6 on the n-butyl or the cyclopentyl.

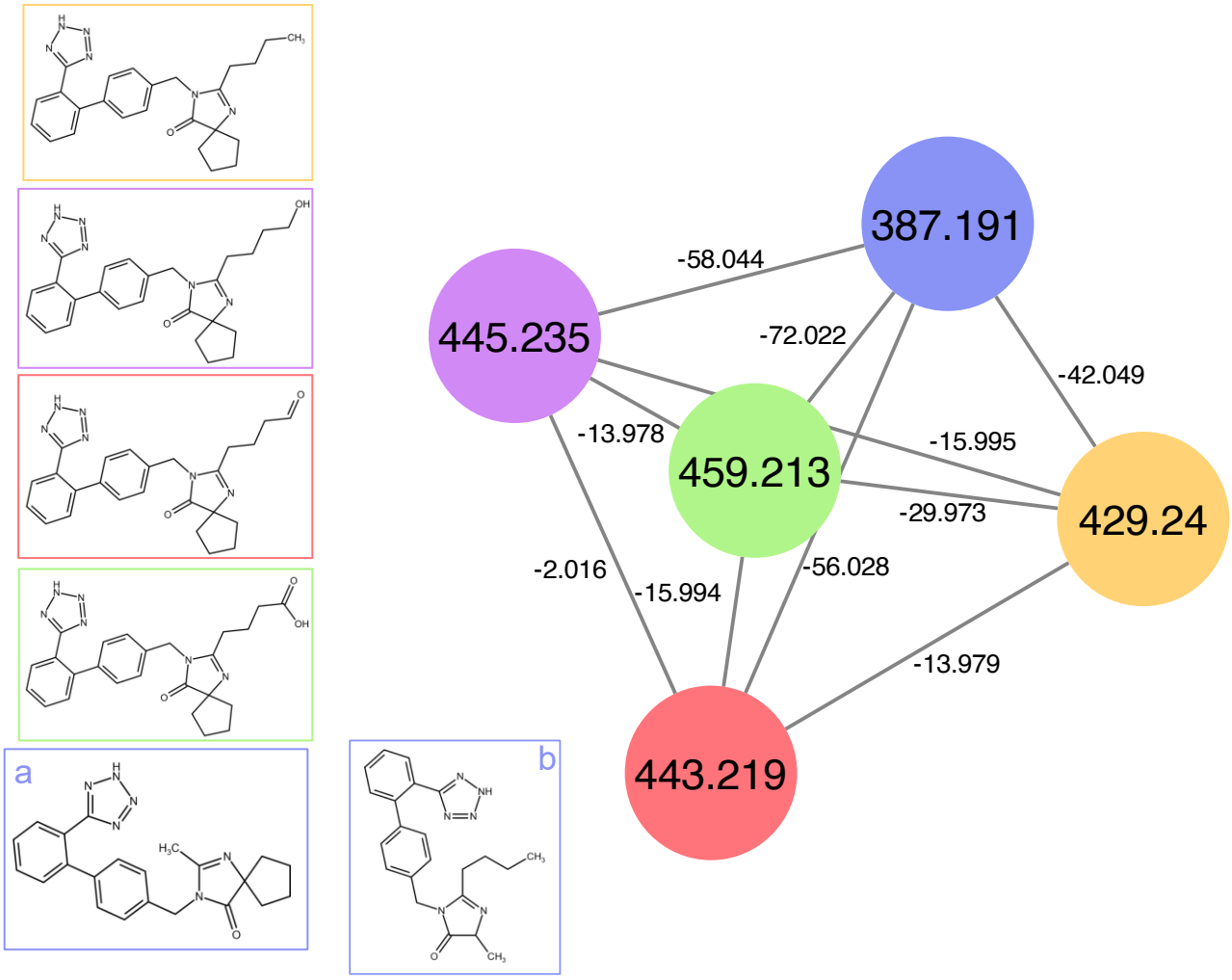
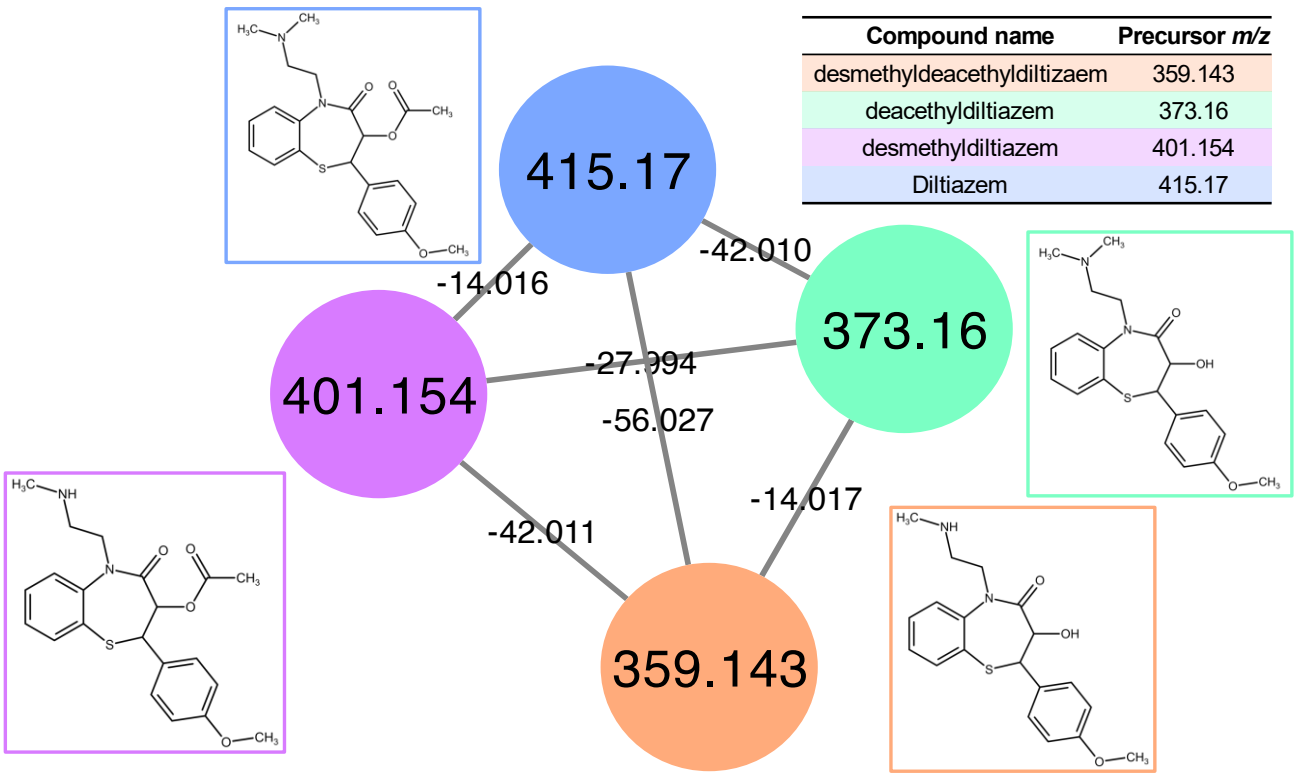


Figure 69. Irbesartan (melon, *m/z* 429.24), along with the proposed structures of irbesartan_TP445 (mauve, *m/z* 445.235), irbesartan_TP443 (amaranth, *m/z* 443.219), irbesartan_TP459 (mint, *m/z* 459.213), and irbesartan_TP387 (cornflower, *m/z* 387.191). Irbesartan_TP387 has two proposed potential transformation products in a and b.



Compound name	Precursor *m/z*
desmethyldeacethyldiltizaem	359.143
deacethyldiltiazem	373.16
desmethyldiltiazem	401.154
Diltiazem	415.17

Figure 70. Molecular network of diltiazem (blue) and its transformation products desmethyldiltiazem (purple), deacethyldiltiazem (turquoise), and desmethyldeacethyldiltizaem (peach).

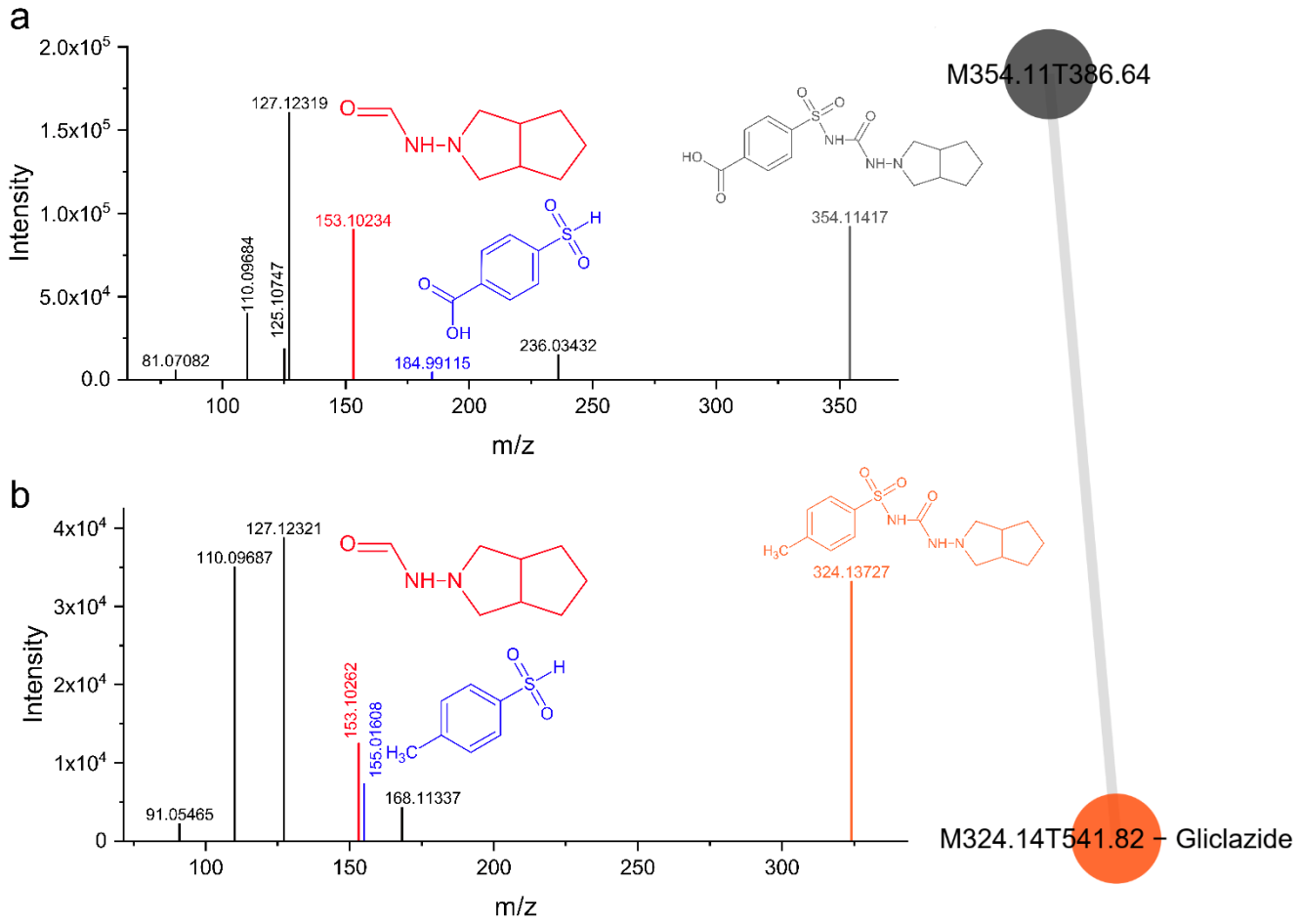


Figure 71. Spectra and structure of the transformation product of gliclazide_TP354 (a) and gliclazide (b) along with their respective product ions. On the left is the molecular cluster generated with CluMSID.

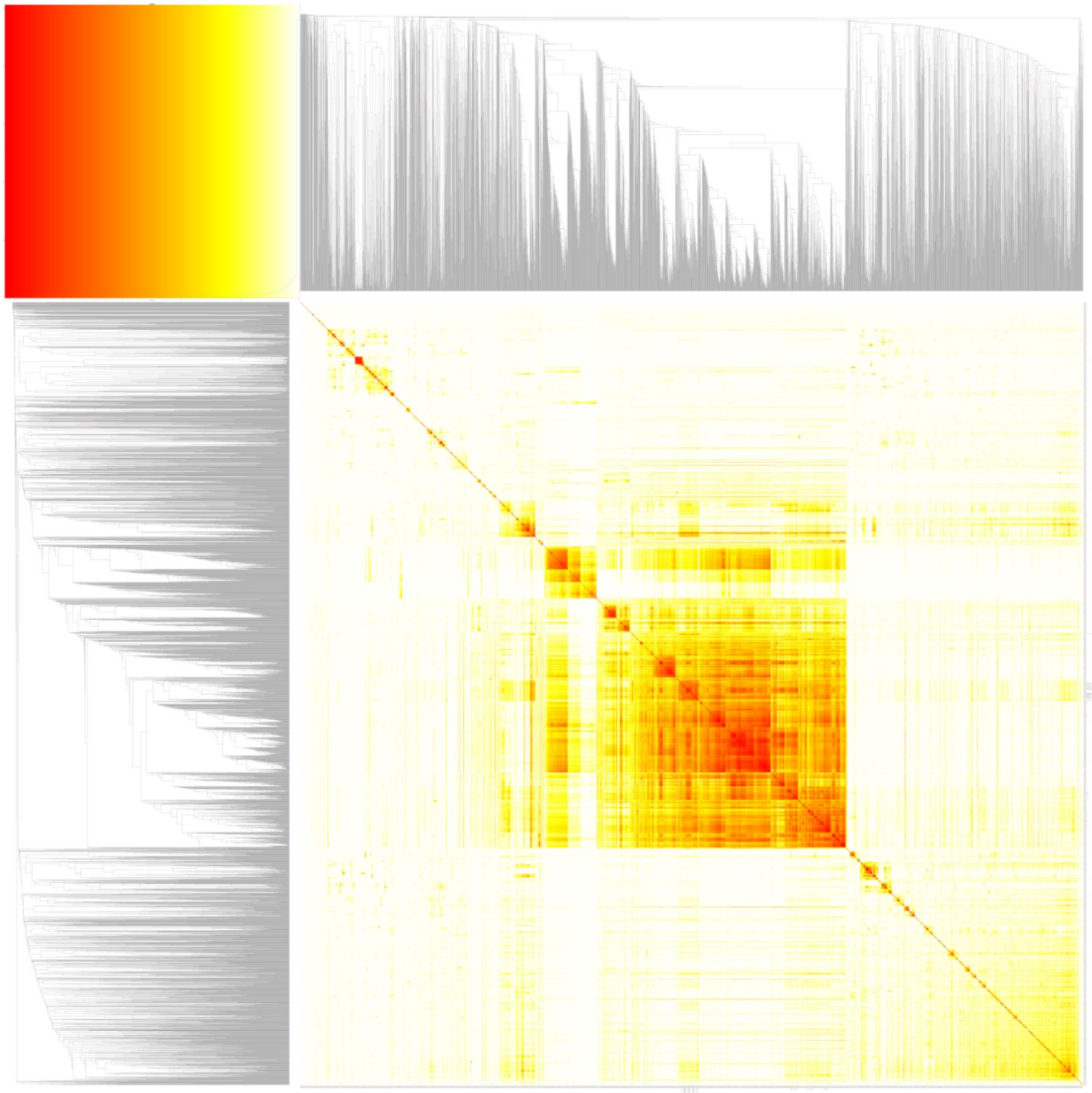
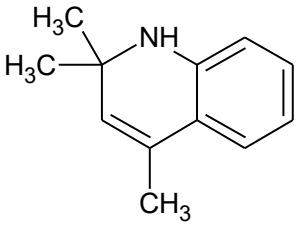


Figure 72. Heatmap and dendrogram of the spectra similarity between the precursors detected in downstream Farnham. Red color means a similarity of one while white means a similarity of zero. The closer to red a color is, the more similar the spectra are. The closer the arborescence of precursors (top and left) to the dendrogram is, the more similar their MS^2 spectra are. In this figure, a very large cluster can be observed in the center of the heatmap and dendrogram.

9.5.2.1 Confirmed compounds

2,2,4-Trimethyl-1,2-dihydroquinoline (TMQ)

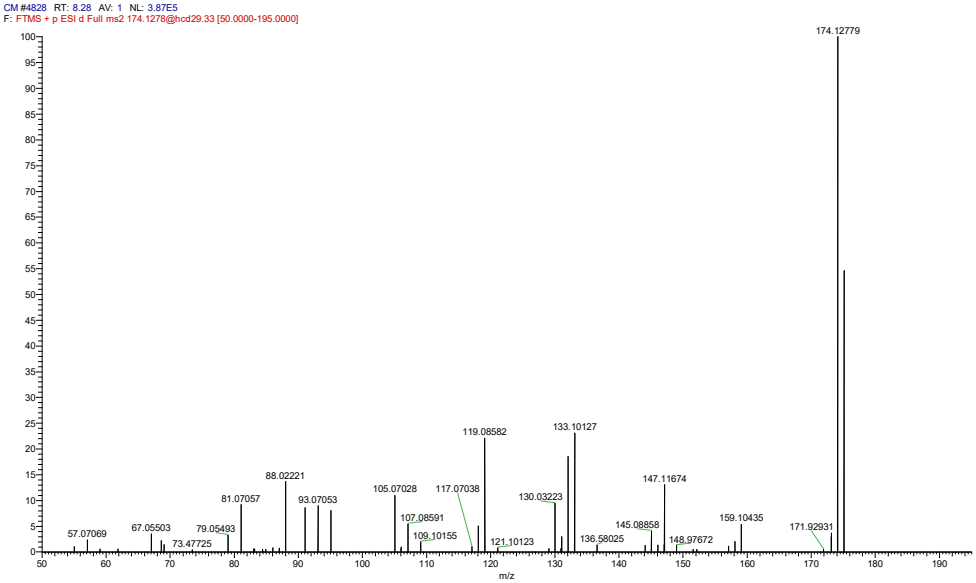


Polymer additive, consumer product additive

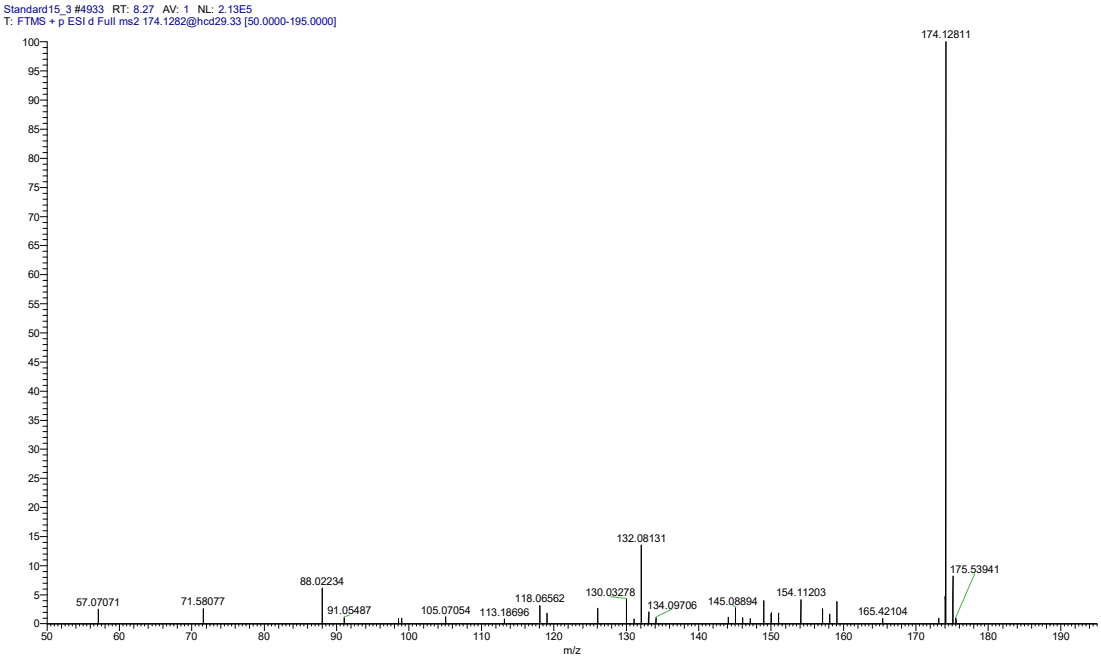
Molecular formula: C12H15N

Monoisotopic mass: 173.120449483

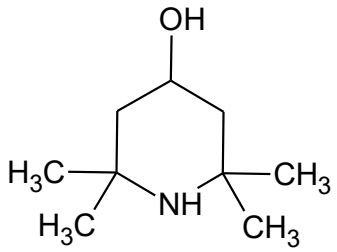
River sample



Reference standard



2,2,6,6-Tetramethyl- 4-piperidinol

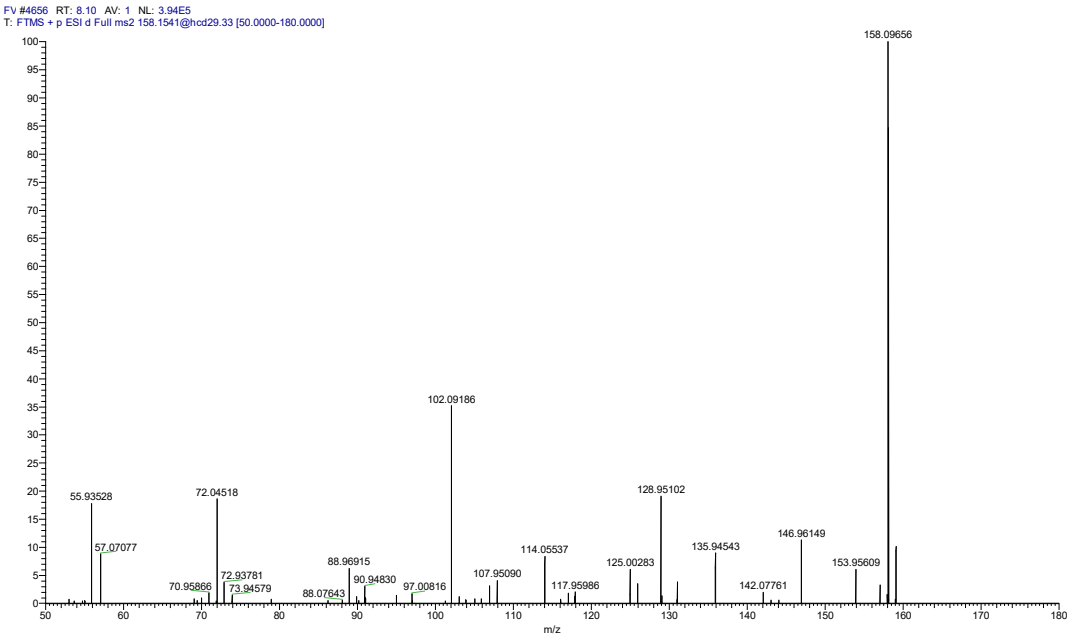


Polymer additive, consumer product additive

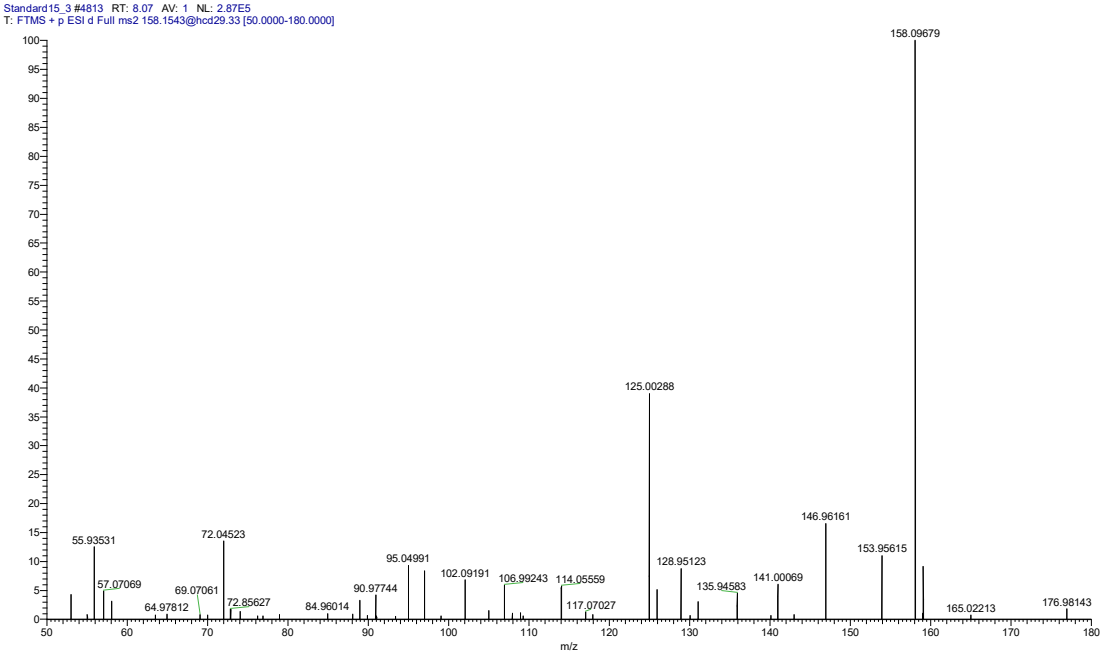
Molecular formula: C9H19NO

Monoisotopic mass: 157.146664230

River sample



Reference standard

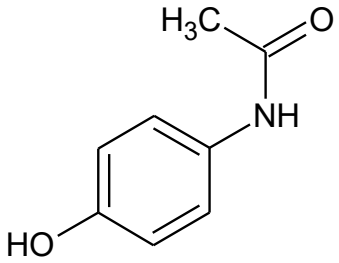


Acetaminophen

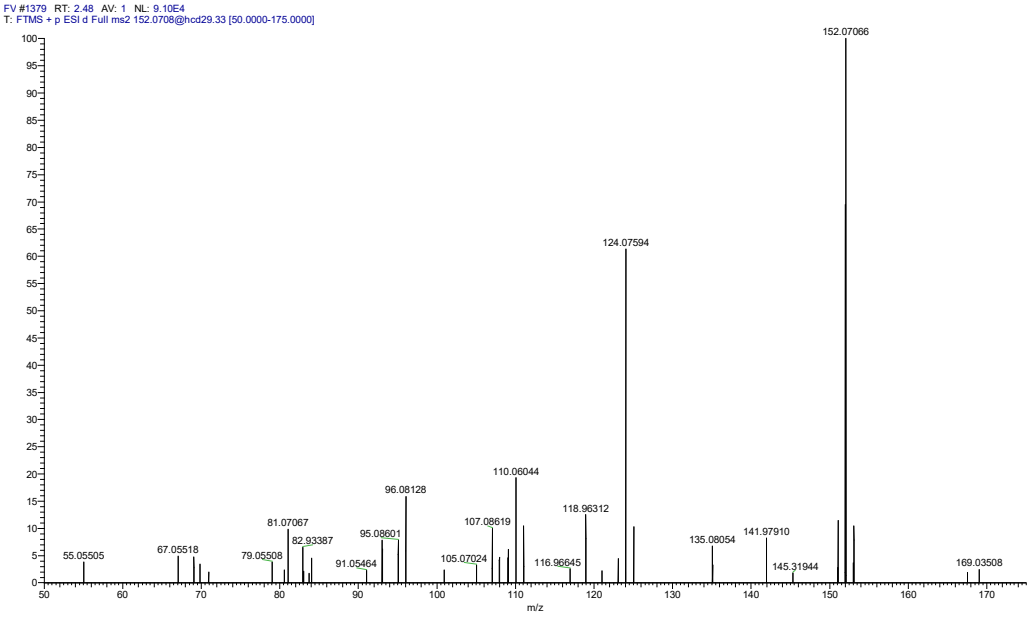
Nervous system, pharmaceutical

Molecular formula: C8H9NO2

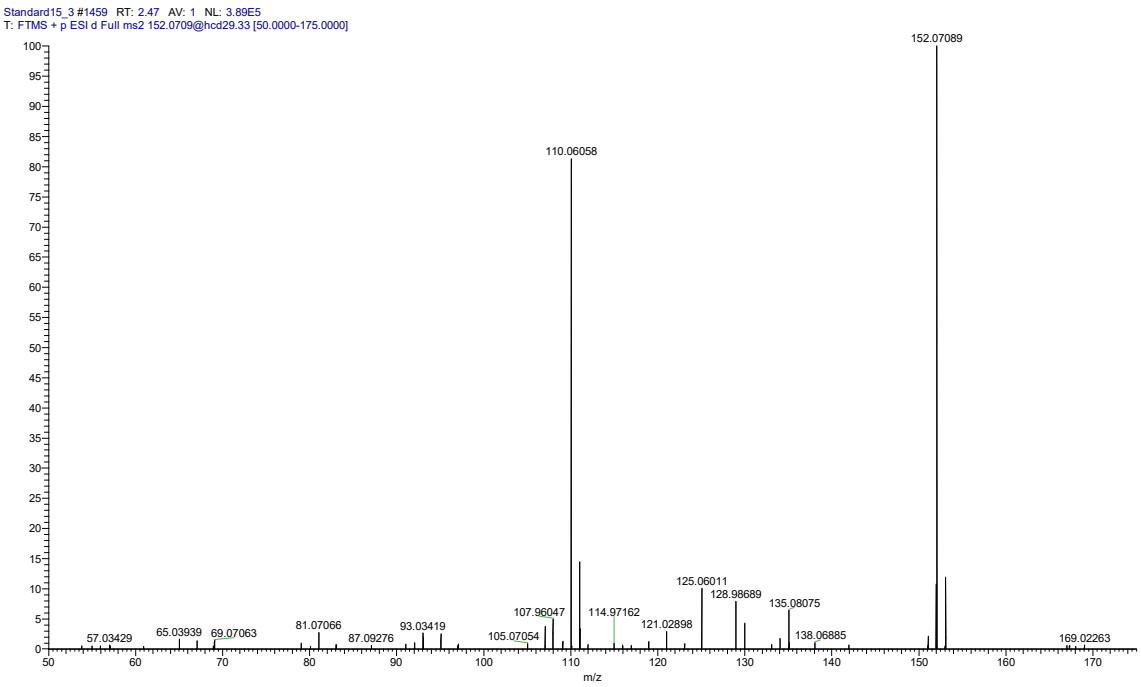
Monoisotopic mass: 151.063328530



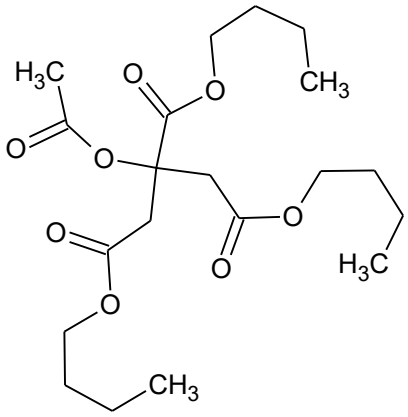
River sample



Reference standard



Acetyltributyl citrate

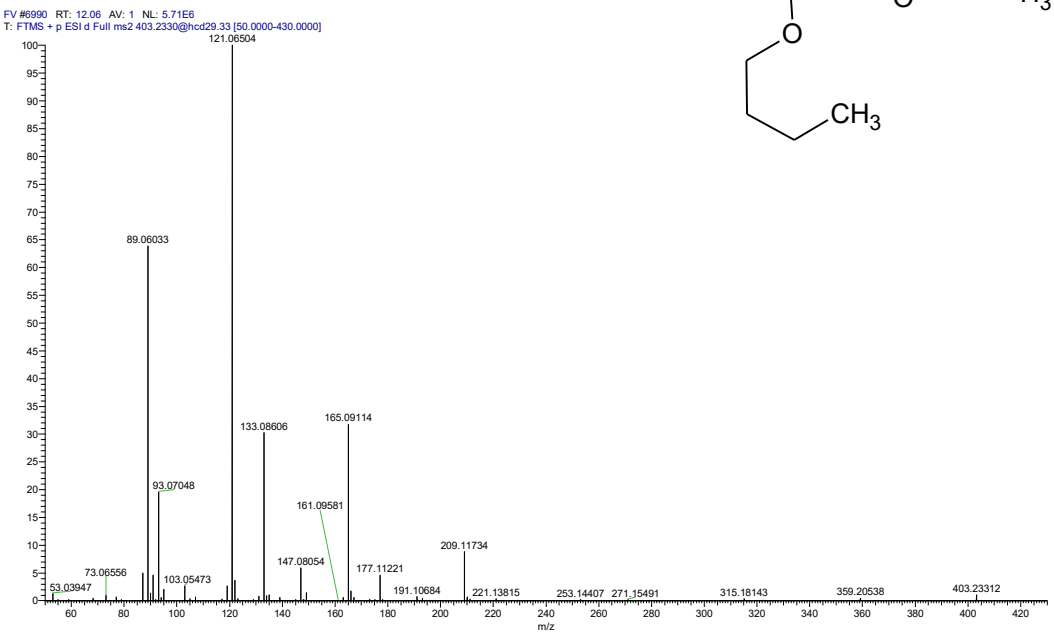


Polymer additive, consumer product additive

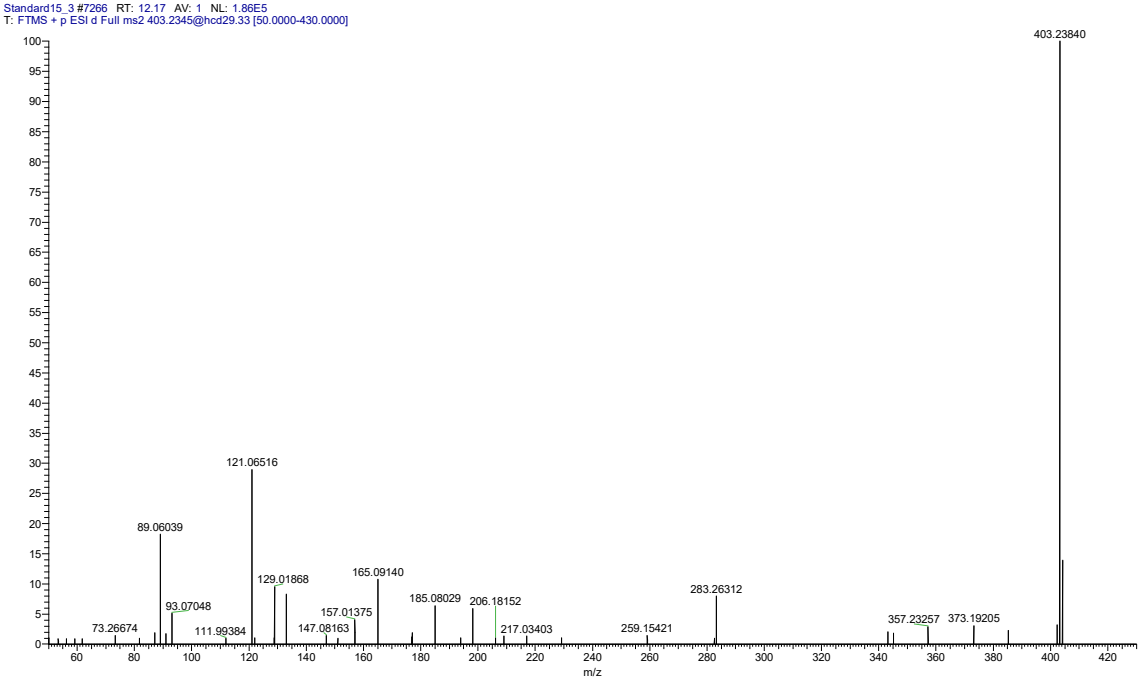
Molecular formula: C20H34O8

Monoisotopic mass: 402.22536804

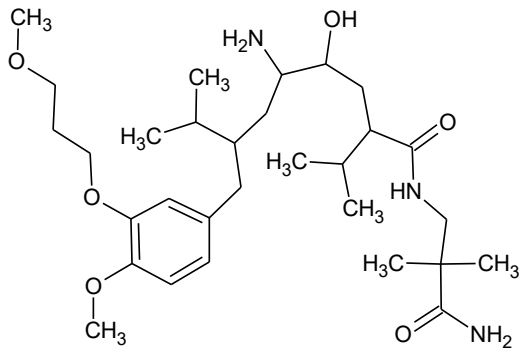
River sample



Reference standard



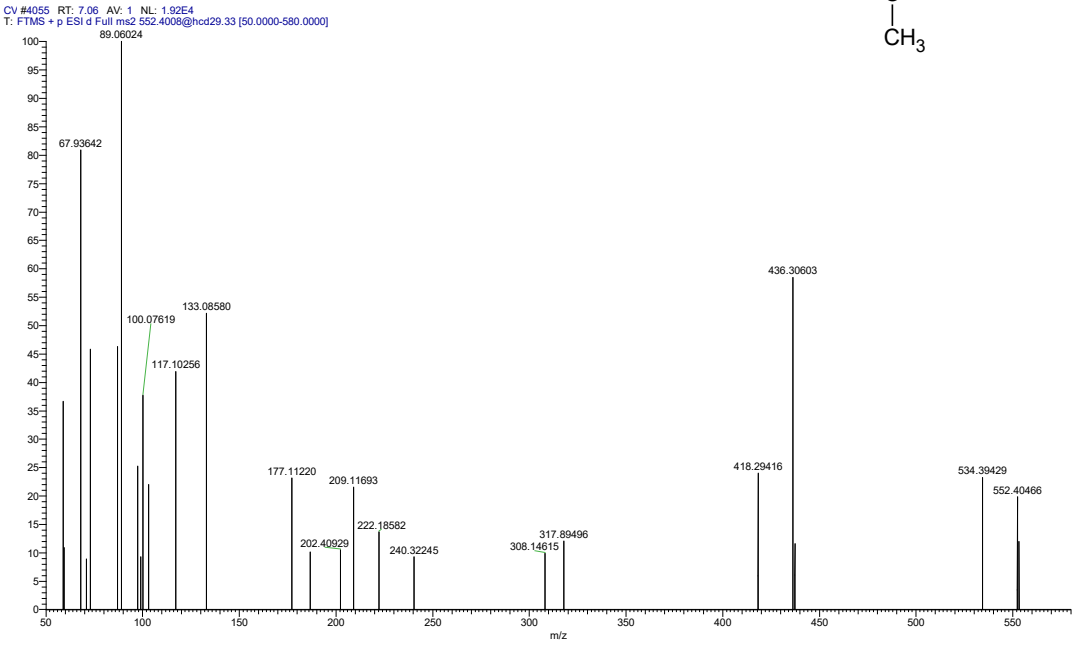
Aliskiren



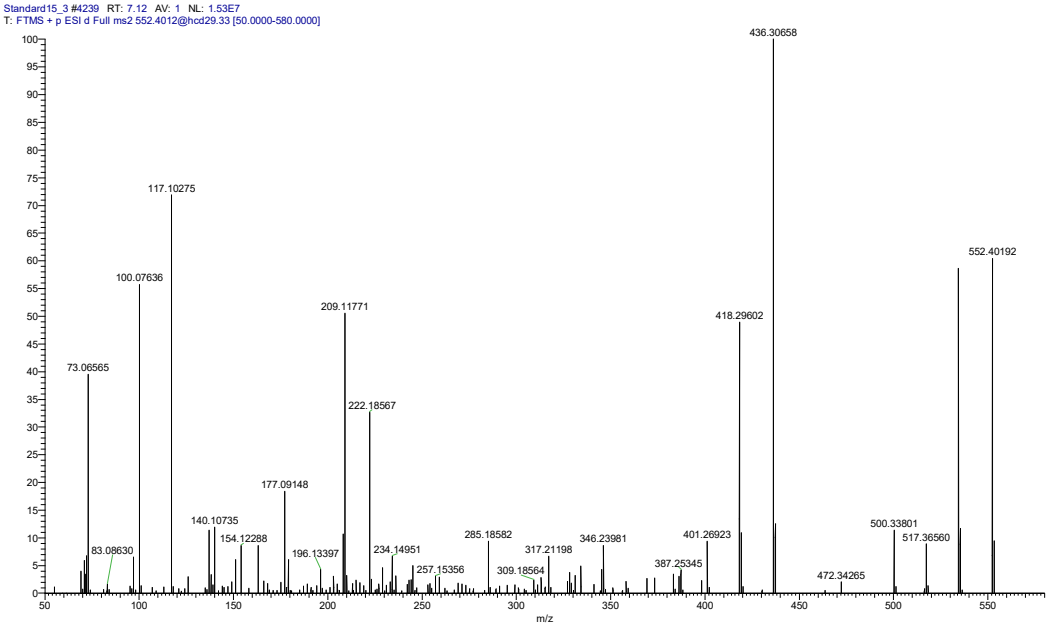
Cardiovascular system, pharmaceutical

Molecular Formula: C30H53N3O6
Monoisotopic mass: 551.39343642

River sample



Reference standard

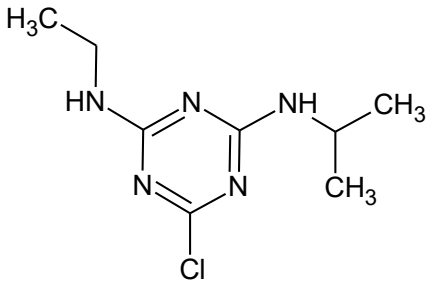


Atrazine

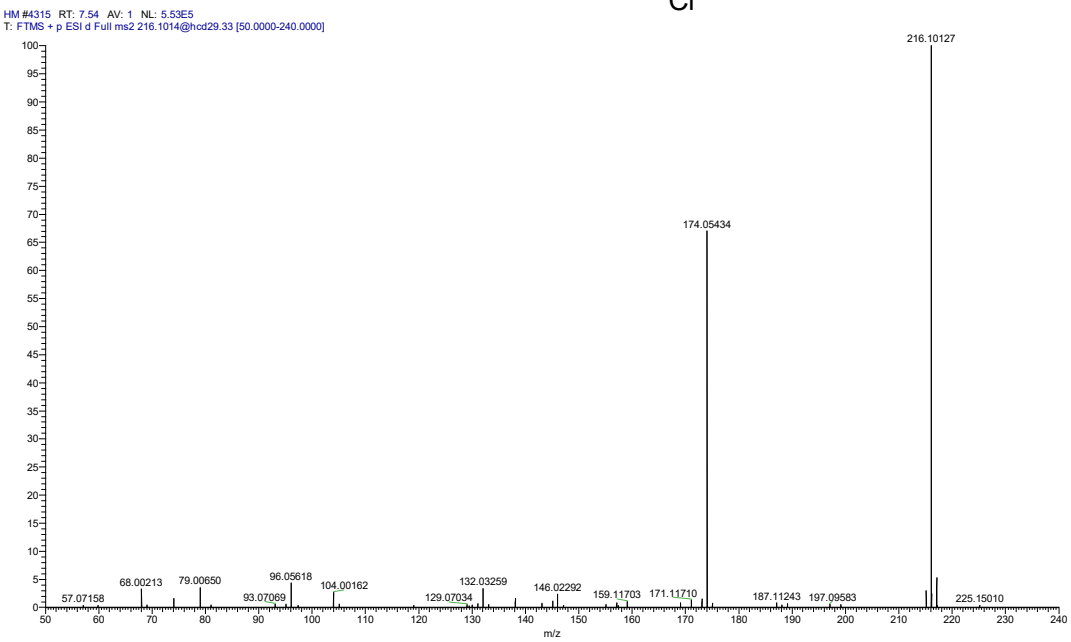
Herbicide, pesticide

Molecular formula: C8H14ClN5

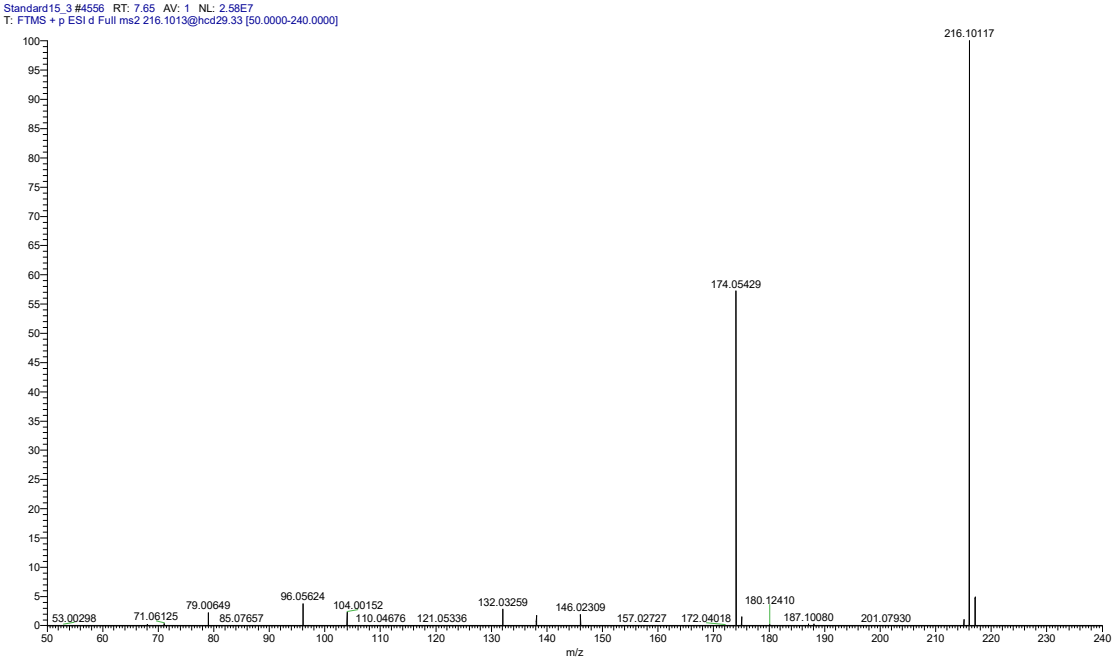
Monoisotopic mass: 215.0937732



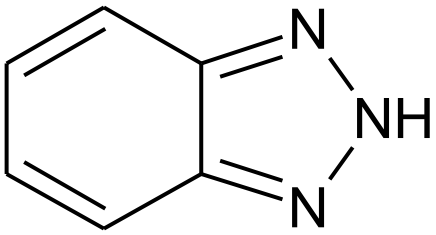
River sample



Reference standard



Benzotriazole-1H

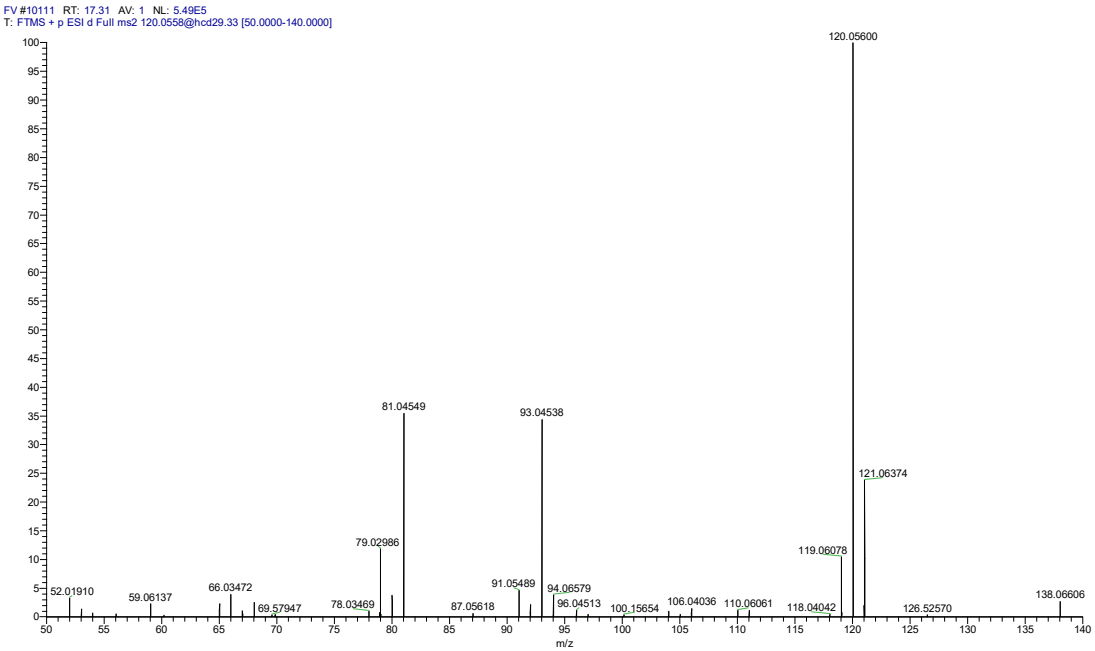


Polymer additive, consumer product additive

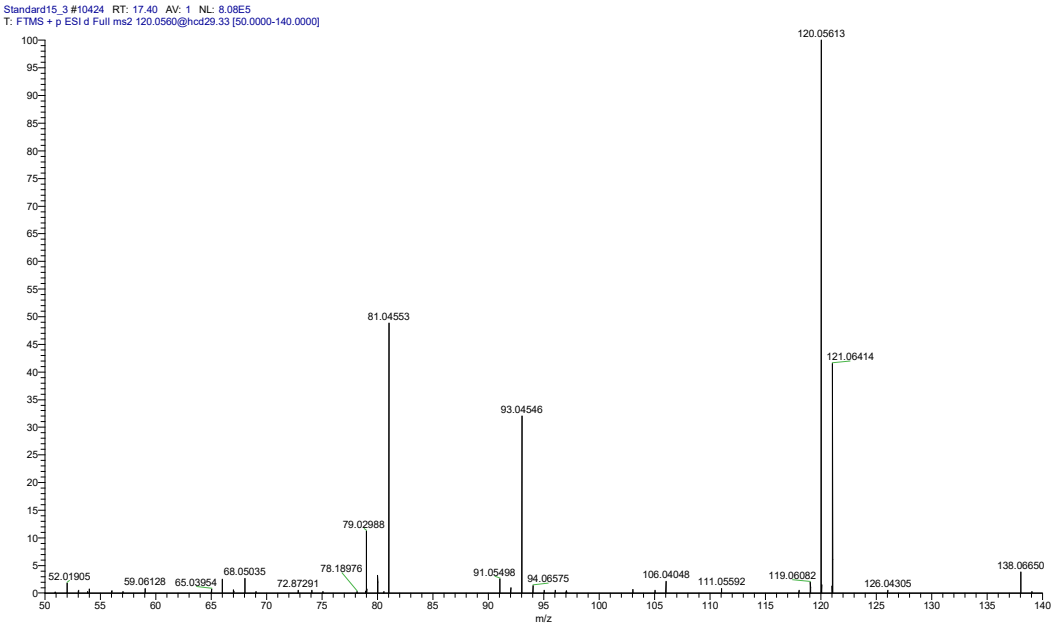
Molecular formula: C6H5N3

Monoisotopic mass: 119.048347172

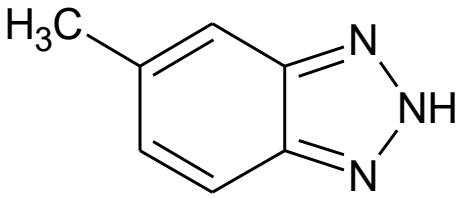
River sample



Reference standard



Benzotriazole-5-methyl-1H

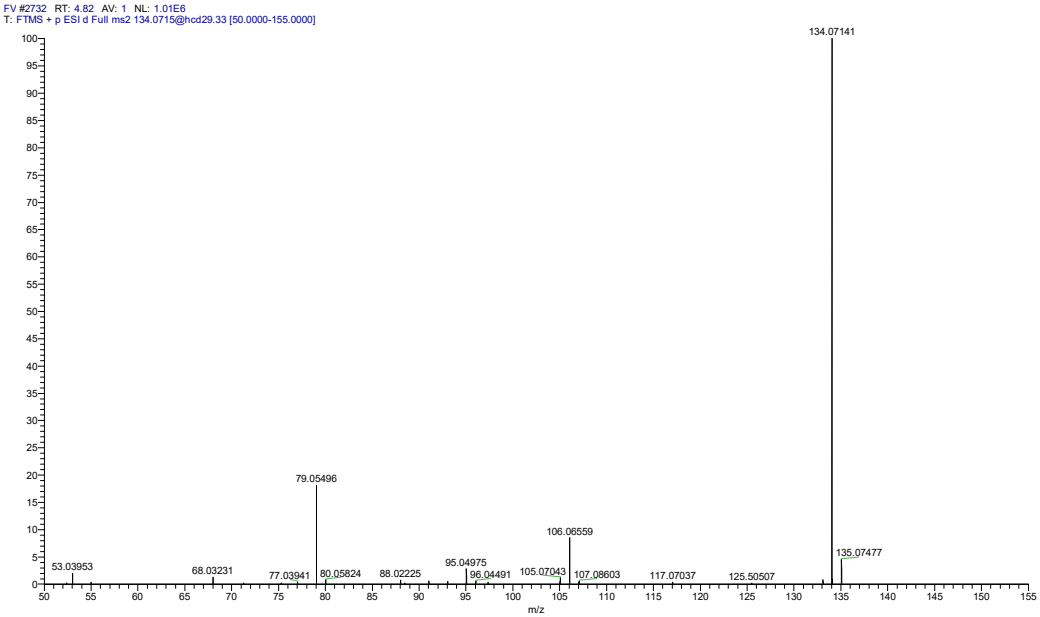


Polymer additive, consumer product additive

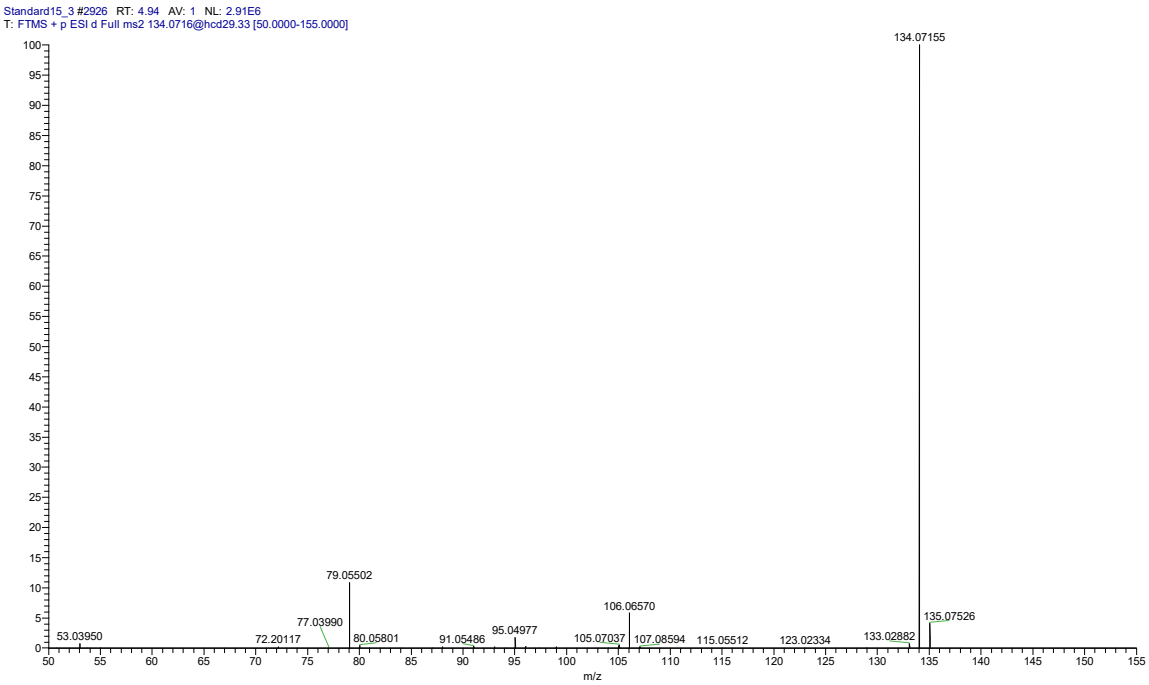
Molecular formula: C7H7N3

Monoisotopic mass: 133.063997236

River sample



Reference standard

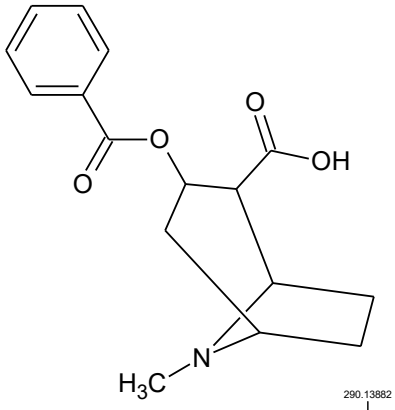


Benzoylecgonine

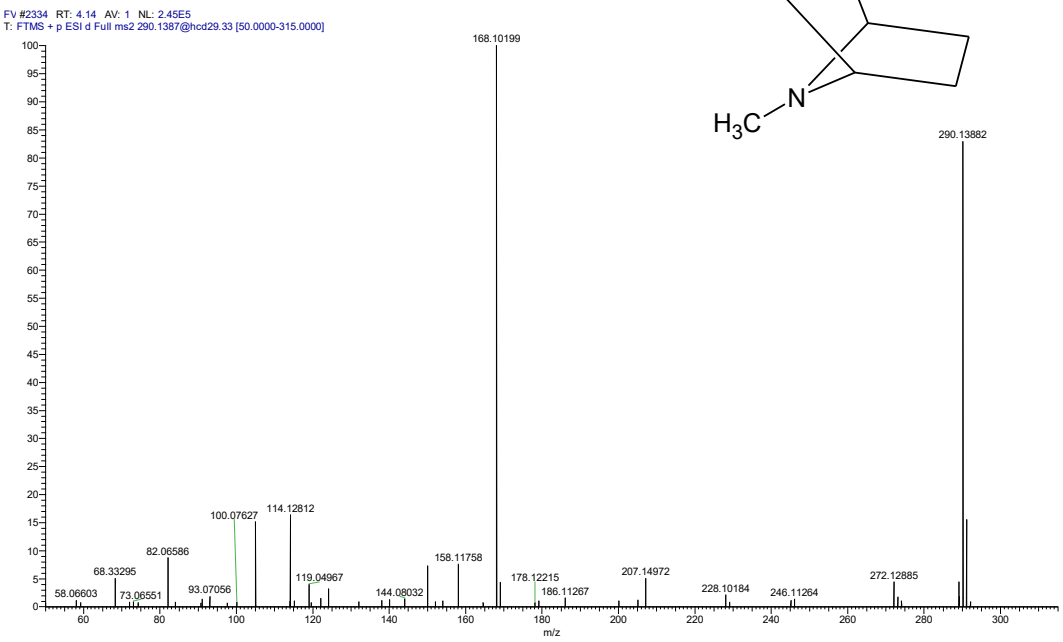
Opioid, illicit drugs

Molecular formula: C16H19NO4

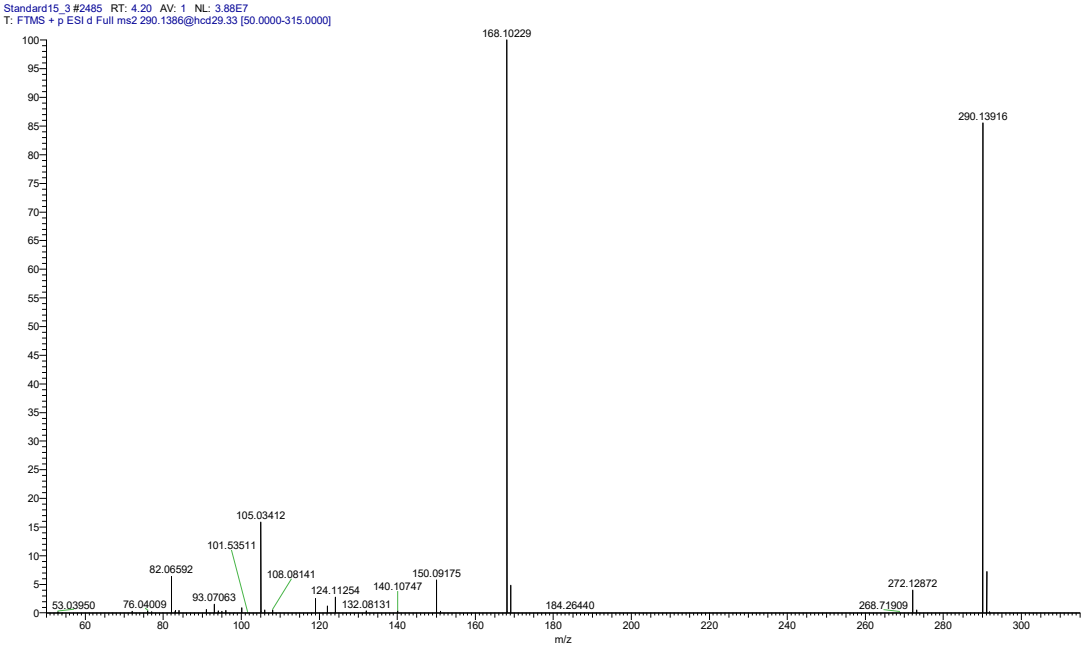
Monoisotopic mass: 289.13140809



River sample



Reference standard

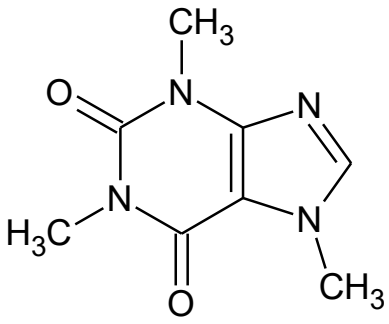


Caffeine

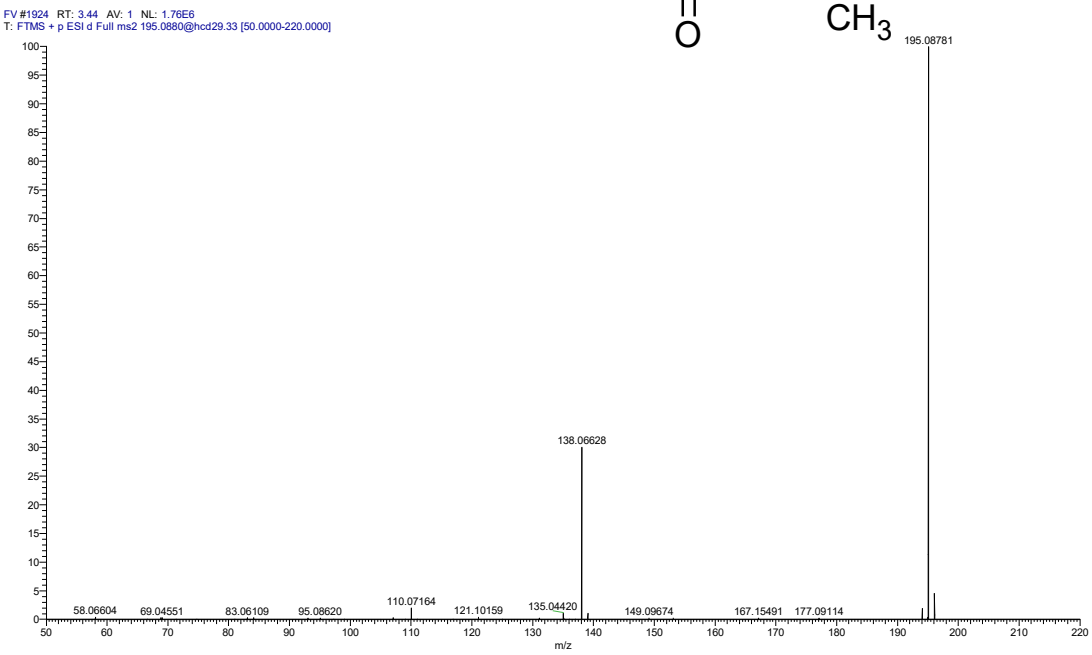
Nervous system, pharmaceutical

Molecular formula: C8H10N4O2

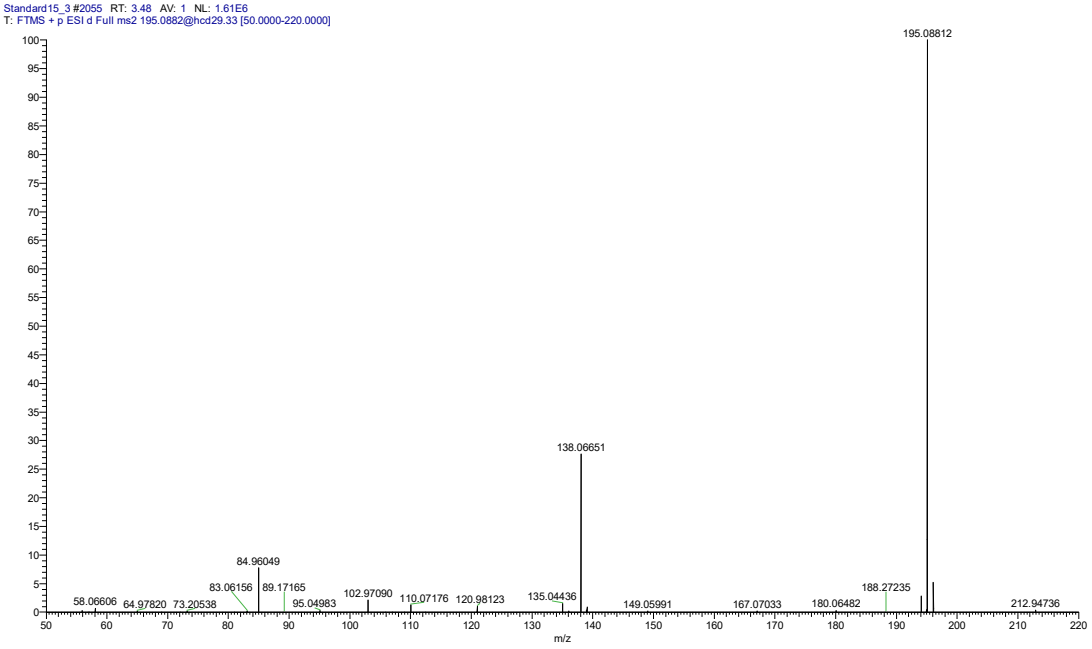
Monoisotopic mass: 194.08037557



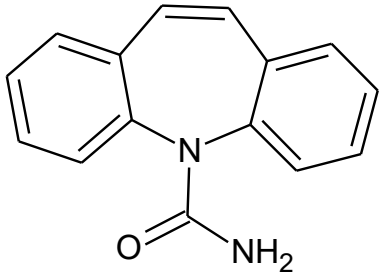
River sample



Reference standard



Carbamazepine

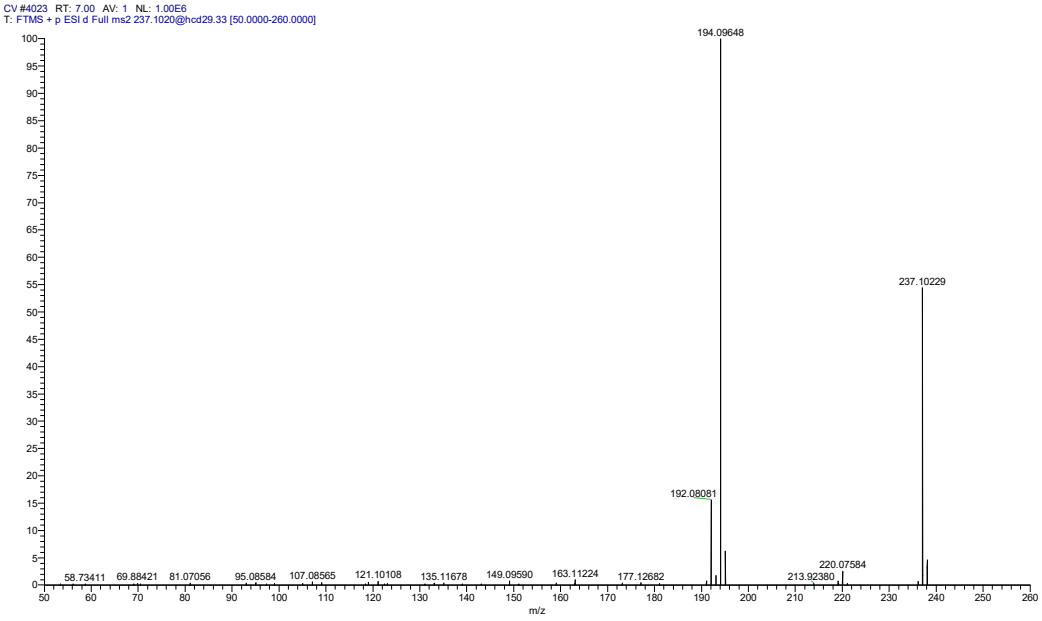


Nervous system, pharmaceutical

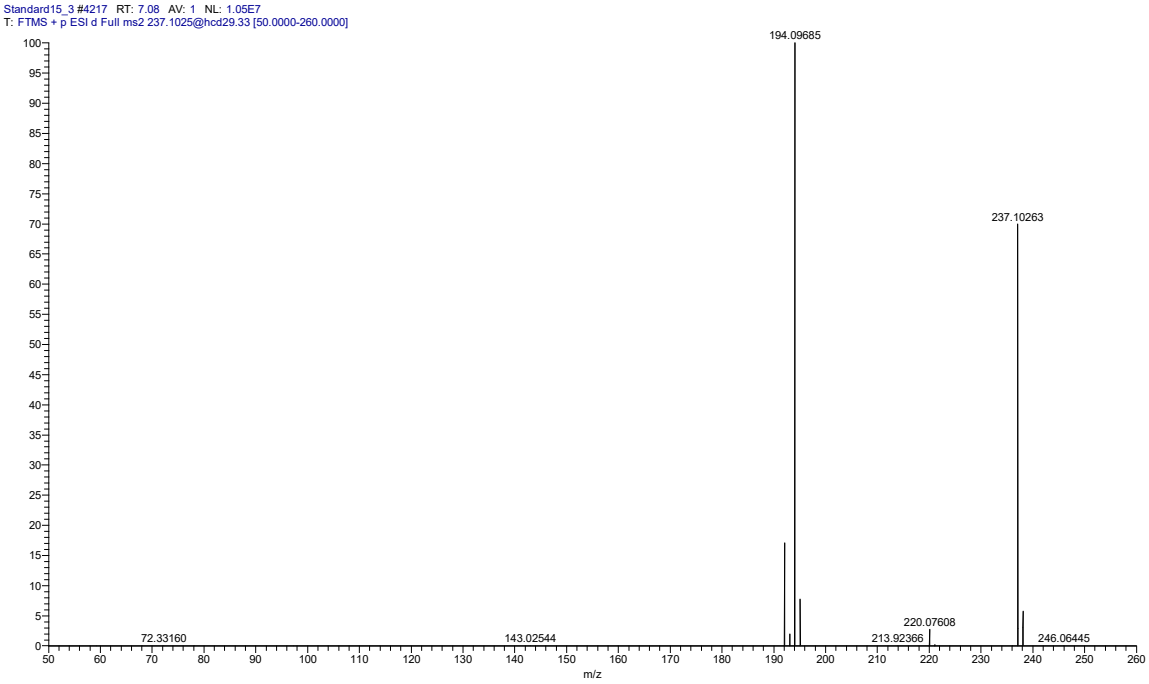
Molecular formula: C15H12N2O

Monoisotopic mass: 236.094963011

River sample



Reference standard

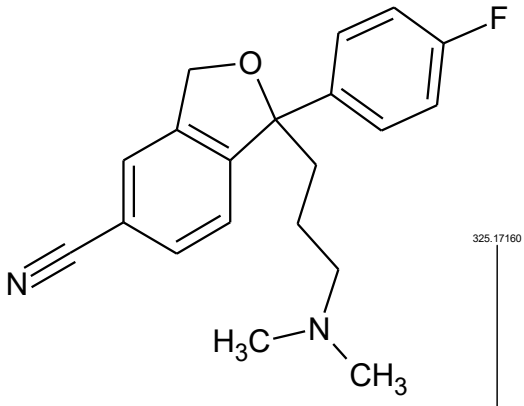


Citalopram

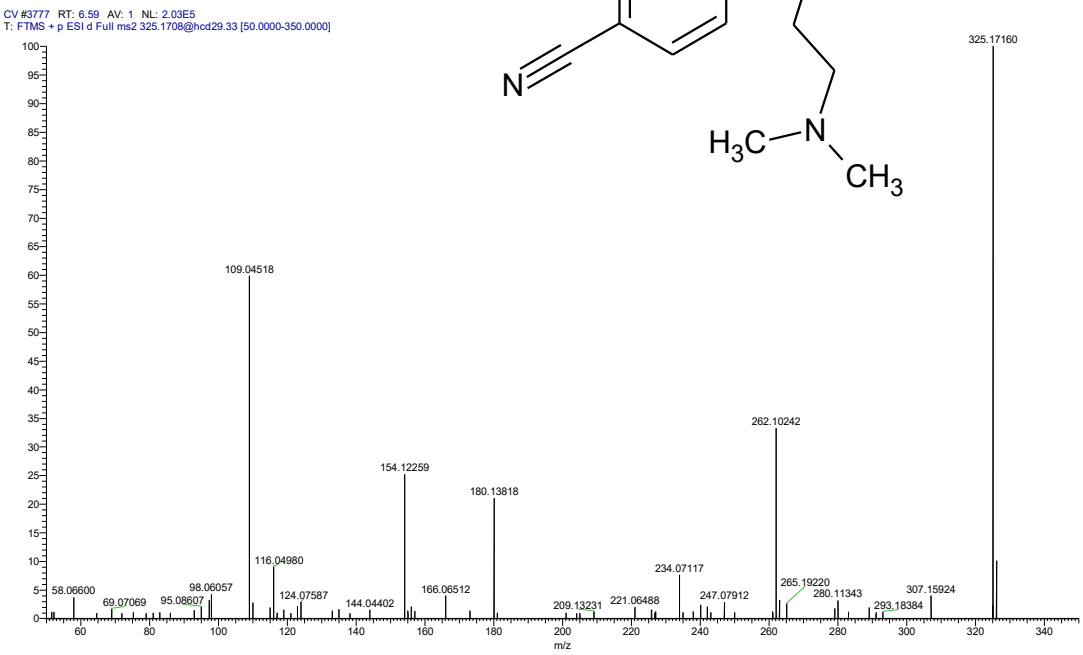
Nervous system, pharmaceutical

Molecular formula: C20H21FN2O

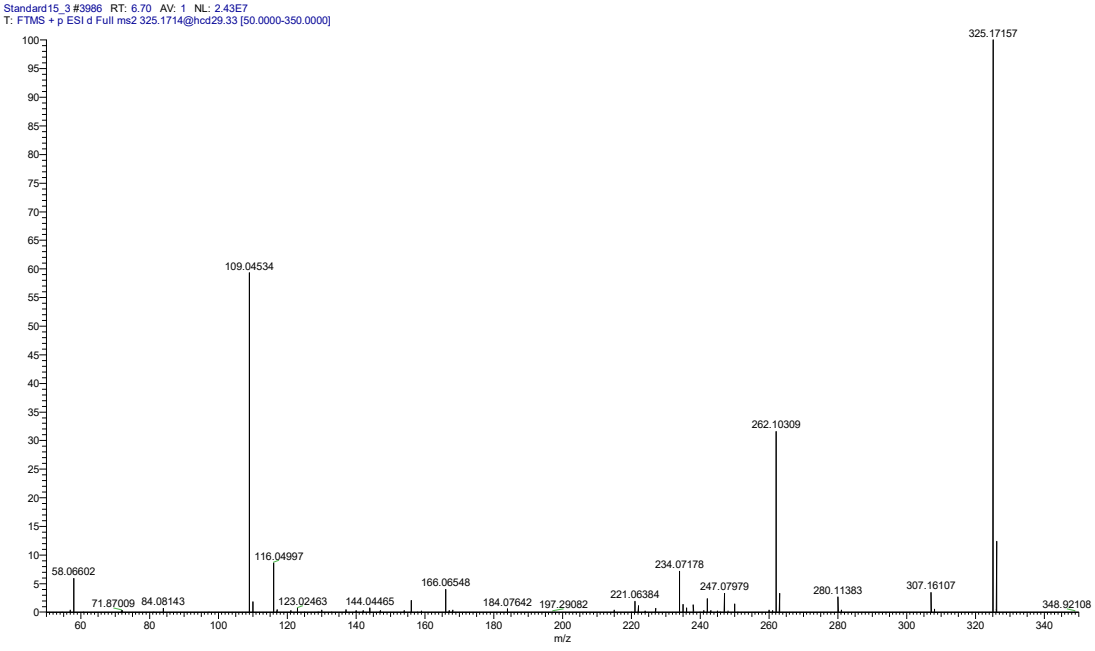
Monoisotopic mass: 324.16379146



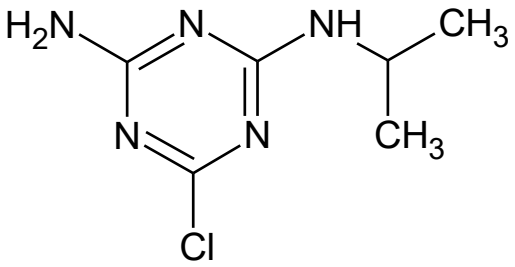
River sample



Reference standard



Desethylatrazine

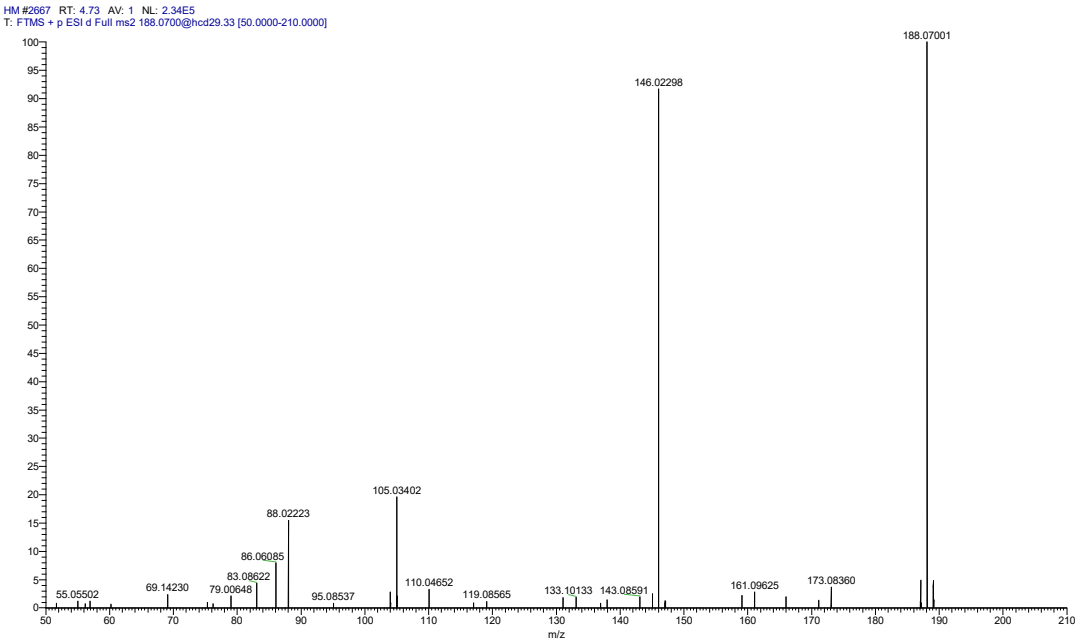


Herbicide, pesticide

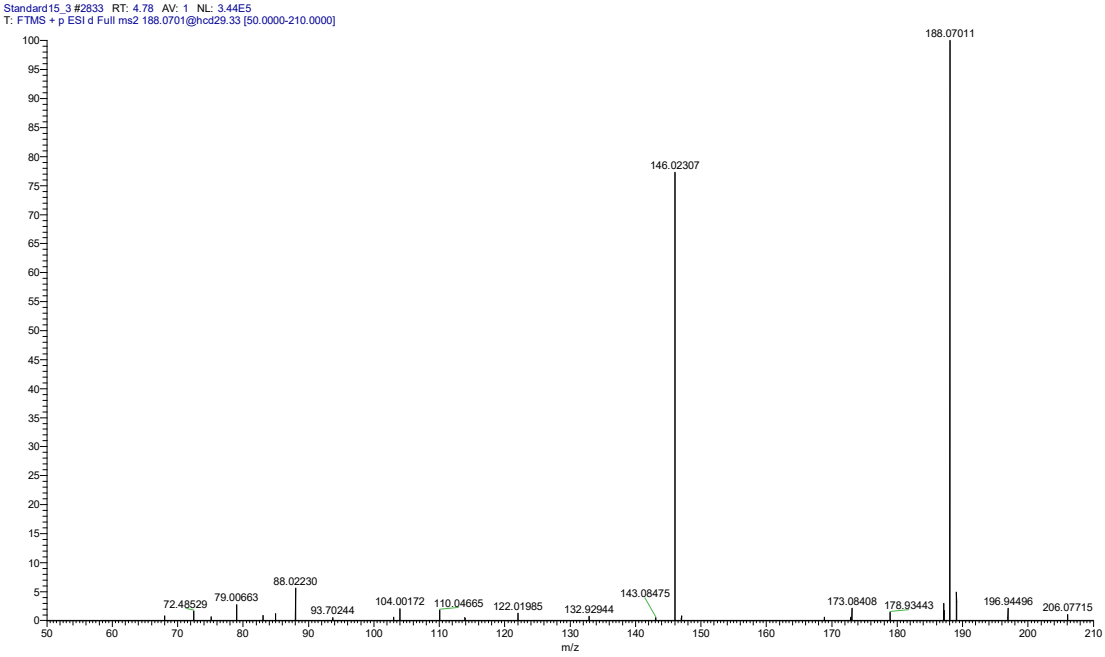
Molecular formula: C6H10ClN5

Monoisotopic mass: 187.0624730

River sample



Reference standard

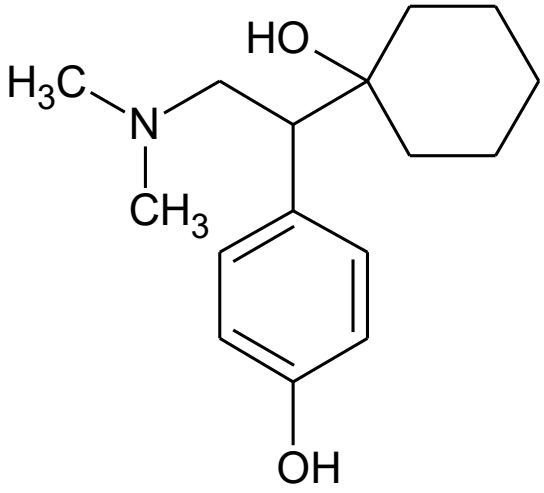


O-Desmethylvenlafaxine

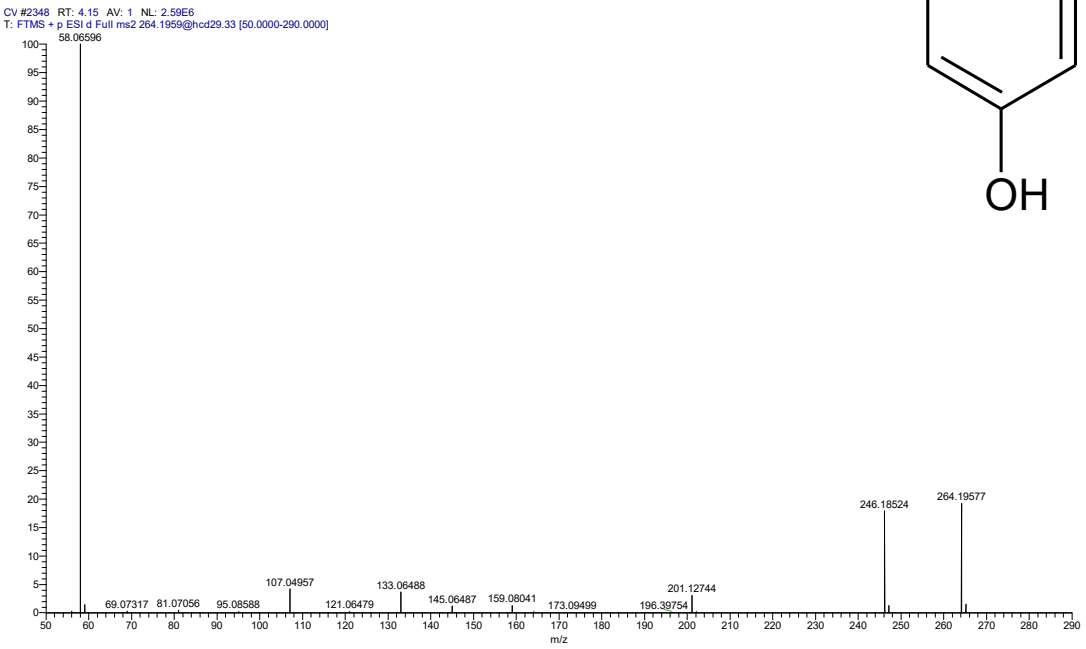
Nervous system, pharmaceutical

Molecular formula: C16H25NO2

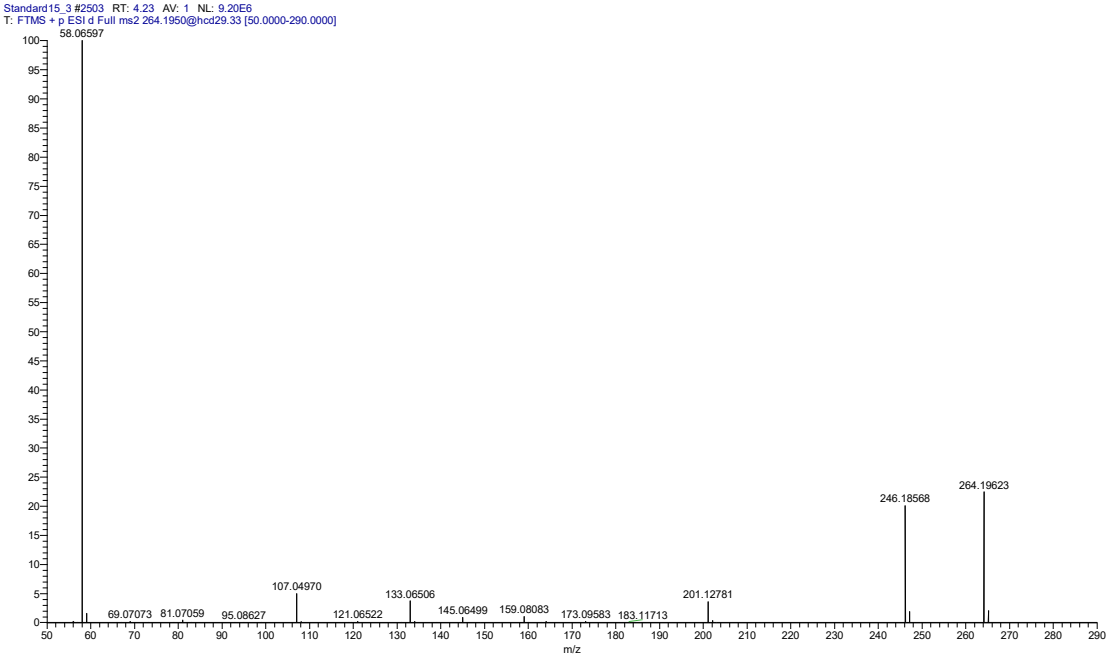
Monoisotopic mass: 263.188529040



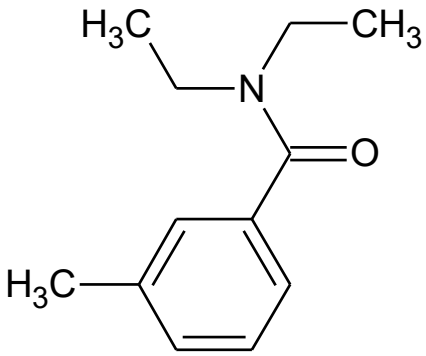
River sample



Reference standard



Diethyltoluamide

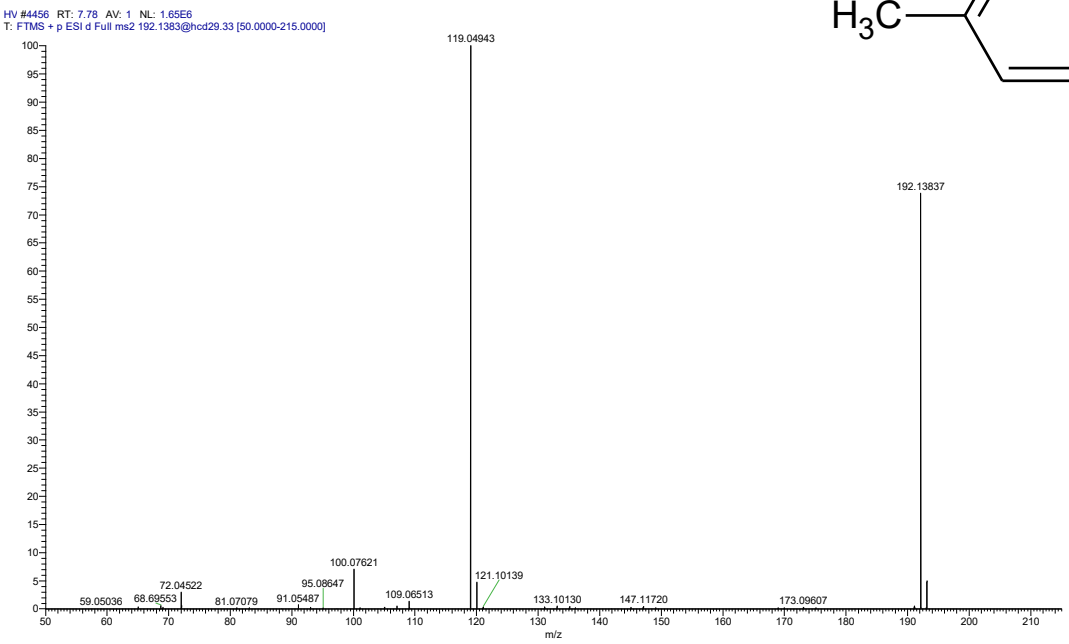


Cosmetic, consumer product additive

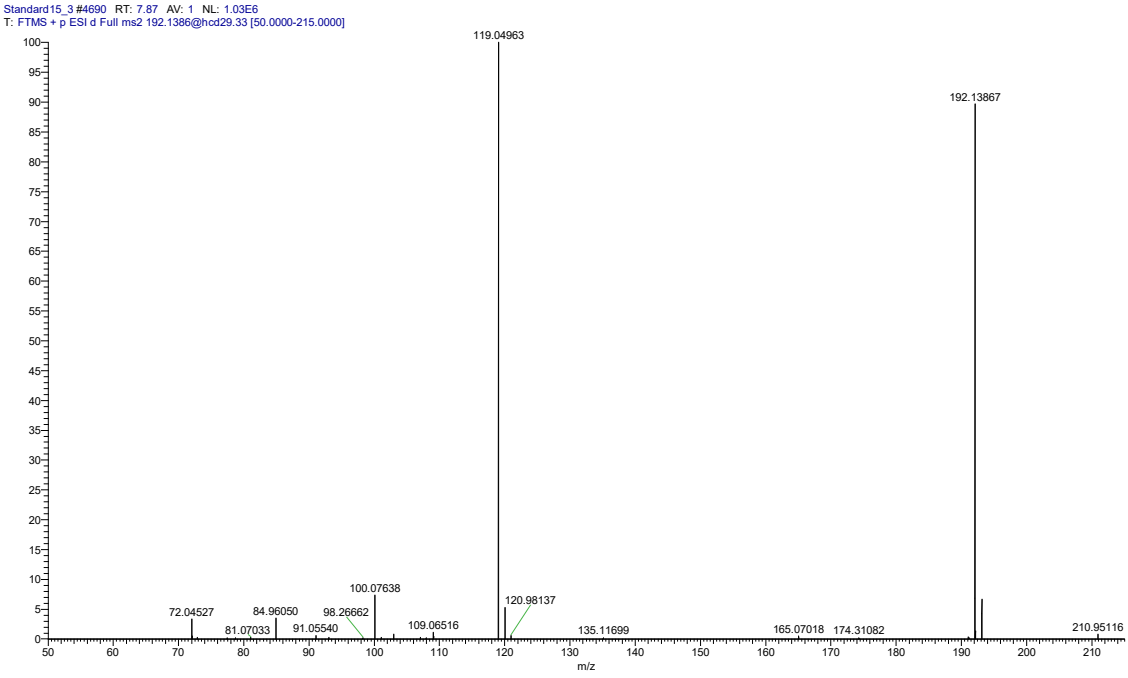
Molecular formula: C12H17NO

Monoisotopic mass: 191.131014166

River sample



Reference standard

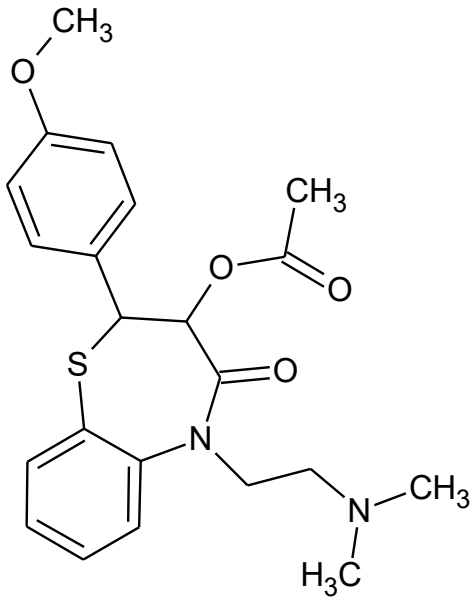


Diltiazem

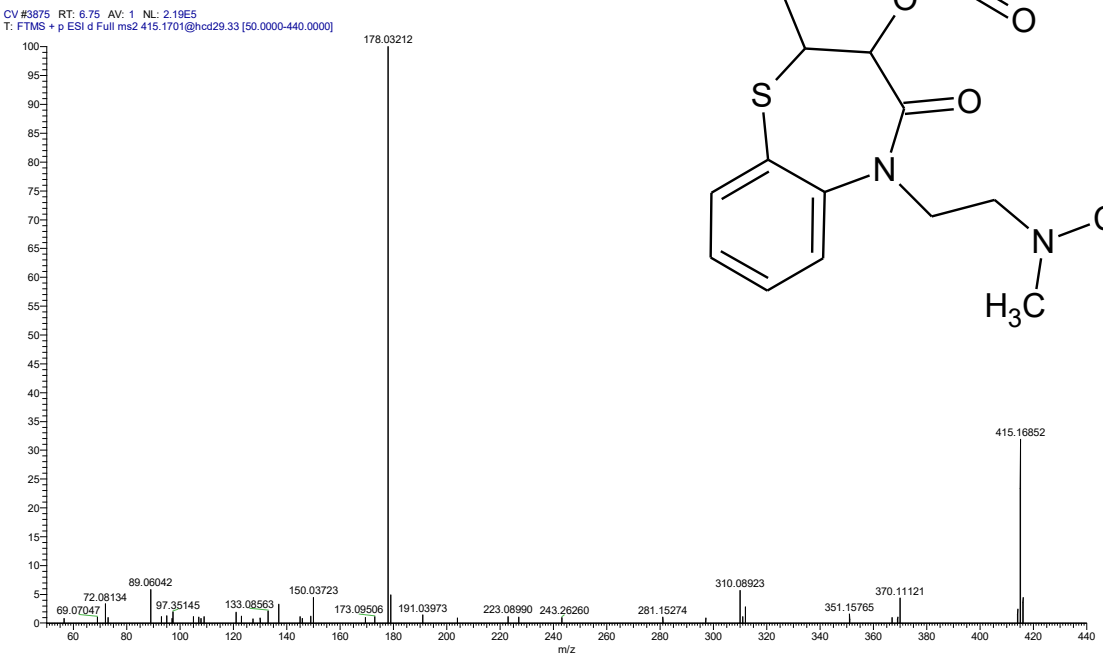
Cardiovascular system, pharmaceutical

Molecular formula: C22H26N2O4S

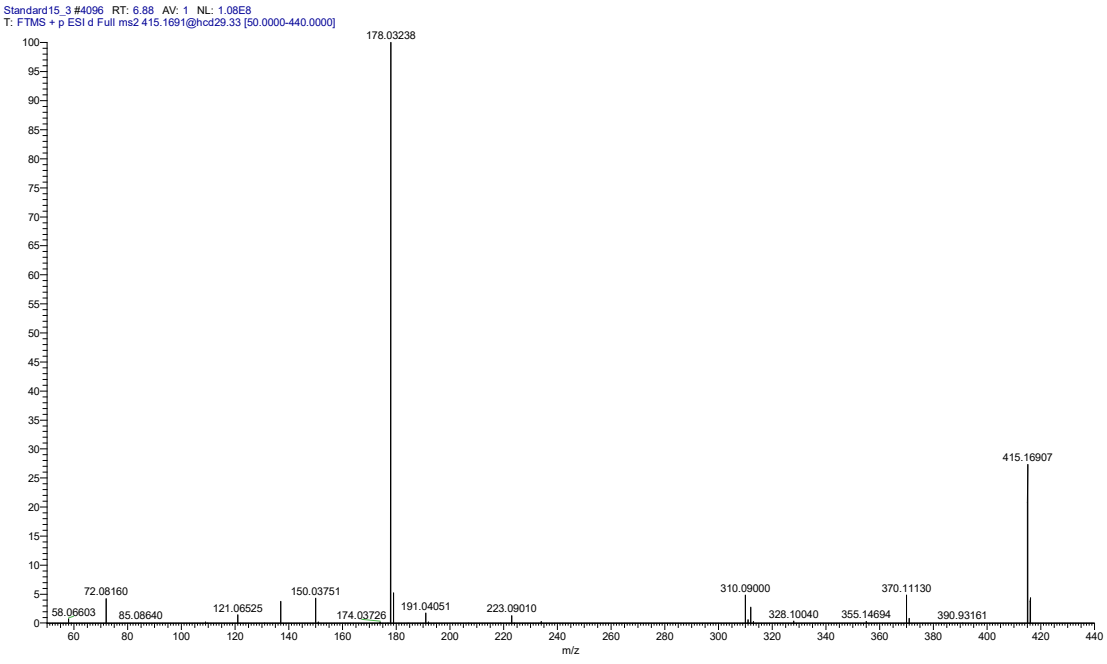
Monoisotopic mass: 414.16132849



River sample



Reference standard

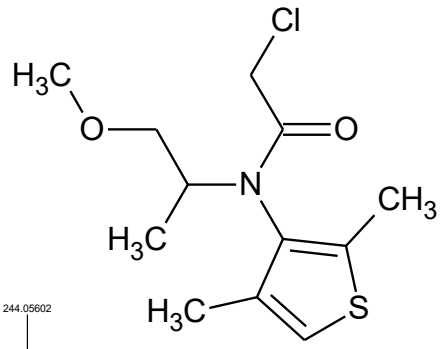


Dimethenamid

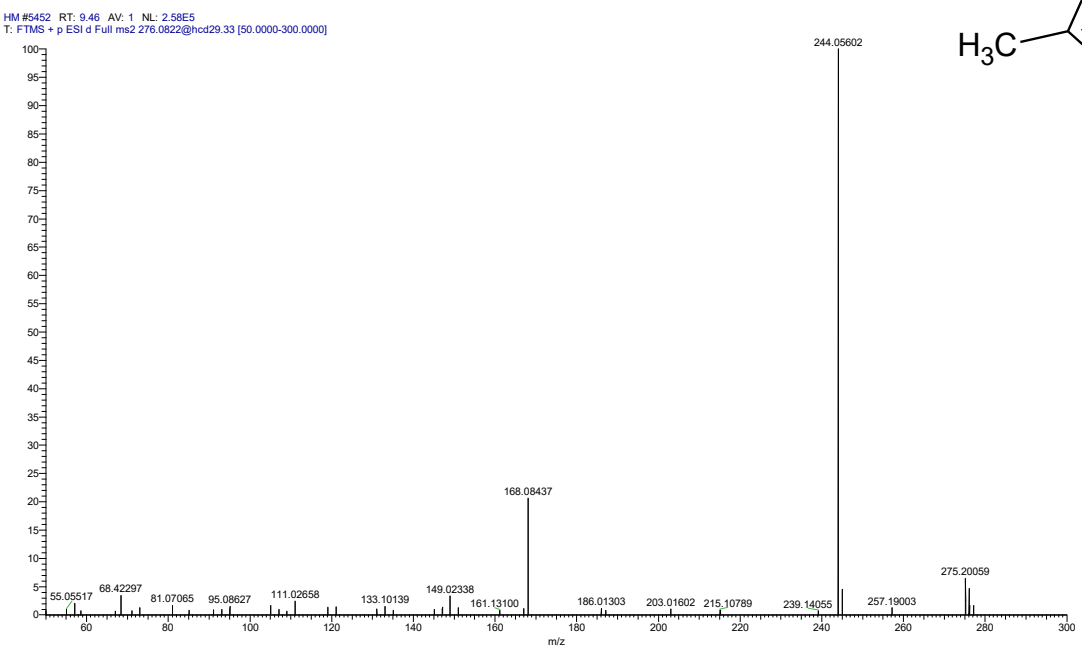
Herbicide, pesticide

Molecular formula: C12H18ClNO2S

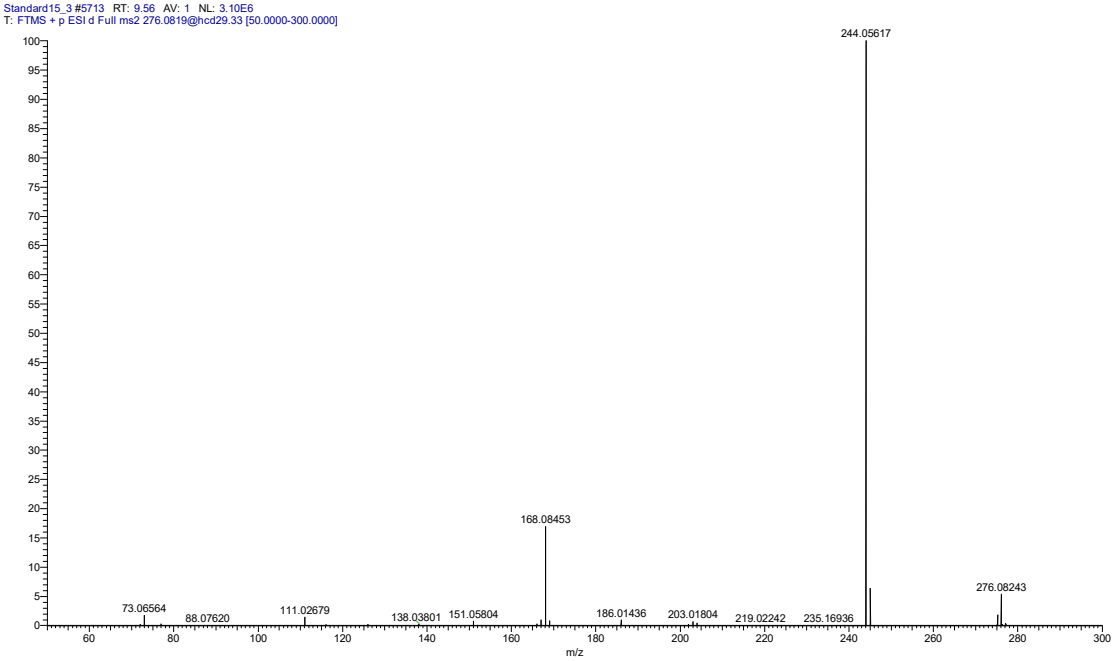
Monoisotopic mass: 275.0746777



River sample



Reference standard

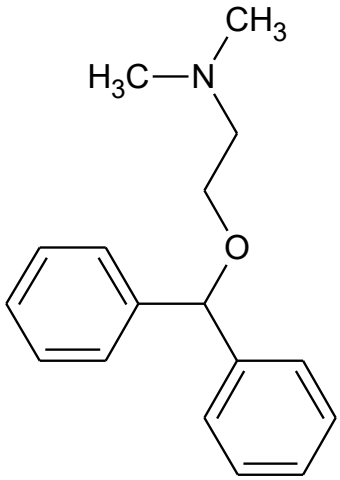


Diphenhydramine

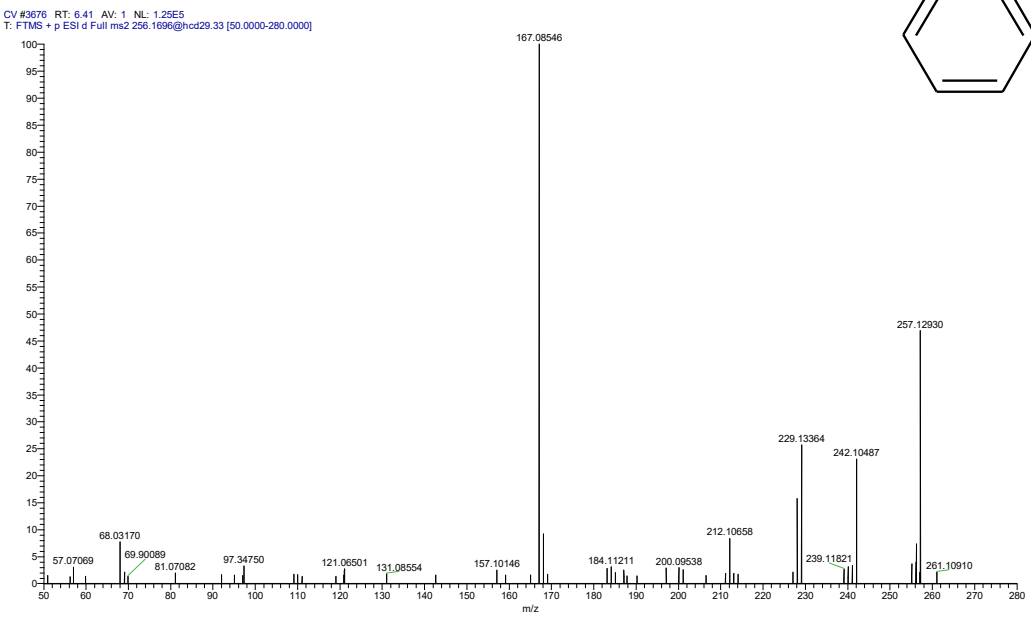
Respiratory system, pharmaceutical

Molecular formula: C17H21NO

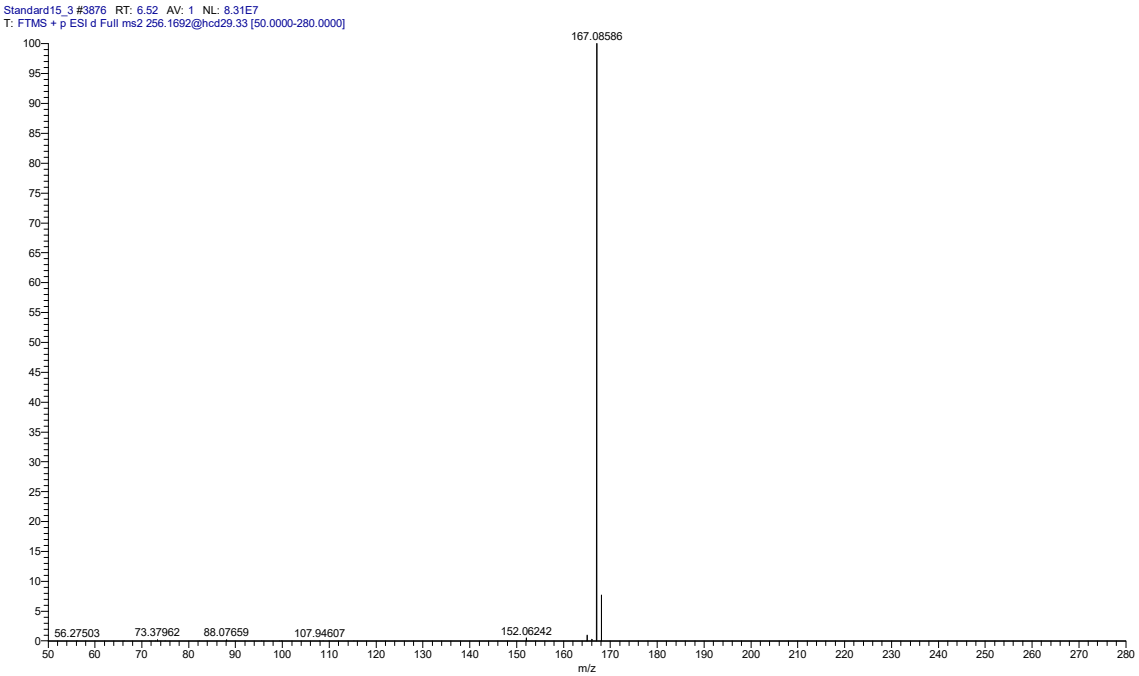
Monoisotopic mass: 255.162314293



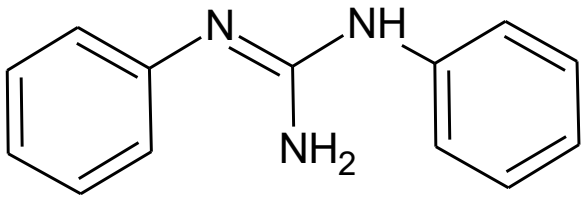
River sample



Reference standard



Diphenylguanidine

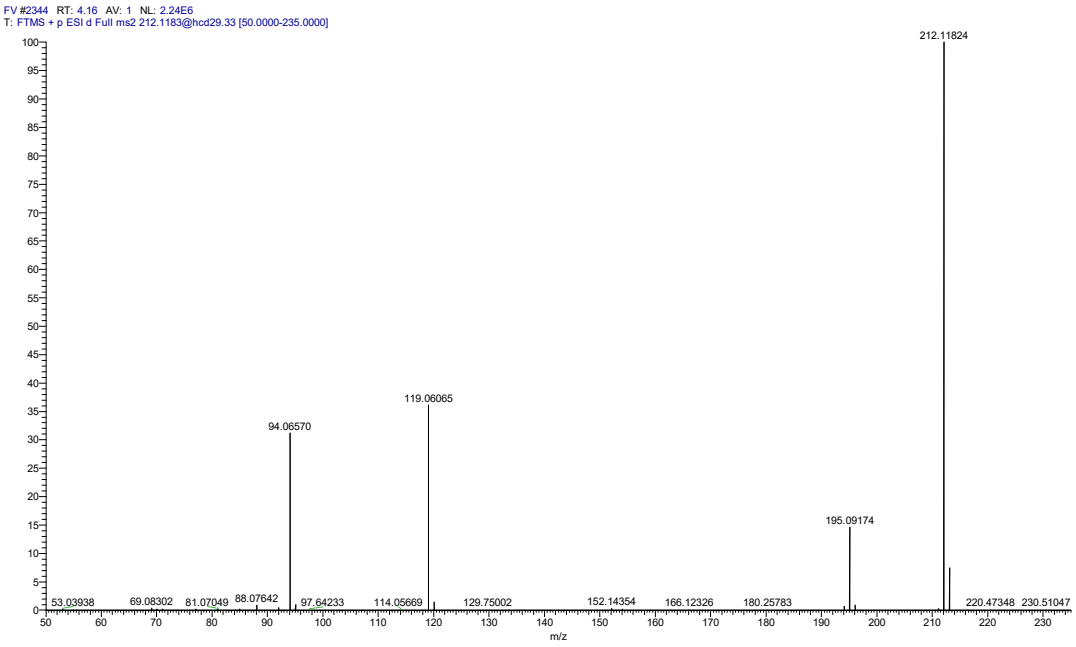


Polymer additive, consumer product additive

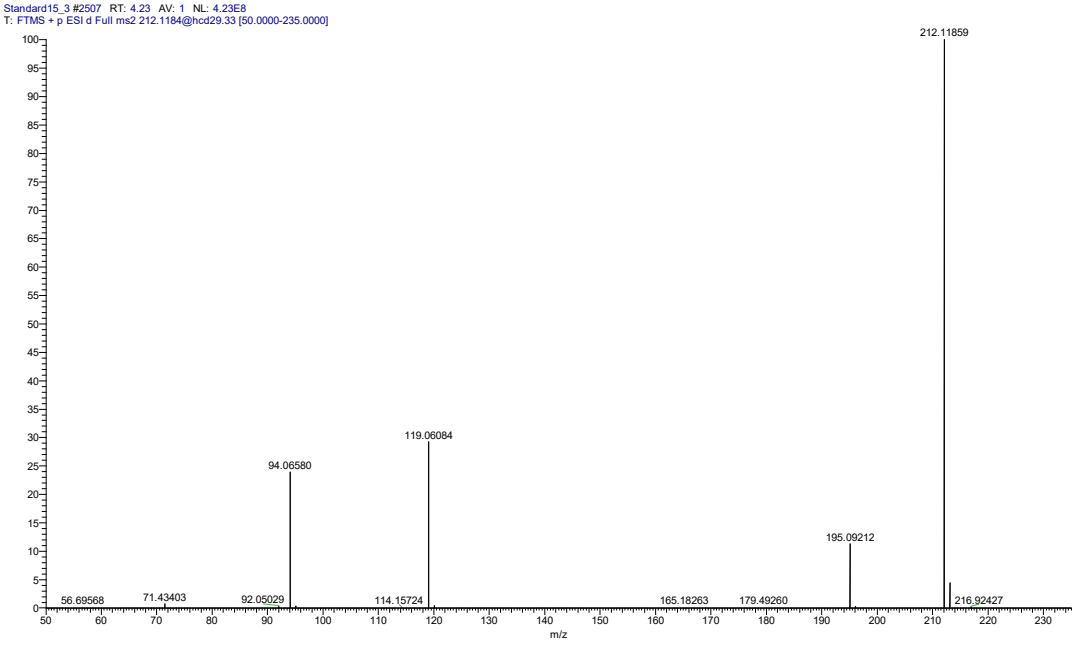
Molecular formula: C13H13N3

Monoisotopic mass: 211.110947427

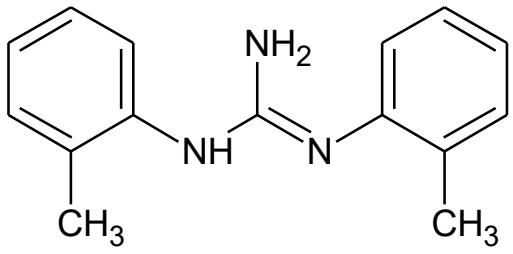
River sample



Reference standard



Ditolylguanidine

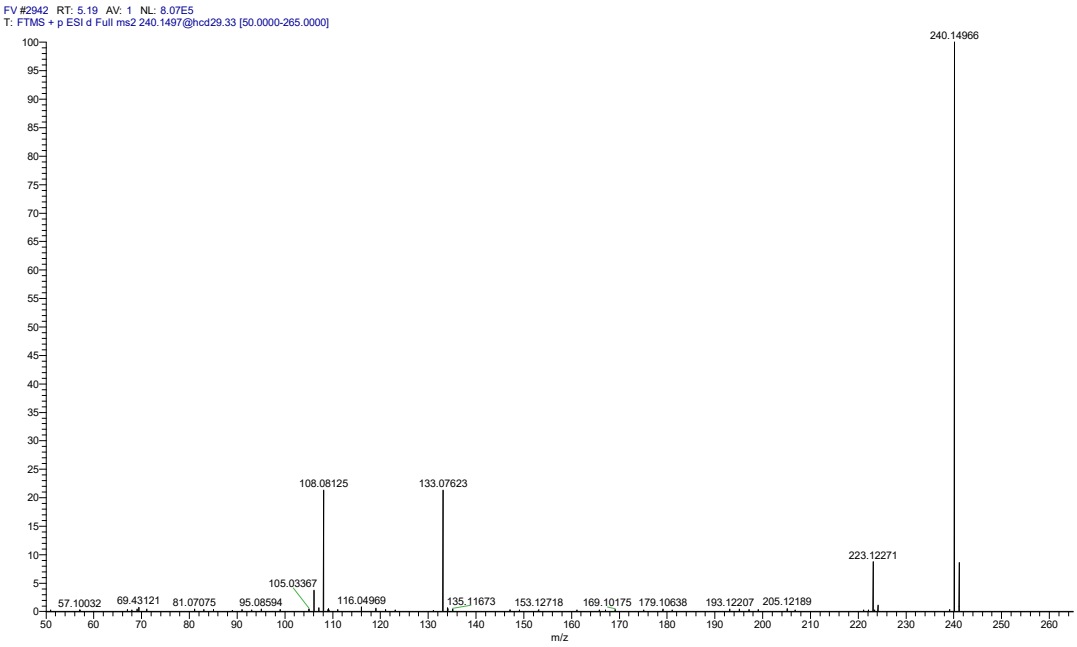


Polymer additive, consumer product additive

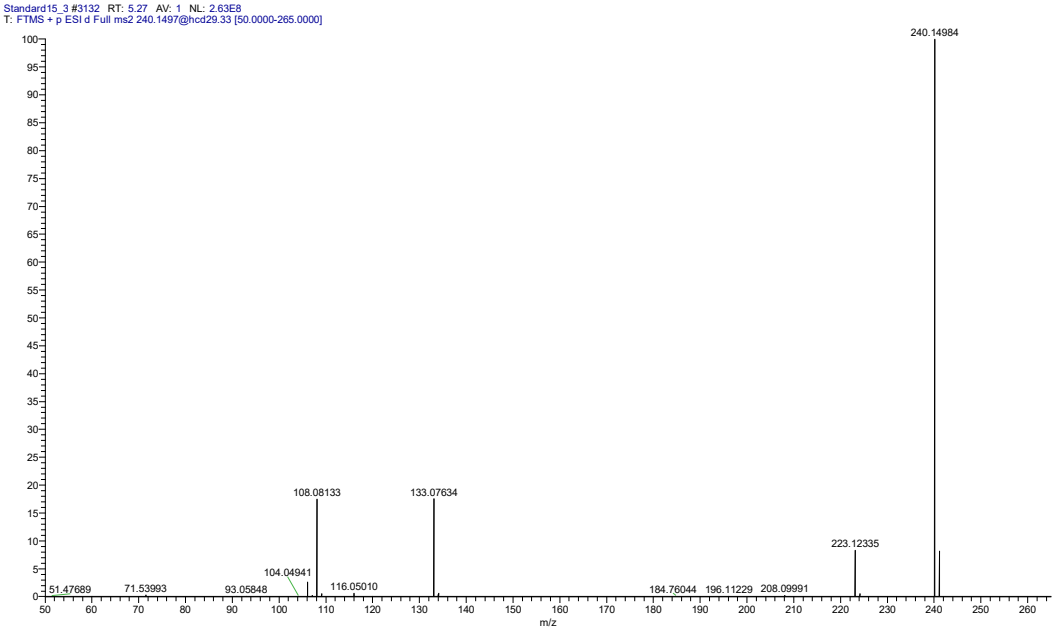
Molecular formula: C15H17N3

Monoisotopic mass: 239.142247555

River sample



Reference standard

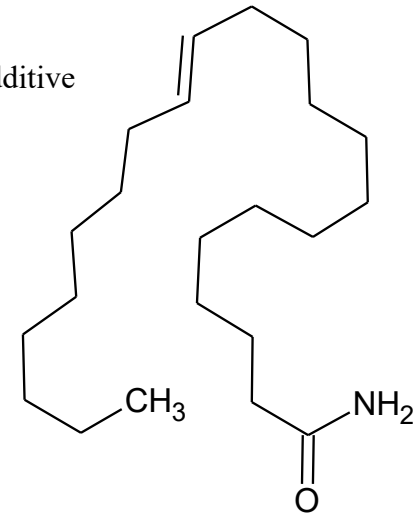


Erucamide

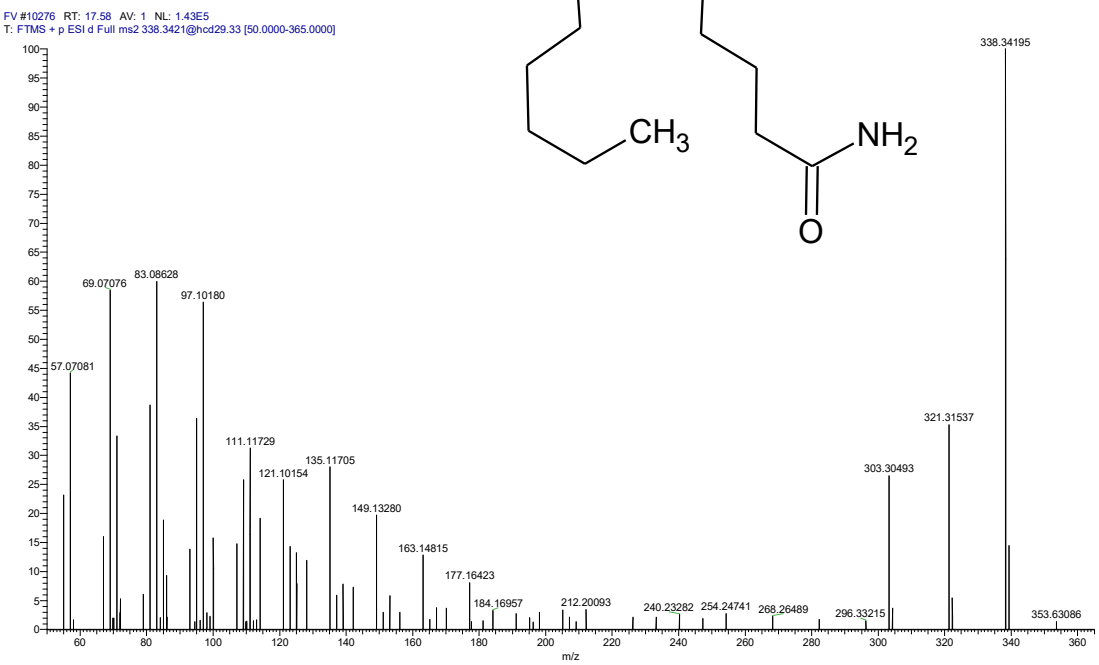
Polymer additive, consumer product additive

Molecular formula: C22H43NO

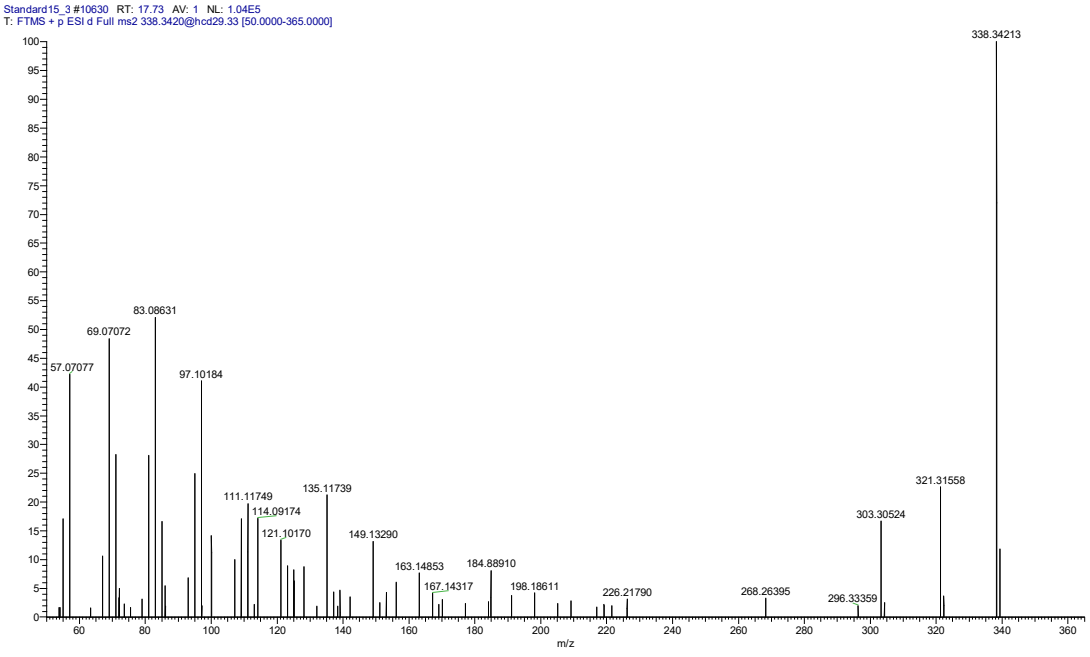
Monoisotopic mass: 337.334464995



River sample



Reference standard

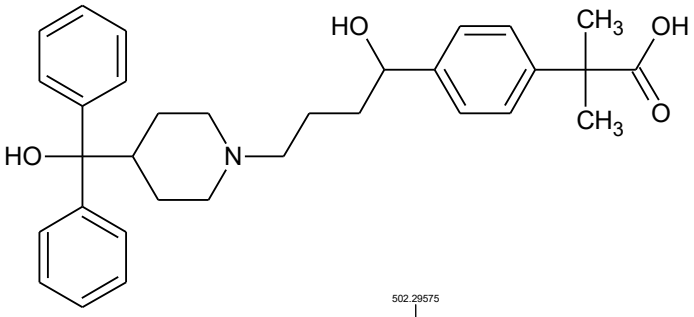


Fexofenadine

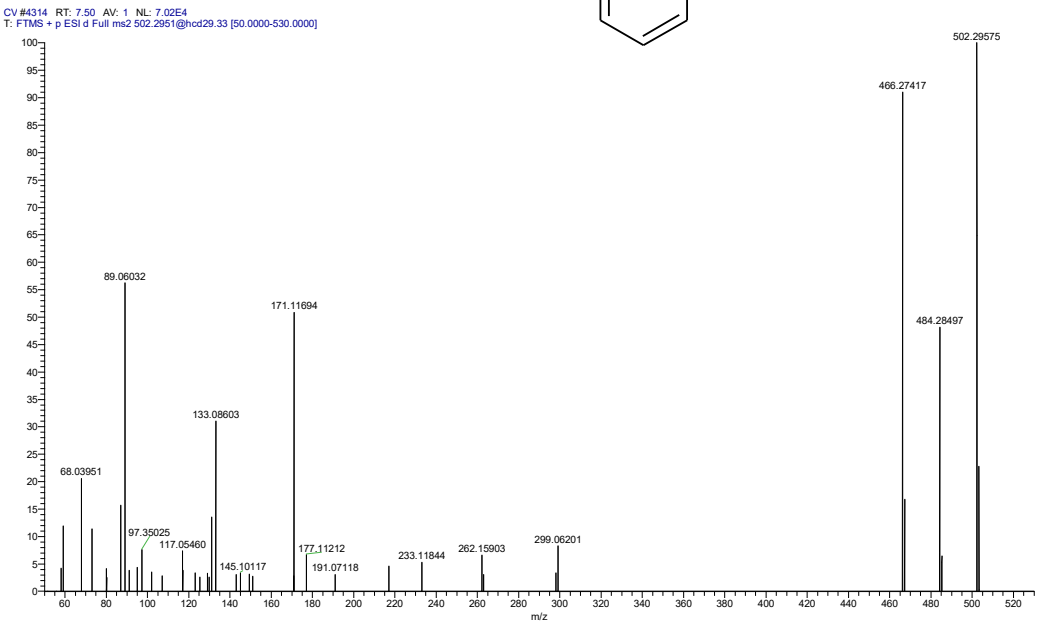
Respiratory system, pharmaceutical

Molecular formula: C32H39NO4

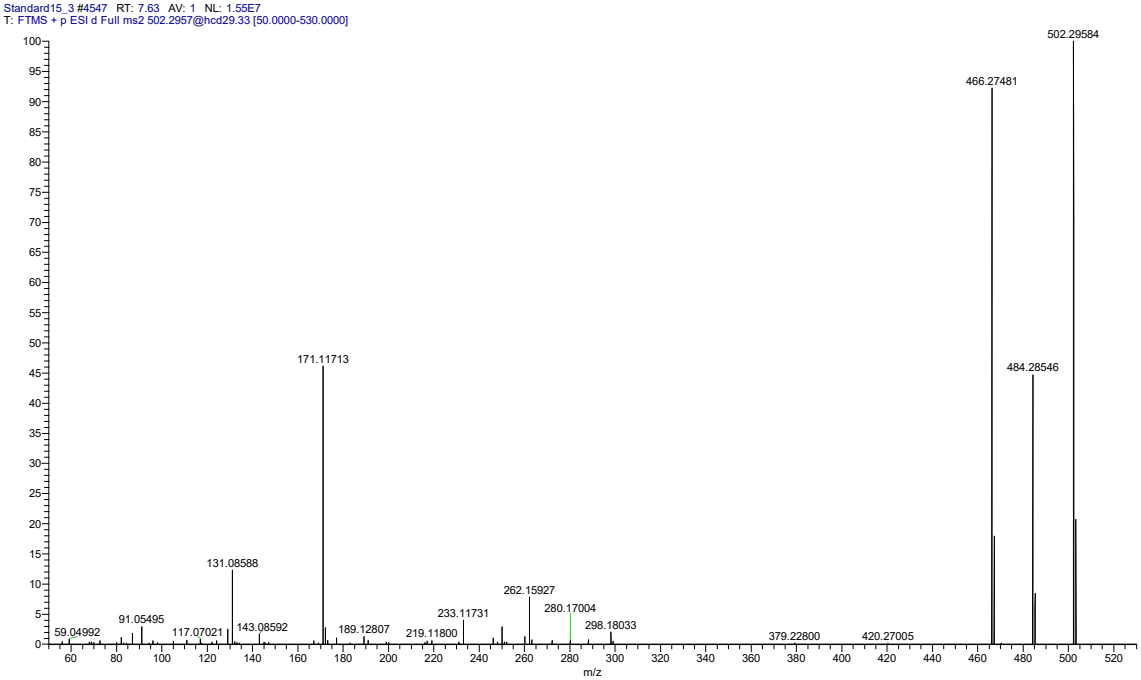
Monoisotopic mass: 501.28790873



River sample



Reference standard

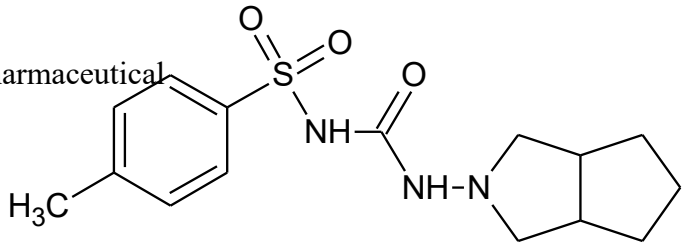


Gliclazide

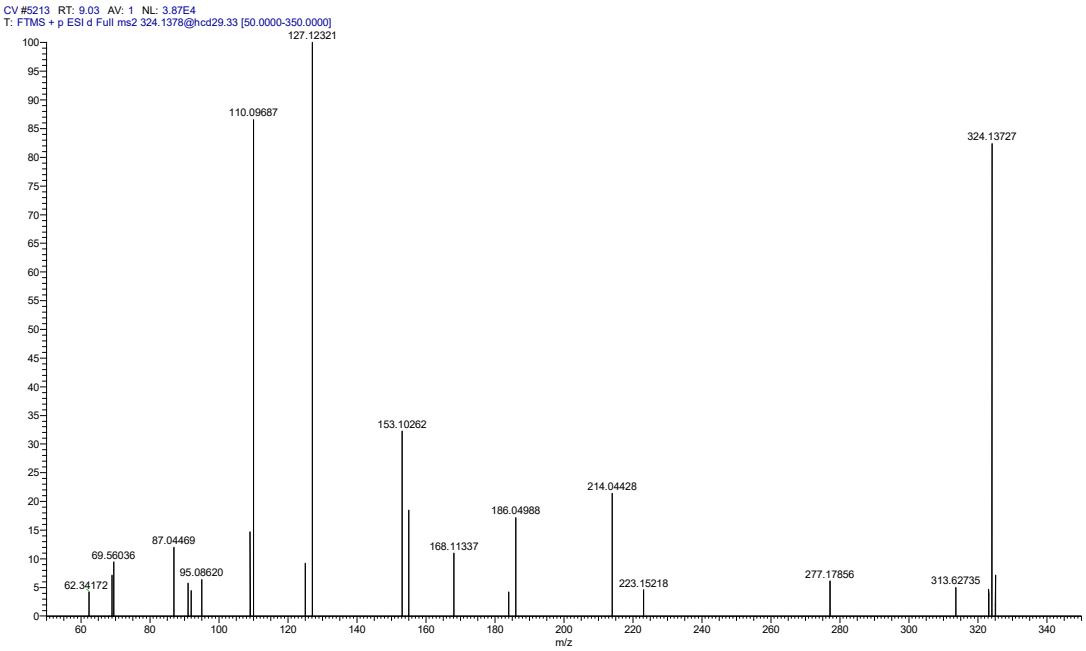
Alimentary tract and metabolism, pharmaceutical

Molecular formula: C15H21N3O3S

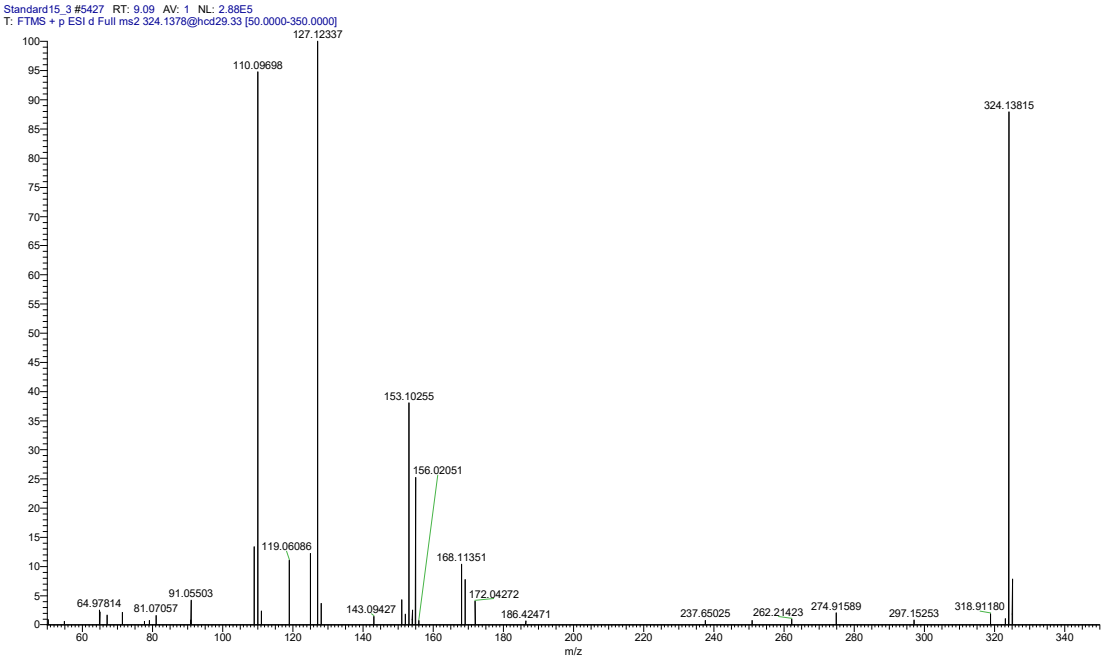
Monoisotopic mass: 323.13036271



River sample



Reference standard

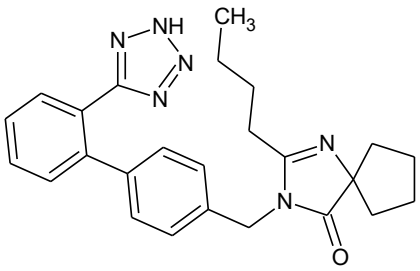


Irbesartan

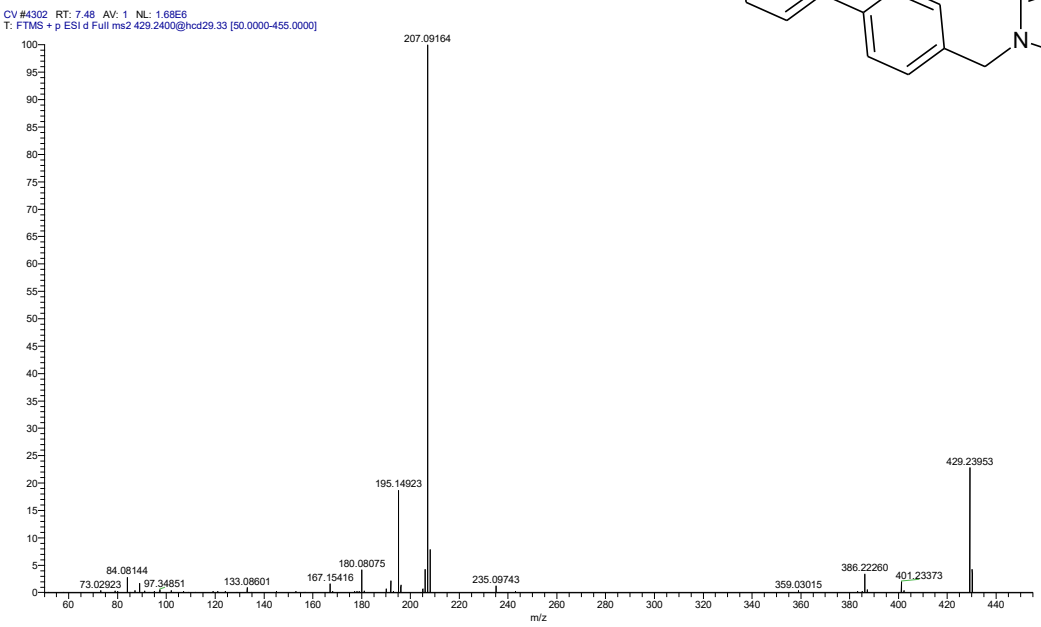
Cardiovascular system, pharmaceutical

Molecular formula: C25H28N6O

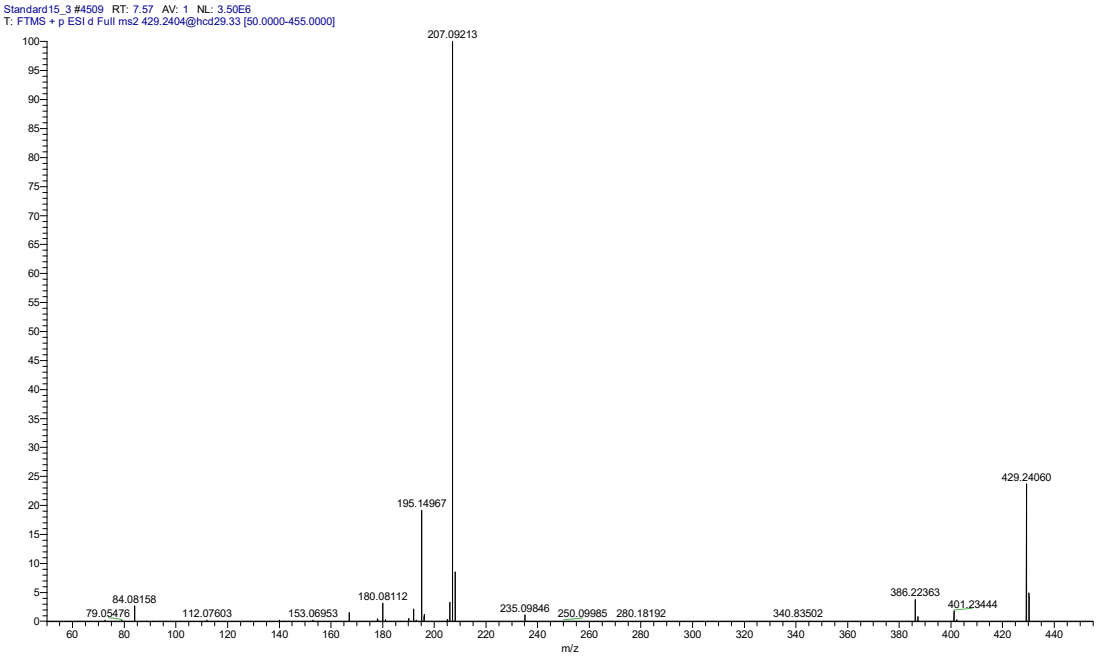
Monoisotopic mass: 428.23245954



River sample



Reference standard

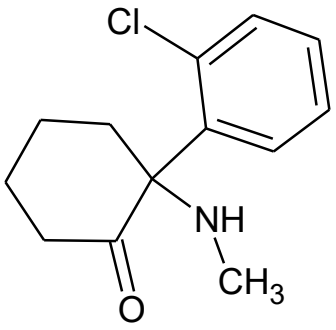


Ketamine

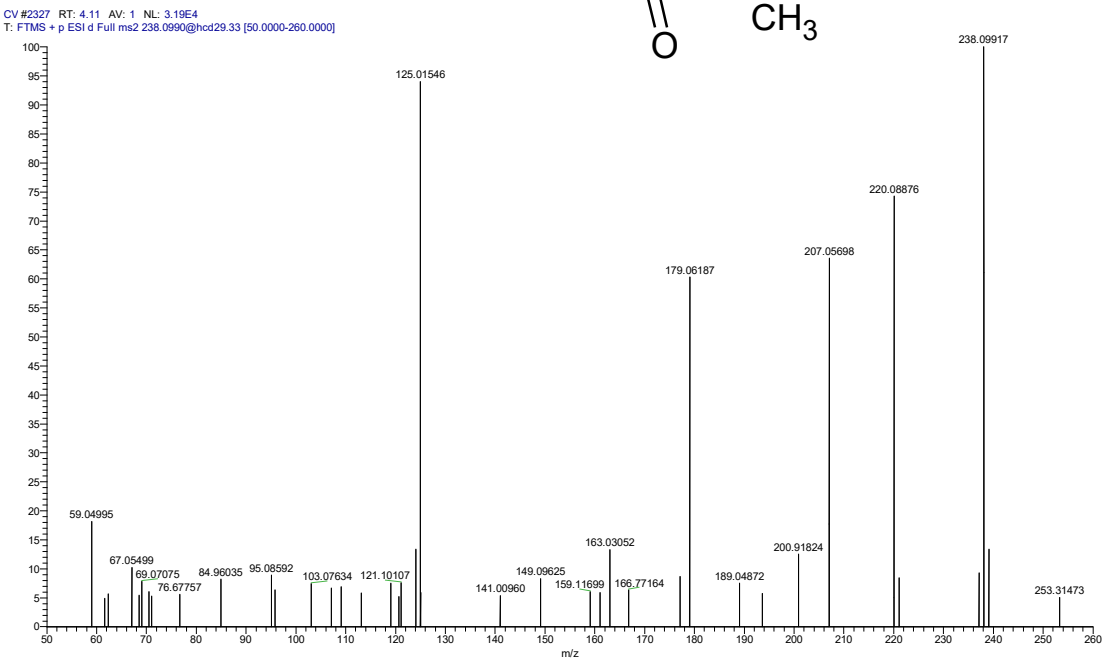
Nervous system, pharmaceutical

Molecular formula: C13H16ClNO

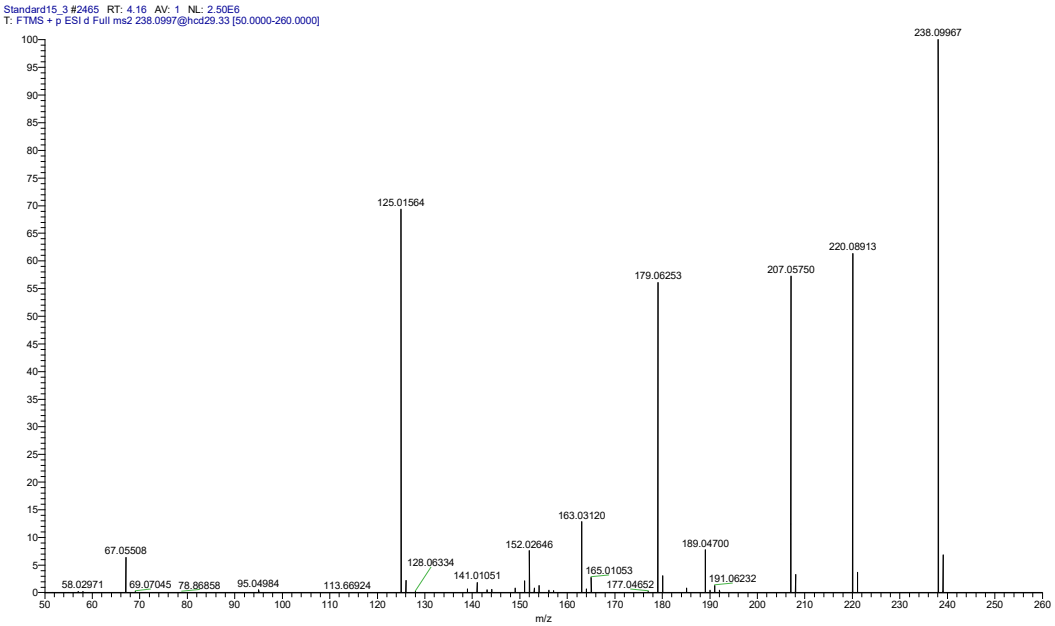
Monoisotopic mass: 237.0920418



River sample



Reference standard

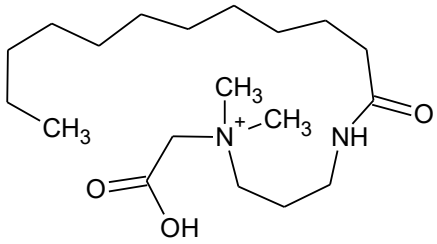


Lauramidopropyl betaine

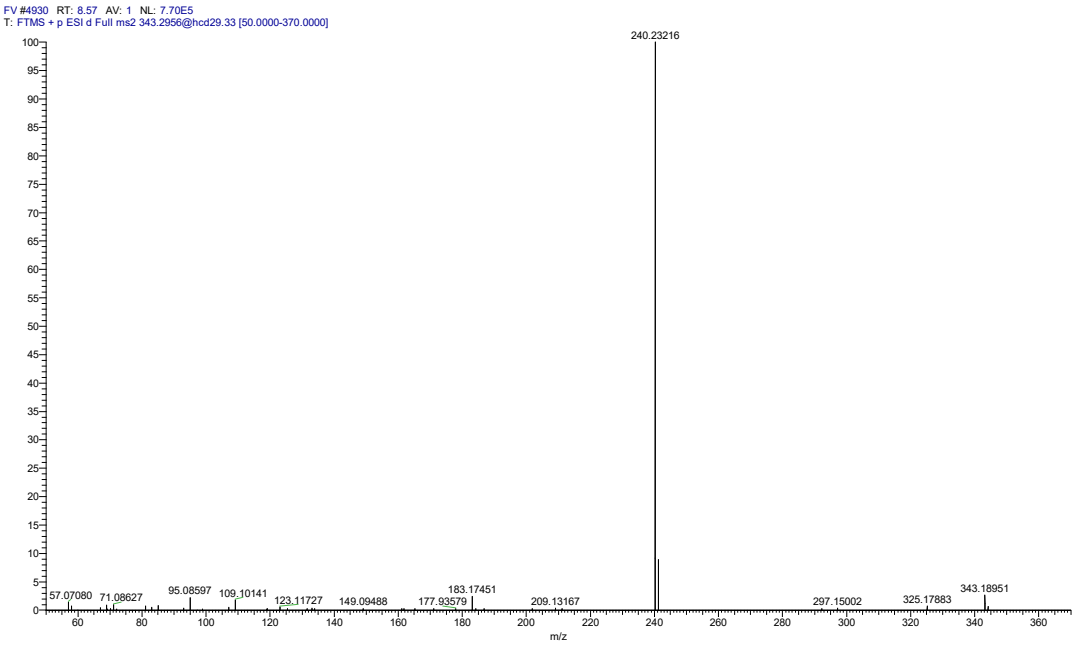
Cosmetic, consumer product additive

Molecular formula: C19H39N2O3+

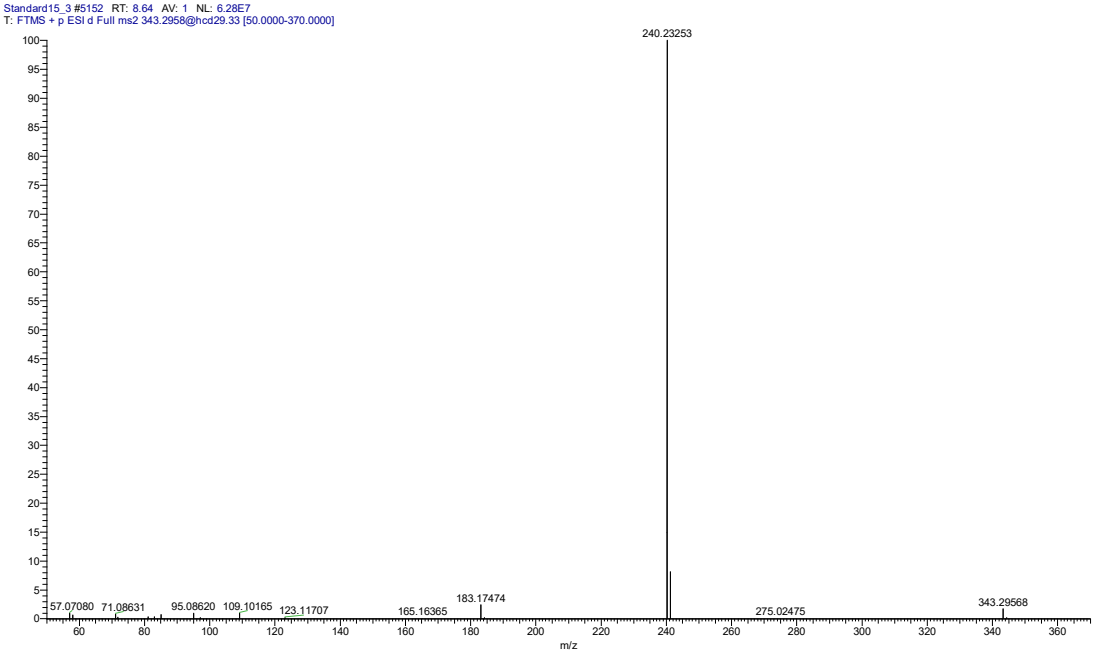
Monoisotopic mass: 343.29606811



River sample



Reference standard

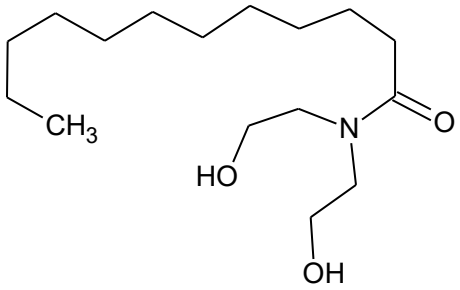


Lauryldiethanolamide

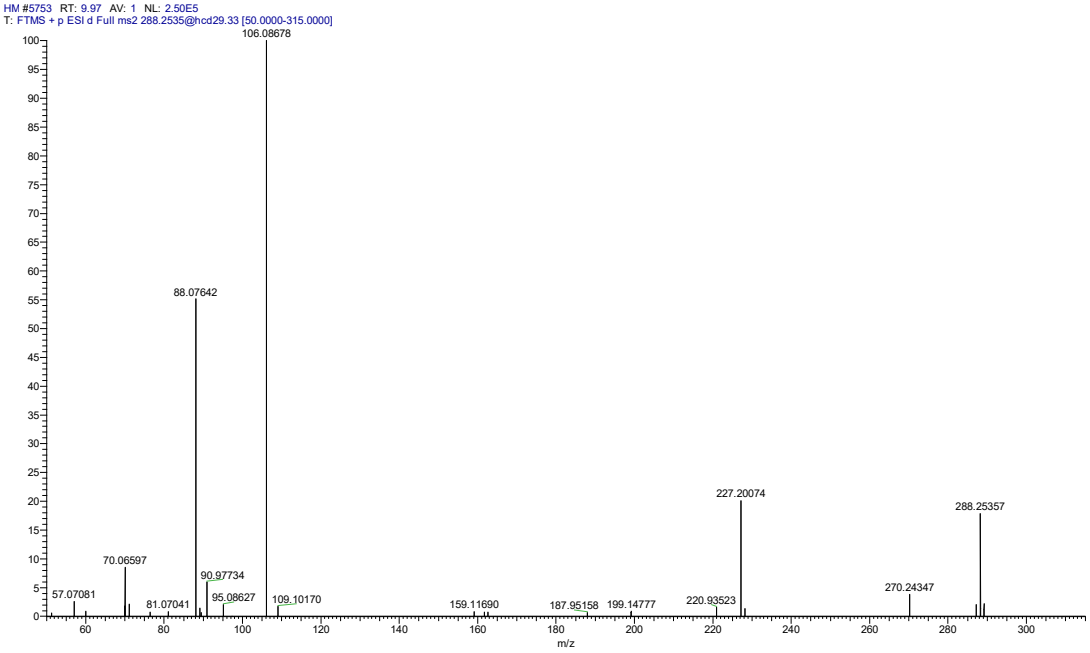
Cosmetic, consumer product additive

Molecular formula: C16H33NO3

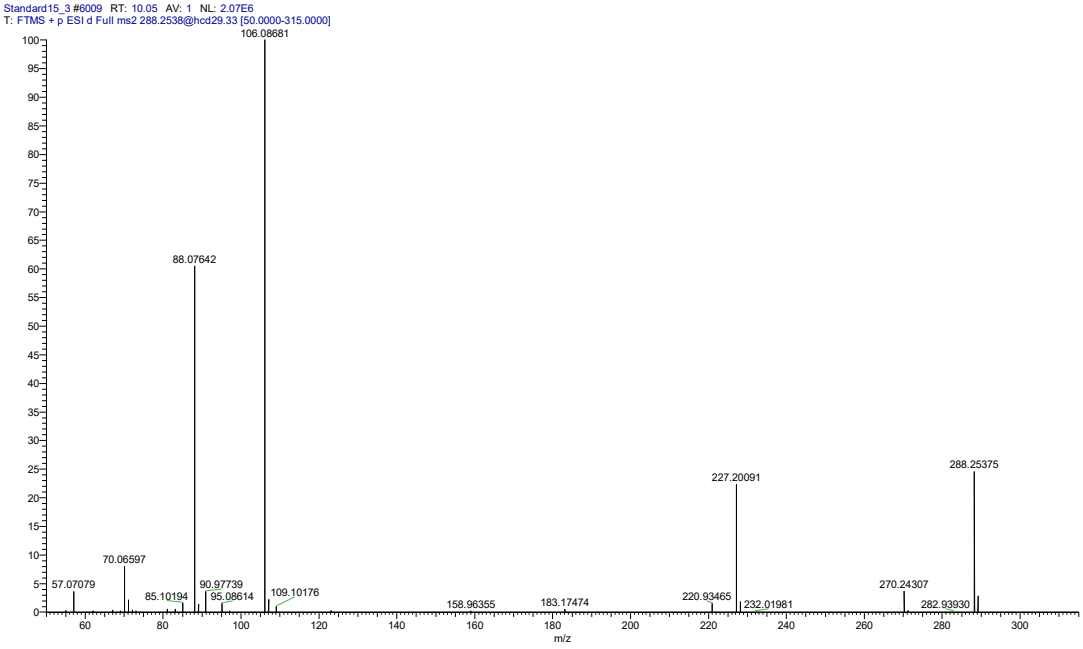
Monoisotopic mass: 287.24604391



River sample



Reference standard

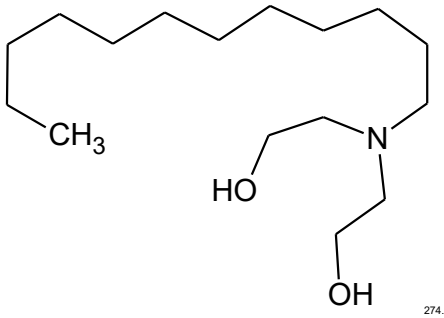


Lauryldiethanolamine

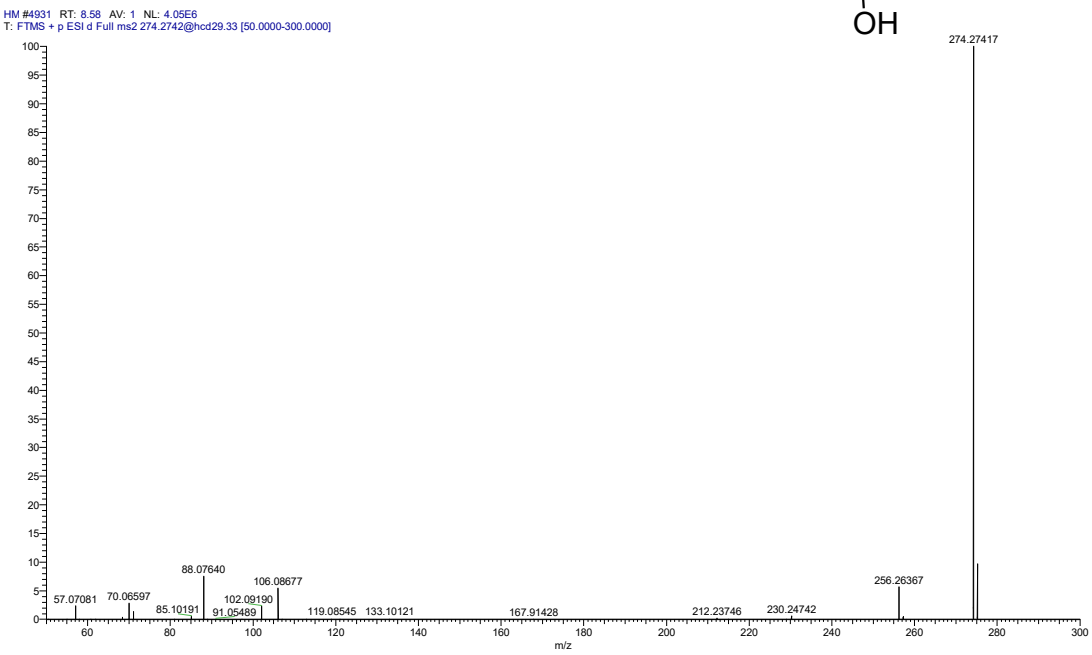
Cosmetic, consumer product additive

Molecular formula: C16H35NO2

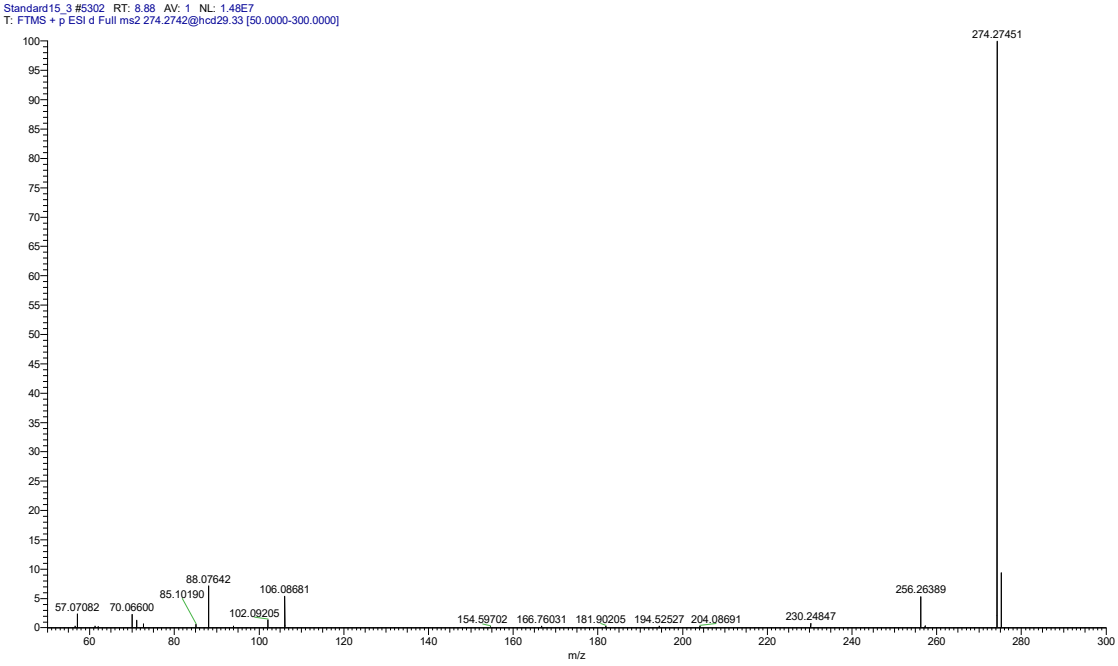
Monoisotopic mass: 273.266779359



River sample



Reference standard

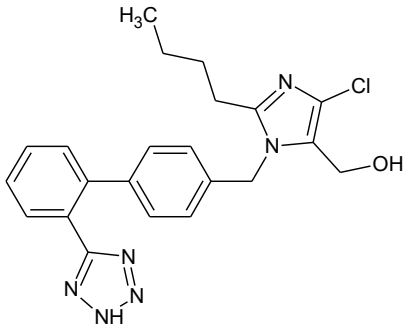


Losartan

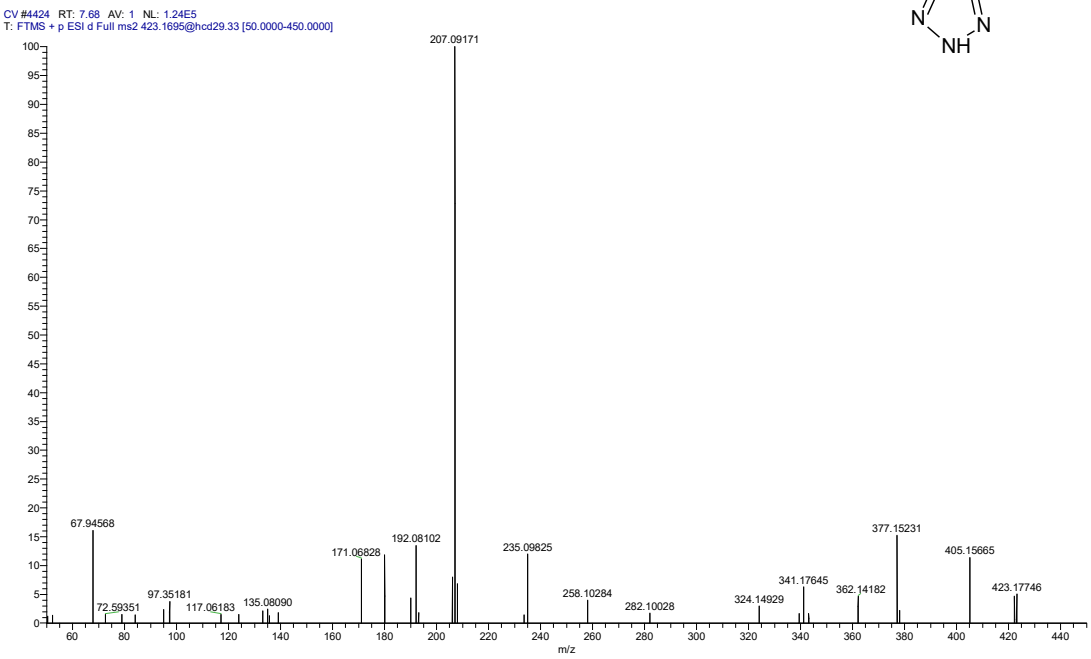
Cardiovascular system, pharmaceutical

Molecular formula: C22H23ClN6O

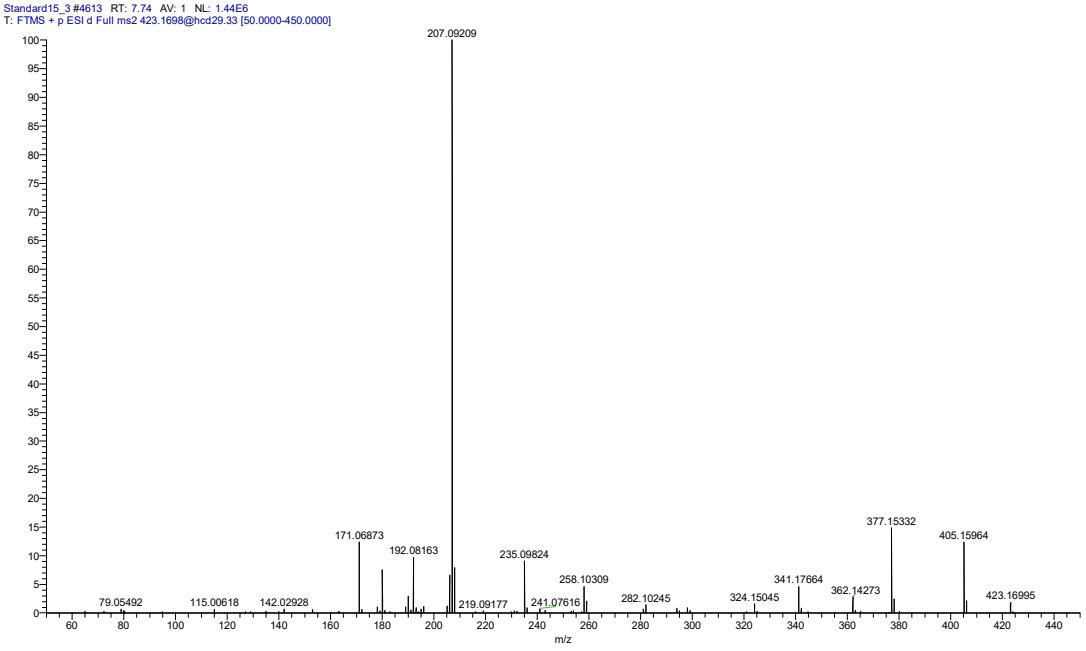
Monoisotopic mass: 422.1621871



River sample



Reference standard

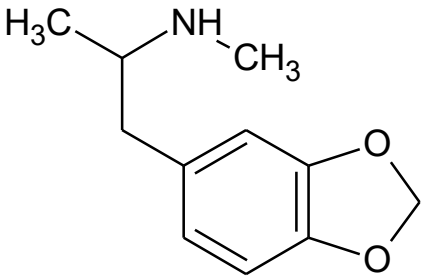


MDMA

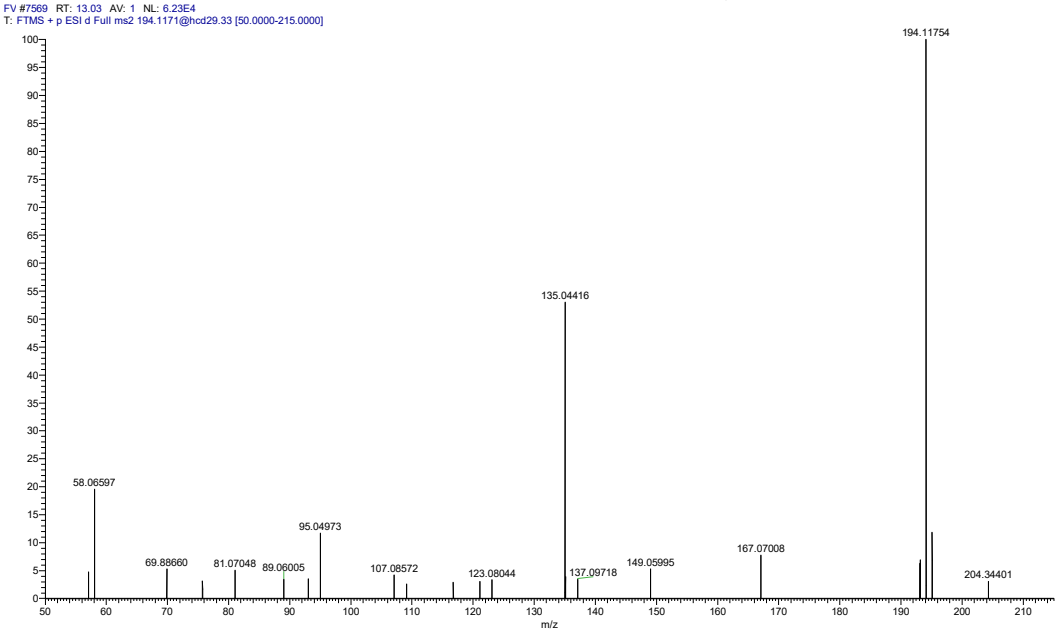
Amphetamine, illicit drug

Molecular formula: C11H15NO2

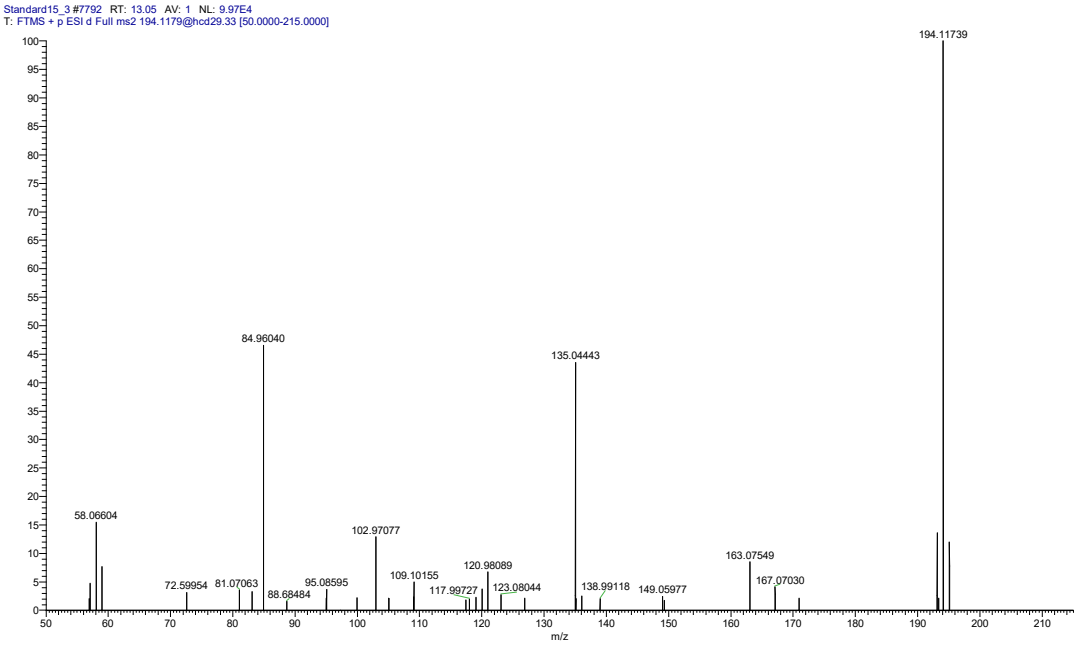
Monoisotopic mass: 193.110278721



River sample



Reference standard

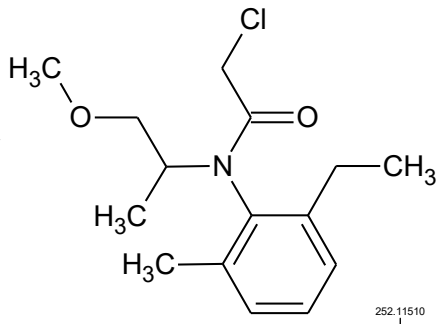


Metolachlor

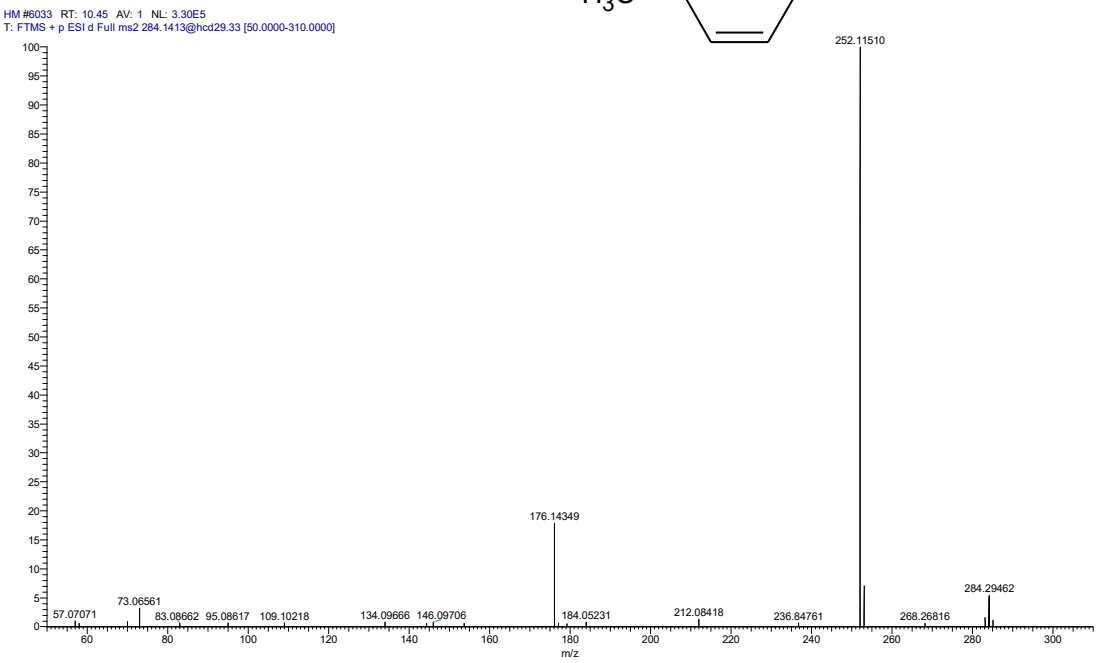
Herbicide, pesticide

Molecular formula: C15H22ClNO2

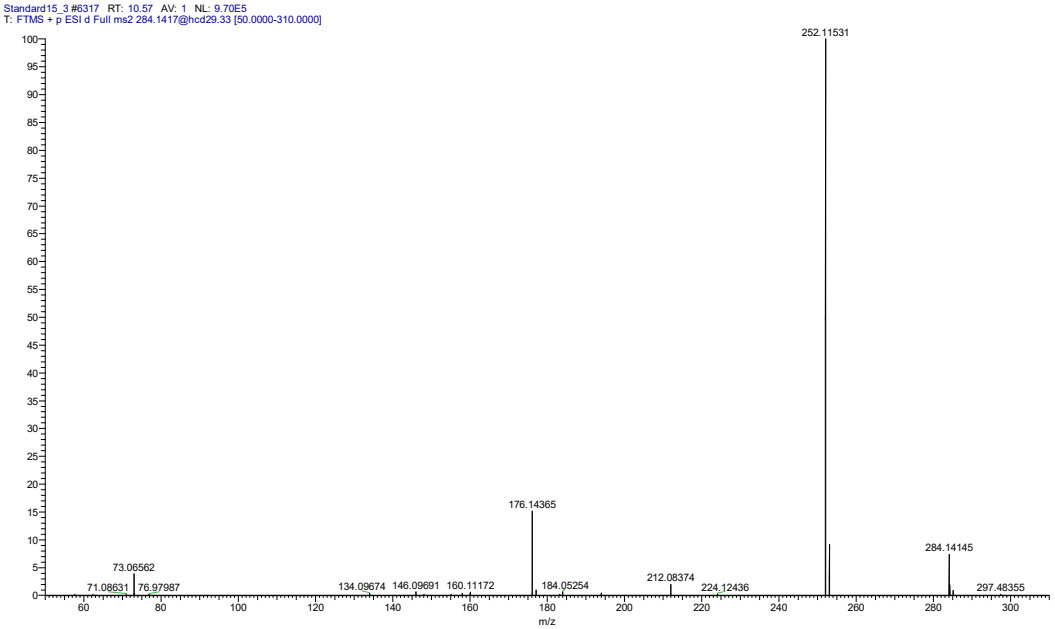
Monoisotopic mass: 283.1339066



River sample



Reference standard

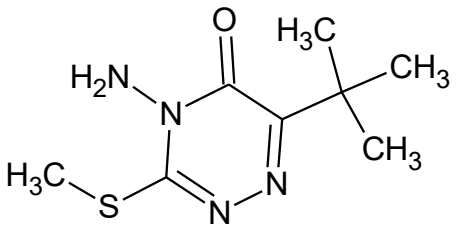


Metribuzin

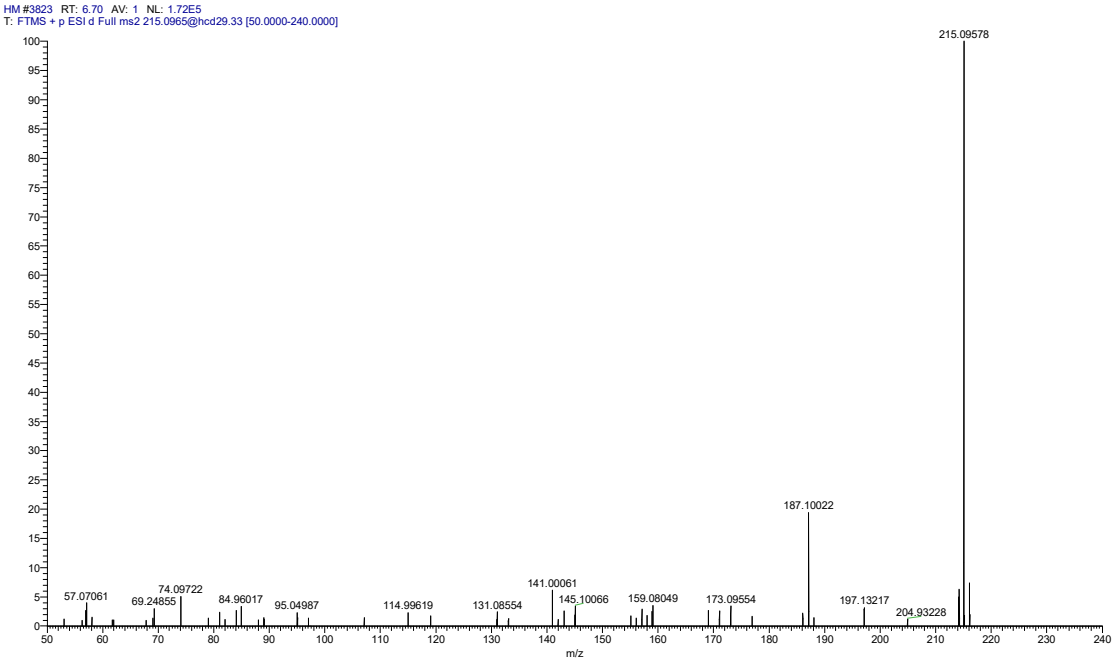
Herbicide, pesticide

Molecular formula: C8H14N4OS

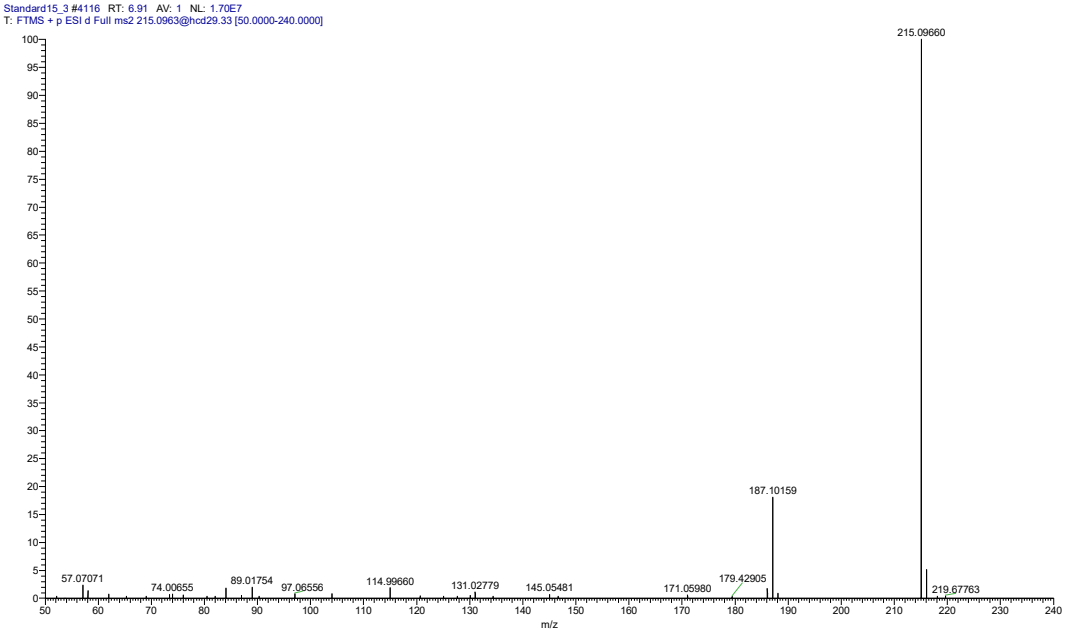
Monoisotopic mass: 214.08883226



River sample



Reference standard

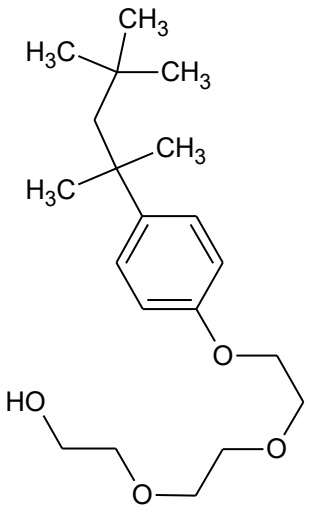


OPEO-03

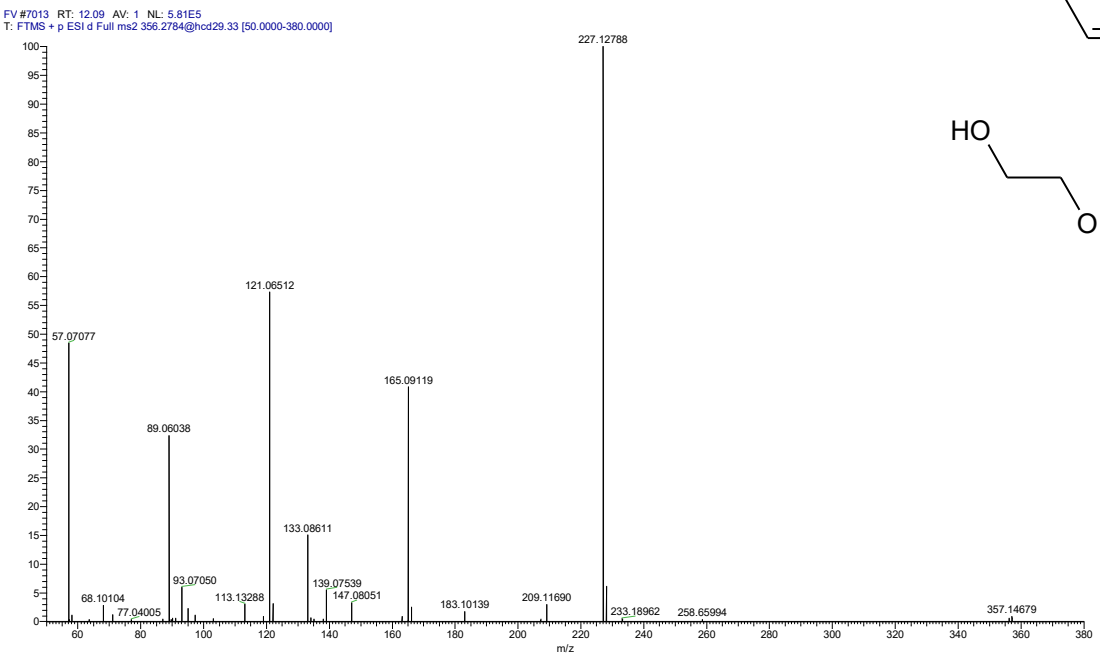
Surfactant, consumer product additive

Molecular formula: C20H34O4

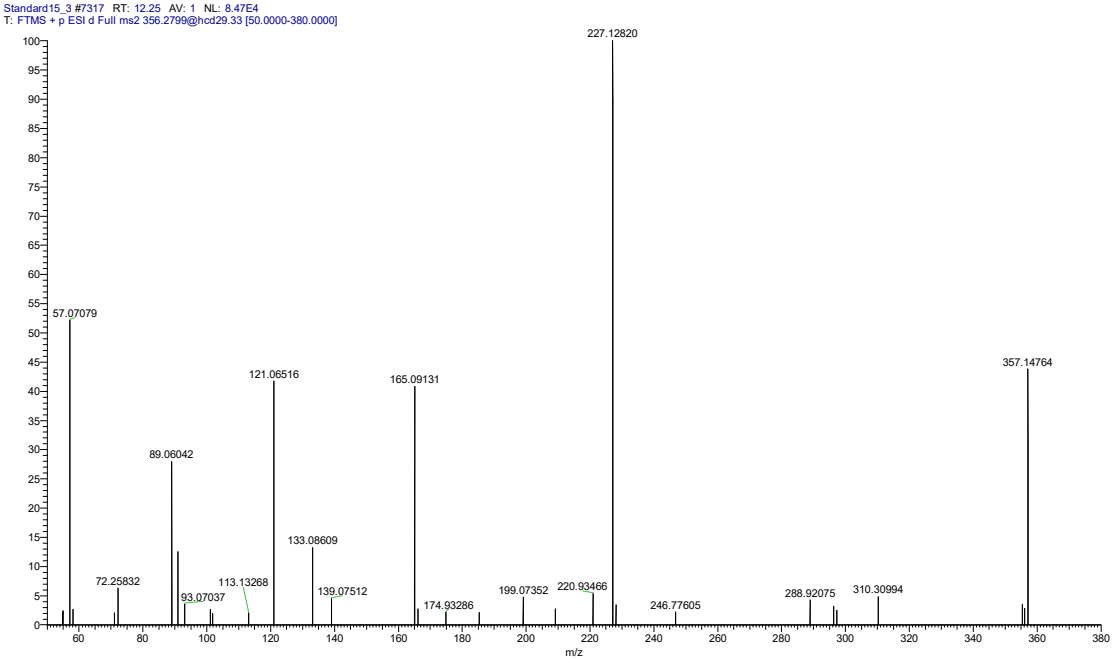
Monoisotopic mass: 338.24571



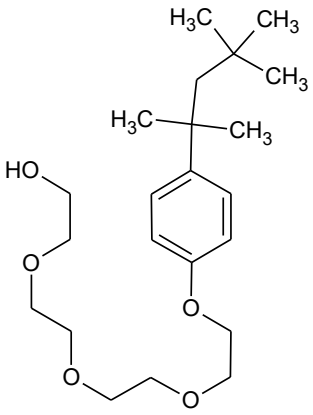
River sample



Reference standard



OPEO-04

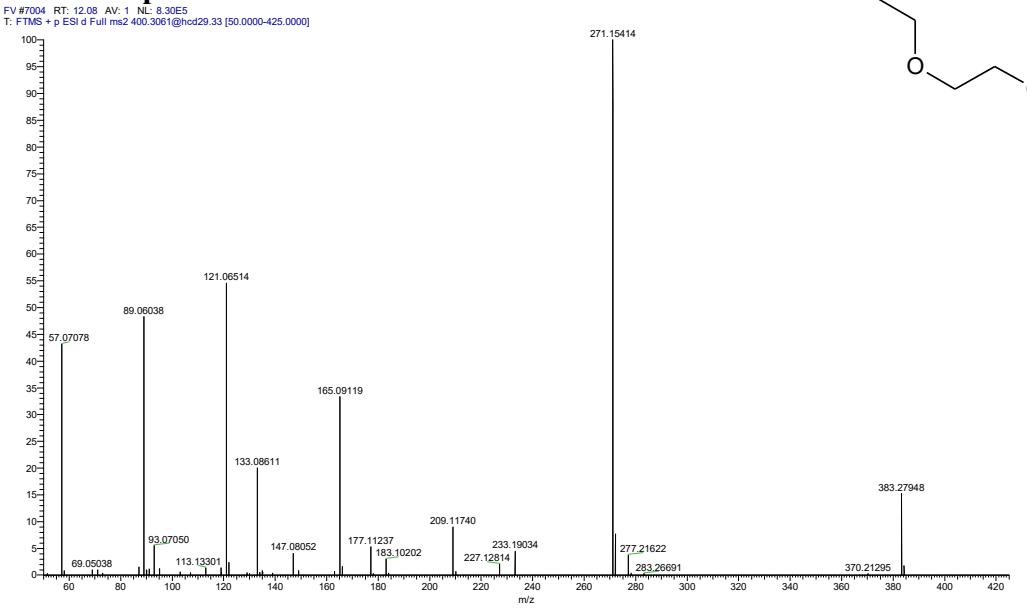


Surfactant, consumer product additive

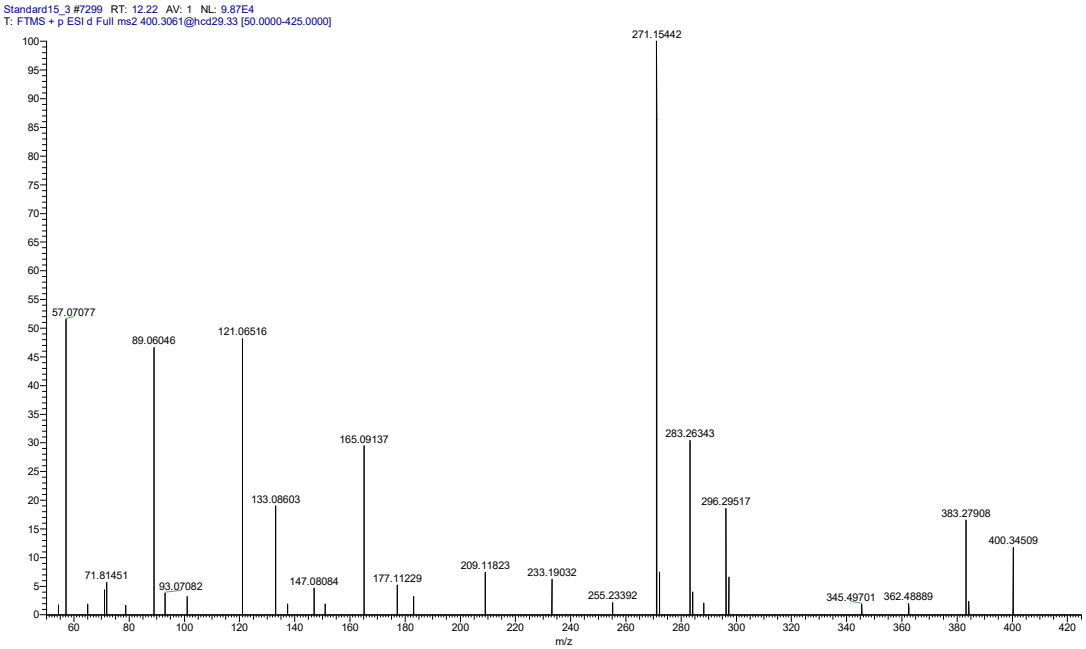
Molecular formula: C22H38O5

Monoisotopic mass: 382.270278

River sample



Reference standard

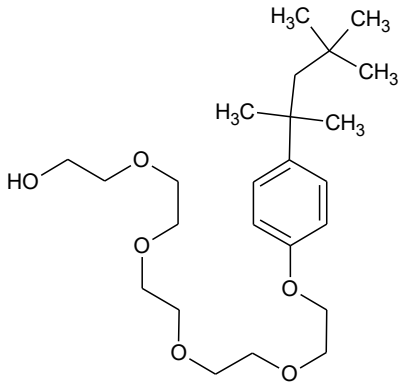


OPEO-05

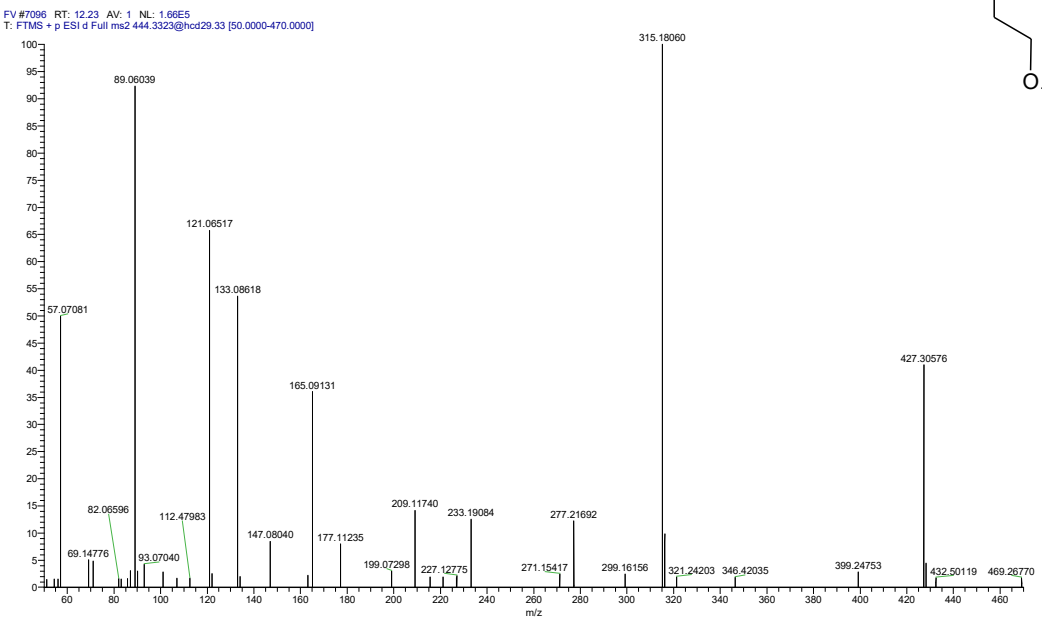
Surfactant, consumer product additive

Molecular formula: C24H42O6

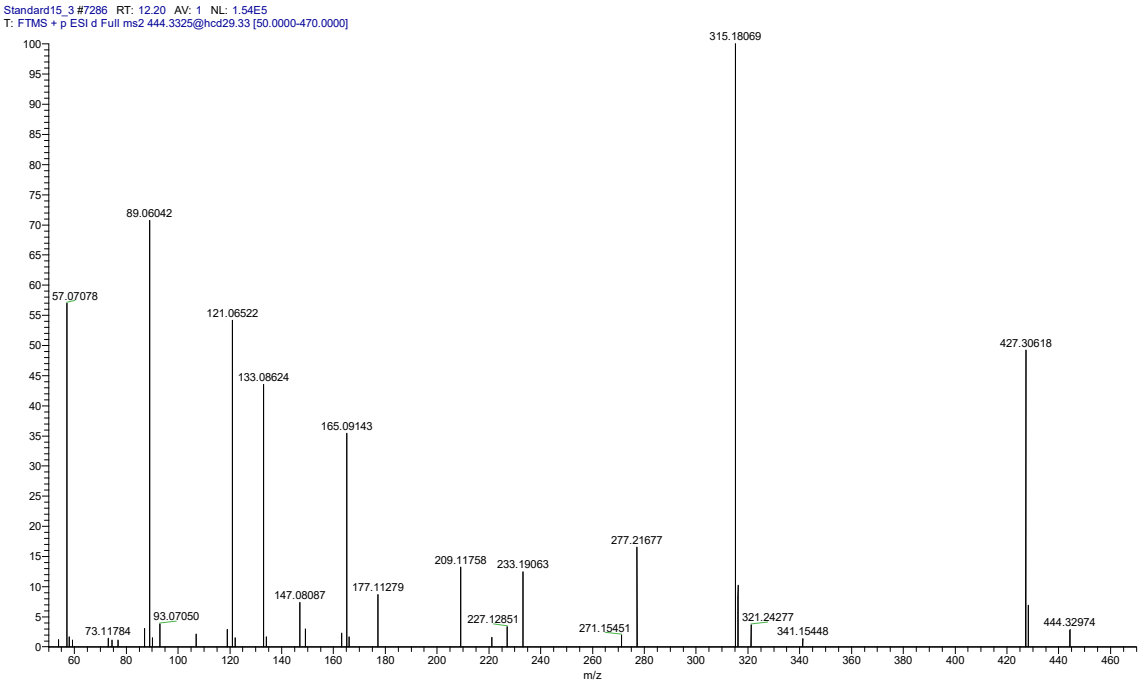
Monoisotopic mass: 426.295944



River sample



Reference standard

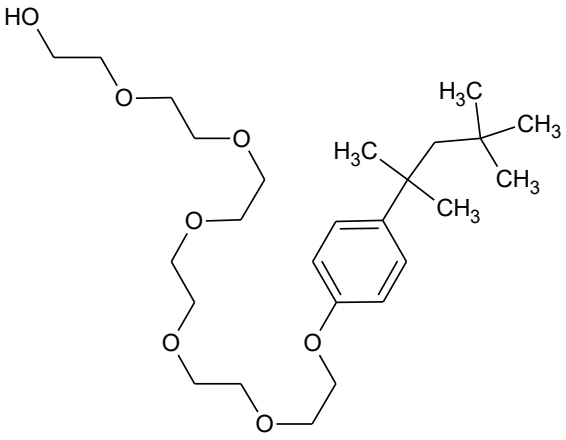


OPEO-06

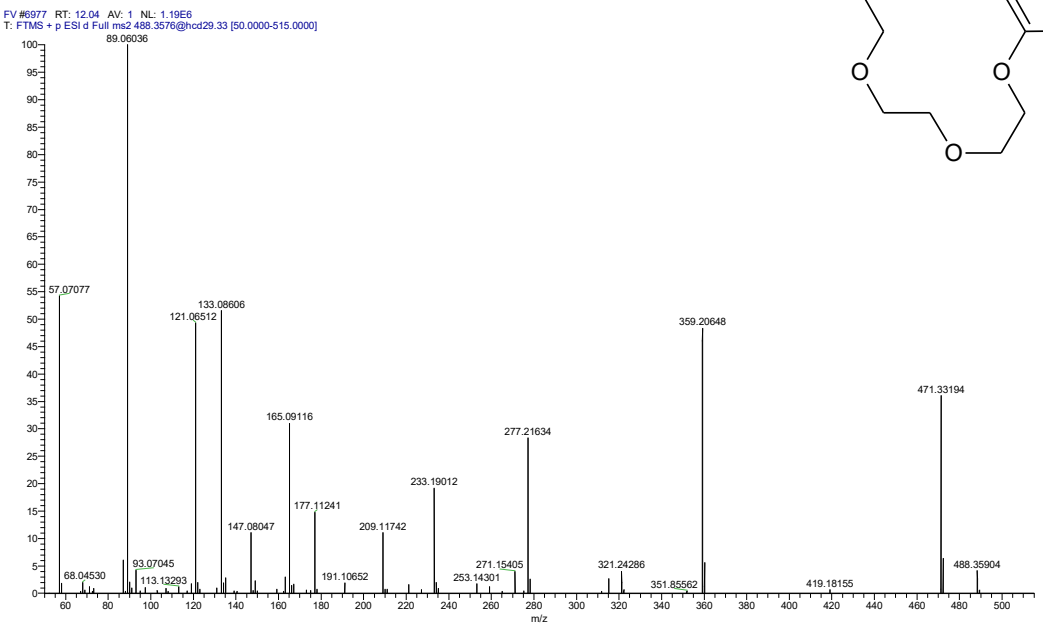
Surfactant, consumer product additive

Molecular formula: C26H46O7

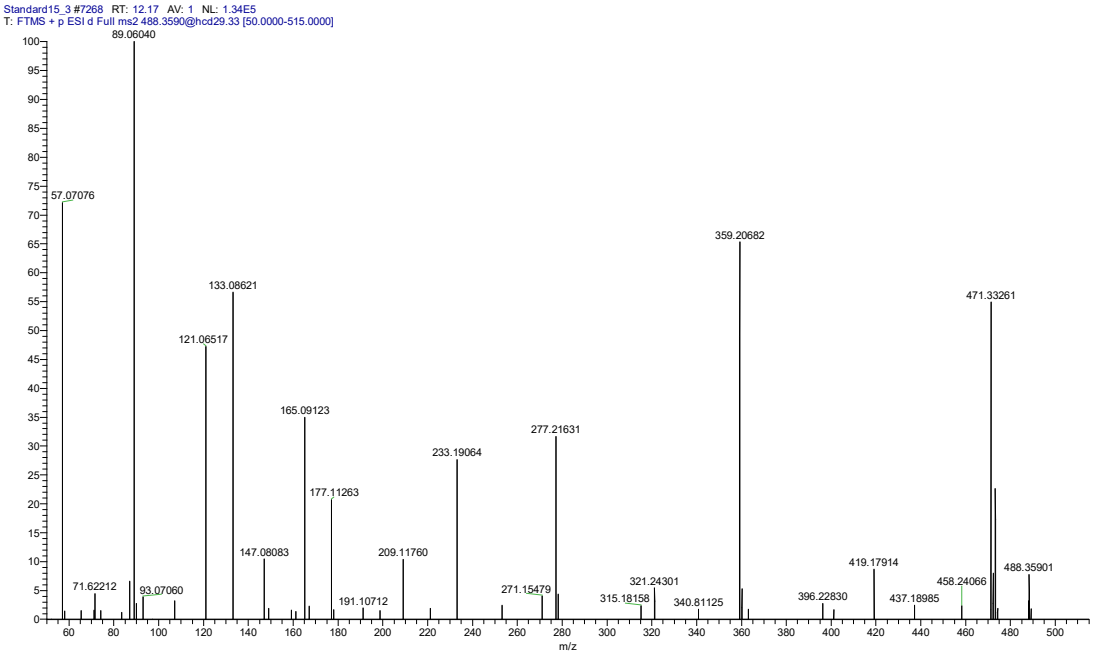
Monoisotopic mass: 470.3237614



River sample



Reference standard

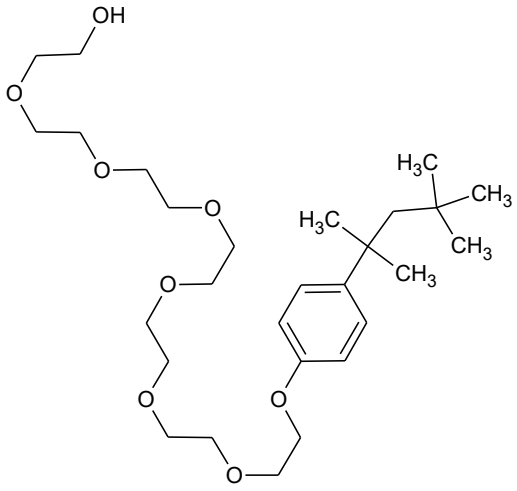


OPEO-07

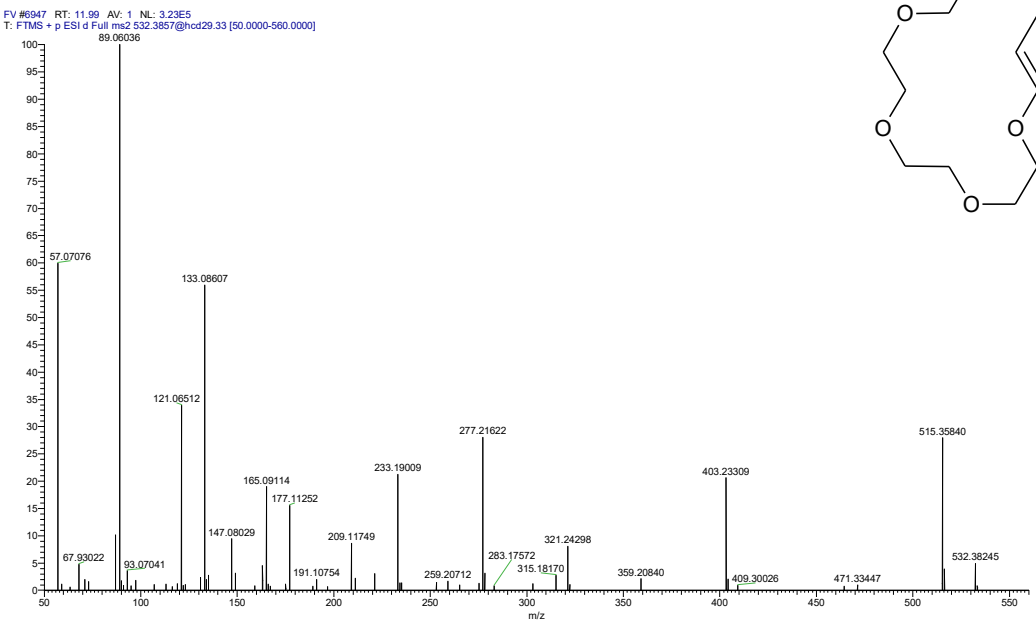
Surfactant, consumer product additive

Molecular formula: C28H50O8

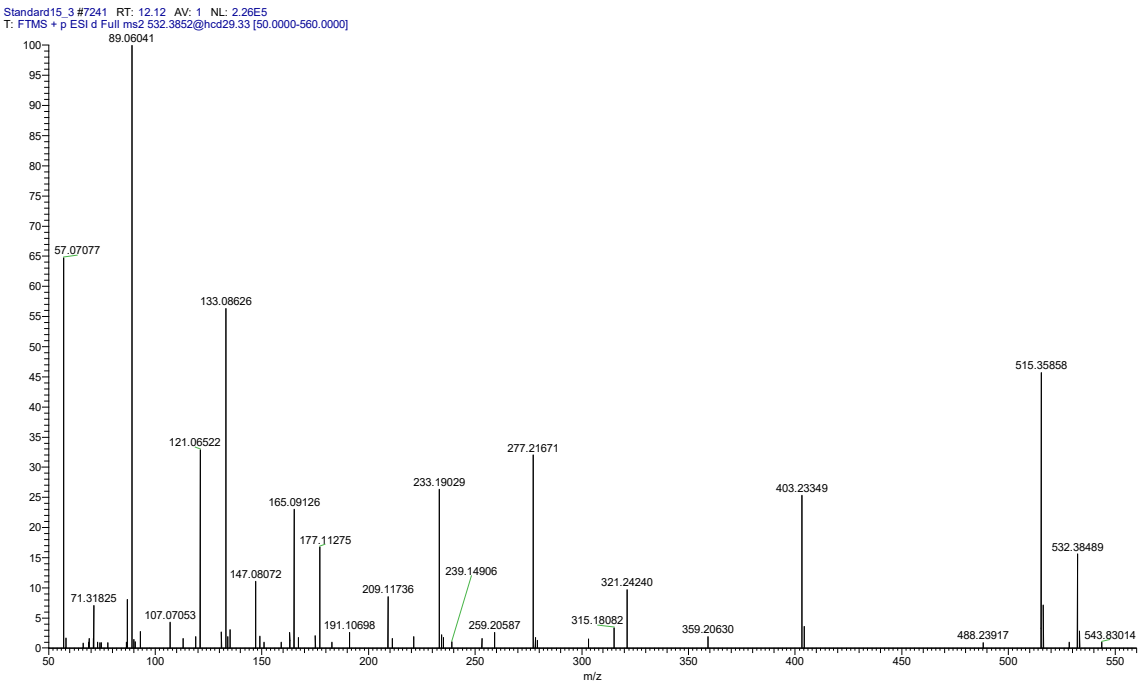
Monoisotopic mass: 514.347276



River sample



Reference standard

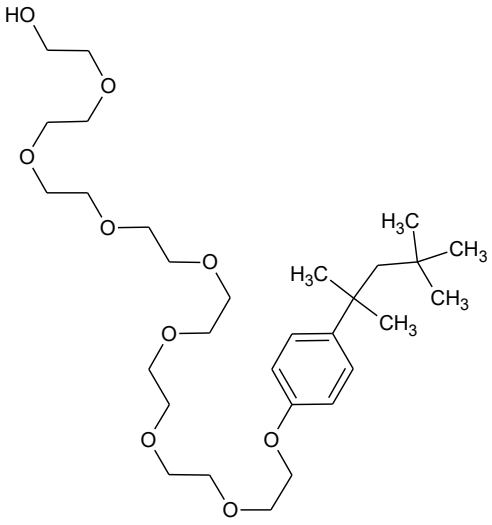


OPEO-08

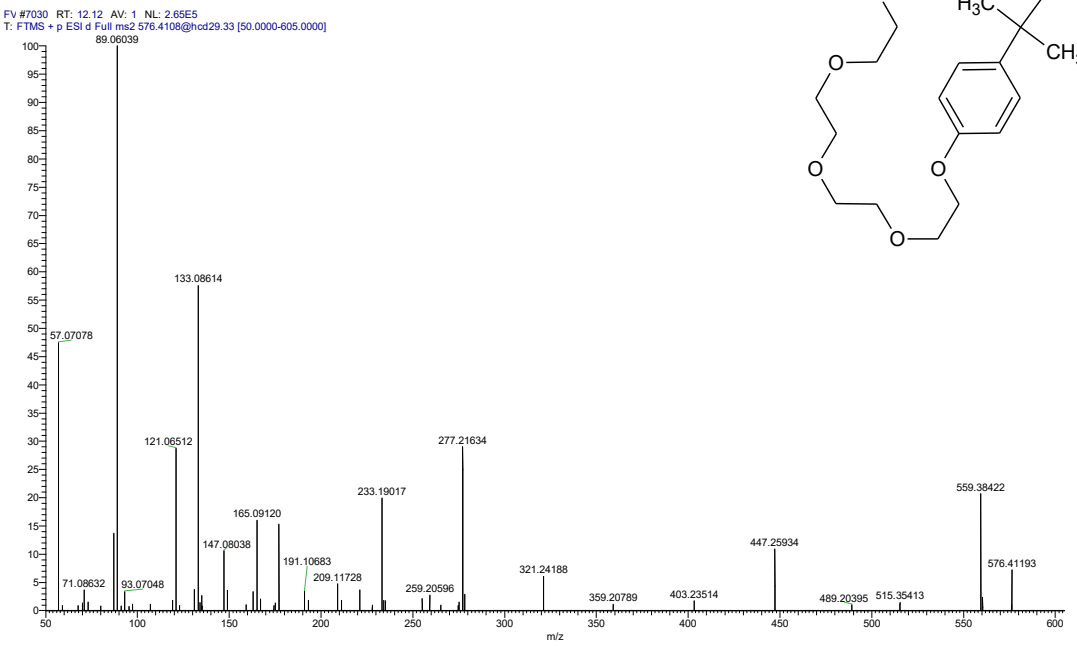
Surfactant, consumer product additive

Molecular formula: C30H54O9

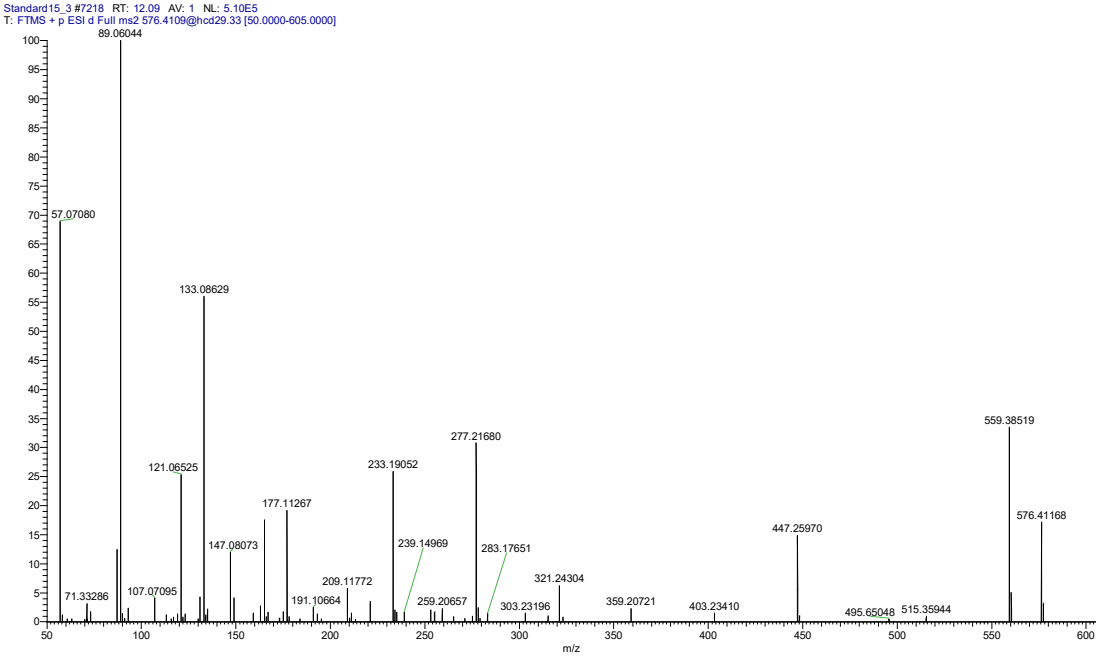
Monoisotopic mass: 558.372942



River sample



Reference standard

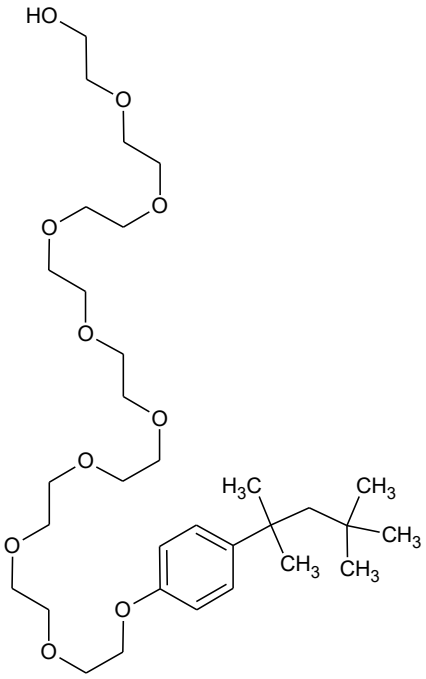


OPEO-09

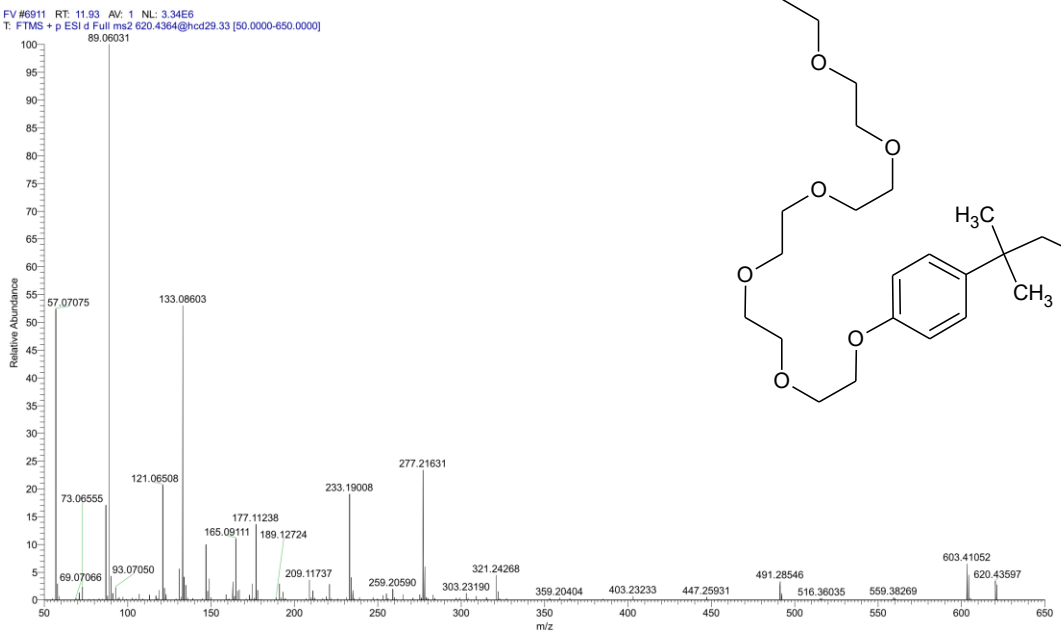
Surfactant, consumer product additive

Molecular formula: C32H58O10

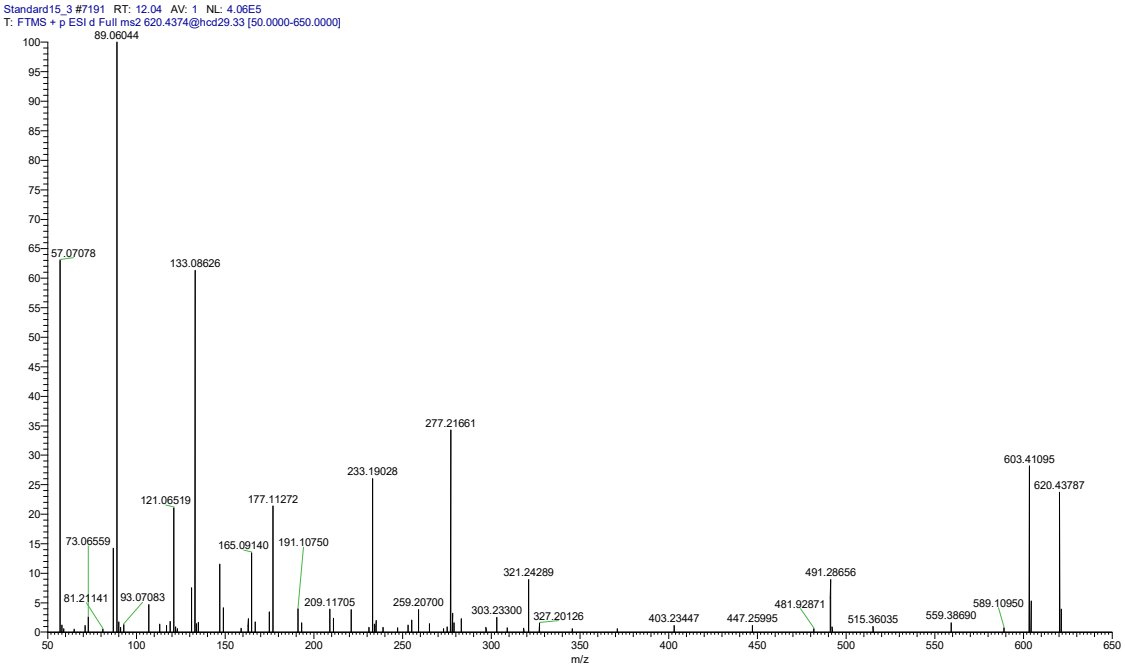
Monoisotopic mass: 602.398608



River sample



Reference standard

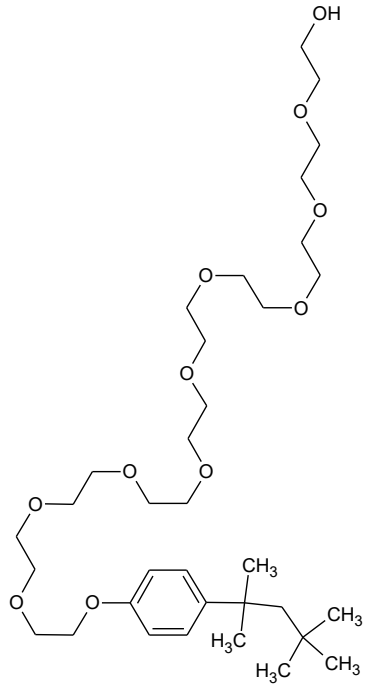


OPEO-10

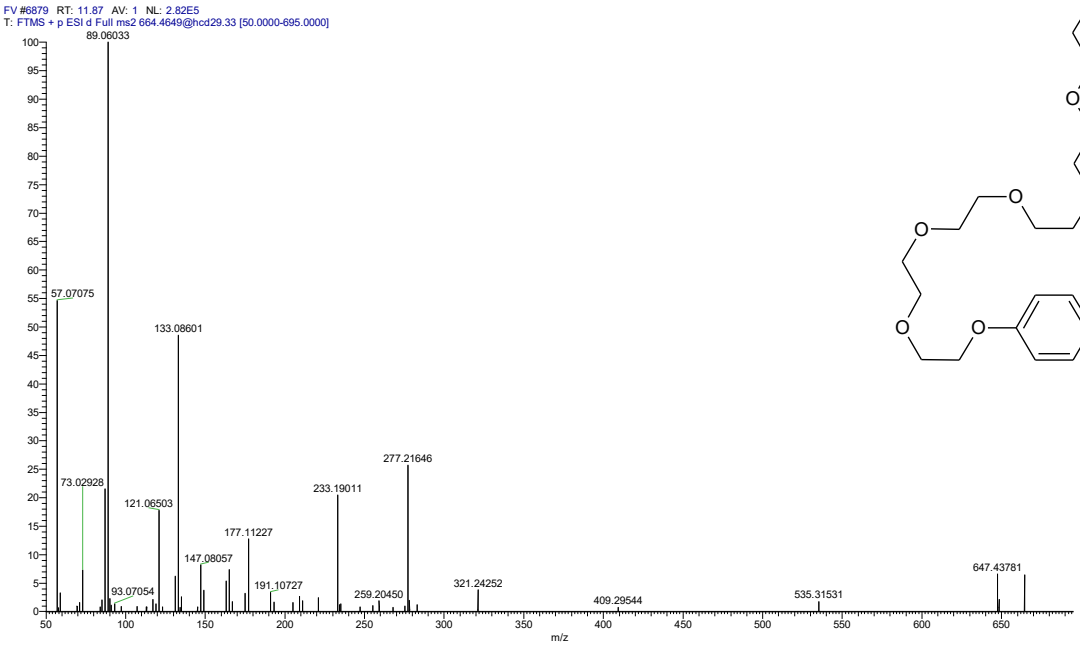
Surfactant, consumer product additive

Molecular formula: C34H62O11

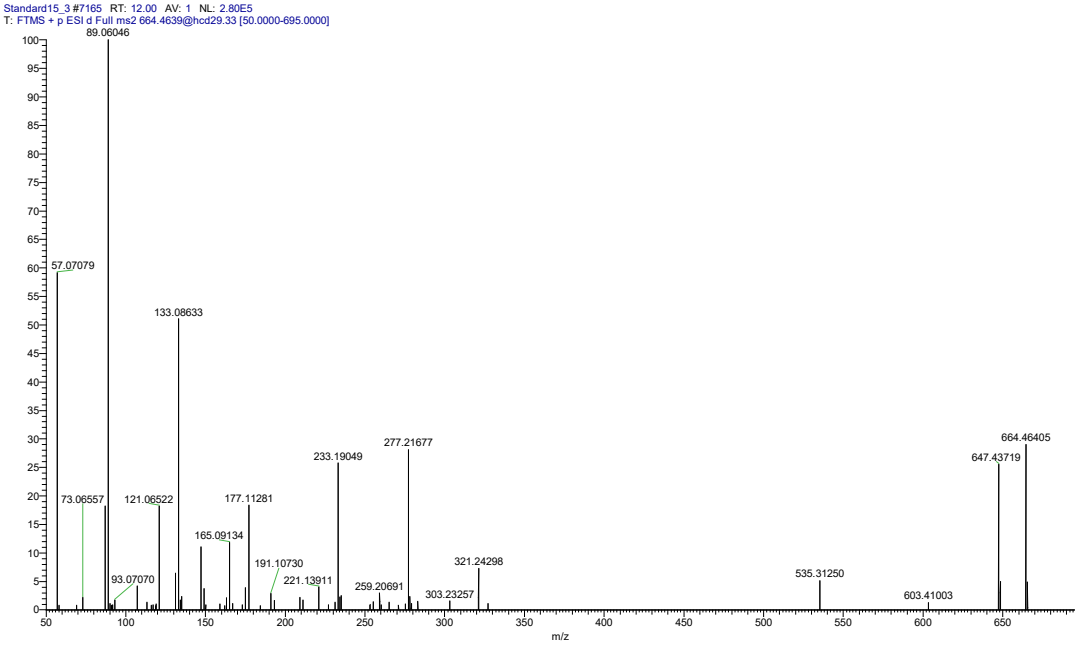
Monoisotopic mass: 646.424274



River sample



Reference standard

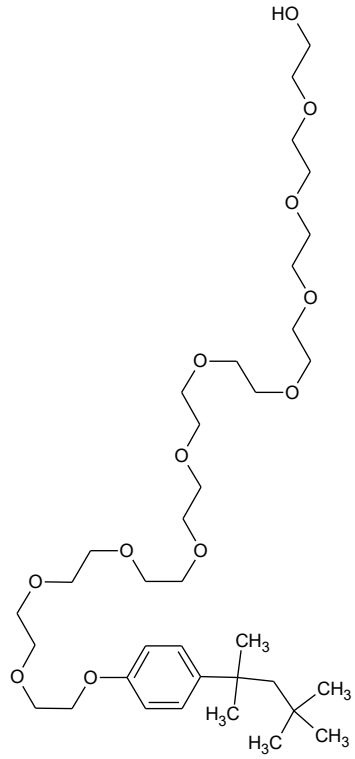


OPEO-11

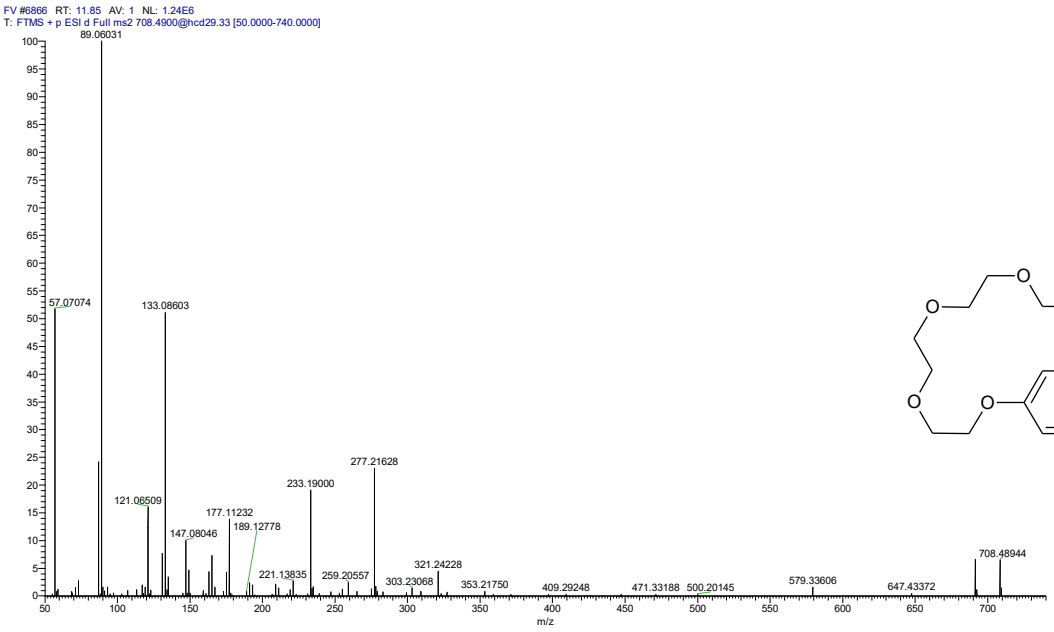
Surfactant, consumer product additive

Molecular formula: C36H66O12

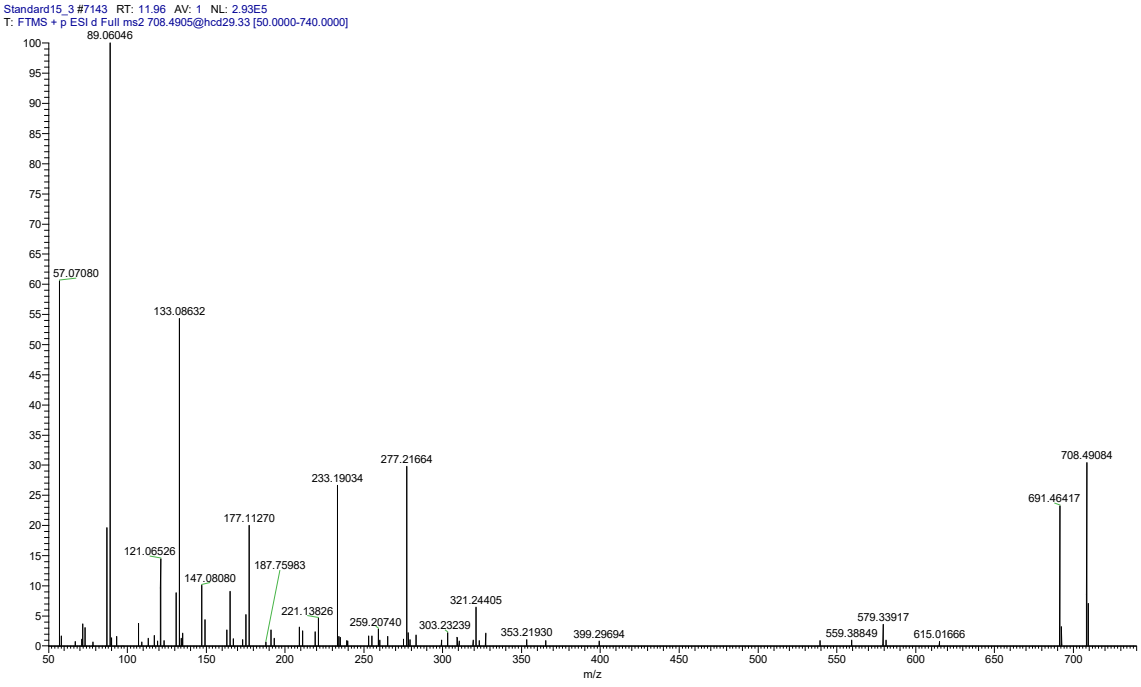
Monoisotopic mass: 690.44994



River sample



Reference standard

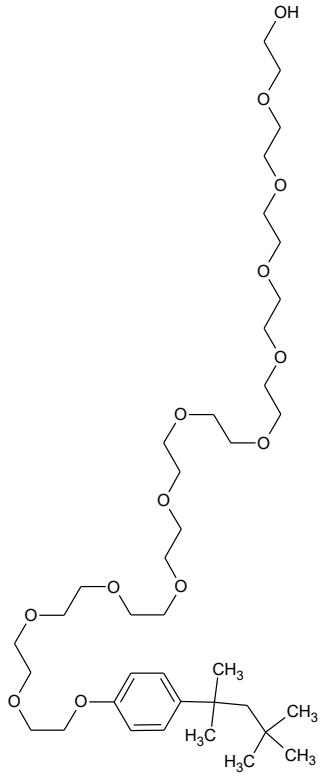


OPEO-12

Surfactant, consumer product additive

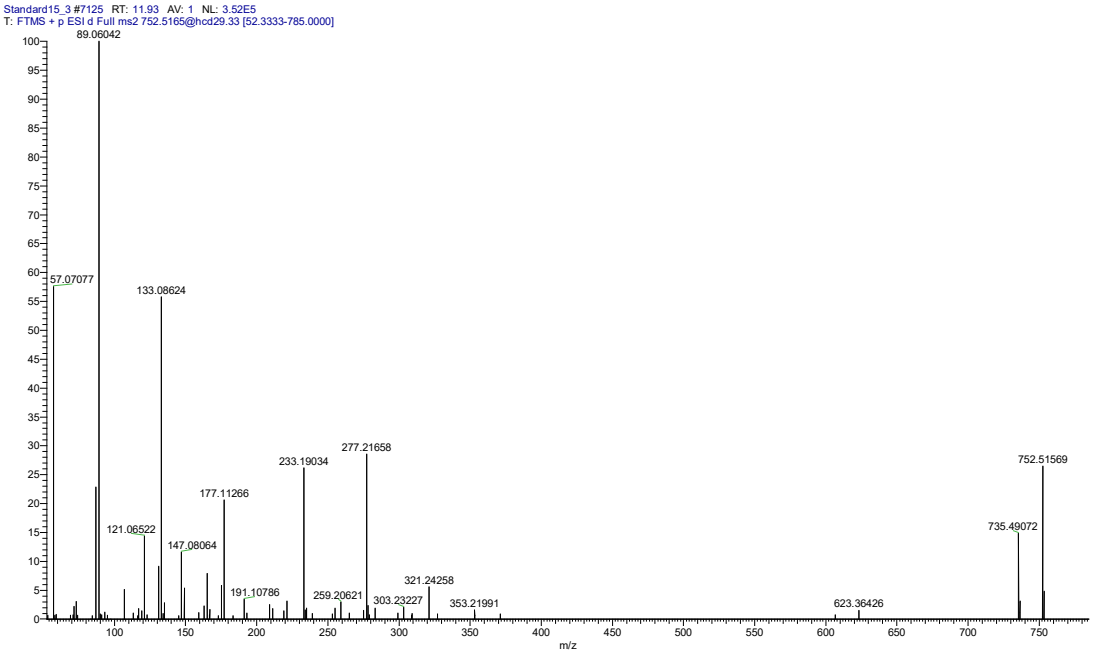
Molecular formula: C38H70O13

Monoisotopic mass: 734.475606



River sample

Reference standard

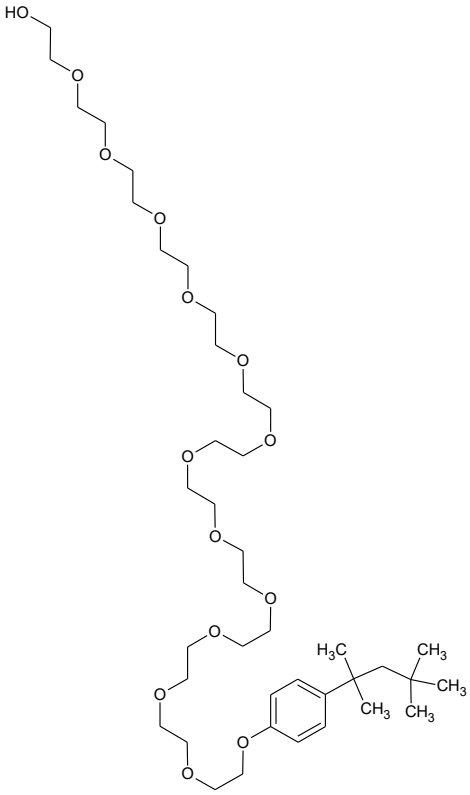


OPEO-13

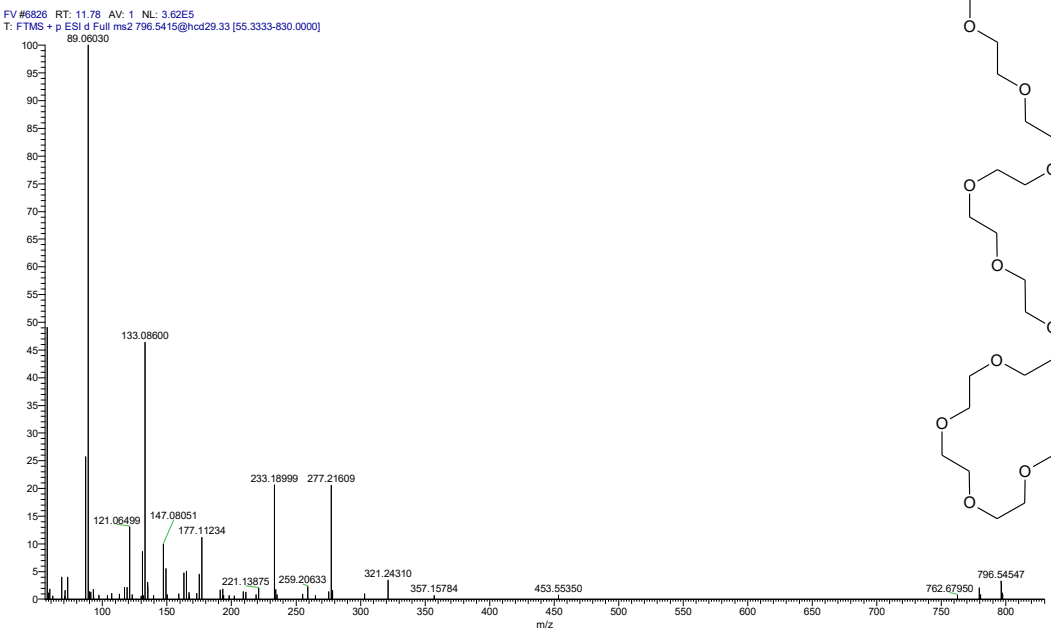
Surfactant, consumer product additive

Molecular formula: C40H74O14

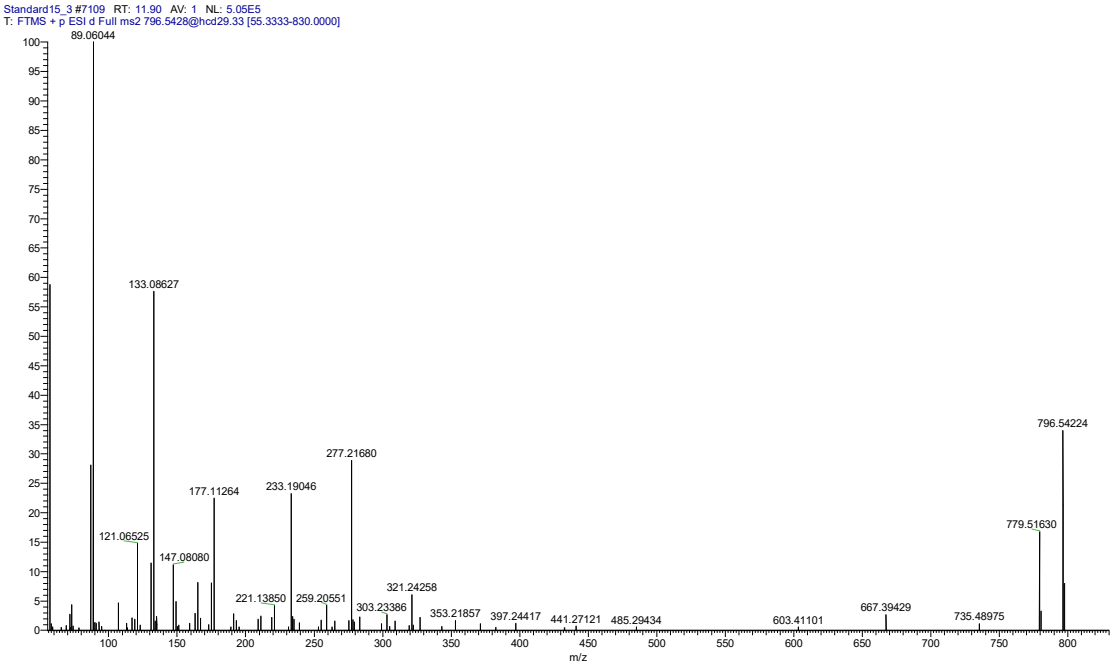
Monoisotopic mass: 778.501272



River sample



Reference standard

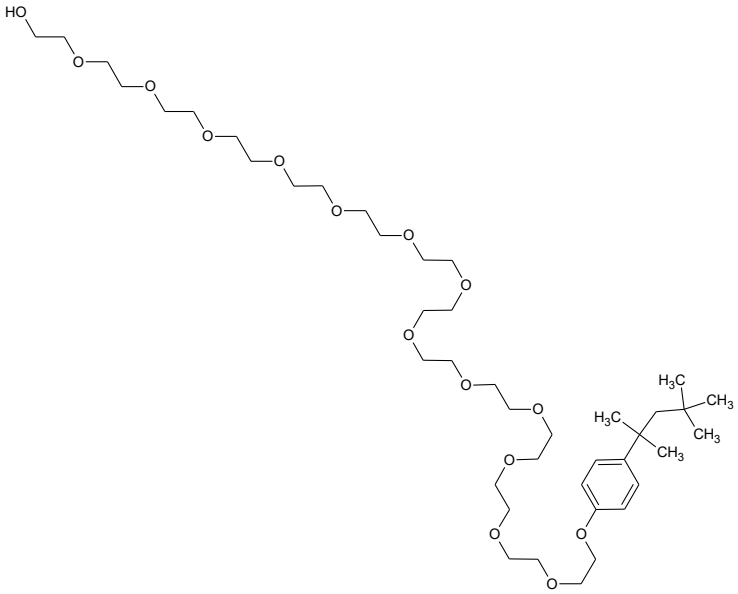


OPEO-14

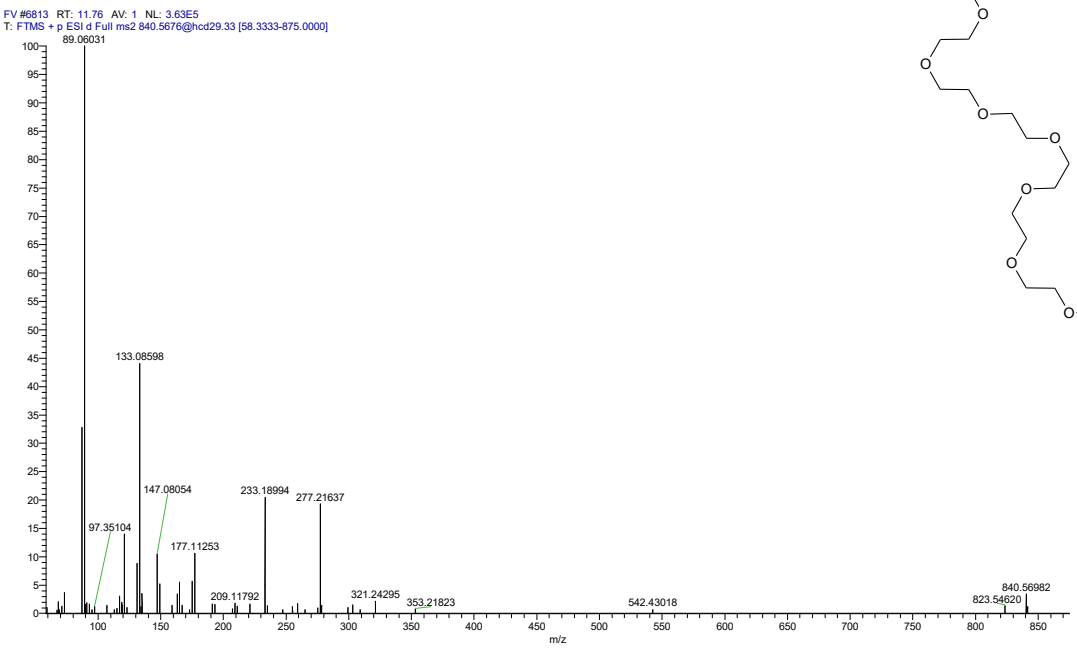
Surfactant, consumer product additive

Molecular formula: C42H78O15

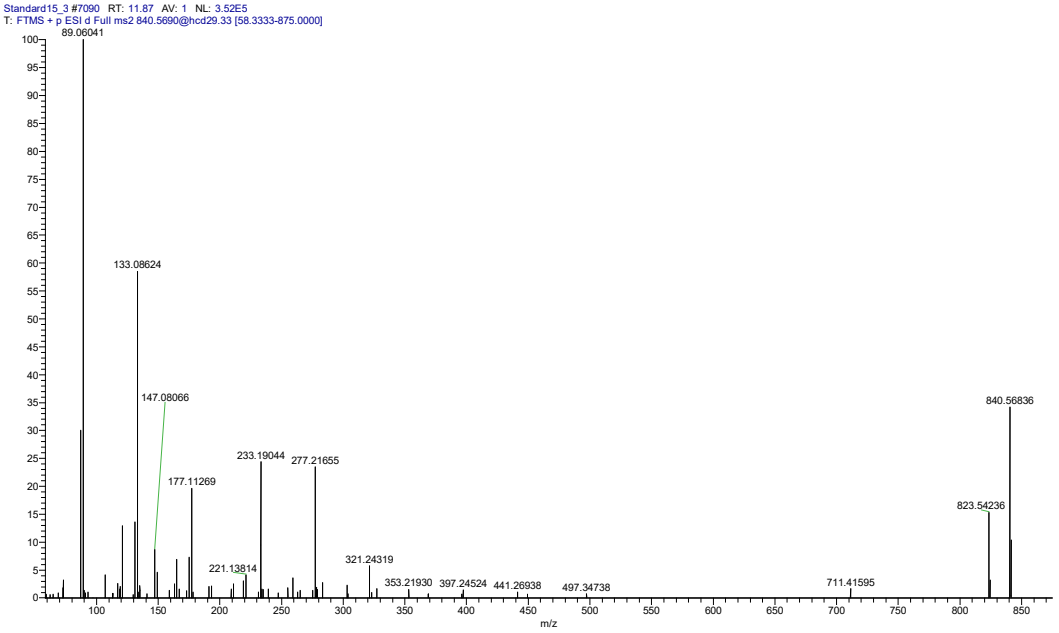
Monoisotopic mass: 822.526938



River sample



Reference standard

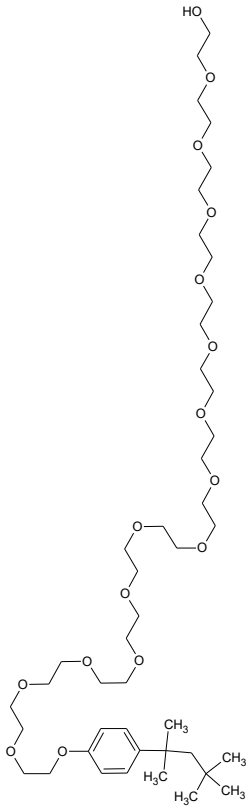


OPEO-15

Surfactant, consumer product additive

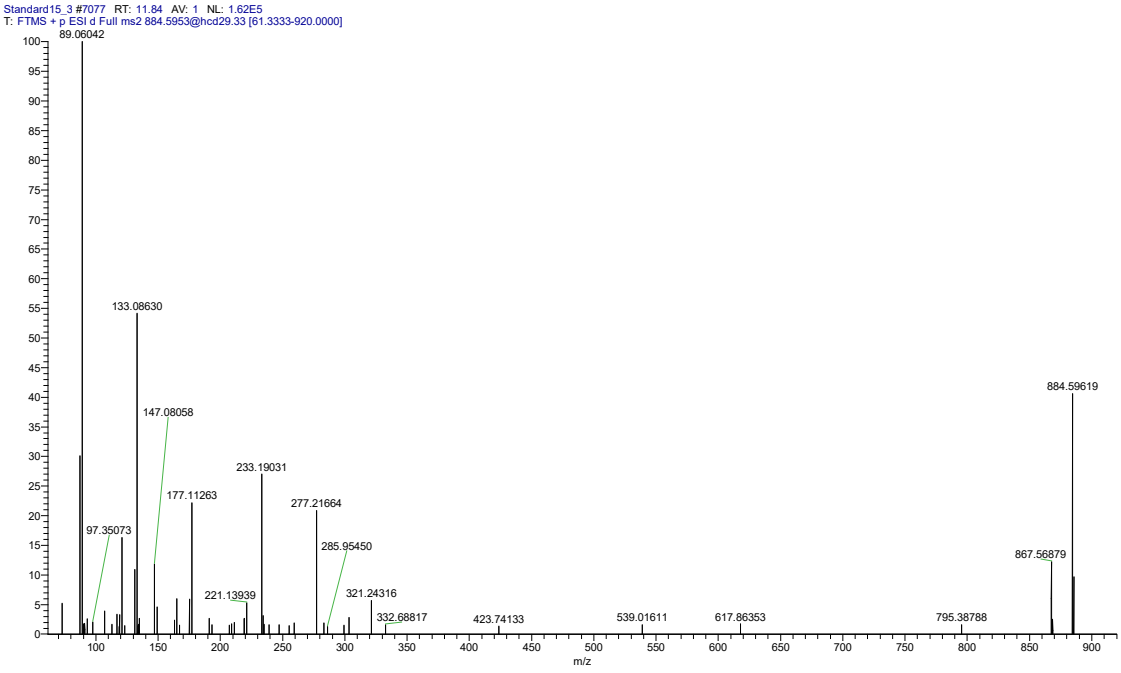
Molecular formula: C44H82O16

Monoisotopic mass: 866.552604



River sample

Reference standard

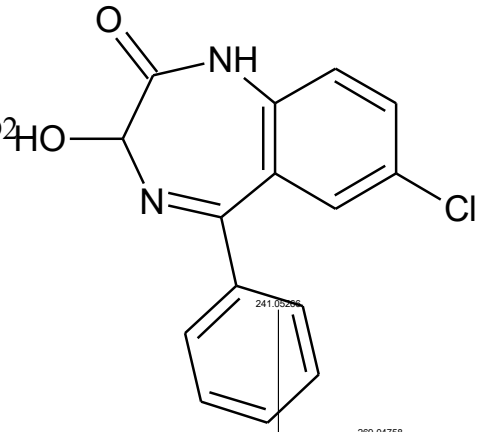


Oxazepam

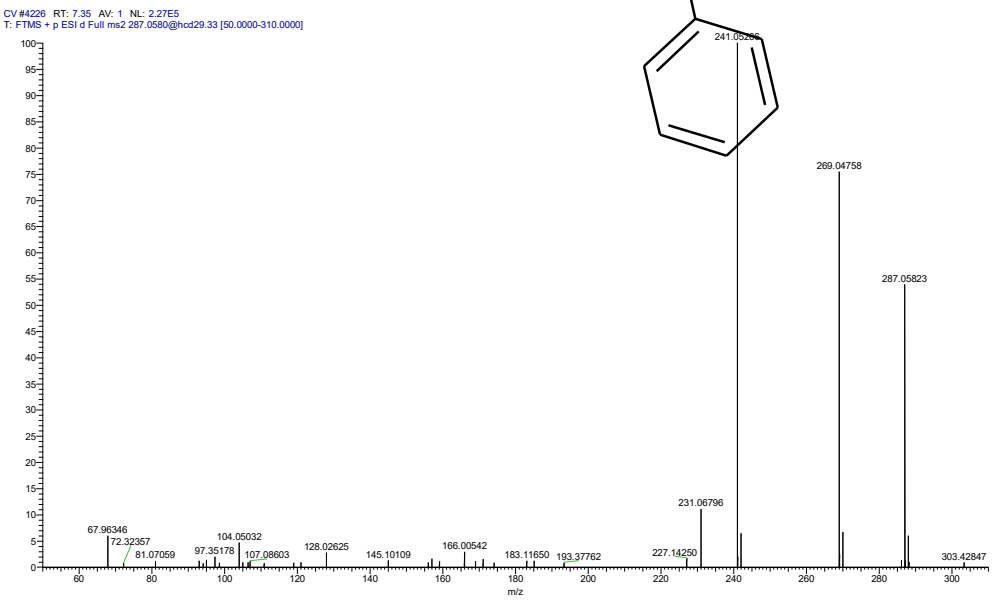
Nervous system, pharmaceutical

Molecular formula: C15H11ClN2O2

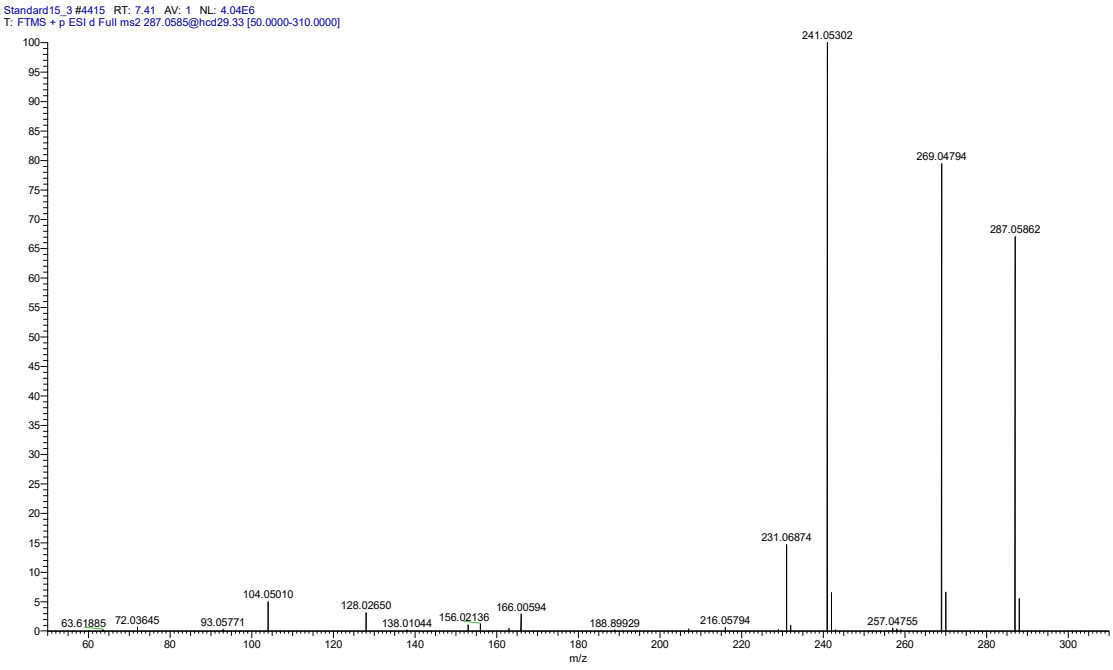
Monoisotopic mass: 286.050905



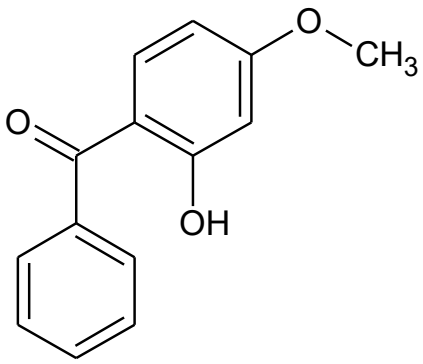
River sample



Reference standard



Oxybenzone

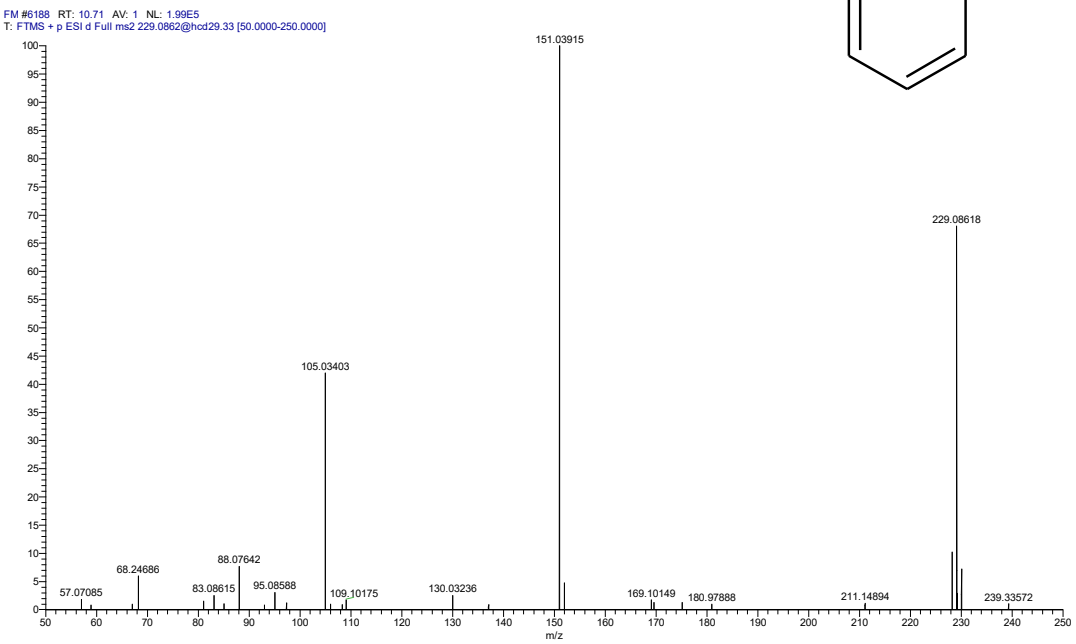


Cosmetic, consumer product additive

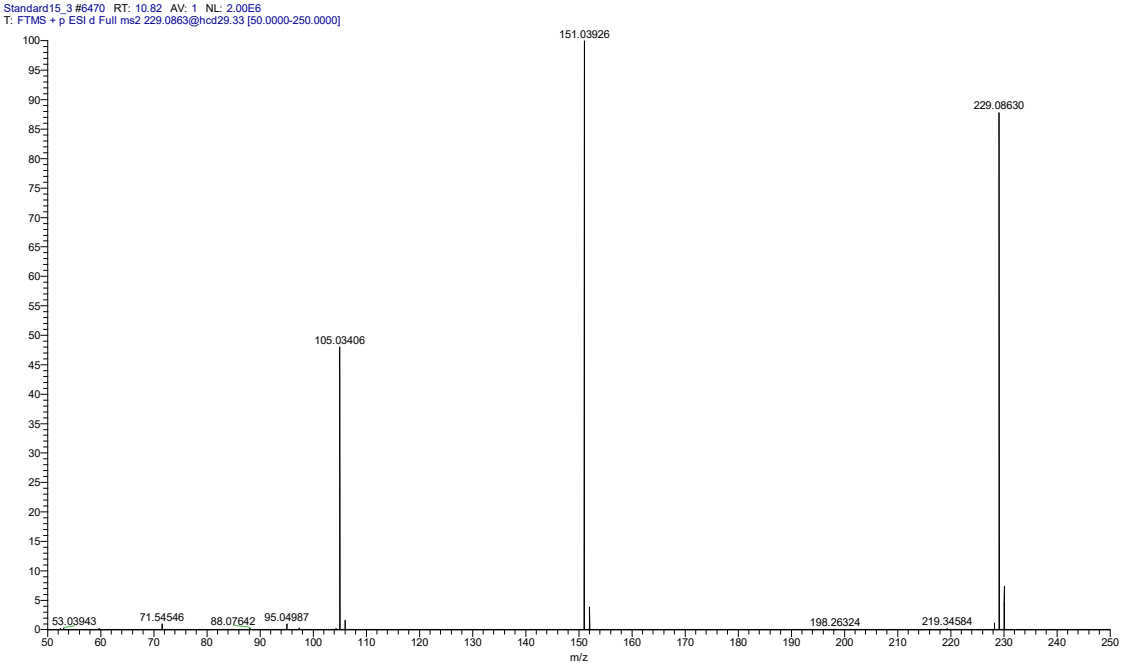
Molecular formula: C14H12O3

Monoisotopic mass: 228.078644

River sample



Reference standard

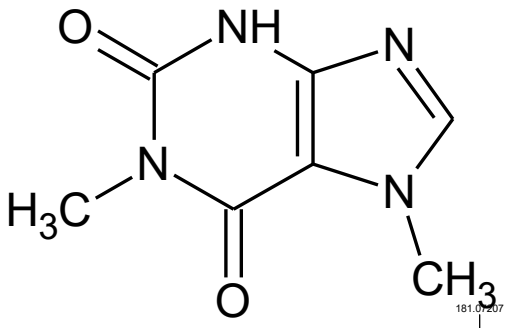


Paraxanthine

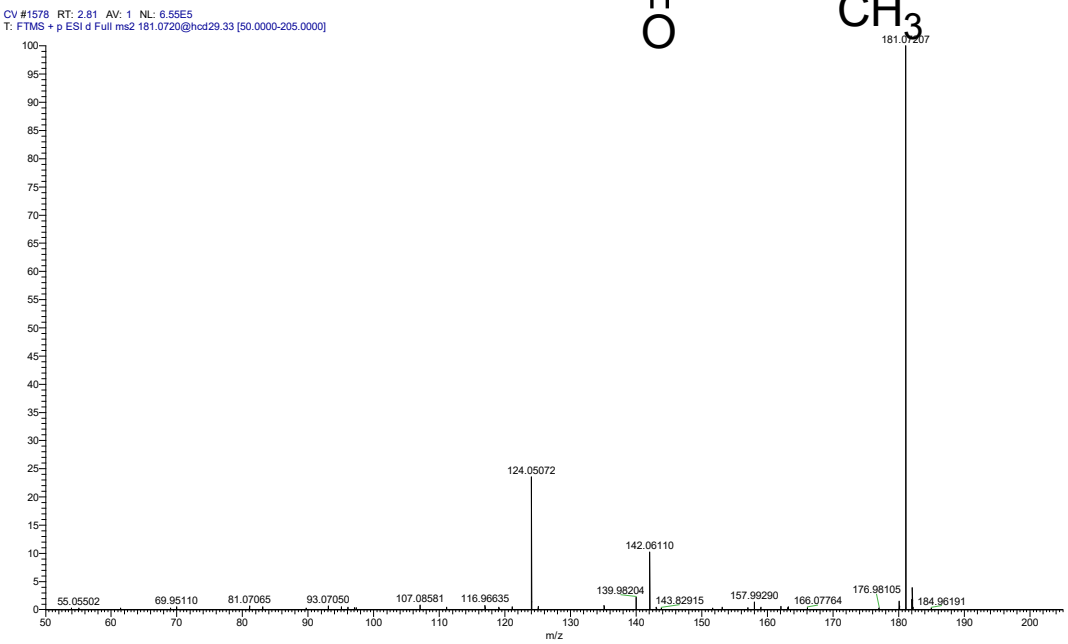
Nervous system, pharmaceutical

Molecular formula: C7H8N4O2

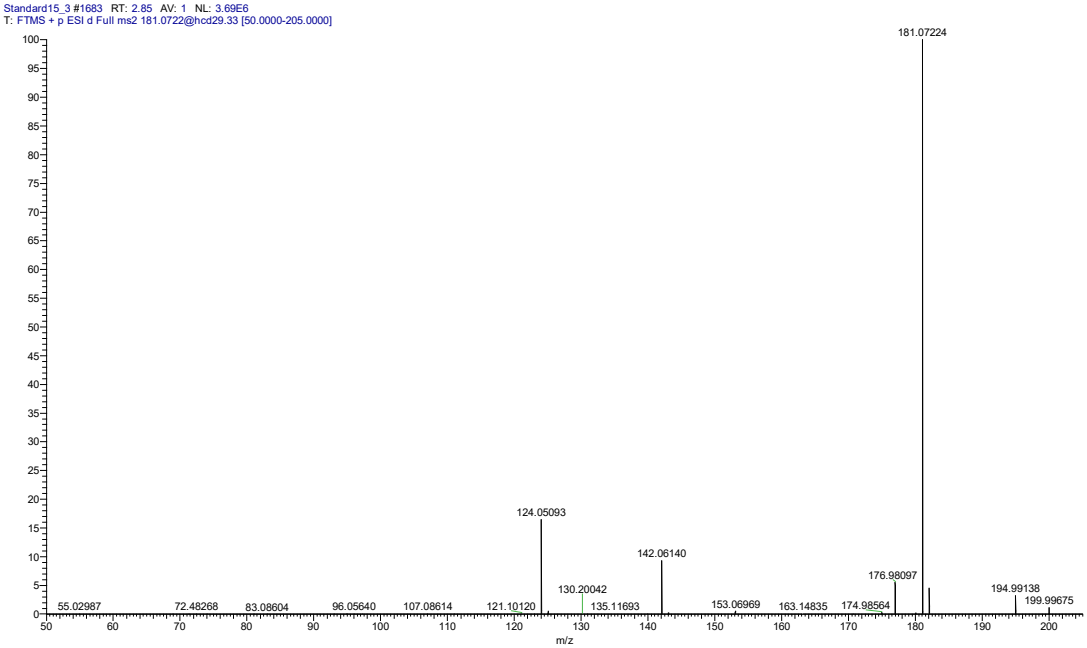
Monoisotopic mass: 180.064726



River sample



Reference standard

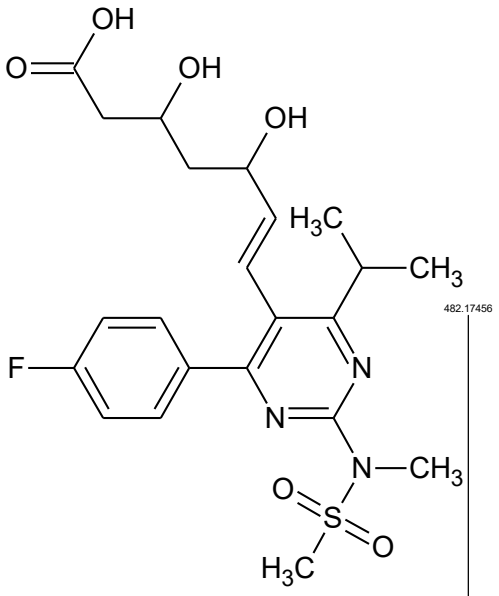


Rosuvastatin

Cardiovascular system, pharmaceutical

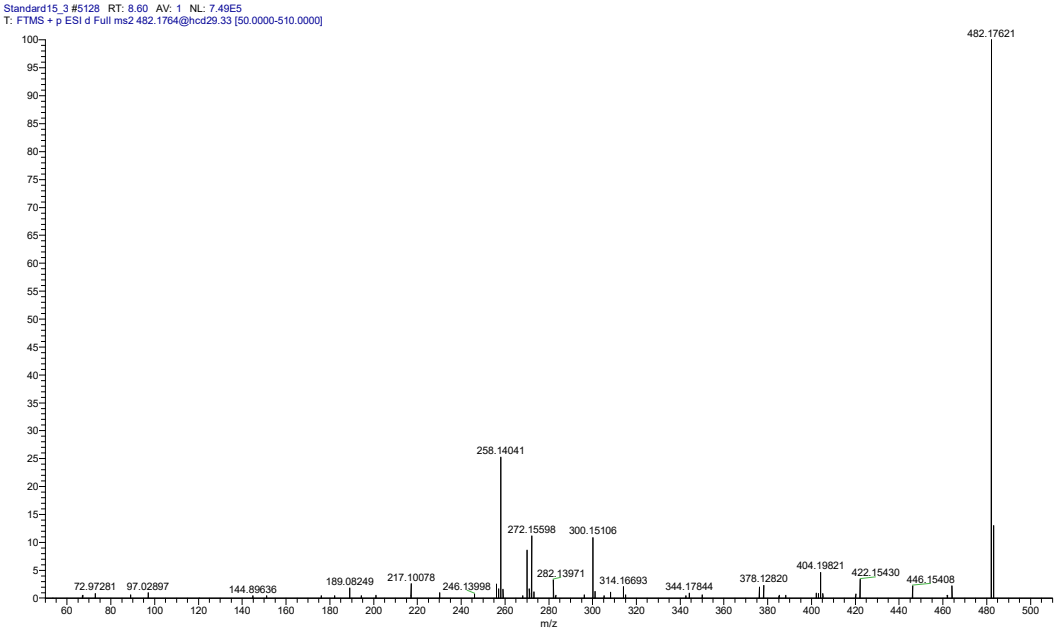
Molecular formula: C22H28FN3O6S

Monoisotopic mass: 481.1678614



River sample

Reference standard

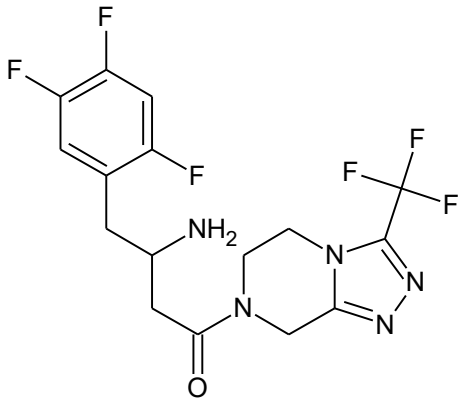


Sitagliptin

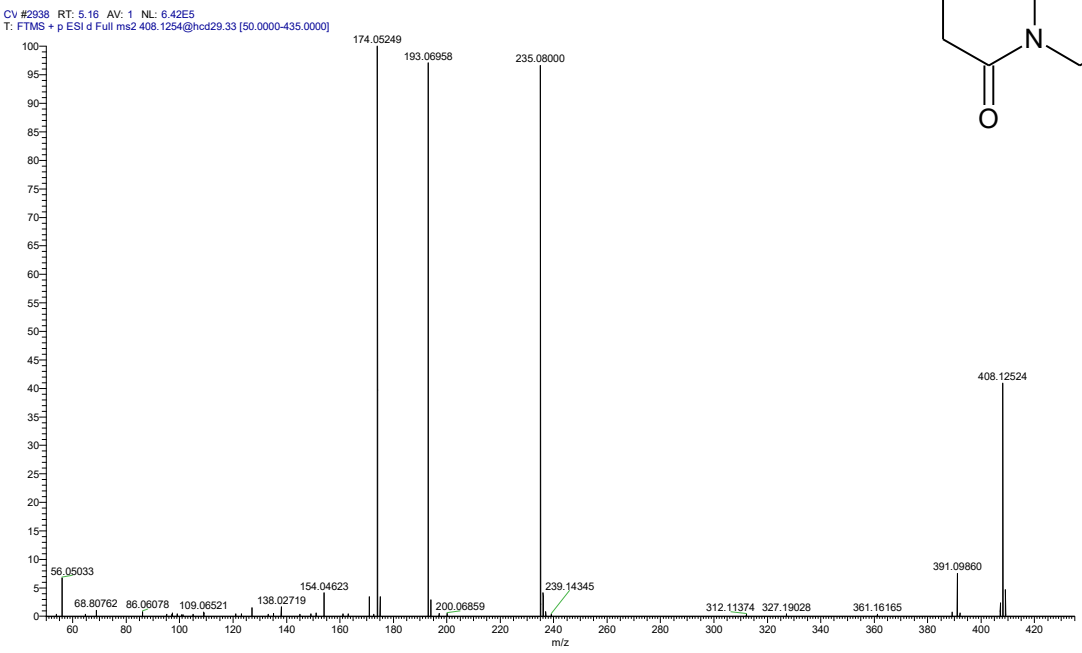
Alimentary tract and metabolism, pharmaceutical

Molecular formula: C16H15F6N5O

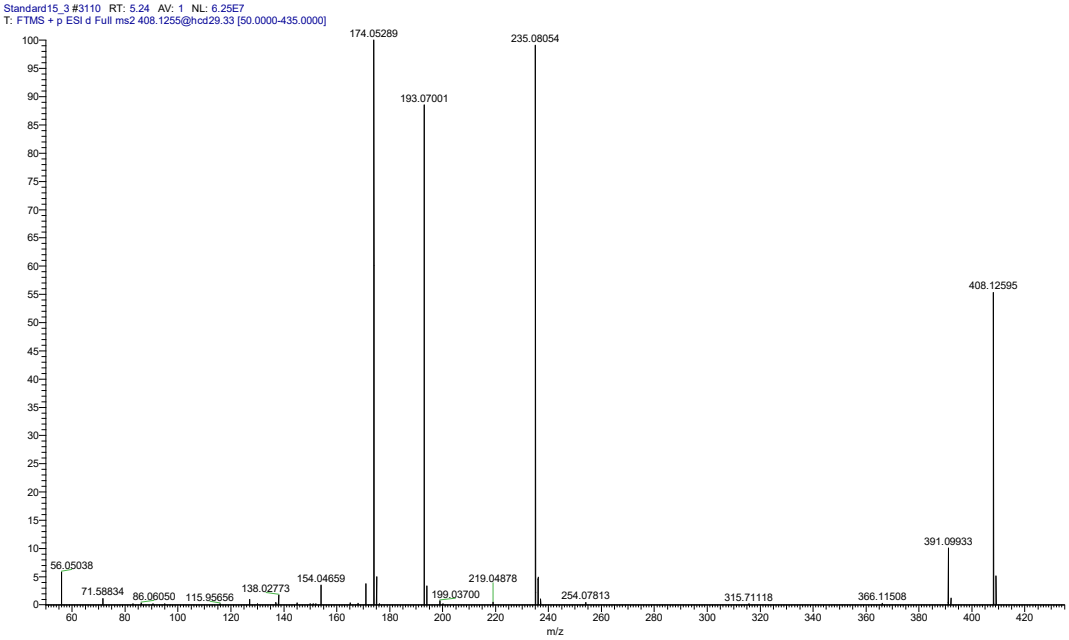
Monoisotopic mass: 407.118079



River sample



Reference standard

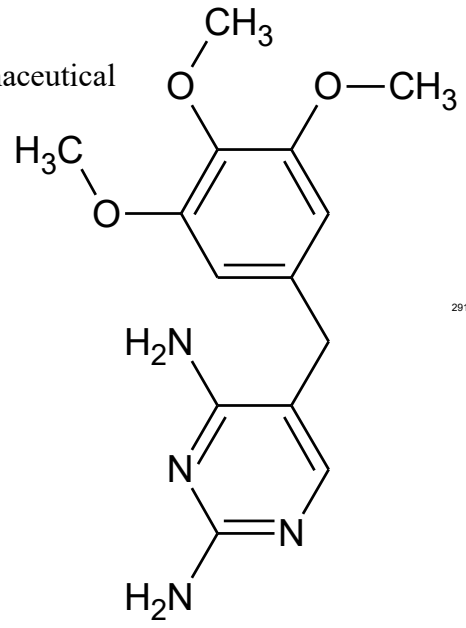


Trimethoprim

Antiinfective for systemic use, pharmaceutical

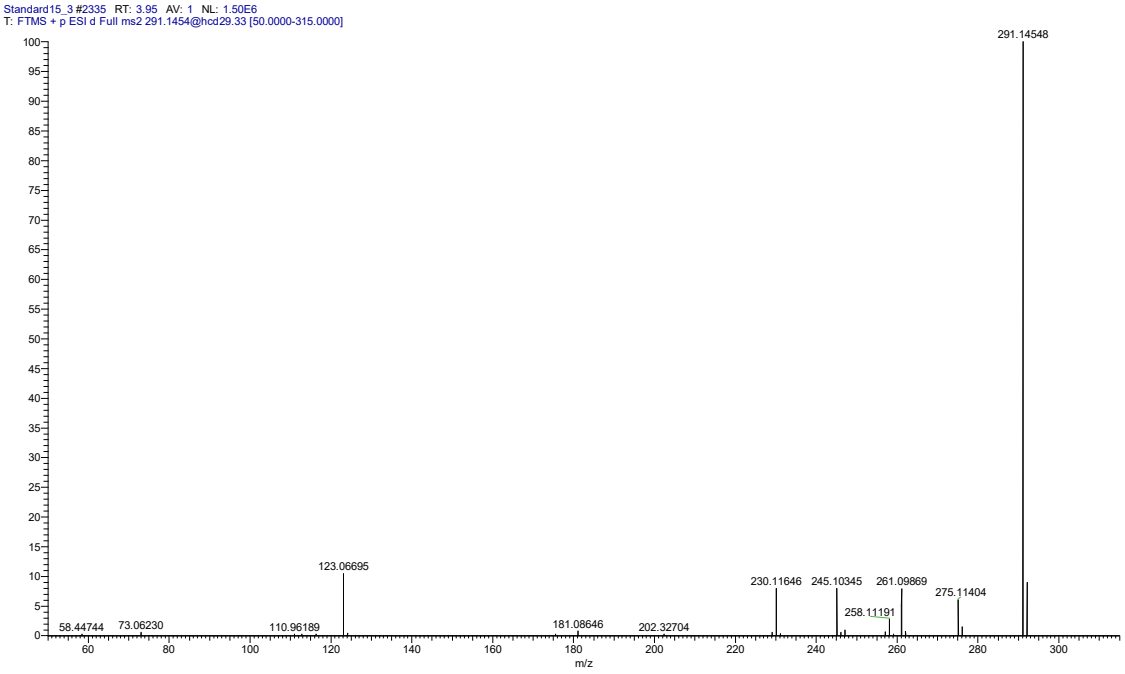
Molecular formula: C14H18N4O3

Monoisotopic mass: 290.13789



River sample

Reference standard

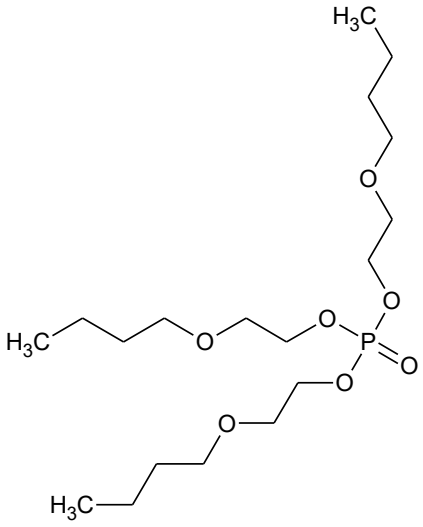


TBEP

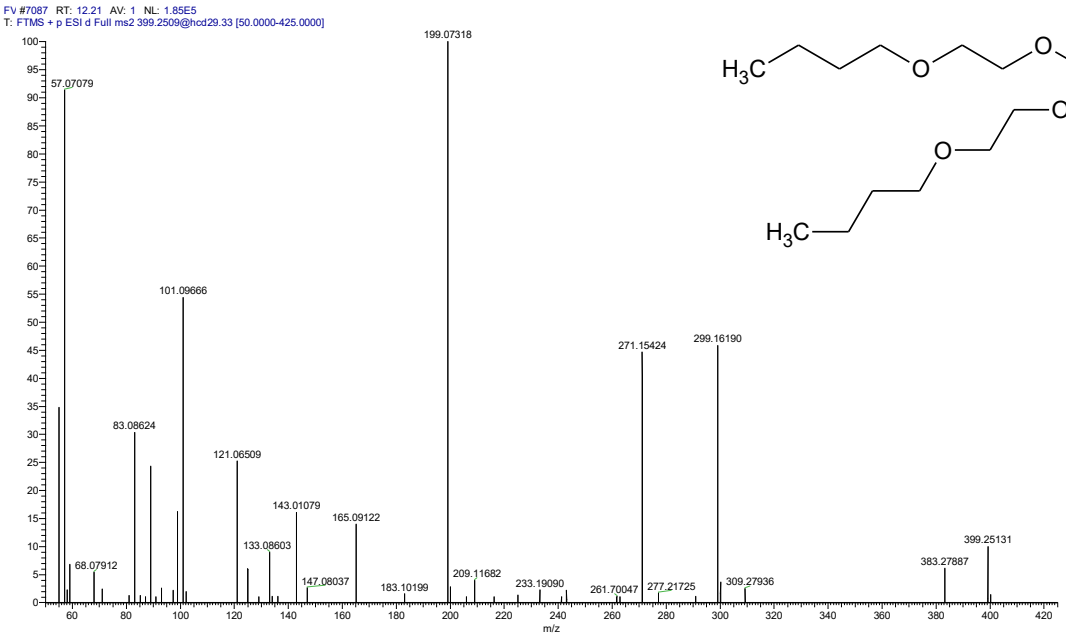
Polymer additive, consumer product additive

Molecular formula: C18H39O7P

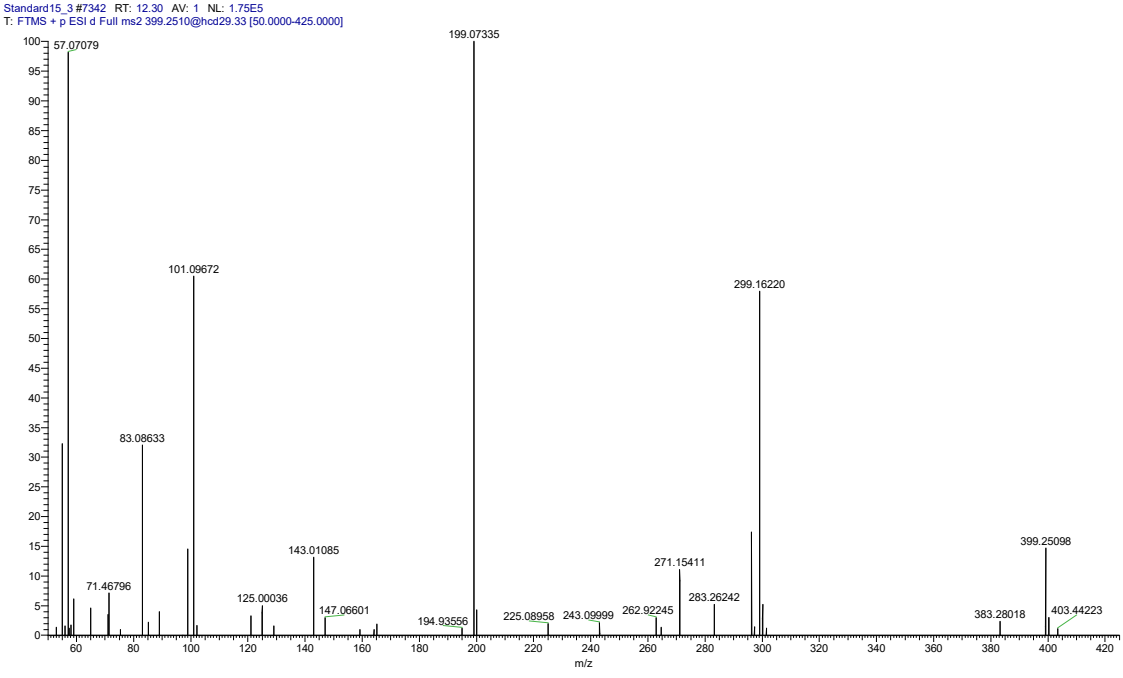
Monoisotopic mass: 398.243341



River sample



Reference standard

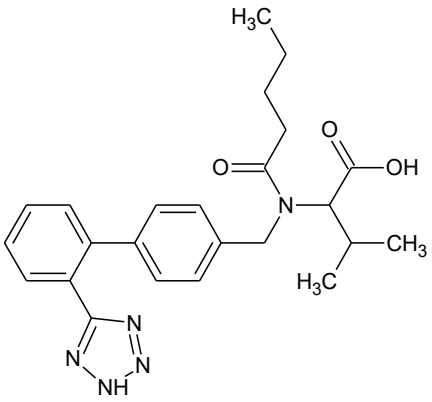


Valsartan

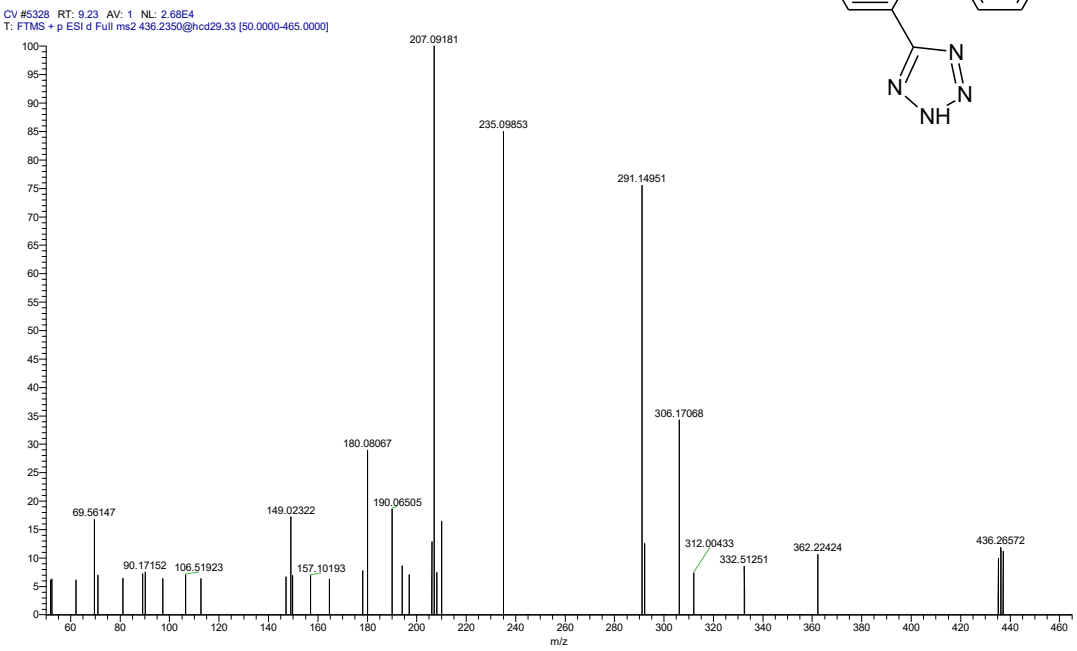
Cardiovascular system, pharmaceutical

Molecular formula: C24H29N5O3

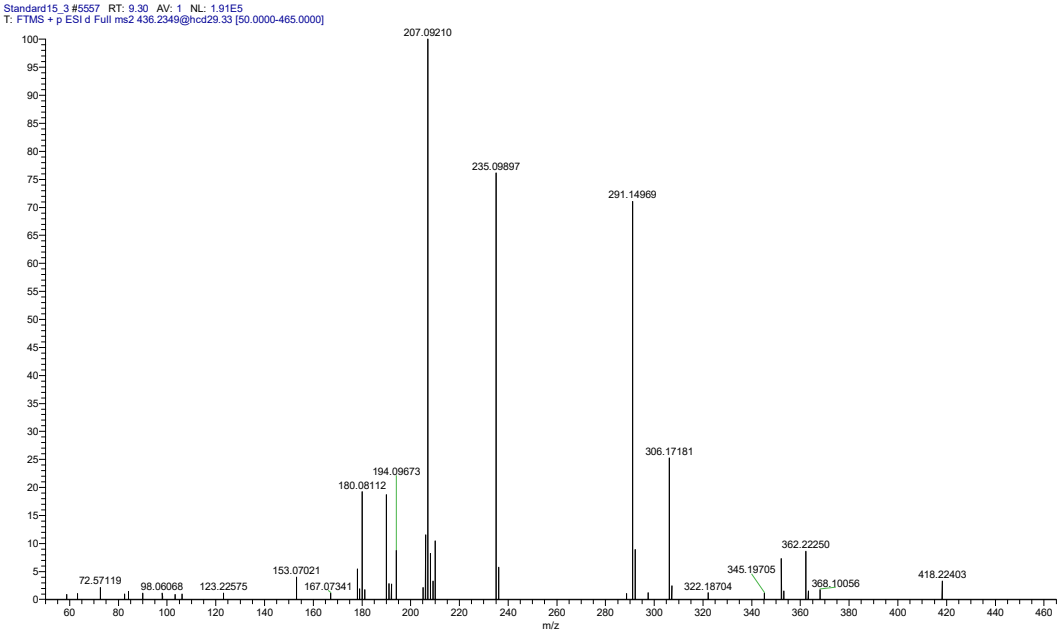
Monoisotopic mass: 435.22704



River sample



Reference standard



CHAPITRE 10. RÉFÉRENCES BIBLIOGRAPHIQUES

(1) Wang, Z.; Walker, G. W.; Muir, D. C. G.; Nagatani-Yoshida, K. Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. *Environ. Sci. Technol.* **2020**, *54* (5), 2575–2584. https://doi.org/10.1021/acs.est.9b06379.

(2) Bernhardt, E. S.; Rosi, E. J.; Gessner, M. O. Synthetic Chemicals as Agents of Global Change. *Front Ecol Environ* **2017**, *15* (2), 84–90. https://doi.org/10.1002/fee.1450.

(3) *Toxic exposure: Chemicals are in our water, food, air and furniture*. University of California. https://www.universityofcalifornia.edu/news/toxic-exposure-chemicals-are-our-water-food-air-and-furniture (accessed 2022-07-22).

(4) Magazine, S.; Daley, J. *Science Is Falling Woefully Behind in Testing New Chemicals*. Smithsonian Magazine. https://www.smithsonianmag.com/smart-news/science-falling-woefully-behind-testing-new-chemicals-180962027/ (accessed 2022-07-22).

(5) Zanolli, L.; Oliver, M. Explained: The Toxic Threat in Everyday Products, from Toys to Plastic. *The Guardian*. May 22, 2019.

(6) *How many chemicals are in use today?*. Chemical & Engineering News. https://cen.acs.org/articles/95/i9/chemicals-use-today.html (accessed 2022-07-22).

(7) Diamond, M. L.; de Wit, C. A.; Molander, S.; Scheringer, M.; Backhaus, T.; Lohmann, R.; Arvidsson, R.; Bergman, Å.; Hauschild, M.; Holoubek, I.; Persson, L.; Suzuki, N.; Vighi, M.; Zetzsch, C. Exploring the Planetary Boundary for Chemical Pollution. *Environment International* **2015**, *78*, 8–15. https://doi.org/10.1016/j.envint.2015.02.001.

(8) Suman, T.-Y.; Kim, S.-Y.; Yeom, D.-H.; Jeon, J. Transformation Products of Emerging Pollutants Explored Using Non-Target Screening: Perspective in the Transformation Pathway and Toxicity Mechanism—A Review. *Toxics* **2022**, *10* (2), 54. https://doi.org/10.3390/toxics10020054.

(9) Daughton, C. G.; Ternes, T. A. Pharmaceuticals and Personal Care Products in the Environment: Agents of Subtle Change? *Environmental Health Perspectives* **1999**, *107*, 32.

(10) Burns, E. E.; Carter, L. J.; Snape, J.; Thomas-Oates, J.; Boxall, A. B. A. Application of Prioritization Approaches to Optimize Environmental Monitoring and Testing of Pharmaceuticals. *Journal of Toxicology and Environmental Health, Part B* **2018**, *21* (3), 115–141. https://doi.org/10.1080/10937404.2018.1465873.

(11) Burns, E. E.; Thomas-Oates, J.; Kolpin, D. W.; Furlong, E. T.; Boxall, A. B. A. Are Exposure Predictions, Used for the Prioritization of Pharmaceuticals in the Environment, Fit for Purpose?: Are Prioritization PECs Fit for Purpose? *Environ Toxicol Chem* **2017**, *36* (10), 2823–2832. https://doi.org/10.1002/etc.3842.

(12) Canada, H. Chemicals Management Plan. **2016**.

(13) Daughton, C. G. The Matthew Effect and Widely Prescribed Pharmaceuticals Lacking Environmental Monitoring: Case Study of an Exposure-Assessment Vulnerability. *Science of The Total Environment* **2014**, *466–467*, 315–325. https://doi.org/10.1016/j.scitotenv.2013.06.111.

(14) Bletsou, A. A.; Jeon, J.; Hollender, J.; Archontaki, E.; Thomaidis, N. S. 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. https://doi.org/10.1016/j.trac.2014.11.009.
(15) Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N. S.; Bletsou, A.; Zwiener, C.; Ibáñez, M.; Portolés, T.; de Boer, R.; Reid, M. J.; Onghena, M.; Kunkel, U.; Schulz, W.; Guillon, A.; Noyon, N.; Leroy, G.; Bados, P.; Bogialli, S.; Stipaničev, D.; Rostkowski, P.; Hollender, J. Non-Target Screening with High-Resolution Mass Spectrometry: Critical Review Using a Collaborative Trial on Water Analysis. *Anal Bioanal Chem* **2015**, *407* (21), 6237–6255. https://doi.org/10.1007/s00216-015-8681-7.
(16) Baena-Nogueras, R. M.; González-Mazo, E.; Lara-Martín, P. A. Photolysis of Antibiotics under Simulated Sunlight Irradiation: Identification of Photoproducts by High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* **2017**, *51* (6), 3148–3156. https://doi.org/10.1021/acs.est.6b03038.
(17) Cermola, M.; DellaGreca, M.; Iesce, M. R.; Previtera, L.; Rubino, M.; Temussi, F.; Brigante, M. Phototransformation of Fibrate Drugs in Aqueous Media. *Environ Chem Lett* **2005**, *3* (1), 43–47. https://doi.org/10.1007/s10311-005-0112-0.
(18) Fatta-Kassinos, D.; Vasquez, M. I.; Kümmerer, K. Transformation Products of Pharmaceuticals in Surface Waters and Wastewater Formed during Photolysis and Advanced Oxidation Processes – Degradation, Elucidation of Byproducts and Assessment of Their Biological Potency. *Chemosphere* **2011**, *85* (5), 693–709. https://doi.org/10.1016/j.chemosphere.2011.06.082.
(19) West, C. E.; Rowland, S. J. Aqueous Phototransformation of Diazepam and Related Human Metabolites under Simulated Sunlight. *Environ. Sci. Technol.* **2012**, *46* (9), 4749–4756. https://doi.org/10.1021/es203529z.
(20) Wilde, M. L.; Menz, J.; Trautwein, C.; Leder, C.; Kümmerer, K. Environmental Fate and Effect Assessment of Thioridazine and Its Transformation Products Formed by Photodegradation. *Environmental Pollution* **2016**, *213*, 658–670. https://doi.org/10.1016/j.envpol.2016.03.018.
(21) Zeng, T.; Arnold, W. A. Pesticide Photolysis in Prairie Potholes: Probing Photosensitized Processes. *Environ. Sci. Technol.* **2013**, *47* (13), 6735–6745. https://doi.org/10.1021/es3030808.
(22) Baena-Nogueras, R. M.; González-Mazo, E.; Lara-Martín, P. A. Degradation Kinetics of Pharmaceuticals and Personal Care Products in Surface Waters: Photolysis vs Biodegradation. *Science of The Total Environment* **2017**, *590–591*, 643–654. https://doi.org/10.1016/j.scitotenv.2017.03.015.
(23) Drouin, G.; Droz, B.; Leresche, F.; Payraudeau, S.; Masbou, J.; Imfeld, G. Direct and Indirect Photodegradation of Atrazine and *S* -Metolachlor in Agriculturally Impacted Surface Water and Associated C and N Isotope Fractionation. *Environ. Sci.: Processes Impacts* **2021**, *23* (11), 1791–1802. https://doi.org/10.1039/D1EM00246E.
(24) Boix, C.; Ibáñez, M.; Sancho, J. V.; Parsons, J. R.; Voogt, P. de; 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. https://doi.org/10.1016/j.jhazmat.2015.09.053.

(25) Beretsou, V. G.; Psoma, A. K.; Gago-Ferrero, P.; Aalizadeh, R.; Fenner, K.; Thomaidis, N. S. Identification of Biotransformation Products of Citalopram Formed in Activated Sludge. *Water Research* **2016**, *103*, 205–214. https://doi.org/10.1016/j.watres.2016.07.029.
(26) Terzic, S.; Senta, I.; Matosic, M.; Ahel, M. Identification of Biotransformation Products of Macrolide and Fluoroquinolone Antimicrobials in Membrane Bioreactor Treatment by Ultrahigh-Performance Liquid Chromatography/Quadrupole Time-of-Flight Mass Spectrometry. *Anal Bioanal Chem* **2011**, *401* (1), 353–363. https://doi.org/10.1007/s00216-011-5060-x.
(27) Poirier-Larabie, S.; Segura, P. A.; Gagnon, C. Degradation of the Pharmaceuticals Diclofenac and Sulfamethoxazole and Their Transformation Products under Controlled Environmental Conditions. *Science of The Total Environment* **2016**, *557–558*, 257–267. https://doi.org/10.1016/j.scitotenv.2016.03.057.
(28) Albergamo, V.; Schollée, J. E.; Schymanski, E. L.; Helmus, R.; Timmer, H.; Hollender, J.; de Voogt, P. Nontarget Screening Reveals Time Trends of Polar Micropollutants in a Riverbank Filtration System. *Environ. Sci. Technol.* **2019**, *53* (13), 7584–7594. https://doi.org/10.1021/acs.est.9b01750.
(29) Chung, I.-Y.; Park, Y.-M.; Lee, H.-J.; Kim, H.; Kim, D.-H.; Kim, I.-G.; Kim, S.-M.; Do, Y.-S.; Seok, K.-S.; Kwon, J.-H. Nontarget Screening Using Passive Air and Water Sampling with a Level II Fugacity Model to Identify Unregulated Environmental Contaminants. *Journal of Environmental Sciences* **2017**, *62*, 84–91. https://doi.org/10.1016/j.jes.2017.06.036.
(30) Ferrer, I.; Sweeney, D. L.; Thurman, E. M.; Zweigenbaum, J. A. Nontargeted Screening of Water Samples Using Data-Dependent Acquisition with Similar Partition Searching. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (6), 1189–1204. https://doi.org/10.1021/jasms.0c00031.
(31) Fu, Y.; Zhou, Z.; Kong, H.; Lu, X.; Zhao, X.; Chen, Y.; Chen, J.; Wu, Z.; Xu, Z.; Zhao, C.; Xu, G. Nontargeted Screening Method for Illegal Additives Based on Ultrahigh-Performance Liquid Chromatography–High-Resolution Mass Spectrometry. *Anal. Chem.* **2016**, *88* (17), 8870–8877. https://doi.org/10.1021/acs.analchem.6b02482.
(32) Hollender, J.; Schymanski, E. L.; Singer, H. P.; Ferguson, P. L. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51* (20), 11505–11512. https://doi.org/10.1021/acs.est.7b02184.
(33) Hug, C.; Ulrich, N.; Schulze, T.; Brack, W.; Krauss, M. Identification of Novel Micropollutants in Wastewater by a Combination of Suspect and Nontarget Screening. *Environmental Pollution* **2014**, *184*, 25–32. https://doi.org/10.1016/j.envpol.2013.07.048.
(34) Kiefer, K.; Du, L.; Singer, H.; Hollender, J. Identification of LC-HRMS Nontarget Signals in Groundwater after Source Related Prioritization. *Water Research* **2021**, *196*, 116994. https://doi.org/10.1016/j.watres.2021.116994.
(35) Lara-Martín, P. A.; Chiaia-Hernández, A. C.; Biel-Maeso, M.; Baena-Nogueras, R. M.; Hollender, J. Tracing Urban Wastewater Contaminants into the Atlantic Ocean by Nontarget Screening. *Environ. Sci. Technol.* **2020**, *54* (7), 3996–4005. https://doi.org/10.1021/acs.est.9b06114.

(36) Zedda, M.; Zwiener, C. Is Nontarget Screening of Emerging Contaminants by LC-HRMS Successful? A Plea for Compound Libraries and Computer Tools. *Anal Bioanal Chem* **2012**, *403* (9), 2493–2502. https://doi.org/10.1007/s00216-012-5893-y.
(37) McCord, J.; Strynar, M. Identification of Per- and Polyfluoroalkyl Substances in the Cape Fear River by High Resolution Mass Spectrometry and Nontargeted Screening. *Environ. Sci. Technol.* **2019**, *53* (9), 4717–4727. https://doi.org/10.1021/acs.est.8b06017.
(38) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105.
(39) Emwas, A.-H. M. The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research. In *Metabonomics*; Bjerrum, J. T., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2015; Vol. 1277, pp 161–193. https://doi.org/10.1007/978-1-4939-2377-9_13.
(40) Pfaendner, R. Plastic Additives. In *Kirk-Othmer Encyclopedia of Chemical Technology*; American Cancer Society, 2019; pp 1–38. https://doi.org/10.1002/0471238961.koe00051.
(41) *Convention de Stockholm Sur Les Polluants Organiques Persistants*; 2004.
(42) *Loi Canadienne Sur La Sécurité Des Produits de Consommation*; 2010.
(43) Soares, A.; Guieysse, B.; Jefferson, B.; Cartmell, E.; Lester, J. N. Nonylphenol in the Environment: A Critical Review on Occurrence, Fate, Toxicity and Treatment in Wastewaters. *Environment International* **2008**, *34* (7), 1033–1049. https://doi.org/10.1016/j.envint.2008.01.004.
(44) Sardar, S. W.; Choi, Y.; Park, N.; Jeon, J. Occurrence and Concentration of Chemical Additives in Consumer Products in Korea. *IJERPH* **2019**, *16* (24), 5075. https://doi.org/10.3390/ijerph16245075.
(45) Priac, A. Alkylphenol and Alkylphenol Polyethoxylates in Water and Wastewater: A Review of Options for Their Elimination. 25.
(46) Acir, I.-H.; Guenther, K. Endocrine-Disrupting Metabolites of Alkylphenol Ethoxylates – A Critical Review of Analytical Methods, Environmental Occurrences, Toxicity, and Regulation. *Science of The Total Environment* **2018**, *635*, 1530–1546. https://doi.org/10.1016/j.scitotenv.2018.04.079.
(47) Monnery, B. D.; Wright, M.; Cavill, R.; Hoogenboom, R.; Shaunak, S.; Steinke, J. H. G.; Thanou, M. Cytotoxicity of Polycations: Relationship of Molecular Weight and the Hydrolytic Theory of the Mechanism of Toxicity. *International Journal of Pharmaceutics* **2017**, *521* (1–2), 249–258. https://doi.org/10.1016/j.ijpharm.2017.02.048.
(48) Tian, Z.; Zhao, H.; Peter, K. T.; Gonzalez, M.; Wetzel, J.; Wu, C.; Hu, X.; Prat, J.; Mudrock, E.; Hettinger, R.; Cortina, A. E.; Biswas, R. G.; Kock, F. V. C.; Soong, R.; Jenne, A.; Du, B.; Hou, F.; He, H.; Lundeen, R.; Gilbreath, A.; Sutton, R.; Scholz, N. L.; Davis, J. W.; Dodd, M. C.; Simpson, A.; McIntyre, J. K.; Kolodziej, E. P. A Ubiquitous Tire Rubber–Derived Chemical Induces Acute Mortality in Coho Salmon. *Science* **2021**, *371* (6525), 185–189. https://doi.org/10.1126/science.abd6951.

(49) Tran, N. H.; Reinhard, M.; Gin, K. Y.-H. Occurrence and Fate of Emerging Contaminants in Municipal Wastewater Treatment Plants from Different Geographical Regions-a Review. *Water Research* **2018**, *133*, 182–207. https://doi.org/10.1016/j.watres.2017.12.029.
(50) Cleuvers, M. Aquatic Ecotoxicity of Pharmaceuticals Including the Assessment of Combination Effects. *Toxicology Letters* **2003**, *142* (3), 185–194. https://doi.org/10.1016/S0378-4274(03)00068-7.
(51) Ferrari, B.; Paxéus, N.; Giudice, R. L.; Pollio, A.; Garric, J. Ecotoxicological Impact of Pharmaceuticals Found in Treated Wastewaters: Study of Carbamazepine, Clofibric Acid, and Diclofenac. *Ecotoxicology and Environmental Safety* **2003**, *55* (3), 359–370. https://doi.org/10.1016/S0147-6513(02)00082-9.
(52) Al Aukidy, M.; Verlicchi, P.; Jelic, A.; Petrovic, M.; Barcelò, D. Monitoring Release of Pharmaceutical Compounds: Occurrence and Environmental Risk Assessment of Two WWTP Effluents and Their Receiving Bodies in the Po Valley, Italy. *Science of The Total Environment* **2012**, *438*, 15–25. https://doi.org/10.1016/j.scitotenv.2012.08.061.
(53) Verlicchi, P.; Al Aukidy, M.; Zambello, E. Occurrence of Pharmaceutical Compounds in Urban Wastewater: Removal, Mass Load and Environmental Risk after a Secondary Treatment—A Review. *Science of The Total Environment* **2012**, *429*, 123–155. https://doi.org/10.1016/j.scitotenv.2012.04.028.
(54) Verlicchi, P.; Zambello, E. Pharmaceuticals and Personal Care Products in Untreated and Treated Sewage Sludge: Occurrence and Environmental Risk in the Case of Application on Soil — A Critical Review. *Science of The Total Environment* **2015**, *538*, 750–767. https://doi.org/10.1016/j.scitotenv.2015.08.108.
(55) Time to Get Clean. *Nature* **2015**, *526* (7572), 164–164. https://doi.org/10.1038/526164a.
(56) Oaks, J. L.; Gilbert, M.; Virani, M. Z.; Watson, R. T.; Meteyer, C. U.; Rideout, B. A.; Shivaprasad, H. L.; Ahmed, S.; Iqbal Chaudhry, M. J.; Arshad, M.; Mahmood, S.; Ali, A.; Ahmed Khan, A. Diclofenac Residues as the Cause of Vulture Population Decline in Pakistan. *Nature* **2004**, *427* (6975), 630–633. https://doi.org/10.1038/nature02317.
(57) Harris, C. A.; Hamilton, P. B.; Runnalls, T. J.; Vinciotti, V.; Henshaw, A.; Hodgson, D.; Coe, T. S.; Jobling, S.; Tyler, C. R.; Sumpter, J. P. The Consequences of Feminization in Breeding Groups of Wild Fish. *Environmental Health Perspectives* **2011**, *119* (3), 306–311. https://doi.org/10.1289/ehp.1002555.
(58) Lange, A.; Katsu, Y.; Ichikawa, R.; Paull, G. C.; Chidgey, L. L.; Coe, T. S.; Iguchi, T.; Tyler, C. R. Altered Sexual Development in Roach (Rutilus Rutilus) Exposed to Environmental Concentrations of the Pharmaceutical 17α-Ethinylestradiol and Associated Expression Dynamics of Aromatases and Estrogen Receptors. *Toxicological Sciences* **2008**, *106* (1), 113–123. https://doi.org/10.1093/toxsci/kfn151.
(59) Länge, R.; Hutchinson, T. H.; Croudace, C. P.; Siegmund, F.; Schweinfurth, H.; Hampe, P.; Panter, G. H.; Sumpter, J. P. Effects of the Synthetic Estrogen 17α-Ethinylestradiol on the Life-Cycle of the Fathead Minnow (Pimephales Promelas). *Environ Toxicol Chem* **2001**, *20* (6), 1216–1227. https://doi.org/10.1002/etc.5620200610.
(60) Hahlbeck, E.; Griffiths, R.; Bengtsson, B. The Juvenile Three-Spined Stickleback (Gasterosteus Aculeatus L.) as a Model Organism for Endocrine DisruptionI. Sexual

Differentiation. *Aquatic Toxicology* **2004**, *70* (4), 287–310. https://doi.org/10.1016/S0166-445X(04)00265-6.
(61) Miller, T. H.; Bury, N. R.; Owen, S. F.; MacRae, J. I.; Barron, L. P. A Review of the Pharmaceutical Exposome in Aquatic Fauna. *Environmental Pollution* **2018**, *239*, 129–146. https://doi.org/10.1016/j.envpol.2018.04.012.
(62) Hashim, N. H. Enantioselective Analysis of Ibuprofen, Ketoprofen and Naproxen in Wastewater and Environmental Water Samples. *J. Chromatogr. A* 9.
(63) Scheurell, M.; Franke, S.; Shah, R. M.; Hühnerfuss, H. Occurrence of Diclofenac and Its Metabolites in Surface Water and Effluent Samples from Karachi, Pakistan. **2009**, 7.
(64) Hanif, H.; Waseem, A.; Kali, S.; Qureshi, N. A.; Majid, M.; Iqbal, M.; Ur-Rehman, T.; Tahir, M.; Yousaf, S.; Iqbal, M. M.; Khan, I. A.; Zafar, M. I. Environmental Risk Assessment of Diclofenac Residues in Surface Waters and Wastewater: A Hidden Global Threat to Aquatic Ecosystem. 1.
(65) Deng, A.; Himmelsbach, M.; Zhu, Q.-Z.; Sengl, M.; Buchberger, W.; Niessner, R.; Knopp, D. Residue Analysis of the Pharmaceutical Diclofenac in Different Water Types Using ELISA and GC-MS. 8.
(66) Buser, H.-R.; Poiger, T.; Müller, M. D. Occurrence and Environmental Behavior of the Chiral Pharmaceutical Drug Ibuprofen in Surface Waters and in Wastewater. *Environ. Sci. Technol.* **1999**, *33* (15), 2529–2535. https://doi.org/10.1021/es981014w.
(67) Contardo-Jara, V.; Lorenz, C.; Pflugmacher, S.; Nützmann, G.; Kloas, W.; Wiegand, C. Exposure to Human Pharmaceuticals Carbamazepine, Ibuprofen and Bezafibrate Causes Molecular Effects in Dreissena Polymorpha. *Aquatic Toxicology* **2011**, *105* (3–4), 428–437. https://doi.org/10.1016/j.aquatox.2011.07.017.
(68) Jakimska, A.; Namie, J. Elucidation of Transformation Pathway of Ketoprofen, Ibuprofen, and Furosemide in Surface Water and Their Occurrence in the Aqueous Environment Using. 14.
(69) López Zavala, M. Á.; Vega, D. A.; Álvarez Vega, J. M.; Castillo Jerez, O. F.; Cantú Hernández, R. A. Electrochemical Oxidation of Acetaminophen and Its Transformation Products in Surface Water: Effect of PH and Current Density. *Heliyon* **2020**, *6* (2), e03394. https://doi.org/10.1016/j.heliyon.2020.e03394.
(70) Phong Vo, H. N.; Le, G. K.; Hong Nguyen, T. M.; Bui, X.-T.; Nguyen, K. H.; Rene, E. R.; Vo, T. D. H.; Thanh Cao, N.-D.; Mohan, R. Acetaminophen Micropollutant: Historical and Current Occurrences, Toxicity, Removal Strategies and Transformation Pathways in Different Environments. *Chemosphere* **2019**, *236*, 124391. https://doi.org/10.1016/j.chemosphere.2019.124391.
(71) Zhou, C.; Zhou, Q.; Zhang, X. Transformation of Acetaminophen in Natural Surface Water and the Change of Aquatic Microbes. *Water Research* **2019**, *148*, 133–141. https://doi.org/10.1016/j.watres.2018.10.042.
(72) Wilkinson, J.; Hooda, P. S.; Barker, J.; Barton, S.; Swinden, J. Occurrence, Fate and Transformation of Emerging Contaminants in Water: An Overarching Review of the Field. *Environmental Pollution* **2017**, *231*, 954–970. https://doi.org/10.1016/j.envpol.2017.08.032.
(73) Zhou, S.; Di Paolo, C.; Wu, X.; Shao, Y.; Seiler, T.-B.; Hollert, H. Optimization of Screening-Level Risk Assessment and Priority Selection of Emerging Pollutants – The Case of

Pharmaceuticals in European Surface Waters. *Environment International* **2019**, *128*, 1–10. https://doi.org/10.1016/j.envint.2019.04.034.
(74) Choi, K.; Kim, Y.; Park, J.; Park, C. K.; Kim, M.; Kim, H. S.; Kim, P. Seasonal Variations of Several Pharmaceutical Residues in Surface Water and Sewage Treatment Plants of Han River, Korea. *Science of The Total Environment* **2008**, *405* (1–3), 120–128. https://doi.org/10.1016/j.scitotenv.2008.06.038.
(75) Vanderford, B. J.; Snyder, S. A. Analysis of Pharmaceuticals in Water by Isotope Dilution Liquid Chromatography/Tandem Mass Spectrometry. *Environ. Sci. Technol.* **2006**, *40* (23), 7312–7320. https://doi.org/10.1021/es0613198.
(76) López-Serna, R.; Jurado, A.; Vázquez-Suñé, E.; Carrera, J.; Petrović, M.; Barceló, D. Occurrence of 95 Pharmaceuticals and Transformation Products in Urban Groundwaters Underlying the Metropolis of Barcelona, Spain. *Environmental Pollution* **2013**, *174*, 305–315. https://doi.org/10.1016/j.envpol.2012.11.022.
(77) Desbiolles, F.; Malleret, L.; Tiliacos, C.; Wong-Wah-Chung, P.; Laffont-Schwob, I. Occurrence and Ecotoxicological Assessment of Pharmaceuticals: Is There a Risk for the Mediterranean Aquatic Environment? *Science of The Total Environment* **2018**, *639*, 1334–1348. https://doi.org/10.1016/j.scitotenv.2018.04.351.
(78) Al-Habsi, A. A.; Massarsky, A.; Moon, T. W. Exposure to Gemfibrozil and Atorvastatin Affects Cholesterol Metabolism and Steroid Production in Zebrafish (Danio Rerio). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **2016**, *199*, 87–96. https://doi.org/10.1016/j.cbpb.2015.11.009.
(79) Brain, R. A.; Reitsma, T. S.; Lissemore, L. I.; Bestari, K. (Jim); Sibley, P. K.; Solomon, K. R. Herbicidal Effects of Statin Pharmaceuticals in *Lemna Gibba*. *Environ. Sci. Technol.* **2006**, *40* (16), 5116–5123. https://doi.org/10.1021/es0600274.
(80) Lee, H.-B.; Peart, T. E.; Lewina Svoboda, M.; Backus, S. Occurrence and Fate of Rosuvastatin, Rosuvastatin Lactone, and Atorvastatin in Canadian Sewage and Surface Water Samples. *Chemosphere* **2009**, *77* (10), 1285–1291. https://doi.org/10.1016/j.chemosphere.2009.09.068.
(81) Santos, M. M.; Ruivo, R.; Lopes-Marques, M.; Torres, T.; de los Santos, C. B.; Castro, L. F. C.; Neuparth, T. Statins: An Undesirable Class of Aquatic Contaminants? *Aquatic Toxicology* **2016**, *174*, 1–9. https://doi.org/10.1016/j.aquatox.2016.02.001.
(82) Sulaiman, S.; Khamis, M.; Nir, S.; Lelario, F.; Scrano, L.; Bufo, S. A.; Mecca, G.; Karaman, R. Stability and Removal of Atorvastatin, Rosuvastatin and Simvastatin from Wastewater. *Environmental Technology* **2015**, *36* (24), 3232–3242. https://doi.org/10.1080/09593330.2015.1058422.
(83) Canesi, L.; Lorusso, L. C.; Ciacci, C.; Betti, M.; Regoli, F.; Poiana, G.; Gallo, G.; Marcomini, A. Effects of Blood Lipid Lowering Pharmaceuticals (Bezafibrate and Gemfibrozil) on Immune and Digestive Gland Functions of the Bivalve Mollusc, Mytilus Galloprovincialis. *Chemosphere* **2007**, *69* (6), 994–1002. https://doi.org/10.1016/j.chemosphere.2007.04.085.
(84) Sui, Q.; Yan, P.; Cao, X.; Lu, S.; Zhao, W.; Chen, M. Biodegradation of Bezafibrate by the Activated Sludge under Aerobic Condition: Effect of Initial Concentration, Temperature and PH. *Emerging Contaminants* **2016**, *2* (4), 173–177. https://doi.org/10.1016/j.emcon.2016.09.001.

(85) Oliveira, M.; Franco, L.; Balasch, J. C.; Fierro-Castro, C.; Tvarijonaviciute, A.; Soares, A. M. V. M.; Tort, L.; Teles, M. Tools to Assess Effects of Human Pharmaceuticals in Fish: A Case Study with Gemfibrozil. *Ecological Indicators* **2018**, *95*, 1100–1107. https://doi.org/10.1016/j.ecolind.2017.12.051.
(86) Garcia-Ac, A.; Segura, P. A.; Gagnon, C.; Sauvé, S. Determination of Bezafibrate, Methotrexate, Cyclophosphamide, Orlistat and Enalapril in Waste and Surface Waters Using on-Line Solid-Phase Extraction Liquid Chromatography Coupled to Polarity-Switching Electrospray Tandem Mass Spectrometry. *J. Environ. Monit.* **2009**, *11* (4), 830. https://doi.org/10.1039/b817570e.
(87) Gallardo-Altamirano, M. J.; Maza-Márquez, P.; Pérez, S.; Rodelas, B.; Pozo, C.; Osorio, F. Fate of Pharmaceutically Active Compounds in a Pilot-Scale A2O Integrated Fixed-Film Activated Sludge (IFAS) Process Treating Municipal Wastewater. *Journal of Environmental Chemical Engineering* **2021**, *9* (4), 105398. https://doi.org/10.1016/j.jece.2021.105398.
(88) Hernández, F.; Ibáñez, M.; Bade, R.; Bijlsma, L.; Sancho, J. V. Investigation of Pharmaceuticals and Illicit Drugs in Waters by Liquid Chromatography-High-Resolution Mass Spectrometry. *TrAC Trends in Analytical Chemistry* **2014**, *63*, 140–157. https://doi.org/10.1016/j.trac.2014.08.003.
(89) Carmona, E.; Andreu, V.; Picó, Y. Occurrence of Acidic Pharmaceuticals and Personal Care Products in Turia River Basin: From Waste to Drinking Water. *Science of The Total Environment* **2014**, *484*, 53–63. https://doi.org/10.1016/j.scitotenv.2014.02.085.
(90) 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. *Anal Bioanal Chem* **2021**, *413* (21), 5291–5300. https://doi.org/10.1007/s00216-021-03500-7.
(91) Wolf, L.; Zwiener, C.; Zemann, M. Tracking Artificial Sweeteners and Pharmaceuticals Introduced into Urban Groundwater by Leaking Sewer Networks. *Science of The Total Environment* **2012**, *430*, 8–19. https://doi.org/10.1016/j.scitotenv.2012.04.059.
(92) Loos, R.; Tavazzi, S.; Mariani, G.; Suurkuusk, G.; Paracchini, B.; Umlauf, G. Analysis of Emerging Organic Contaminants in Water, Fish and Suspended Particulate Matter (SPM) in the Joint Danube Survey Using Solid-Phase Extraction Followed by UHPLC-MS-MS and GC–MS Analysis. *Science of The Total Environment* **2017**, *607–608*, 1201–1212. https://doi.org/10.1016/j.scitotenv.2017.07.039.
(93) Zuccato, E.; Castiglioni, S.; Fanelli, R. Identification of the Pharmaceuticals for Human Use Contaminating the Italian Aquatic Environment. *Journal of Hazardous Materials* **2005**, *122* (3), 205–209. https://doi.org/10.1016/j.jhazmat.2005.03.001.
(94) Rúa-Gómez, P. C.; Püttmann, W. Degradation of Lidocaine, Tramadol, Venlafaxine and the Metabolites O-Desmethyltramadol and O-Desmethylvenlafaxine in Surface Waters. *Chemosphere* **2013**, *90* (6), 1952–1959. https://doi.org/10.1016/j.chemosphere.2012.10.039.
(95) Lin, W.; Huang, Z.; Gao, S.; Luo, Z.; An, W.; Li, P.; Ping, S.; Ren, Y. Evaluating the Stability of Prescription Drugs in Municipal Wastewater and Sewers Based on Wastewater-Based Epidemiology. *Science of The Total Environment* **2021**, *754*, 142414. https://doi.org/10.1016/j.scitotenv.2020.142414.

(96) Ben, Y.; Fu, C.; Hu, M.; Liu, L.; Wong, M. H.; Zheng, C. Human Health Risk Assessment of Antibiotic Resistance Associated with Antibiotic Residues in the Environment: A Review. *Environmental Research* **2019**, *169*, 483–493. https://doi.org/10.1016/j.envres.2018.11.040.
(97) Kovalakova, P.; Cizmas, L.; McDonald, T. J.; Marsalek, B.; Feng, M.; Sharma, V. K. Occurrence and Toxicity of Antibiotics in the Aquatic Environment: A Review. *Chemosphere* **2020**, *251*, 126351. https://doi.org/10.1016/j.chemosphere.2020.126351.
(98) Carson, R. *Silent Spring*; Houghton Mifflin Harcourt, 2002.
(99) Krebs, J. R.; Wilson, J. D.; Bradbury, R. B.; Siriwardena, G. M. The Second Silent Spring? *Nature* **1999**, *400* (6745), 611–612. https://doi.org/10.1038/23127.
(100) de Albuquerque, F. P.; de Oliveira, J. L.; Moschini-Carlos, V.; Fraceto, L. F. An Overview of the Potential Impacts of Atrazine in Aquatic Environments: Perspectives for Tailored Solutions Based on Nanotechnology. *Science of The Total Environment* **2020**, *700*, 134868. https://doi.org/10.1016/j.scitotenv.2019.134868.
(101) Van Bruggen, A. H. C.; He, M. M.; Shin, K.; Mai, V.; Jeong, K. C.; Finckh, M. R.; Morris, J. G. Environmental and Health Effects of the Herbicide Glyphosate. *Science of The Total Environment* **2018**, *616–617*, 255–268. https://doi.org/10.1016/j.scitotenv.2017.10.309.
(102) Gutowski, L.; Olsson, O.; Leder, C.; Kümmerer, K. A Comparative Assessment of the Transformation Products of S-Metolachlor and Its Commercial Product Mercantor Gold® and Their Fate in the Aquatic Environment by Employing a Combination of Experimental and in Silico Methods. *Science of The Total Environment* **2015**, *506–507*, 369–379. https://doi.org/10.1016/j.scitotenv.2014.11.025.
(103) Fenner, K.; Canonica, S.; Wackett, L. P.; Elsner, M. Evaluating Pesticide Degradation in the Environment: Blind Spots and Emerging Opportunities. *Science* **2013**, *341* (6147), 752–758. https://doi.org/10.1126/science.1236281.
(104) Stara, A.; Kubec, J.; Zuskova, E.; Buric, M.; Faggio, C.; Kouba, A.; Velisek, J. Effects of S-Metolachlor and Its Degradation Product Metolachlor OA on Marbled Crayfish (Procambarus Virginalis). *Chemosphere* **2019**, *224*, 616–625. https://doi.org/10.1016/j.chemosphere.2019.02.187.
(105) Recommandations pour la qualité de l'eau potable au Canada - Tableau sommaire, 2020.
(106) Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M. *Environmental Organic Chemistry*; John Wiley & Sons, 2016.
(107) Rossini, D.; Ciofi, L.; Ancillotti, C.; Checchini, L.; Bruzzoniti, M. C.; Rivoira, L.; Fibbi, D.; Orlandini, S.; Del Bubba, M. Innovative Combination of QuEChERS Extraction with On-Line Solid-Phase Extract Purification and Pre-Concentration, Followed by Liquid Chromatography-Tandem Mass Spectrometry for the Determination of Non-Steroidal Anti-Inflammatory Drugs and Their Metabolites in Sewage Sludge. *Analytica Chimica Acta* **2016**, *935*, 269–281. https://doi.org/10.1016/j.aca.2016.06.023.
(108) Miao, X.-S.; Metcalfe, C. D. Determination of Carbamazepine and Its Metabolites in Aqueous Samples Using Liquid Chromatography−Electrospray Tandem Mass Spectrometry. *Anal. Chem.* **2003**, *75* (15), 3731–3738. https://doi.org/10.1021/ac030082k.
(109) Rubirola, A.; Llorca, M.; Rodriguez-Mozaz, S.; Casas, N.; Rodriguez-Roda, I.; Barceló, D.; Buttiglieri, G. Characterization of Metoprolol Biodegradation and Its Transformation

Products Generated in Activated Sludge Batch Experiments and in Full Scale WWTPs. *Water Research* **2014**, *63*, 21–32. https://doi.org/10.1016/j.watres.2014.05.031.
(110) SUGIHARA, J.; SUGAWARA, Y.; ANDO, H.; HARIGAYA, S.; ETOH, A.; KOHNO, K. Studies on the Metabolism of Diltiazem in Man. *Journal of pharmacobio-dynamics* **1984**, *7* (1), 24–32.
(111) Evgenidou, E. N.; Konstantinou, I. K.; Lambropoulou, D. A. Occurrence and Removal of Transformation Products of PPCPs and Illicit Drugs in Wastewaters: A Review. *Science of The Total Environment* **2015**, *505*, 905–926. https://doi.org/10.1016/j.scitotenv.2014.10.021.
(112) Dong, Z.; Senn, D. B.; Moran, R. E.; Shine, J. P. Prioritizing Environmental Risk of Prescription Pharmaceuticals. *Regulatory Toxicology and Pharmacology* **2013**, *65* (1), 60–67. https://doi.org/10.1016/j.yrtph.2012.07.003.
(113) Boxall, A. B. A.; Rudd, M. A.; Brooks, B. W.; Caldwell, D. J.; Choi, K.; Hickmann, S.; Innes, E.; Ostapyk, K.; Staveley, J. P.; Verslycke, T.; Ankley, G. T.; Beazley, K. F.; Belanger, S. E.; Berninger, J. P.; Carriquiriborde, P.; Coors, A.; DeLeo, P. C.; Dyer, S. D.; Ericson, J. F.; Gagné, F.; Giesy, J. P.; Gouin, T.; Hallstrom, L.; Karlsson, M. V.; Larsson, D. G. J.; Lazorchak, J. M.; Mastrocco, F.; McLaughlin, A.; McMaster, M. E.; Meyerhoff, R. D.; Moore, R.; Parrott, J. L.; Snape, J. R.; Murray-Smith, R.; Servos, M. R.; Sibley, P. K.; Straub, J. O.; Szabo, N. D.; Topp, E.; Tetreault, G. R.; Trudeau, V. L.; Van Der Kraak, G. Pharmaceuticals and Personal Care Products in the Environment: What Are the Big Questions? *Environmental Health Perspectives* **2012**, *120* (9), 1221–1229. https://doi.org/10.1289/ehp.1104477.
(114) Guengerich, F. P. Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* **2001**, *14* (6), 611–650. https://doi.org/10.1021/tx0002583.
(115) Liston, H. L.; Markowitz, J. S.; DeVane, C. L. Drug Glucuronidation in Clinical Psychopharmacology: *Journal of Clinical Psychopharmacology* **2001**, *21* (5), 500–515. https://doi.org/10.1097/00004714-200110000-00008.
(116) Commandeur, J. N. Enzymes and Transport Systems Involved in the Formation and Disposition of Glutathione S-Conjugates. Role in Bioactivation and Detoxication Mechanisms of Xenobiotics. *Pharmacol Rev.* **1995**, *47*, 271–330.
(117) Brown, A. K.; Wong, C. S. Distribution and Fate of Pharmaceuticals and Their Metabolite Conjugates in a Municipal Wastewater Treatment Plant. *Water Research* **2018**, *144*, 774–783. https://doi.org/10.1016/j.watres.2018.08.034.
(118) Escher, B. I.; Fenner, K. Recent Advances in Environmental Risk Assessment of Transformation Products. *Environ. Sci. Technol.* **2011**, *45* (9), 3835–3847. https://doi.org/10.1021/es1030799.
(119) Lin, A. Y.-C.; Wang, X.-H.; Lee, W.-N. Phototransformation Determines the Fate of 5-Fluorouracil and Cyclophosphamide in Natural Surface Waters. *Environ. Sci. Technol.* **2013**, *47* (9), 4104–4112. https://doi.org/10.1021/es304976q.
(120) Wang, M.; Li, J.; Shi, H.; Miao, D.; Yang, Y.; Qian, L.; Gao, S. Photolysis of Atorvastatin in Aquatic Environment: Influencing Factors, Products, and Pathways. *Chemosphere* **2018**, *212*, 467–475. https://doi.org/10.1016/j.chemosphere.2018.08.086.
(121) Gross, J. H. *Mass Spectrometry: A Textbook*; Springer Science & Business Media, 2006.

(122) Guo, Z.; Huang, S.; Wang, J.; Feng, Y.-L. Recent Advances in Non-Targeted Screening Analysis Using Liquid Chromatography - High Resolution Mass Spectrometry to Explore New Biomarkers for Human Exposure. *Talanta* **2020**, *219*, 121339. https://doi.org/10.1016/j.talanta.2020.121339.

(123) Agier, L.; Basagaña, X.; Maitre, L.; Granum, B.; Bird, P. K.; Casas, M.; Oftedal, B.; Wright, J.; Andrusaityte, S.; de Castro, M.; Cequier, E.; Chatzi, L.; Donaire-Gonzalez, D.; Grazuleviciene, R.; Haug, L. S.; Sakhi, A. K.; Leventakou, V.; McEachan, R.; Nieuwenhuijsen, M.; Petraviciene, I.; Robinson, O.; Roumeliotaki, T.; Sunyer, J.; Tamayo-Uria, I.; Thomsen, C.; Urquiza, J.; Valentin, A.; Slama, R.; Vrijheid, M.; Siroux, V. Early-Life Exposome and Lung Function in Children in Europe: An Analysis of Data from the Longitudinal, Population-Based HELIX Cohort. *The Lancet Planetary Health* **2019**, *3* (2), e81–e92. https://doi.org/10.1016/S2542-5196(19)30010-5.

(124) Ccanccapa-Cartagena, A.; Pico, Y.; Ortiz, X.; Reiner, E. J. Suspect, Non-Target and Target Screening of Emerging Pollutants Using Data Independent Acquisition: Assessment of a Mediterranean River Basin. *Science of The Total Environment* **2019**, *687*, 355–368. https://doi.org/10.1016/j.scitotenv.2019.06.057.

(125) Bouslimani, A.; Melnik, A. V.; Xu, Z.; Amir, A.; da Silva, R. R.; Wang, M.; Bandeira, N.; Alexandrov, T.; Knight, R.; Dorrestein, P. C. Lifestyle Chemistries from Phones for Individual Profiling. *Proc Natl Acad Sci USA* **2016**, *113* (48), E7645–E7654. https://doi.org/10.1073/pnas.1610019113.

(126) Jiang, C.; Wang, X.; Li, X.; Inlora, J.; Wang, T.; Liu, Q.; Snyder, M. Dynamic Human Environmental Exposome Revealed by Longitudinal Personal Monitoring. *Cell* **2018**, *175* (1), 277-291.e31. https://doi.org/10.1016/j.cell.2018.08.060.

(127) Krauss, M.; Singer, H.; Hollender, J. LC–High Resolution MS in Environmental Analysis: From Target Screening to the Identification of Unknowns. *Anal Bioanal Chem* **2010**, *397* (3), 943–951. https://doi.org/10.1007/s00216-010-3608-9.

(128) Brack, W.; Dulio, V.; Slobodnik, J. The NORMAN Network and Its Activities on Emerging Environmental Substances with a Focus on Effect-Directed Analysis of Complex Environmental Contamination. *Environ Sci Eur* **2012**, *24* (1), 29. https://doi.org/10.1186/2190-4715-24-29.

(129) Williams, A. J.; Grulke, C. M.; Edwards, J.; McEachran, A. D.; Mansouri, K.; Baker, N. C.; Patlewicz, G.; Shah, I.; Wambaugh, J. F.; Judson, R. S.; Richard, A. M. The CompTox Chemistry Dashboard: A Community Data Resource for Environmental Chemistry. *J Cheminform* **2017**, *9* (1), 61. https://doi.org/10.1186/s13321-017-0247-6.

(130) McEachran, A. D.; Sobus, J. R.; Williams, A. J. Identifying Known Unknowns Using the US EPA's CompTox Chemistry Dashboard. *Anal Bioanal Chem* **2017**, *409* (7), 1729–1735. https://doi.org/10.1007/s00216-016-0139-z.

(131) Thurman, E. M.; Ferrer, I.; Blotevogel, J.; Borch, T. Analysis of Hydraulic Fracturing Flowback and Produced Waters Using Accurate Mass: Identification of Ethoxylated Surfactants. *Anal. Chem.* **2014**, *86* (19), 9653–9661. https://doi.org/10.1021/ac502163k.

(132) Baena-Nogueras, R. M.; González-Mazo, E.; Lara-Martín, P. A. Determination and Occurrence of Secondary Alkane Sulfonates (SAS) in Aquatic Environments. *Environmental Pollution* **2013**, *176*, 151–157. https://doi.org/10.1016/j.envpol.2013.01.020.

(133) Baena-Nogueras, R. M.; Rojas-Ojeda, P.; Sanz, J. L.; González-Mazo, E.; Lara-Martín, P. A. Reactivity and Fate of Secondary Alkane Sulfonates (SAS) in Marine Sediments. *Environmental Pollution* **2014**, *189*, 35–42. https://doi.org/10.1016/j.envpol.2014.02.019.
(134) Singer, H. P.; Wössner, A. E.; McArdell, C. S.; Fenner, K. Rapid Screening for Exposure to "Non-Target" Pharmaceuticals from Wastewater Effluents by Combining HRMS-Based Suspect Screening and Exposure Modeling. *Environ. Sci. Technol.* **2016**, *50* (13), 6698–6707. https://doi.org/10.1021/acs.est.5b03332.
(135) Kosma, C. I.; Lambropoulou, D. A.; Albanis, T. A. Photochemical Transformation and Wastewater Fate and Occurrence of Omeprazole: HRMS for Elucidation of Transformation Products and Target and Suspect Screening Analysis in Wastewaters. *Science of The Total Environment* **2017**, *590–591*, 592–601. https://doi.org/10.1016/j.scitotenv.2017.02.233.
(136) Picardo, M.; Núñez, O.; Farré, M. A Data Independent Acquisition All Ion Fragmentation Mode Tool for the Suspect Screening of Natural Toxins in Surface Water. *MethodsX* **2021**, *8*, 101286. https://doi.org/10.1016/j.mex.2021.101286.
(137) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag Relaunched: Incorporating Strategies beyond in Silico Fragmentation. *J Cheminform* **2016**, *8* (1), 3. https://doi.org/10.1186/s13321-016-0115-9.
(138) Sweeney, D. L. A Data Structure for Rapid Mass Spectral Searching. *Mass Spectrometry* **2014**, *3* (Special_Issue_2), S0035–S0035. https://doi.org/10.5702/massspectrometry.S0035.
(139) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(140) 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* (17), 3196–3198. https://doi.org/10.1093/bioinformatics/btz005.
(141) Quinn, R. A.; Nothias, L.-F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P. C. Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends in Pharmacological Sciences* **2017**, *38* (2), 143–154. https://doi.org/10.1016/j.tips.2016.10.011.
(142) Napoletani, D.; Panza, M.; Struppa, D. C. Agnostic Science. Towards a Philosophy of Data Analysis. *Found Sci* **2011**, *16* (1), 1–20. https://doi.org/10.1007/s10699-010-9186-7.
(143) Dennis, K. K.; Marder, E.; Balshaw, D. M.; Cui, Y.; Lynes, M. A.; Patti, G. J.; Rappaport, S. M.; Shaughnessy, D. T.; Vrijheid, M.; Barr, D. B. Biomonitoring in the Era of the Exposome. *Environmental Health Perspectives* **2017**, *125* (4), 502–510. https://doi.org/10.1289/EHP474.
(144) Hemmer, S.; Manier, S. K.; Fischmann, S.; Westphal, F.; Wagmann, L.; Meyer, M. R. Comparison of Three Untargeted Data Processing Workflows for Evaluating LC-HRMS Metabolomics Data. *Metabolites* **2020**, *10* (9), 378. https://doi.org/10.3390/metabo10090378.
(145) Zahn, D.; Mucha, P.; Zilles, V.; Touffet, A.; Gallard, H.; Knepper, T. P.; 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. https://doi.org/10.1016/j.watres.2018.11.042.
(146) Wolf, S.; Schmidt, S.; Müller-Hannemann, M.; Neumann, S. In Silico Fragmentation for Computer Assisted Identification of Metabolite Mass Spectra. *BMC Bioinformatics* **2010**, *11* (1), 148. https://doi.org/10.1186/1471-2105-11-148.
(147) Hill, D. W.; Kertesz, T. M.; Fontaine, D.; Friedman, R.; Grant, D. F. *Mass Spectral Metabonomics beyond Elemental Formula: Chemical Database Querying by Matching Experimental with Computational Fragmentation Spectra*. ACS Publications. https://pubs.acs.org/doi/pdf/10.1021/ac800548g (accessed 2022-04-06). https://doi.org/10.1021/ac800548g.
(148) Frank, A. M.; Bandeira, N.; Shen, Z.; Tanner, S.; Briggs, S. P.; Smith, R. D.; Pevzner, P. A. Clustering Millions of Tandem Mass Spectra. *J. Proteome Res.* **2008**, *7* (1), 113–122. https://doi.org/10.1021/pr070361e.
(149) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. Mass Spectral Molecular Networking of Living Microbial Colonies. *Proceedings of the National Academy of Sciences* **2012**, *109* (26), E1743–E1752. https://doi.org/10.1073/pnas.1203689109.
(150) Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo-Rodriguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kameník, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf, A.; Le Gouellec, A.; Ludwig, M.; Martin H., C.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.;

Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nat Methods* **2020**, *17* (9), 905–908. https://doi.org/10.1038/s41592-020-0933-6.

(151) Smoot, M. E.; Ono, K.; Ruscheinski, J.; Wang, P.-L.; Ideker, T. Cytoscape 2.8: New Features for Data Integration and Network Visualization. *Bioinformatics* **2011**, *27* (3), 431–432. https://doi.org/10.1093/bioinformatics/btq675.

(152) Su, G.; Morris, J. H.; Demchak, B.; Bader, G. D. Biological Network Exploration with Cytoscape 3. *Current Protocols in Bioinformatics* **2014**, *47* (1). https://doi.org/10.1002/0471250953.bi0813s47.

(153) Depke, T.; Franke, R.; Brönstrup, M. Clustering of MS2 Spectra Using Unsupervised Methods to Aid the Identification of Secondary Metabolites from Pseudomonas Aeruginosa. *Journal of Chromatography B* **2017**, *1071*, 19–28. https://doi.org/10.1016/j.jchromb.2017.06.002.

(154) Hao, C.; Helm, P. A.; Morse, D.; Reiner, E. J. Liquid Chromatography-Tandem Mass Spectrometry Direct Injection Analysis of Organophosphorus Flame Retardants in Ontario Surface Water and Wastewater Effluent. *Chemosphere* **2018**, *191*, 288–295. https://doi.org/10.1016/j.chemosphere.2017.10.060.

(155) 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* (11), 27–33. https://doi.org/10.2166/wst.2006.334.