



HAL
open science

Caractérisation des eaux résiduaires urbaines par spectrométrie de masse haute résolution : influence de la stratégie analytique, limitations et perspectives

Tinh Nghi Nina Huynh

► To cite this version:

Tinh Nghi Nina Huynh. Caractérisation des eaux résiduaires urbaines par spectrométrie de masse haute résolution : influence de la stratégie analytique, limitations et perspectives. Ingénierie de l'environnement. Université Paris-Est Créteil Val-de-Marne - Paris 12, 2022. Français. NNT : 2022PA120026 . tel-04191790

HAL Id: tel-04191790

<https://theses.hal.science/tel-04191790v1>

Submitted on 30 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT
UNIVERSITÉ PARIS-EST CRÉTEIL

Spécialité Sciences et Techniques de l'Environnement

**Caractérisation des eaux résiduaires urbaines par
spectrométrie de masse haute résolution :
influence de la stratégie analytique, limitations
et perspectives**

Laboratoire Eau Environnement et Systèmes Urbains

présentée par

Tinh Nghi Nina HUYNH

Thèse défendue le 8 avril 2022 à l'Université Paris-Est Créteil devant le jury
composé de :

Serge CHIRON	Directeur de recherche, Université de Montpellier	Rapporteur Président du jury
Emmanuelle VULLIET	Directrice de recherche, Université de Lyon	Rapportrice
Christelle MARGOUM	Ingénieure de recherche, INRAE-UR Riverly	Examinatrice
Valeria DULIO	Chargée de mission, INERIS	Invitée
Hervé GALLARD	Professeur, Université de Poitiers	Invité
Régis MOILLERON	Professeur, Université Paris-Est Créteil	Directeur de thèse
Julien LE ROUX	Maître de conférences, Université Paris-Est Créteil	Co-encadrant

Résumé

De nombreux micropolluants sont présents dans les eaux urbaines mais leur comportement et leurs produits de transformations dans les stations de traitement des eaux usées et dans l'environnement sont encore peu connus. Les méthodes d'analyse ciblées conventionnelles utilisées pour l'analyse de ces substances permettent de les détecter et de les quantifier à l'état de traces dans différentes matrices environnementales. Cependant, ces techniques s'appuient sur l'utilisation d'étalons de référence qui ne sont pas disponibles pour tous les composés, encore moins pour les produits de transformation qui sont souvent mal caractérisés ou tout simplement inconnus. Avec l'émergence de la spectrométrie de masse haute résolution (HRMS), de nouvelles stratégies d'analyses dites en mode suspect ou non-ciblées se sont développées. Elles permettent la détection d'un grand nombre de signaux et contribuent ainsi à l'analyse de composés connus ou pas. Cependant, ces nouvelles méthodes d'analyse sont longues et complexes et beaucoup d'études se concentrent sur un nombre limité des signaux prioritaires. Ce travail a eu pour objectif de caractériser les méthodes analytiques disponibles pour ce type d'analyse, de la préparation d'échantillon à l'analyse des données en passant par la séparation chromatographique, avec des approches globales, afin de "traquer" les polluants organiques dans les eaux urbaines.

La première partie de ce travail s'est concentrée sur la comparaison de méthodes de traitement de données en analyse non-ciblée. Différents logiciels ont été utilisés et comparés afin de tirer le maximum d'informations des données acquises par HRMS. Leur capacité à identifier précisément des composés a été évaluée. Cette stratégie d'exploitation des données étant longue, des méthodes de caractérisation plus globales ont ensuite été mises en place afin d'obtenir des résultats plus rapides à partir de données acquises par HRMS, sans diminuer la précision des informations obtenues. Ce travail a ainsi permis la mise en place d'un *workflow* permettant de traiter les données contenant des informations de mobilité ionique qui ne sont pas encore prise en charge par la majorité des algorithmes existant actuellement.

La seconde partie a consisté à évaluer l'influence des méthodes, à la fois de préparation d'échantillon et d'analyse, de micropolluants organiques dans les eaux urbaines. Différents protocoles ont été appliqués à un même échantillon afin de comparer le type de composés retenus et détectés par les méthodes employées. Celles-ci ont été comparées en termes de

richesse d'information et de gammes de propriétés physico-chimiques couvertes. De nombreuses différences ont été mises en évidence en fonction de la stratégie appliquée, montrant qu'il est impossible de pouvoir caractériser un échantillon dans son ensemble en utilisant un protocole unique. Notre souhait était de caractériser au mieux les méthodes employées afin d'apporter des informations sur le type de composés qui sont potentiellement analysables par chacune d'elles et donc que les utilisateurs le fassent en connaissance de cause.

Ces approches ont été testées à travers l'évaluation des performances de procédés de traitement avancés pour l'élimination de micropolluants organiques dans les filières de traitement des eaux usées. Différents procédés d'oxydation (ozonation, photolyse UV, chloration, peracides) ont été étudiés afin d'évaluer leur influence sur les mécanismes d'élimination des micropolluants organiques, notamment en termes de produits de transformation générés. La valeur ajoutée des stratégies d'analyse en mode suspect et en non-ciblé a été mise en évidence. En particulier, le suivi de l'évolution globale des échantillons (en lien avec le nombre de signaux observés) au cours du processus de traitement a montré que l'on pouvait aboutir à des conclusions parfois contradictoires par rapport à celles obtenues lors de l'analyse ciblée de composés connus ou réglementés.

Mots-clés : analyse non-ciblée, eaux résiduaires urbaines, micropolluants organiques, oxydation, spectrométrie de masse haute résolution

Abstract

Many emerging micropollutants (such as personal care products, antibiotics, endocrine disruptors, etc.) are present in urban waters, but little is known about their behaviour and transformation products in wastewater treatment plants and in the environment. Conventional targeted analytical methods used for the analysis of these substances allow them to be detected and quantified at trace levels in different environmental matrices. However, these techniques rely on the use of reference standards that are not available for all compounds, let alone for transformation products that are often poorly characterized or simply unknown. With the emergence of high-resolution mass spectrometry (HRMS), new analysis strategies (called suspect or non-targeted approaches) have been developed in recent years. They allow the detection of a large number of signals and thus contribute to the analysis of known or unknown compounds. However, these new analytical methods are time consuming and complex and many studies focus on a limited number of prioritized signals. The objective of this work was to characterize the analytical methods available for this type of analysis, from sample preparation to data analysis through chromatographic separation, with global approaches, in order to "track" organic pollutants in urban waters by HRMS.

Therefore, the first part of our work focused on the comparison of data processing methods in non-targeted analysis. Different softwares have been used and compared in order to get the maximum information from the data acquired by HRMS. Firstly, their capacity to precisely identify compounds was evaluated. As this data exploitation strategy is time consuming, more global characterization methods were then implemented in order to obtain faster results from data acquired by HRMS, without reducing the accuracy of the information obtained. This work has allowed the implementation of a workflow to process data containing ion mobility information that are not yet supported by the majority of existing algorithms.

The second part of this work consisted in evaluating the influence of methods, both of sample preparation and analysis, of organic micropollutants in urban waters. Different protocols were applied to the same sample in order to compare the type of compounds retained and detected by the methods used. These methods were compared in terms of richness of information and range of physico-chemical properties covered. Numerous differences were highlighted depending on the strategy applied, showing that it is impossible to characterize a sample as a

whole using a single protocol. Our wish was to characterize as well as possible the methods used in order to bring information on the type of compounds which are potentially analysable by each one of them and thus that the users do it in knowledge of the facts.

These approaches were then tested through the evaluation of the performance of advanced treatment processes for the removal of organic micropollutants in wastewater treatment plants. Different oxidation processes (ozonation, UV photolysis, chlorination, use of peracids) were studied in order to evaluate their influence on the removal mechanisms of organic micropollutants, especially in terms of transformation products generated. The added value of suspect and non-targeted analysis strategies was highlighted. In particular, the monitoring of the overall evolution of the samples (in relation to the number of signals observed) during the treatment process showed that different conclusions could be reached than those obtained during the targeted analysis of known or regulated compounds.

Keywords: non-targeted analysis, urban wastewater, organic micropollutants, oxidation, high-resolution mass spectrometry

Remerciements

Ce travail a été réalisé au sein du Laboratoire Eau Environnement et Systèmes Urbains (LEESU, UMR MA 102) à l'Université Paris-Est Créteil, dans le cadre du projet WaterOmics (ANR-17-CE34-0009-01) et en association avec la phase 5 du programme OPUR. A l'issue de ces trois années et demie, je souhaite adresser ces quelques mots à toutes les personnes qui ont pu contribuer, de près ou de loin, à ce projet de recherche.

J'adresse tout d'abord mes sincères remerciements à mes encadrants pour la confiance qu'ils ont placée en moi. À mon directeur de thèse, Régis Moilleron, pour avoir toujours su trouver du temps à m'accorder malgré ses nombreuses responsabilités. Merci pour toutes les anecdotes, connaissances et leçons de vie que tu as partagées avec moi. À mon co-encadrant de thèse, Julien Le Roux, pour m'avoir accordé cette liberté de travail tout en gardant un œil sur ce que je faisais au quotidien. Merci pour ton implication et ton dynamisme, qui m'ont permis de pouvoir participer à tant de projets différents et d'apprendre tellement. Merci à tous les deux pour votre bienveillance et pour tous les échanges que nous avons pu avoir. J'ai appris énormément à vos côtés et travailler avec vous a été un réel plaisir. J'aimerais ensuite remercier Serge Chiron et Emmanuelle Vulliet d'avoir accepté de rapporter cette thèse, Christelle Margoum pour son rôle d'examinatrice lors du jury, et particulièrement Valeria Dulio et Hervé Gallard, invités du jury et membres de mon comité de suivi de thèse, pour leur disponibilité, leur bienveillance et leurs conseils.

Cette étude a été réalisée en partenariat avec le Syndicat Interdépartemental pour l'Assainissement de l'Agglomération Parisienne (SIAAP) qui a permis l'accès aux différents échantillons. À ce titre, j'aimerais remercier Romain Mailler, Mélissa Lopez Viveros, Marie Ponce de Léon et Perrine Mèche pour l'organisation des campagnes et leur réalisation à Seine Centre et Christophe Descamps et Mathieu Houmault pour les prélèvements à Seine Amont. Les analyses effectuées au cours de ce travail ont été rendues possibles grâce à la plateforme d'analyse Prammics de l'Osue-Fluve. Je remercie particulièrement Emmanuelle Mebold pour son aide précieuse, sa disponibilité et tous les échanges qu'on a pu avoir. Sans ton soutien technique, mais surtout moral, je n'aurais jamais pu affronter tous les caprices de l'HRMS. Merci également à Isabelle Schmitz-Afonso et Carlos Afonso du laboratoire COBRA (UMR CNRS 6014) avec qui nous avons collaboré pour les analyses complémentaires en UHRMS dans le cadre d'un projet Infranalytics. Je n'oublie évidemment pas la cellule technique chimie du

Leesu, avec Emilie, Lila et Chandirane. Merci à vous pour toutes les astuces et les conseils que vous m'avez donnés. Vous avez rendu mes journées au labo beaucoup plus agréables ! Merci également aux personnes qui ont contribué à la réalisation de ce projet : Jérémy et Caroline qui ont débroussaillé un peu le terrain avant mon arrivée, Ronan pour les premiers échantillons et mes stagiaires, Fatoumata et Lamyae, pour le travail qu'elles ont effectué.

Cette thèse n'aurait pas été la même sans tous les membres du Leesu qui font de ce laboratoire un endroit où il fait bon venir travailler. Merci à Laure (qui a été comme une seconde maman pendant cette thèse), Rachid, Angélique, Gilles, Adèle, Bertille, Françoise, Christophe, Claire, Romain, Maolida, My Dung, Julia, Sadia, Guilherme, Max et tous les autres que je ne peux pas citer. J'ai une pensée particulière pour les résidents, anciens comme nouveaux, du bureau 309 : Claudia, Soline, Robin, Christelle, Nadège et Nadia, avec qui nous avons partagé énormément de bons moments, mais avec qui nous sommes aussi soutenus dans les moins bons ! Mention spéciale pour mes deux compères : Maxime avec qui j'ai commencé cette grande aventure qu'est la thèse, et Fidji qui nous a rejoint seulement quelques mois plus tard. Merci à vous deux pour votre soutien quotidien, pour avoir toujours trouvé de supers idées pour occuper mes pauses et pour m'avoir motivée à retourner bosser quand le canapé ne voulait plus me laisser partir. Nous avons affronté de nombreux obstacles ensemble (le déménagement, la MSE, Aspergillus ou le COVID, pour ne citer que ceux-là), mais c'est aussi dans l'adversité nous avons lancé de grandes traditions (le café sciences, les soirées jeux, le mur de la honte ...) qui, je l'espère, nous survivrons.

J'aimerais également remercier toutes ces personnes qui m'ont sorti la tête du guidon pendant toute cette période. Je pense aux Primus (Gwenny, Léa, Marion, Rémi et Roxanne), aux grosses (Ambre, Cons' et Sophie), mais aussi à Mallaury, Sophie et Yasmina.

Pour finir, un énorme merci à ma famille. A mon frère et ma sœur pour m'avoir supportée et soutenue toutes ces années. A Romu, pour ta patience et ta compréhension. Merci d'avoir toujours cru en moi, même lorsque je n'y croyais plus, de m'avoir toujours poussée à dépasser mes limites et plus simplement, d'avoir toujours été là pour moi. Enfin, merci à mes parents sans qui, rien de cela n'aurait été possible. Merci d'avoir toujours tout fait pour que je puisse faire mes études, et en particulier cette thèse, dans les meilleures conditions possibles. Merci de m'avoir toujours écoutée, encouragée et poussée à aller au bout des choses.

Cette thèse, c'est la mienne, mais quelque part, c'est aussi un peu la vôtre... Alors, encore une fois, merci à tous !

Nina

Table des matières

Résumé	I
Abstract	III
Remerciements	V
Table des matières	VII
Liste des Figures	XII
Liste des Tableaux	XVI
Liste de abréviations	XX
Glossaire	XXIII
Liste des publications et communications	XXIV
INTRODUCTION GENERALE	1
I. Contexte	2
I.1. Les micropolluants organiques dans les eaux urbaines	2
I.2. L'analyse de micropolluants organiques dans des échantillons aqueux	5
I.3. L'analyse non-ciblée en analyse environnementale	10
II. Objectifs de la thèse et organisation du manuscrit	13
PARTIE A : EVALUATION DE METHODES DE TRAITEMENT DE DONNEES	16
I. Introduction et objectifs	17
II. État de l'art	19
II.1. Introduction	19
II.2. Méthodologie pour le traitement de données	24
II.2.1. <i>Workflow</i> général pour le pré-traitement	24
II.2.2. Stratégies de traitement des données	27
II.3. Difficultés d'identification	32
II.3.1. Stratégies pour l'identification	32
II.3.2. Calcul de formules brutes	33
	VII

II.3.3. Différents types de bases de données	42
II.3.3.1. Librairies internes	42
II.3.3.2. Fragmentation in-silico	43
II.3.3.3. Librairies “fournisseur”	44
II.3.3.4. Bases de données en ligne	44
II.4. Partage de données et plateformes communautaires	46
III. Matériels et méthodes	47
III.1. Présentation des outils	47
III.1.1. Acquisition des données	47
III.1.2. Logiciels de traitements : UNIFI et EZInfo	50
III.1.3. Interface de programmation de UNIFI et MSConvert	51
III.1.4. R et ses différents packages	51
III.2. Présentation des bases de données	52
III.2.1. Construction d’une librairie interne	52
III.2.2. Description des bases de données externes	54
III.2.2.1. Librairies Waters	54
III.2.2.2. Bases de données reliées à UNIFI	55
III.2.2.3. Bases de données externes importées	56
III.2.2.4. Bases de données en ligne	57
III.3. Acquisition des données	57
III.3.1. Contrôle qualité	57
III.3.2. Choix des critères d’identification en TS et SS	58
III.3.3. Création des <i>marker table</i> pour le NTS	59
III.3.3.1. Par UNIFI	59
III.3.3.2. Par patRoom	60
IV. Résultats	62
IV.1. Etude des données obtenues avec le logiciel fournisseur et optimisations	62
IV.1.1. Méthode employée et limites observées	62
IV.1.2. Optimisation de la méthode employée	65
IV.1.2.1. Identification des points faibles	65
IV.1.2.2. Optimisation du calcul des formules moléculaires	71
IV.1.2.3. Optimisation de la sélection des marqueurs	74
IV.2. Comparaison avec des méthodes développées sur d’autres outils	79
IV.2.1. Méthode hybride : logiciel fournisseur et libre	79
IV.2.1.1. Sélection par des méthodes statistiques	79
IV.2.1.2. Sélection par des méthodes de comparaison simples	81
IV.2.1.3. Comparaison des méthodes hybrides	82

IV.2.2. Méthode entièrement sous logiciels libres	85
IV.3. Comparaison des méthodes appliquées	89
IV.3.1. Avantages et inconvénients des différentes voies possibles	89
IV.3.1.1. Comparaison d'ordre pratique	90
IV.3.1.2. Comparaison en NTS	91
IV.3.1.3. Comparaison en SS	93
IV.3.2. Voies de développement	94
IV.4. Conclusion sur le développement de méthodes de traitement de données	99
PARTIE B : COMPARAISON DE METHODES ANALYTIQUES	101
I. Introduction et objectifs	102
II. Etat de l'art	103
II.1. Préparation d'échantillon	103
II.1.1. Injection directe	104
II.1.2. Extraction sur phase solide	105
II.1.3. Fractionnement d'échantillon	110
II.1.4. Études comparatives des méthodes de préparation	111
II.2. Analyse d'échantillon	112
II.2.1. HRMS couplée à la chromatographie liquide	114
II.2.1.1. Séparation	114
II.2.2. Analyse et détection	116
II.2.2. HRMS couplée à la chromatographie gazeuse	121
II.2.3. HRMS sans couplage chromatographique	122
III. Matériel et Méthodes	124
III.1. Matériaux utilisés	124
III.1.1. Etalons internes	124
III.1.2. Choix des méthodes et supports de fractionnement	124
III.1.3. Colonnes chromatographiques	126
III.2. Echantillonnage et réactifs	127
III.3. Prétraitement des échantillons	128
III.3.1. Extraction sur phase solide (SPE)	128
III.3.2. Fractionnement sur résines	129
III.3.3. Fractionnement par chromatographie d'exclusion stérique (SEC)	130
III.4. Procédures analytiques	131
III.4.1. Analyses par HRMS	131
III.4.2. Analyses par UHRMS	133

IV. Résultats	135
IV.1. Comparaison des méthodes chromatographiques	135
IV.2. Comparaison des méthodes de préparation	142
IV.2.1. Comparaison des techniques de préparation	142
IV.2.2. Comparaison de différents supports pour l'extraction sur phase solide	145
IV.3. Influence du type d'analyseurs	153
IV.4. Caractérisation de l'espace chimique couvert par ces méthodes	159
IV.4.1. Comparaison en SS des colonnes	159
IV.4.2. Comparaison en SS entre cartouches SPE	166
IV.4.3. Comparaison en SS sur l'ensemble des cartouches SPE et colonnes chromatographiques	168
IV.4.4. Comparaison en NTS entre cartouches SPE	174
IV.5. Conclusion sur la comparaison de méthodes analytiques	177

PARTIE C : APPLICATION DES METHODES POUR LA COMPARAISON D'ECHANTILLONS OXYDES

179

I. Introduction et objectifs

180

II. Etat de l'art

181

II.1. Traitement des eaux usées	181
II.1.1. Filières conventionnelles et limites des traitements face aux micropolluants	181
II.1.2. Traitements avancés des micropolluants	183
II.1.3. Problématique des produits de transformation	188
II.2. Le NTS au service de l'évaluation des traitements d'élimination des MP organiques en STEU	190
II.2.1. Comparaisons basées sur les intensités des signaux	194
II.2.2. Comparaisons basées sur les masses détectées des signaux	195

III. Applications de méthodes à l'évaluation de procédés d'oxydation et de la formation de TP

197

III.1. Cas de l'acide performique	197
III.1.1. Introduction	198
III.1.2. Matériel et méthodes	200
III.1.2.1. Réactifs	200
III.1.2.2. Traitement des rejets de STEU à l'acide performique	201
III.1.2.3. Oxydation de molécules individuelles par l'acide performique	201
III.1.2.4. Analyse et traitement de données	202
III.1.3. Résultats	203
III.1.3.1. NTS des rejets de STEU soumis à l'oxydation par PFA	203
III.1.3.2. SS de N-oxydes et de leurs précurseurs dans les rejets de STEU soumis à l'oxydation au PFA	206

III.1.3.3. Etude des mécanismes de transformation de MP soumis au PFA	208
III.1.3.4. Impact environnemental	212
III.2. Cas de l'ozonation	213
III.2.1. Introduction	213
III.2.2. Matériel et méthodes	215
III.2.2.1. Echantillonnage	215
III.2.2.2. Analyses	215
III.2.3. Résultats	216
III.2.3.1. Caractérisation globale	216
III.2.3.2. Recherche de N-oxydes en mode suspect	219
III.2.3.3. Recherche de N-oxydes par approches non-ciblées	221
III.3. Conclusion	225
IV. Influence de la stratégie analytique sur l'évaluation des procédés d'oxydation	226
IV.1. Introduction	226
IV.2. Matériel et méthodes	226
IV.2.1. Stratégie d'échantillonnage	226
IV.2.2. Oxydation	228
IV.2.3. Analyses et traitement de données	232
IV.3. Résultats	235
IV.3.1. Comparaison des colonnes chromatographiques	235
IV.3.2. Comparaison des méthodes de préparation	240
V. Conclusions sur les applications des approches non ciblées pour l'analyse des procédés d'oxydation	246
CONCLUSION ET PERSPECTIVES	248
Conclusion	249
Perspectives	253
RÉFÉRENCES	256
ANNEXES	284

Liste des Figures

Figure 1 : <i>Workflow</i> pour l'analyse ciblée, l'analyse en mode suspect et l'analyse non-ciblée, selon (Krauss et al., 2010)	6
Figure 2 : Classification en niveaux permettant de décrire la fiabilité des résultats obtenus en analyse en mode suspect et en analyse non-ciblée, selon (Schymanski et al., 2015).....	8
Figure 3 : Description du <i>workflow</i> général pour le prétraitement de données, selon (Smith et al., 2006).....	24
Figure 4 : Principe des réseaux moléculaires, selon (Fisher et al., 2021)	29
Figure 5 : Stratégie d'échantillonnage de l'EIL organisé dans le réseau NORMAN (2019).....	48
Figure 6 : Exemple de résultat d'identification d'un marqueur inconnu dans Unifi (fonction <i>Elucidate</i>).....	56
Figure 7 : Structure d'une <i>marker table</i>	60
Figure 8 : Graphique des échantillons obtenu par ACP à partir des marqueurs issus de la séquence d'analyse de l'EIL (STEP = station de traitement d'eau potable, RTI = mélange de standards analytiques)	62
Figure 9 : Résultats de a) l'OPLS-DA et du b) s-plot pour la comparaison des marqueurs des échantillonneurs passifs (4 jours de contact)	64
Figure 10 : Sélection des marqueurs discriminants (carrés rouges) sur le <i>loadings plot</i>	76
Figure 11 : Description de la sélection de marqueurs discriminants sur le <i>bi-plot</i>	77
Figure 12 : Comparaison de la spécificité des méthodes de sélection des 100 marqueurs les plus discriminants.....	78
Figure 13 : Comparaison de différentes méthodes de sélection en employant les logiciels fournisseur ou libre	83
Figure 14 : Comparaison des spectres d'un même marqueur ($m/z = 267,16980$, RT = 3,0 min) A) MS avec DT, B) MS sans DT, C) MS ^E avec DT, D) MS ^E sans DT	87
Figure 15 : Voies de traitement des données distinguées	89
Figure 16 : Comparaison d'empreintes de deux échantillons	95
Figure 17 : Comparaison d'échantillon par a) Diagramme d'Euler et b) Distribution en m/z . 96	
Figure 18 : Diagramme de Kendrick (base C ₂ H ₄ O). Les points alignés sur la ligne rouge appartiennent à la même famille d'homologues.....	98
Figure 19 : Diagramme de van Krevelen	99

Figure 20 : Occurrence des différents analyseurs sur <i>Web of Sciences</i> a) associés à HRMS puis à b) NTS (consulté le 12/10/2021)	112
Figure 21 : Diagramme de polarité-volatilité pour le choix de la méthode de chromatographie, adapté de (Giger, 2008).....	113
Figure 22 : Procédure de fractionnement successif sur résines	130
Figure 23 : Déviation des TR (par rapport à la moyenne) inter et intra-séquence des quatre étalons deutérés visibles en ESI+ pour les 3 colonnes chromatographiques) (n > 300 dans des matrices de complexité différentes)	136
Figure 24 : Distribution des valeurs de (a) m/z et (b) DT des marqueurs séparés sur les 3 colonnes	137
Figure 25 : Chromatogrammes d'un même extrait séparé sur a) PGC et b) HILIC	138
Figure 26 : Comparaison des marqueurs <i>communs</i> entre les 3 colonnes en fonction des différentes étapes d'alignement.....	140
Figure 27 : Nombre de marqueurs issus des différentes méthodes de préparation (analyse sur colonne C18 en mode ESI+).....	144
Figure 28 : Classification des marqueurs pour comparer les différents extraits d'un même échantillon, après séparation sur colonne C18.....	146
Figure 29 : Comparaison des distributions de a) m/z entre l'injection directe et l'extrait HLB en ESI+, b) TR entre l'injection directe et l'extrait DP en ESI-	150
Figure 30 : ACP (Scores plot) obtenue à partir des marqueurs issus de l'analyse HRMS d'effluent de STEU après SPE sur différentes phases solides (HLB, C18, XAW, XCW, ENV, Carb, Coco, DP) ou injection directe et analyse après séparation sur colonne C18 en ESI+.....	151
Figure 31 : Classification des marqueurs pour comparer les différents extraits d'un même échantillon après séparation sur (a) HILIC et (b) PGC.....	152
Figure 32 : Comparaison des proportions de marqueurs <i>uniques</i> entre les différents extraits pour les quatre méthodes d'analyse.....	153
Figure 33 : Diagrammes d'Euler donnant le nombre de composés retrouvés dans les trois colonnes en SS.....	159
Figure 34 : Heatmap des composés suspects (dendrogramme vertical) discriminés par leurs aires relatives dans chaque extrait provenant des différentes méthodes de préparation (dendrogramme horizontal), après séparation sur colonne C18 et analyse en ESI+ (les couleurs les plus sombres indiquent des aires plus importantes).....	166

Figure 35 : ACP réalisée sur les aires des composés suspects détectés dans les extraits des différentes méthodes de préparation, après séparation sur colonne C18 et analyse en mode ESI+	167
Figure 36 : <i>Heatmap</i> des composés retrouvés en SS par chacune des combinaison cartouche SPE / colonne chromatographique. L'intensité de couleur est proportionnelle à l'aire du pic analysé par UPLC-IMS-QToF.....	169
Figure 37 : ACP réalisée sur les aires des composés suspects détectés avec différentes combinaisons analytiques (cartouche SPE / colonne chromatographique)	171
Figure 38 : Comparaison de la gamme de a) masses molaires b) log Kow, c) pKa, d) CCS et e) surface polaire topologique couverte par chaque groupe de suspects responsables de la discrimination des trois clusters définis en Figure 36. Les barres d'erreur représentent les valeurs minimales et maximales de chaque groupe.	173
Figure 39 : S-plot correspondant à l'OPLS-DA entre les rejets de STEU (n = 6, injectés en triplicat) et les rejets oxydés (n = 6, injectés en triplicat) soumis à des doses comprises entre 0,8 et 2,5 mg/L de PFA pendant 10 min. Les marqueurs les plus discriminants sélectionnés apparaissent en orange et sont disponibles en Annexe 5A.	204
Figure 40 : Evolution des aires de la lidocaïne et de la lidocaïne N-oxyde dans les rejets de STEU soumis à différentes doses de PFA. Les barres d'erreur correspondent à l'écart-type entre les triplicats d'injection.....	205
Figure 41 : Cinétique de dégradation de la lidocaïne par oxydation au PFA (concentration initiale de lidocaïne à 5 µM, PFA entre 5 et 50 mg/L dans un tampon phosphate à 50 mM pH 8).....	208
Figure 42 : Concentration de lidocaïne et de sa N-oxyde au bout de 10 min de réaction dans un tampon phosphate à 50 mM à pH 8	209
Figure 43 : Mécanisme de dégradation de la lidocaïne par action du PFA.....	210
Figure 44 : Mécanismes de transformation de l'amisulpride et du tramadol par oxydation par le PFA.....	211
Figure 45 : Empreintes de l'échantillon à différentes étapes de traitement (seuls les marqueurs d'aires supérieures à 10 000 ont été représentés)	216
Figure 46 : Diagramme d'Euler pour la comparaison du nombre de marqueurs de l'échantillon à la suite du traitement par ozonation et CAP	218
Figure 47 : Recherche non-ciblée de N-oxydes basée sur la méthode de Kendrick.....	222

Figure 48 : Montage du pilote de photolyse UV	229
Figure 49 : Montage pour la production d'ozone en laboratoire	231
Figure 50 : Classification des marqueurs suivants 5 catégories en fonction de leur comportement à la suite des 3 procédés d'oxydation et détectés par les quatre méthodes d'analyse, dans les deux modes d'ionisation (ESI+ et ESI-).....	236
Figure 51 : Comparaison des différents types de marqueurs relevés pour différents extraits dans la manipulation d'ozonation après analyse en mode ESI+.....	241

Liste des Tableaux

Tableau 1 : Description des logiciels fournisseurs courants	21
Tableau 2 : Isotopes et abondances des principaux atomes étudiés	35
Tableau 3 : Description de différents logiciels permettant l'assignation de formules brutes	36
Tableau 4 : Principales bases de données en ligne	45
Tableau 5 : Principales plateformes de partage de données pour le NTS	46
Tableau 6 : Liste des molécules présentes dans le mélange de standards analytiques de l'EIL, fournie par les organisateurs	47
Tableau 7 : Protocoles de séparation chromatographique utilisés pour l'analyse des échantillons de l'EIL.....	49
Tableau 8 : Récapitulatif des composés présents dans la librairie interne (janvier 2022)	54
Tableau 9 : Groupes établis pour la comparaison par OPLS-DA des échantillons EIL	63
Tableau 10 : Comparaison de l'identification de composés d'un mélange de standards analytiques par les approches SS et NTS, avec description de l'étape limitante empêchant éventuellement l'identification.....	67
Tableau 11 : Optimisation des paramètres pour le calcul des formules moléculaires via UNIFI	72
Tableau 12 : Amélioration des identifications grâce à l'optimisation de l'étape de calcul de formules moléculaires, pour l'analyse de standards analytiques avec la <i>méthode LEESU</i>	73
Tableau 14 : Comparaison des identifications correctes pour les 3 méthodes de sélection de marqueurs. RTI = mélange de standards analytiques.....	79
Tableau 15 : Comparaison des outils pour l'identification après sélection statistique des marqueurs	80
Tableau 16 : Comparaison des outils pour l'identification après sélection des marqueurs par comparaison simple	82
Tableau 17 : Comparaison des paramètres de détection et alignement de <i>features</i> par les 3 méthodes.....	86
Tableau 18 : Comparaison du calcul de formule moléculaire par <i>Genform</i> à partir des spectres avec et sans mobilité ionique.....	88
Tableau 19 : Comparaison des méthodes de traitement des données de l'EIL en NTS	93

Tableau 20 : Comparaison des résultats d'identification par SS avec logiciel fournisseur (UNIFI) et logiciel libre (<i>patRoon</i>)	94
Tableau 21 : Présentation de différents groupes de phases stationnaire disponibles pour la SPE	106
Tableau 22 : Exemples de méthodes de préparation d'échantillon par SPE pour l'analyse non-ciblée de matrices aqueuses. N.A. = pas spécifié.....	108
Tableau 23 : Comparaison des méthodes analytiques pour le NTS d'échantillons aqueux environnementaux.....	119
Tableau 24 : Classification établie pour catégoriser la MOD selon les ratios O/C et H/C (Minor et al., 2014; Sanchís et al., 2020).....	123
Tableau 25 : Propriétés des étalons internes sélectionnés	124
Tableau 26 : Liste des cartouches SPE sélectionnées et leurs propriétés selon les fabricants	125
Tableau 27 : Réactifs utilisés dans les différentes méthodes analytiques.....	128
Tableau 28 : Protocoles détaillés de conditionnement et d'élution pour les 8 cartouches SPE	129
Tableau 29 : Gamme de TR utilisées pour la collection de fractions après séparation par HPLC-SEC.....	131
Tableau 30 : Conditions chromatographiques des 3 colonnes pour les analyses HRMS	132
Tableau 31 : Paramètres d'ionisation pour les analyses en HRMS.....	132
Tableau 32 : Paramètres d'acquisition en FT-ICR-MS.....	133
Tableau 33 : Caractérisation globales de trois méthodes de préparation.....	145
Tableau 34 : Caractérisation globale du total des marqueurs détectés et des marqueurs <i>uniques</i> à chaque extrait d'un même échantillon après séparation sur colonne C18. Les valeurs extrêmes, par propriété, ont été mises en évidence pour faciliter la lecture du tableau : Vert - moyennes les plus grandes, Rouge - moyennes les plus petites	149
Tableau 35 : Comparaison des moyennes de <i>m/z</i> totaux et <i>uniques</i> obtenus sur chaque cartouche par les différentes méthodes d'analyse en ESI+	157
Tableau 36 : Comparaison des proportions de formules moléculaires calculées après analyse par différentes méthodes en ESI+	158
Tableau 37 : Comparaison des masses molaires, CCS et structures 3D des molécules retenues sur la HILIC et la C18 en SS.	162

Tableau 38 : Composés suspects permettant de discriminer les clusters définis par l'ACP en Figure 36.....	172
Tableau 39 : Comparaison des ratios atomiques, pondérés par les intensités, des marqueurs retrouvés dans chaque extrait après analyse par FT-ICR-MS. Les valeurs extrêmes sont indiquées en rouge (minimum) et vert (maximum).....	175
Tableau 40 : Exemples d'amélioration des taux d'abattelements de différents MP organiques par implémentation d'un traitement complémentaire	185
Tableau 41 : Comparaison des méthodes de traitement de données en NTS pour l'évaluation des procédés de traitement en STEU.....	191
Tableau 42 : Conditions pour l'analyse des composés pharmaceutiques et leurs <i>N</i> -oxydes par HPLC-DAD	202
Tableau 43 : Classification des marqueurs selon leur évolution après oxydation par PFA ...	203
Tableau 44 : Détection de sept amines tertiaires et de leurs <i>N</i> -oxydes respectives en SS dans les rejets de STEU traités par le PFA (2 mg/L, 20 min). Les aires sont moyennées sur les triplicats d'injection, normalisées par un étalon interne (propylparaben-d4)	207
Tableau 45 : Caractérisation globale des rejets de STEU soumis à ozonation et adsorption sur CAP	217
Tableau 46 : SS de 5 amines tertiaires et de leur <i>N</i> -oxyde dans un effluent de STEU soumis à ozonation puis adsorption sur CAP. Les valeurs correspondent aux aires détectées corrigées par l'aire totale de l'échantillon considéré. ND : Non Détecté.....	220
Tableau 47 : Tentative d'identification de précurseurs de <i>N</i> -oxyde par des méthodes non-ciblées.....	224
Tableau 48 : Paramètre globaux des échantillons réels	227
Tableau 49 : Etablissement de liens de transformation entre marqueurs <i>éliminés</i> et <i>formés</i>	233
Tableau 50 : Proportion de marqueurs <i>formés</i> (%) pouvant être expliqués par des mécanismes de transformation usuels pour chaque procédé (par colonne chromatographique selon le mode d'ionisation et indépendamment du support d'extraction employé).....	238
Tableau 51 : Proportion de marqueurs <i>formés</i> (%) pouvant être expliqués par des mécanismes de transformation usuels pour l'échantillon d'effluent d'eau usée chloré, analysé en mode ESI+ après séparation sur PGC, selon les différentes méthodes de préparation	245

Liste de abréviations

AcEt	Acétate d'éthyle
ACN	Acétonitrile
AF	Acide Formique
ACP	Analyse par Composante Principale
AMV	Analyses Multivariées
API	Interface de programmation d'application (de l'anglais <i>Application Programming Interface</i>)
C18	Cartouche d'extraction Strata C18 (Phenomenex)
Carb	Cartouche d'extraction Envi-Carb (Supelclean)
CAP	Charbon Actif en Poudre
CAS	<i>Chemical Abstracts Service</i>
CCS	Section efficace de collision (de l'anglais <i>Collisional Cross Section</i>)
Coco	Cartouche d'extraction Coconut Charcoal
COT	Carbone Organique Total
COD	Carbone Organique Dissous
DAD	Détecteur à barrette de diode (de l'anglais <i>Diode Array Detector</i>)
DCM	Dichlorométhane
DDA	Data-Dependent Analysis
DIA	Data-Independent Analysis
DP	Cartouche d'extraction DPA-6S (Supelclean)
DT	Temps de dérive (de l'anglais <i>Drift Time</i>)
EIL	Essai Inter-Laboratoire

EUP	Eau Ultra Pure
ESI	Source d'ionisation par électrospray (de l'anglais <i>Electrospray Ionisation</i>)
FT-ICR	Détecteur à résonance cyclotronique et transformation de Fourier (de l'anglais <i>Fourier Transform Ion Cyclotron Resonance</i>)
GC	Chromatographie gazeuse (de l'anglais <i>Gas Chromatography</i>)
HILIC	Colonne chromatographique d'interaction hydrophilique
HLB	Cartouche d'extraction Oasis HLB (Waters)
HPLC	Chromatographie liquide haute performance (de l'anglais <i>High Performance Liquid Chromatography</i>)
HRMS	Spectrométrie de masse haute résolution (de l'anglais <i>High Resolution Mass Spectrometry</i>)
IMS	Spectrométrie de mobilité ionique (de l'anglais <i>Ion Mobility Spectrometry</i>)
InChI	<i>International Chemical Identifier</i>
IUPAC	Union internationale de chimie pure et appliquée (de l'anglais <i>International Union of Pure and Applied Chemistry</i>)
LC	Chromatographie liquide (de l'anglais <i>Liquid Chromatography</i>)
MeOH	Méthanol
MOD	Matière Organique Dissoute
MON	Matière Organique Naturelle
MP	Micropolluant
MS	Spectrométrie de masse (de l'anglais <i>Mass Spectrometry</i>)
MS/MS	Spectrométrie de masse en tandem
NTS	Analyse non ciblée (de l'anglais <i>Non-Target Screening</i>)
PAA	Acide peracétique

PC	Composante principale (de l'anglais <i>Principal Component</i>)
PFA	Acide performique
PGC	Colonne chromatographique en carbone graphitique poreux
QToF	Analyseur en tandem quadripolaire-temps de vol
SEC	Chromatographie d'exclusion stérique (de l'anglais <i>Size Exclusion Chromatography</i>)
SIAAP	Syndicat Interdépartemental pour l'Assainissement de l'Agglomération Parisienne
SPE	Extraction sur phase solide (de l'anglais <i>Solid Phase Extraction</i>)
SS	Analyse en mode suspect (de l'anglais <i>Suspect Screening</i>)
STEP	Station de Traitement des Eaux Potables
STEU	Station de Traitement des Eaux Usées
TP	Produits de transformation (de l'anglais <i>Transformation Products</i>)
TR	Temps de Rétention
TS	Analyse ciblée (de l'anglais <i>Target Screening</i>)
UHPLC	Chromatographie liquide ultra haute performance (de l'anglais <i>Ultra High Performance Liquid Chromatography</i>)
UV	Ultra-Violet
XAW	Cartouche d'extraction Strata X-AW (Phenomenex)
XCW	Cartouche d'extraction Strata X-CW (Phenomenex)

Glossaire

Feature	Association d'un pic chromatographique, de son pic de mobilité ionique et de ses spectres de masses, combinée à son aire relative
Marker table	Matrice regroupant les marqueurs détectés dans chaque injection et leurs propriétés relatives
Marqueur	Association d'un pic chromatographique, de son pic de mobilité ionique et de ses spectres de masses, combinée à son aire relative, après alignement sur toutes les injections d'une même séquence
Masse exacte	Somme des masses exactes des isotopes les plus abondant de chaque atome de la molécule
Masse molaire moyenne	Somme des masses de tous les isotopes de chaque atome, pondérée par l'abondance naturelle de chaque isotope
Masse monoisotopique	Masse mesurée
Masse nominale	Somme des masses entières des isotopes les plus abondants de chaque atome de la molécule. Pour un marqueur, la masse nominale sera la masse détectée arrondie à l'entier le plus proche

Liste des publications et communications

Publications internationales

Articles publiés

Fidji Sandré, Nina Huynh, Marie-Christine Gromaire, Gilles Varrault, Christophe Morin, Régis Moilleron, Julien Le Roux, Laure Garrigue-Antar. Road Runoff characterization: ecotoxicological assessment combined with (non-)target screenings of micropollutants for the identification of relevant toxicants in the dissolved phase. *Water*, MDPI, 14 (4), pp.511. <10.3390/w14040511>

Johnny Gasperi, Julien Le Roux, Steven Deshayes, Sophie Ayrault, Louise Bordier, Lila Boudahmane, Hélène Budzinski, Emilie Caupos, Nadège Caubrière, Kelsey Flanagan, Martin Guillon, Nina Huynh, Pierre Labadie, Laurent Meffray, Pascale Neveu, Chandirane Partibane, Julien Paupardin, Mohamed Saad, Lucie Varnede, Marie-Christine Gromaire. Micropollutants in Urban Runoff from Traffic Areas: Target and Non-Target Screening on Four Contrasted Sites. *Water*, MDPI, 14 (3), pp.394. <10.3390/w14030394>

Nina Huynh, Emilie Caupos, Caroline Soares Peirera, Julien Le Roux, Adèle Bressy, Régis Moilleron. Evaluation of Sample Preparation Methods for Non-Target Screening of Organic Micropollutants in Urban Waters Using High-Resolution Mass Spectrometry. *Molecules*, MDPI, 2021, 26 (23), pp.7064. <10.3390/molecules26237064>

Bastian Schulze, Denice van Herwerden, Ian Allan, Lubertus Bijlsma, Nestor Etxebarria, Martin Hansen, Sylvain Merel, Branislav Vrana, Reza Aalizadeh, Bernard Bajema, Florian Dubocq, Gianluca Coppola, Aurélie Fildier, Pavla Fialová, Emil Frøkjær, Roman Grabic, Pablo Gago-Ferrero, Thorsten Gravert, Juliane Hollender, Nina Huynh, Griet Jacobs, Tim Jonkers, Sarit Kaserzon, Marja Lamoree, Julien Le Roux, Teresa Mairinger, Christelle Margoum, Giuseppe Mascolo, Emmanuelle Mebold, Frank Menger, Cécile Miège, Jeroen Meijer, Régis Moilleron, Sapia Murgolo, Massimo Peruzzo, Martijn Pijnappels, Malcolm Reid, Claudio Roscioli, Coralie Soulier, Sara Valsecchi, Nikolaos Thomaidis, Emmanuelle Vulliet, Robert Young, Saer Samanipour. Inter-laboratory mass spectrometry dataset based on passive sampling of drinking water for non-target analysis. *Scientific Data*, Nature Publishing Group, 2021, 8 (1), pp.223. <10.1038/s41597-021-01002-w>

Article soumis

Maolida Nihemaiti, Nina Huynh, Romain Mailler, Perrine Mèche-Ananit, Vincent Rocher, Rachid Barhdadi, Régis Moilleron, Julien Le Roux. High-resolution mass spectrometry screening of wastewater effluent for micropollutants and their transformation products during disinfection with performic acid. – Soumis à ACS ES&T Water.

Articles en préparation

Nina Huynh, Julien Le Roux, Emmanuelle Mebold, Régis Moilleron. Comparison of analytical strategies using the advantage of ion mobility for non-target screening.

Fidji Sandre, Nina Huynh, Emilie Caupos, Lamyae El-Mrabet, Michael Rivard, Isabelle Lachaise, Christophe Pommier, Christophe Morin, Régis Moilleron, Julien Le Roux, Laure Garrigue-Antar. Comparative study of oxidative treatment of a pollutant of emerging interest: the furosemide.

Chapitre d'ouvrage

Vincent Rocher, Sam Azimi, Romain Mailler, Sabrina Rechdaoui-Guérin, Perrine Mèche, Sébastien Pichon, Angélique Goffin, Jean Bernier, Jérôme Roy, Gilles Varrault, Julien Le Roux, Nina Huynh, Maolida Nihemaiti, Thierry Pigot, Thomas Paulin, Jean-Marie Mouchel, David Du Pasquier, Laurent Paulic, Dan Angelescu, Vaizanne Huynh, Andreas Hausot, Patrizia Ragazzo, Nicoletta Chiucchini. Effectiveness of Disinfecting Wastewater Treatment Plant Discharges: Case of chemical disinfection using performic acid. IWA Publishing, 2021, 9781789062090. (10.2166/9781789062106_0039)

Communications internationales

Présentations orales

Nina Huynh, Julien Le Roux, Jérémy Guyot, Régis Moilleron. Development of a non-target screening workflow for the evaluation of micropollutants fate during oxidation processes. 17th International Conference on Environmental Science & Technology (CEST2021), Sep 2021, Athens, Greece.

Nina Huynh, Emmanuelle Mebold, Régis Moilleron, Julien Le Roux. Comparison of three chromatographic columns using the advantage of ion mobility for non-target analysis of

organic micropollutants in wastewater samples. SETAC North America 41st Annual Meeting, Nov 2020, Online Meeting, United States.

Nina Huynh, Sandre Fidji, Laure Garrigue-Antar, Christophe Morin, Emilie Caupos, Marie-Christine Gromaire, Régis Moilleron, Julien Le Roux. Evaluation of road runoff treatment systems performance by non-target screening combined with ecotoxicological studies. 16th Annual Workshop On Emerging High-Resolution Mass Spectrometry And LC-MS/MS Applications In Environmental Analyses And Food Safety, Oct 2020, Online meeting, Spain.

Maolida Nihemaiti, Nina Huynh, Romain Mailler, Perrine Mèche, Vincent Rocher, Julien Le Roux. Screening of micropollutants and their transformation products in performic acid treated wastewater. 11th Micropol & Ecohazard Conference 2019, Oct 2019, Seoul, South Korea.

Maolida Nihemaiti, Nina Huynh, Romain Mailler, Perrine Mèche, Vincent Rocher, Julien Le Roux. Screening of micropollutants and their transformation products in performic acid treated wastewater. Wasserchemischen Gesellschaft 8th Late Summer Workshop "Chemical and biological transformation processes and tools for their investigation", Sep 2019, Haltern am See, Germany.

Présentations par affiche

Nina Huynh, Emilie Caupos, Caroline Soares Pereira, Julien Le Roux, Adèle Bressy, Régis Moilleron. How important is the sample preparation step for non-target screening of micropollutants in urban waters?. International Conference on Non-Target Screening (ICNTS21), Oct 2021, Erding, Germany.

Nina Huynh, Julien Le Roux, Régis Moilleron. Fractionation of wastewater effluent for the characterization of organic micropollutants by high-resolution mass spectrometry. SETAC Europe 30th Annual Meeting, May 2020, Online Meeting, Ireland.

Communication nationale

Nina Huynh, Julien Le Roux, Régis Moilleron. Fractionnement d'une eau résiduaire urbaine pour la caractérisation de micropolluants organiques par spectrométrie de masse haute résolution. GRUTTEE2020 : 13^{ème} congrès international du GRUTTEE, Feb 2020, Rennes, France

Introduction générale

I. Contexte

I.1. Les micropolluants organiques dans les eaux urbaines

Selon une définition du ministère de la Transition écologique, le terme de micropolluant (MP) désigne « *une substance indésirable détectable dans l'environnement à très faible concentration (microgramme par litre voire nanogramme par litre). Sa présence est, au moins en partie, due à l'activité humaine (procédés industriels, pratiques agricoles ou activités quotidiennes) et peut, à ces très faibles concentrations engendrer des effets négatifs sur les organismes vivants en raison de sa toxicité, de sa persistance et de sa bioaccumulation* » (Plan micropolluants 2016-2021 pour préserver la qualité des eaux et la biodiversité). Ainsi de nombreuses substances peuvent entrer dans cette catégorie. Du fait des nombreuses activités dans les zones urbaines, celles-ci sont sans aucun doute génératrices d'un grand nombre de MP organiques. Certaines grandes familles de MP organiques sont d'ailleurs très médiatisées, comme les biocides, les résidus pharmaceutiques ou les plastifiants. Cependant, il existe différentes classifications des MP qui peuvent reposer sur leurs propriétés physico-chimiques, leurs effets, ou encore sur l'utilisation qui en est faite. Certains composés peuvent donc se retrouver classés dans plusieurs familles.

La majorité de ces MP transitent par les stations d'épuration (STEU), de manière directe (réseau des eaux usées) ou de manière indirecte (ruissellement) dans le cas des réseaux d'assainissement unitaires. Dans les filières de traitement conventionnelles, il a été reporté une élimination des MP par les traitements primaires via des processus d'adsorption, qui pouvaient différer grandement selon la nature des composés étudiés, allant de 10 à 70 % d'élimination (Carballa et al., 2005). Concernant les traitements secondaires, une revue résume différentes études qui montrent la mise en jeu de plusieurs processus d'élimination incluant des mécanismes de transformation physico-chimiques, d'adsorption ou de biodégradation (Jelić et al., 2012, et références associées). Cependant, ces traitements n'étant pas initialement conçus pour éliminer ce type de composés, l'élimination observée se traduit rarement par une minéralisation complète du composé initial. Dans certaines stations, un traitement tertiaire est mis en place afin de réduire la charge de MP rejetée dans l'environnement (Eggen et al., 2014), mais il peut s'avérer être la source de nouveaux composés appelés produits de transformation (TP) (Ike et al., 2019; Merel and Zwiener, 2016).

Ces TP constituent également un domaine de recherche grandissant. D'une part, car ils sont, pour la majorité, encore inconnus et, d'autre part, car de nombreuses études ont montrés que ces composés sont souvent plus polaires (Ferrer et al., 2016), plus persistants (Hübner et al., 2015) ou plus toxiques (Magdeburg et al., 2014) que les composés parents.

Les MP sont introduits dans l'environnement par différentes voies. En effet, les sources de MP en milieu urbain sont multiples et, en raison des différentes utilisations qui peuvent exister d'un même composé, il est souvent difficile de déterminer une source unique. Certaines études ont tenté de modéliser la dynamique des MP dans le cycle de l'eau en prenant en compte de nombreuses sources possibles (Mitchell and Diaper, 2006). Néanmoins, des données décrivant la nature des polluants et leurs concentrations respectives dans chaque type d'eau manquent encore. A l'échelle nationale, le réseau Urbis a pour objectif de coordonner trois observatoires régionaux en hydrologie urbaine incluant des actions sur les MP : pour l'agglomération parisienne (OPUR), la métropole lyonnaise (OTHU) et la métropole nantaise (ONEVU). Ces trois observatoires visent à apporter des connaissances concernant la production, le transfert et la dynamique des polluants dans les eaux urbaines afin de proposer des solutions pour une meilleure gestion. Ces études sont importantes puisque les MP sont susceptibles de produire des effets néfastes sur les organismes présents dans les milieux récepteurs. Pour certains composés organiques, ces effets ont été évalués à l'aide de tests toxicologiques résultants en une valeur qui permet d'évaluer le risque associé à une substance donnée : la *Predicted No Effect Concentration*. Cette valeur représente la plus forte concentration de la substance considérée pour laquelle celle-ci ne présente pas de risque pour l'environnement. Ces valeurs sont obtenues par différentes expérimentations sur des organismes donnés, pour une substance particulière. Outre les études réalisées sur les MP, certains auteurs se sont aussi penchés sur la question des TP montrant que, dans certains cas, ceux-ci pouvaient être plus toxiques que le composé d'origine (Richard et al., 2014). Enfin, il est important de rappeler que les MP ne sont pas isolés lorsqu'ils se retrouvent dans l'environnement. Ainsi, leurs effets sur les organismes peuvent être exacerbés ou atténués en fonction de la présence d'autres molécules dans le milieu à différents niveaux de concentration (Celander, 2011; Vasquez et al., 2014). Ainsi, des études plus larges consistant à estimer l'effet toxicologique d'un échantillon dans son ensemble ont été réalisées (Ocampo-Duque et al., 2008). De la même manière, le développement de méthodes comme les analyses

dirigées par les effets consistent à mener ces tests d'écotoxicologie dans un premier temps sur l'échantillon initial puis de fractionner ce dernier successivement et de répéter la procédure sur les fractions afin d'isoler et d'identifier des molécules qui seraient responsables de la toxicité observée (Brack, 2003).

En France, la Loi sur l'eau et les milieux aquatiques (2006), transposition en droit français de la Directive Cadre sur l'Eau (DCE, 2000), constitue le texte de référence en matière législative concernant la gestion de l'eau. Elle regroupe toutes les directives mises en place concernant les eaux continentales et maritimes. Plus particulièrement, en termes de MP, l'article 16 de la DCE propose une liste de 45 substances prioritaires et dangereuses prioritaires dans le domaine de l'eau incluant, entre autres, des biocides, des hydrocarbures aromatiques polycycliques, des métaux et des dérivés benzéniques (directive 2013/39/UE). En 2010, l'arrêté du 25 janvier établit le programme de surveillance de l'état des eaux, complété plus tard par l'arrêté du 7 août 2015. Ce programme fournit une liste de substances relatives de l'état chimique des eaux de surface (45 substances) et polluants spécifiques de l'état écologique des eaux de surface (31 substances), une liste de substances pertinentes à surveiller dans les eaux de surface (35 substances) et une liste complémentaire spécifique à la métropole (63 substances). Par la suite, différents arrêtés nationaux ont permis de mettre en place une surveillance réglementaire pour la collecte, le traitement et l'évacuation d'effluents en assainissement collectif (21 juillet 2015), complétés par la note technique du 12 août 2016 pour les STEU urbaines (96 substances surveillées dans les eaux brutes et 89 dans les eaux traitées) et l'arrêté du 24 août 2017 qui concerne les installations classées pour la protection de l'environnement. Ces différentes listes combinées résultent en environ 250 MP pertinents à surveiller dans les eaux superficielles, dont 114 sont inclus dans le programme national de réduction des MP. Ainsi le plan micropolluants 2016-2021 constitue la feuille de route de lutte contre les MP avec l'objectif de préserver la qualité des eaux et la biodiversité. Il fait suite à trois plans nationaux préliminaires : un plan pour la lutte contre les polychlorobiphényles, un premier plan sur les micropolluants (2010-2013) et un plan sur les résidus de médicaments. Le plan 2016-2021 a pour objectif d'inclure toutes les molécules susceptibles de polluer les ressources en eau. Il s'articule en trois parties : réduire les émissions des polluants déjà connus, apporter des connaissances afin d'identifier les MP présents dans le milieu et le risque associé, puis, prioriser les MP nécessitant une action immédiate.

1.2. L'analyse de micropolluants organiques dans des échantillons aqueux

Dans le domaine de l'analyse environnementale, la chromatographie gazeuse (GC) couplée à la spectrométrie de masse (MS) a longtemps été la méthode de référence utilisée pour l'analyse de contaminants à l'état de traces (Giger, 2008). Cependant, concernant les composés plus polaires et amphiphiles, des étapes de préparation supplémentaires (comme la dérivation) étaient nécessaires afin de pouvoir utiliser cette technique. Avec le développement de la chromatographie liquide haute performance (HPLC), puis de la chromatographie liquide ultra-haute performance (UHPLC), l'analyse de ces composés a été facilitée, et dès le milieu des années 1990, la chromatographie liquide couplée à la spectrométrie de masse (LC-MS) a été appliquée en routine pour l'analyse de contaminants émergents en matrice aqueuse environnementale (Richardson and Ternes, 2005). La spectrométrie de masse (MS) est une technique d'analyse qui permet de mesurer la masse moléculaire d'un composé donné et permet aussi d'obtenir des informations quant à sa structure. Elle peut être utilisée seule, ou être couplée à une séparation chromatographique préalable pour les échantillons complexes.

Récemment, la spectrométrie de masse haute résolution (HRMS) s'est développée et repose sur le même principe que la MS classique (ou basse résolution). Elle possède cependant un pouvoir de résolution supérieur et une plus grande précision en masse, rendant ainsi possible la séparation de masses très proches et l'obtention de masses avec des valeurs à plusieurs décimales permettant de déterminer une formule brute associée. Avec cette nouvelle technologie, le développement de nouvelles techniques d'analyse dites en mode suspect et les analyses non-ciblées ont pu être mises en place en complément des analyses ciblées réalisées jusqu'à présent. Ces trois types d'analyses ont été décrits par (Krauss et al., 2010) et repris dans la Figure 1.

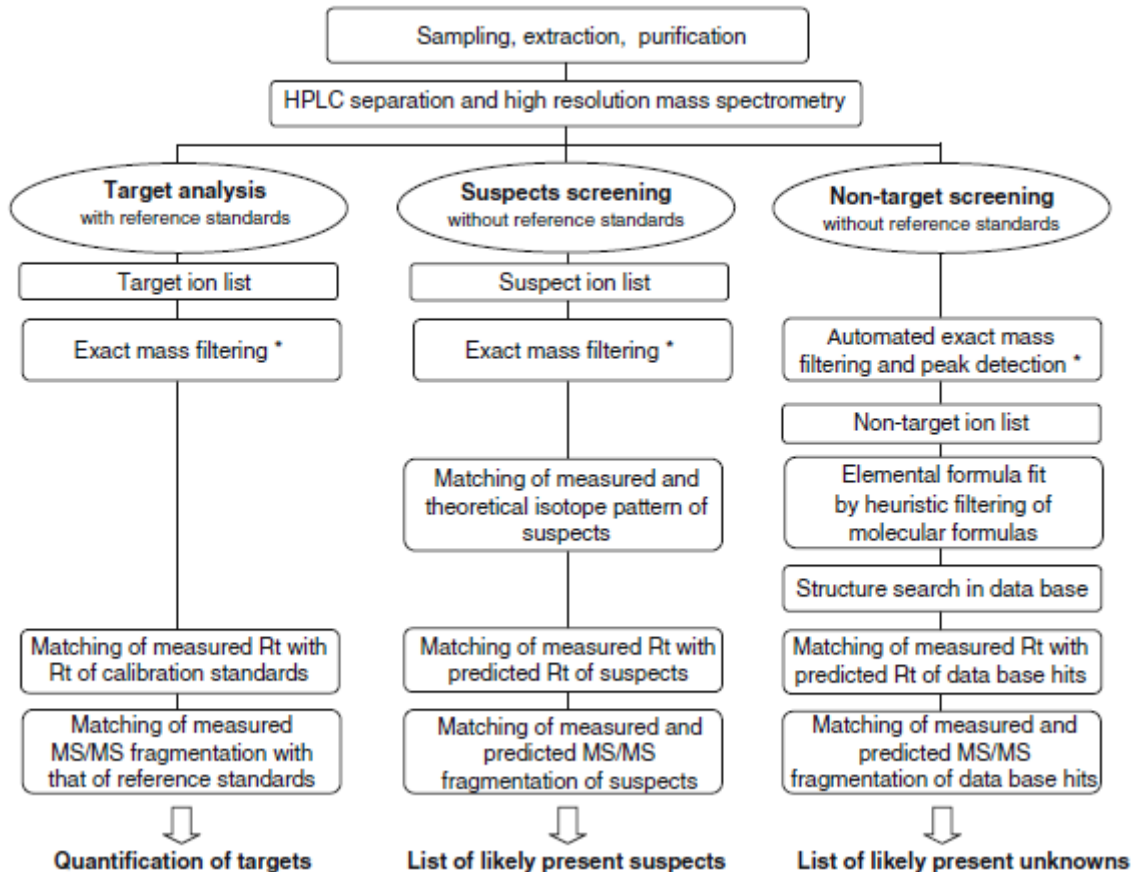


Figure 1 : *Workflow* pour l'analyse ciblée, l'analyse en mode suspect et l'analyse non-ciblée, selon (Krauss et al., 2010)

L'analyse ciblée ou *target screening* (TS) est couramment utilisée notamment pour la quantification de composés. Elle se base sur une liste prédéfinie de molécules à rechercher dans les échantillons. Le TS suppose un travail préalable de priorisation qui aura permis d'identifier des molécules d'intérêt. Un développement analytique à partir d'étalons de référence correspondant à ces molécules est ensuite réalisé afin d'optimiser les paramètres pour identifier et quantifier de manière certaine les composés en question à partir de leur masse. L'analyse d'échantillons est ensuite effectuée dans les conditions prédéfinies et permet de retrouver ou non les molécules de la liste en fonction de leur concentration. Les molécules priorisées dépendent de l'objectif fixé par l'étude et peuvent appartenir à une même famille chimique (Gasperi et al., 2014), appartenir à la même famille d'usage (Lardy-Fontan et al., 2017), présenter des propriétés particulières (Huntscha et al., 2012) ou encore faire partie d'une liste réglementaire (Lopez et al., 2015). Cette technique présente le désavantage de ne prendre en compte qu'un nombre restreint de substances présentes dans

l'échantillon. En effet, tout le protocole d'analyse étant optimisé de façon à retrouver les molécules de la liste prédéfinie, le TS ne permet de trouver que ce qui est recherché.

En analyse en mode suspect ou *suspect screening* (SS), l'opérateur dispose aussi d'une liste limitée de molécules dont la présence dans l'échantillon est attendue. Cependant, la méthode d'acquisition utilisée lors de l'analyse balaye de manière complète l'échantillon et ne cible pas une ou plusieurs masses en particulier comme c'est le cas en TS. A partir des données complètes acquises, les masses des molécules d'intérêt sont alors recherchées *a posteriori* parmi toutes les masses présentes, et un travail sur les données (comme par exemple l'écart par rapport à la masse exacte, la présence de fragments dans les spectres MS...) est réalisé pour déterminer s'il s'agit bien de la molécule recherchée ou non. La principale différence entre le TS et le SS repose sur les informations dont l'opérateur dispose sur les molécules recherchées. En effet, en TS, les composés recherchés doivent être disponibles en étalons de référence permettant d'avoir accès à toutes les propriétés et paramètres nécessaires pour les identifier sans ambiguïté (temps de rétention, masse exacte, fragments...). En SS, ces informations ne sont pas forcément disponibles. Il est même possible que les seules informations disponibles soient la structure de la molécule et donc sa formule brute et sa masse exacte. Le spectre de molécules analysées est ainsi élargi par rapport au TS puisque la disponibilité et l'achat d'étalons ne sont pas nécessaires. Des produits de dégradation, déjà observés en conditions contrôlées ou pouvant résulter de mécanismes connus mais jamais observés, peuvent donc par exemple être intégrés à ces listes de suspects (Deeb et al., 2017; Zonja et al., 2015). Néanmoins, il existe toujours un *a priori* sur ce qui est présent dans l'échantillon, par la sélection de molécules d'intérêt. Des molécules non soupçonnées ou inconnues sont donc ignorées lors du processus de traitement de la donnée.

L'analyse non-ciblée ou *non-target screening* (NTS) permet alors de s'intéresser à ces molécules inconnues. En effet, comme pour le SS, un balayage complet de l'échantillon est réalisé, mais aucune liste préalable de molécules n'est établie. Le but n'est plus de retrouver les masses de molécules spécifiques à partir d'une liste, mais de prendre en compte toutes les masses analysées et donc de regarder l'échantillon dans sa globalité afin d'identifier des molécules inconnues. Un nombre d'étapes plus important est donc souvent nécessaire en NTS afin d'identifier précisément une molécule et confirmer sa présence de manière certaine ; il est d'ailleurs souvent très difficile voire impossible d'arriver jusque-là. Un système a été défini

pour établir le niveau de confiance qui peut être attribué pour l'identification d'une masse d'intérêt (Figure 2). Ce niveau représente le nombre et la fiabilité des informations disponibles pour associer une masse détectée à une identité précise (Schymanski et al., 2014a). En TS (c'est-à-dire avec l'injection d'étalons de référence), toutes les informations relatives à l'identité de la molécule (à savoir les temps de rétention, masse et intensité des fragments) sont disponibles (niveau 1). En SS, la structure de la molécule, et donc sa masse exacte sont généralement connues (niveau 3). Il s'agit alors de retrouver cette masse parmi les masses détectées et ensuite de prouver que les informations supplémentaires recueillies (comme les fragments) peuvent correspondre à la molécule en question (niveau 2). Enfin, en NTS, seules les masses d'intérêt sont relevées et connues au départ (niveau 5). L'objectif sera alors d'associer une formule brute unique à cette masse (niveau 4), accessible grâce à la haute résolution. Il s'agira ensuite de déterminer une structure moléculaire pouvant correspondre (niveau 3) avant de pouvoir vérifier si les données associées sont susceptibles de décrire ce candidat (niveau 2), et si possible de confirmer son identité avec un étalon (niveau 1). En SS et en NTS, il est parfois impossible d'atteindre le niveau 1 de confirmation puisque les molécules peuvent ne pas être disponibles commercialement.

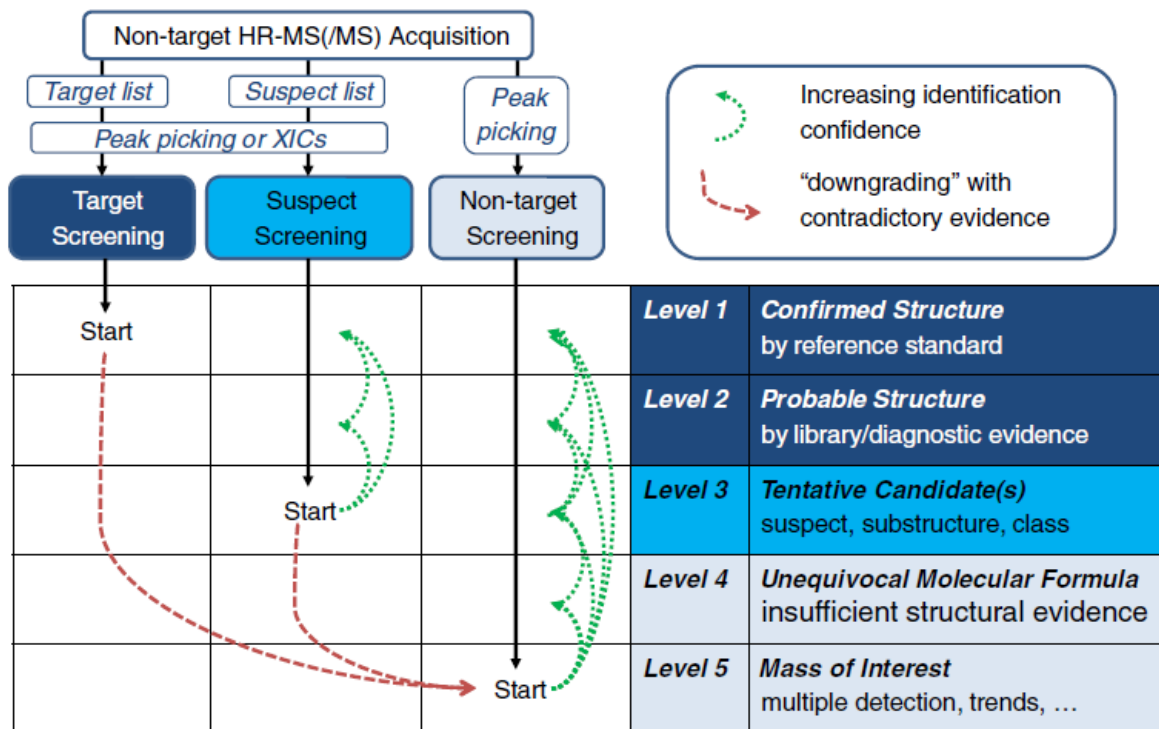


Figure 2 : Classification en niveaux permettant de décrire la fiabilité des résultats obtenus en analyse en mode suspect et en analyse non-ciblée, selon (Schymanski et al., 2015)

En complément de la haute résolution, une technique de séparation supplémentaire s'est également développée récemment : la spectrométrie de mobilité ionique (*ion mobility spectrometry* - IMS). Cette dernière repose sur la séparation des ions dans une phase gazeuse. Une fois formés au niveau de la source d'ionisation, les ions sont envoyés dans une cellule de mobilité ionique dans laquelle ils sont séparés en fonction de leur taille, de leur forme et leur charge. Différentes technologies de mobilité ionique existent. La première est la mobilité ionique dans un tube de dérive appelée *Drift Tube Ion Mobility*. Dans ce cas, les ions se déplaceront à l'aide d'un champ électrique. Un gaz neutre y circule également à contrecourant créant ainsi une résistance à la traversée des ions. Plus les ions seront petits et planaires, moins ils subiront cette résistance. La seconde technologie repose sur le principe inverse, elle est appelée *Trapped Ion Mobility*, où les ions sont cette fois transportés par un gaz avec un contre-courant électrique. Enfin, la dernière séparation repose sur le même principe que la première, seulement, le champ électrique est appliqué sous forme d'impulsions créant des « vagues » ce qui permet une séparation plus rapide des ions : c'est la *Travelling Wave Ion Mobility*. Quelle que soit la technique utilisée, elle permet de déterminer le temps de dérive des ions ou *drift time* (DT), qui décrit le temps que ces ions mettent à parcourir la cellule de mobilité ionique. Une valeur complémentaire décrivant la section de collision (CCS) de l'ion avec le gaz peut être déterminée à partir du *drift time* (Gabelica et al., 2019). Cette valeur est une propriété de l'ion étudié, qui apporte donc une information supplémentaire pour faciliter son identification. Différentes études ont permis de montrer la reproductibilité de cette valeur entre différents appareils ou différentes techniques de mobilité ionique (Hinnenkamp et al., 2018).

L'IMS peut être à la fois couplée à une séparation chromatographique et à la spectrométrie de masse, permettant ainsi d'ajouter une dimension supplémentaire de séparation. De nombreuses études ont montré l'intérêt de l'utilisation de cette technique, en montrant notamment que les valeurs de CCS sont indépendantes de la matrice étudiée (Bijlsma et al., 2017), qu'elles sont plus stables et plus fiables que le temps de rétention (TR), permettant ainsi la transposition directe de ces valeurs d'un laboratoire à un autre (Paglia et al., 2014; Tejada-Casado et al., 2018). D'autres avantages de la mobilité ionique sont la séparation des composés isomériques qui ne sont généralement pas séparés par la chromatographie (Levy et al., 2019) ou l'affranchissement de l'étape de chromatographie qui permet un gain de temps,

tout en gardant une étape de séparation (Lazzaro et al., 2019). La mobilité ionique permet aussi d'améliorer l'attribution de structure lors des étapes d'identification (Celma et al., 2020; Feuerstein et al., 2021).

I.3. L'analyse non-ciblée en analyse environnementale

Le NTS, par son côté exhaustif permet d'avoir une vision plus complète des échantillons analysés. Dans le domaine de l'analyse environnementale, il a été employé notamment pour évaluer les procédés de traitement des eaux en termes d'efficacité d'élimination des MP organiques. Habituellement, cette évaluation est réalisée en TS sur un nombre limité de molécules. Les concentrations en entrée et en sortie de station sont mesurées et le taux d'abattement de la station pour la molécule considérée est calculé. Seulement, les MP sont de natures très diverses et ont, de ce fait, un comportement très différent dans les filières de traitement. Il n'est donc pas possible de conclure sur l'efficacité des stations d'épuration à éliminer les MP de façon générale en se basant uniquement sur les informations obtenues pour un nombre limité de composés. Le NTS permet d'apporter des informations plus complètes puisqu'il s'intéresse à l'ensemble des molécules présentes. Ces études peuvent avoir un caractère très global, permettant de comparer les signatures de divers échantillons à différentes étapes de traitement pour voir la proportion de MP éliminés ou formés (Bergé et al., 2018; Nürenberg et al., 2015, 2019) ou être plus fins afin d'identifier de nouveaux contaminants ou des TP (Merel et al., 2015; Schollée et al., 2018).

Outre ces applications, le NTS a aussi permis d'améliorer la surveillance des milieux. Par exemple, le *Joint Danube Survey*, un programme de recherche européen lancé au début des années 2000 dont le but est de rassembler des données le long du Danube pour déterminer la qualité de l'eau, a mis en œuvre ce type d'analyse. En effet, introduite dans la phase 2 de ce programme lancé en 2007, en utilisant la chromatographie gazeuse, le NTS a permis de détecter la présence de 158 molécules d'intérêt non recherchées précédemment (JDS2 final report). Plus tard, dans sa phase 3, lancée en 2013, le NTS a permis de différencier les différents points de prélèvement. En effet, les résultats ont montré des signaux plus intenses en Allemagne et moins intenses en Roumanie, avec des empreintes de pollution similaires en Serbie, Roumanie et Bulgarie et, entre l'Ukraine et la Croatie. L'Allemagne, l'Autriche et la Slovaquie possèdent, quant à elles, des empreintes de pollution bien spécifiques (JDS3 final report). De la même manière, la commission internationale pour la protection du Rhin a lancé

en 2013 un programme d'analyse spéciale introduisant le NTS comme une méthode prometteuse. Par la suite, une étude menée sur le Rhin (Ruff et al., 2015) a également utilisé le NTS combiné au TS, afin d'identifier des sources de pollution tout le long du Rhin. Par suivi d'une masse d'eau de la source jusqu'à l'embouchure du fleuve, les auteurs ont pu mettre en évidence que les STEU étaient responsables du rejet de 80 % des MP organiques polaires retrouvés dans le Rhin, ce qui a conduit à des discussions avec les municipalités avoisinantes pour déterminer des stratégies de réduction de ces rejets. Cette étude a également permis la détection de plusieurs déversements accidentels qui ont été responsables de 5 % de la pollution totale. Enfin, à partir des données acquises en NTS sur des composés nouvellement identifiés ou détectés, l'idée de mettre en place une analyse rétrospective en SS a émergé (Alygizakis et al., 2018). Ainsi, l'objectif est d'archiver les données acquises en NTS pour pouvoir les retravailler ultérieurement si de nouvelles données pertinentes concernant un composé venaient à être découvertes. L'historique de ce contaminant pourrait alors être retracé, permettant ainsi de mettre en place des mesures de surveillance et de traitement adaptés si le besoin s'en fait ressentir.

Cependant, le NTS environnemental rencontre encore certains obstacles. Récemment, plusieurs essais inter-laboratoires (EIL) ont été réalisés afin de comparer les méthodes employées par différents laboratoires et les résultats qui en découlent. Le premier en date a été organisé par le réseau NORMAN en Europe avec 18 laboratoires participants, rassemblant 12 pays différents (Schymanski et al., 2015). Cet essai consistait à analyser un même extrait du Danube selon la méthode propre à chaque laboratoire participant, et de fournir le nombre de substances détectées et le nombre de substances identifiées en TS, SS ou NTS. Il a notamment permis de mettre en évidence que les principales différences observées au niveau analytique étaient dues aux choix de paramètres spécifiques aux différents instruments utilisés. La principale source de différences entre les résultats obtenus concernait l'étape de traitement de données. Suite aux observations réalisées sur son premier EIL, le réseau NORMAN a lancé un second essai reposant cette fois sur la comparaison des méthodes de traitement de données. Chaque participant devait fournir une liste de 50 composés d'intérêt détectés dans chaque échantillon fourni, après analyse avec une méthode analytique imposée et une méthode propre au participant. Les premiers résultats visaient à présenter l'ensemble de données et les informations associées à chaque méthode analytique employée. Une

évaluation des données a été réalisée à partir de la détection des étalons d'extraction présents dans les échantillons (Schulze et al., 2021). Les futurs résultats s'appuieront d'une part sur les résultats fournis par chaque laboratoire et, d'autre part, sur la comparaison des sets de données en utilisant une unique méthode de traitement de données.

Un autre essai a également été organisé par l'*Environmental Protection Agency* aux Etats-Unis regroupant près de 30 laboratoires participants (Ulrich et al., 2018). Cet essai s'est déroulé en trois parties. La première reposait sur l'analyse de différents mélanges synthétiques de composés connus par les organisateurs afin d'estimer le nombre de composés dopés qui ont été identifiés correctement par chaque participant dans des matrices simples. La deuxième partie utilisait des extraits de différentes matrices dopées avec des mélanges de composés connus par les organisateurs et différents pour chaque participant. Cette seconde phase permettait d'estimer l'effet matrice en termes d'impact sur le nombre de composés correctement identifiés. Enfin, la dernière partie consistait à analyser des étalons de référence, transmis aux laboratoires dans des plaques multi-puits. Le nombre de composés supplémentaires pouvant être identifiés *a posteriori* (lorsque le laboratoire possède toutes les informations nécessaires sur les composés potentiellement présents) a été estimé par chaque laboratoire à partir de ces étalons. Cet essai a montré, dans le meilleur des cas pour les deux premières phases, un pourcentage de vrais positifs variant entre 19 et 46 %, ce qui signifie une identification correcte pour moins de la moitié des composés présents. Suite à la troisième phase, 60 % des composés dopés ont pu être retrouvés (Sobus et al., 2019), montrant ainsi qu'il subsiste encore des difficultés pour l'identification.

Ces différents EIL et leurs objectifs respectifs montrent qu'il existe encore de nombreuses questions concernant le NTS dans le domaine de l'analyse environnementale. Bien qu'un effort de la communauté soit réalisé afin de fournir des directives générales pour l'application du NTS (Schulz et al., 2019; Soulier et al., 2021), le manque d'uniformité dans les méthodes utilisées est encore souvent mis en avant, notamment à cause d'un manque de connaissances quant à la portée de chaque méthode.

II. Objectifs de la thèse et organisation du manuscrit

Les études en analyse non-ciblée pour l'analyse environnementale montrent que de nombreux efforts sont encore nécessaires, notamment en ce qui concerne les méthodes employées qui restent longues, complexes et difficiles à comparer. L'identification de composés inconnus nécessite de multiples étapes de vérification qui, dans certains cas, ne peuvent aboutir. Ces efforts sont souvent concentrés sur un nombre limité de signaux, laissant le reste inexploité. Pour avoir une vision plus large des molécules analysées dans des échantillons, il est crucial d'évaluer la qualité des informations pouvant être obtenues par les méthodes analytiques choisies.

C'est dans ce contexte qu'a été initié le projet de recherche WaterOmics intitulé « Traquer les micropolluants organiques dans les eaux urbaines par spectrométrie de masse haute résolution : approches omiques, empreintes et indices » et financé par l'Agence Nationale de la Recherche (2017-2022). Cette thèse, réalisée au Laboratoire Eau, Environnement et Systèmes Urbains (Leesu), était intégrée à ce projet et consistait à développer et évaluer différentes méthodes de NTS pour l'analyse d'échantillons d'eaux urbaines. Pour cela, trois grands objectifs ont été définis :

- 1) Développer des méthodes de traitement des données pour faciliter la comparaison d'empreintes.** Pour cela, différents logiciels et algorithmes ont été comparés afin de construire un *workflow* fiable, rapide et adapté aux données issues de l'appareil utilisé. Les limitations et possibilités qu'offrent les différents logiciels fournisseurs et logiciels libres ont ainsi été comparées, mettant en avant le besoin d'une méthode hybride optimisée. Ces points ont été développés dans la **Partie A - Evaluation de méthodes de traitement de données**, en s'appuyant sur la comparaison d'approches SS et NTS, et l'utilisation de différentes bases de données, dans le cadre de l'essai-interlaboratoire organisé par le réseau NORMAN en 2019. Des méthodes de caractérisation globales basées sur les empreintes et les indices dérivés de la HRMS ont ensuite été mises en place afin de permettre la comparaison rapide d'échantillons. Enfin, différentes méthodes statistiques pour la discrimination de marqueurs ont été étudiées dans le but d'identifier des marqueurs de pollution.

- 2) Evaluer des protocoles analytiques pour la caractérisation non-ciblée de diverses fractions de micropolluants organiques.** De nombreuses études en NTS sont appliquées avec un protocole donné. Cependant, les études décrivant la comparaison de différents protocoles analytiques en NTS sont rares dans la littérature et s'appuient généralement sur une approche du type SS. La difficulté de ces comparaisons réside essentiellement sur le manque de critères disponibles pour évaluer les différentes méthodes. En s'appuyant sur la spectrométrie de mobilité ionique, les objectifs de la **Partie B – Comparaison de méthodes analytiques**, étaient de fournir des méthodes de comparaison et de caractériser différents protocoles analytiques. Pour cela, plusieurs méthodes de préparation ont été utilisées afin de diviser l'échantillon en diverses fractions de propriétés physico-chimiques différentes. Celles-ci ont ensuite été analysées en utilisant plusieurs méthodes de séparation chromatographique et différents instruments de spectrométrie de masse à haute résolution. Cette comparaison de protocoles visait à apporter des connaissances quant à l'influence du choix de la méthode mais également quant au comportement des micropolluants organiques.
- 3) Appliquer les méthodes développées à la caractérisation d'eaux urbaines soumises à différents traitements d'oxydation.** Les micropolluants organiques étant relativement mal éliminés dans les eaux usées par les filières de traitement conventionnelles, des procédés complémentaires ont été développés mais nécessitent encore une évaluation. Ainsi les différentes approches et méthodes développées dans les parties précédentes ont été appliquées à la comparaison d'échantillons soumis à des traitements d'oxydation. La **Partie C - Application des méthodes pour la comparaison d'échantillons oxydés** développe ces points qui permettent de déterminer des marqueurs caractéristiques du traitement pour suivre l'évolution des micropolluants organiques et en particulier la formation de produits d'oxydation. Dans un premier temps, deux traitements complémentaires des micropolluants (acide performique et ozonation suivie d'une adsorption sur charbon actif) dans les eaux résiduaires urbaines ont été étudiés, notamment en mettant en place des outils complémentaires de traitement de données appliqués aux procédés d'oxydation. Ces outils ont ensuite été réutilisés afin de comparer différentes stratégies analytiques sur

la base du type et de la quantité d'informations qu'elles apportaient concernant la problématique des produits de transformation.

Partie A : Evaluation de méthodes de traitement de données

I. Introduction et objectifs

Un des objectifs principaux du projet ANR WaterOmics (2017-2022) était de développer des méthodes de traitement de données pour faciliter la comparaison d'empreintes HRMS. Suite à l'installation de l'instrument de HRMS (UPLC-IMS-QToF, Vion, Waters) fin 2017 au sein de la Plateforme Régionale d'Analyse Multi-milieux des Micro-Contaminants (PRAMMICS) de l'OSU Efluve (UMS 3563), les premières investigations menées dans le cadre de cette thèse ont été consacrées au développement de méthodes de traitement de données, principalement en s'appuyant sur les logiciels fournis avec l'instrument.

Une étude a, en particulier, été consacrée à la caractérisation par NTS d'eaux de ruissellement urbain et des micropolluants associés au trafic automobile (projet Roulépur, Agence de l'Eau Seine Normandie et Office Français de la Biodiversité). Certains marqueurs (ex. diphénylguanidine) ont ainsi été identifiés dans les eaux de ruissellement et dans des extraits de pneumatiques, et les résultats ont été intégrés dans deux publications :

Fidji Sandré, Nina Huynh, Marie-Christine Gromaire, Gilles Varrault, Christophe Morin, Régis Moillon, Julien Le Roux, Laure Garrigue-Antar. Road Runoff characterization: ecotoxicological assessment combined with (non-)target screenings of micropollutants for the identification of relevant toxicants in the dissolved phase. *Water*, MDPI, 14 (4), pp.511. <10.3390/w14040511>

Johnny Gasperi, Julien Le Roux, Steven Deshayes, Sophie Ayrault, Louise Bordier, Lila Boudahmane, Hélène Budzinski, Emilie Caupos, Nadège Caubrière, Kelsey Flanagan, Martin Guillon, Nina Huynh, Pierre Labadie, Laurent Meffray, Pascale Neveu, Chandirane Partibane, Julien Paupardin, Mohamed Saad, Lucie Varnede, Marie-Christine Gromaire. Micropollutants in Urban Runoff from Traffic Areas: Target and Non-Target Screening on Four Contrasted Sites. *Water*, MDPI, 14 (3), pp.394. <10.3390/w14030394>

En 2019, le réseau NORMAN, à l'initiative du *Norwegian Institute for Water Research* (NIVA), a organisé un essai-interlaboratoire (EIL) à l'échelle européenne, intitulé "*NORMAN Collaborative Trial: Algorithms for deconvolution and library-searching in non-target analysis*".

L'objectif de cet EIL était d'identifier 50 marqueurs représentatifs de chaque échantillon fournis. Pour cela, les laboratoires participants devaient analyser les échantillons suivant deux méthodes : la première correspondait à la méthode NTS interne du laboratoire et la seconde

était une méthode imposée par les organisateurs de l'EIL. Toutefois, en ce qui concerne la méthode de traitement de données, le choix était laissé aux participants d'utiliser les outils et les méthodes qu'ils jugeaient les plus adaptés. Les méthodes employées et les données associées ont fait l'objet d'un article :

Bastian Schulze, Denice van Herwerden, Ian Allan, Lubertus Bijlsma, Nestor Etxebarria, Martin Hansen, Sylvain Merel, Branislav Vrana, Reza Aalizadeh, Bernard Bajema, Florian Dubocq, Gianluca Coppola, Aurélie Fildier, Pavla Fialová, Emil Frøkjær, Roman Grabic, Pablo Gago-Ferrero, Thorsten Gravert, Juliane Hollender, Nina Huynh, Griet Jacobs, Tim Jonkers, Sarit Kaserzon, Marja Lamoree, Julien Le Roux, Teresa Mairinger, Christelle Margoum, Giuseppe Mascolo, Emmanuelle Mebold, Frank Menger, Cécile Miège, Jeroen Meijer, Régis Moilleron, Sapia Murgolo, Massimo Peruzzo, Martijn Pijnappels, Malcolm Reid, Claudio Roscioli, Coralie Soulier, Sara Valsecchi, Nikolaos Thomaidis, Emmanuelle Vulliet, Robert Young, Saer Samanipour. Inter-laboratory mass spectrometry dataset based on passive sampling of drinking water for non-target analysis. *Scientific Data*, Nature Publishing Group, 2021, 8 (1), pp.223. (10.1038/s41597-021-01002-w)

Un objectif majeur de la thèse était de développer des outils complémentaires de traitement de données à partir de logiciels libres et de comparer les possibilités et résultats potentiels avec ceux des logiciels fournis avec l'instrument. L'EIL et les données acquises dans ce cadre ont apporté une occasion intéressante pour la comparaison d'outils de traitement de données. En effet, les résultats initialement obtenus à partir des logiciels fournis avec l'instrument n'étaient pas satisfaisants (molécules censées être présentes parmi les 50 marqueurs sélectionnés) et le besoin de méthodes complémentaires et plus fiables a rapidement émergé.

Dans cette partie, une revue de la littérature décrit dans un premier temps la diversité des méthodes de traitement de données en NTS environnemental, la multiplicité des possibilités et approches (ex. en termes de méthodes de priorisation, bases de données, logiciels) et les différents résultats qui peuvent être obtenus en fonction des choix effectués.

Nous nous attacherons ensuite à décrire la méthode de traitement de données initialement utilisée afin de fournir les résultats aux organisateurs de l'EIL. Par la suite, les limites identifiées de cette méthode seront expliquées et les moyens mis en œuvre pour optimiser la méthode seront détaillés. Cette analyse fine des points faibles de la méthode issue du logiciel fournisseur aura permis le développement d'un *workflow* et d'une application de traitement

de données applicables aux échantillons analysés au laboratoire et utilisables avec les données acquises par le logiciel fournisseur et des données au format universel. En particulier, un mélange de standards analytiques fourni pour l'EIL a été employé afin de comparer la fiabilité des identifications effectuées en NTS par le logiciel fournisseur avec celles des logiciels libres. Cette comparaison avait pour finalité d'identifier les sources de différences entre ces méthodes. Par ailleurs, l'apport de la mobilité ionique dans ce processus a aussi été évaluée afin de confirmer l'utilité de ce paramètre de séparation complémentaire.

II. État de l'art

II.1. Introduction

Les données de NTS sont acquises par des appareils de HRMS, couplés ou non à des technologies séparatives en amont de l'introduction de l'échantillon. Ces appareils peuvent être vendus par différents fabricants, qui, lors de l'achat, fournissent également un logiciel pour piloter l'appareil. Ce logiciel, qui sera par la suite appelé « logiciel fournisseur », sert à gérer l'acquisition des données, mais peut également servir à traiter les données.

Parmi les logiciels fournisseurs régulièrement retrouvés dans la littérature pour traiter les données de NTS, on peut notamment citer *MassHunter* d'Agilent (Causon et al., 2018; Sobus et al., 2019), *Data Analysis* de Bruker (Bergé et al., 2018; Gago-Ferrero et al., 2015), *SIEVE* (Moschet et al., 2013; Zonja et al., 2015), *Compound discoverer* (Huysman et al., 2018) ou *XCalibur* de ThermoFisher et *Chromalynx XS* (Diaz et al., 2013; Nurmi et al., 2012), *UNIFI* (Baz-Lomba et al., 2016; Bijlsma et al., 2017) ou *Progenesis QI* (Lazzaro et al., 2019) de Waters. Ces logiciels, bien que différents les uns des autres, permettent de réaliser directement toutes les étapes de prétraitement des données acquises. Certains proposent également des options pour le traitement et la visualisation des données.

Le prétraitement des données a pour but de rendre les données lisibles par l'utilisateur et passe notamment par la détection des pics et des étapes permettant de filtrer les données pour ne garder que des données exploitables. Le traitement des données désigne, quant à lui, les étapes suivantes qui servent à donner un sens à ces données. Ces étapes regroupent à la fois différents tests statistiques pour la comparaison d'échantillons ou la priorisation de signaux, mais aussi l'utilisation d'outils permettant l'identification de pics d'intérêt (comme

par exemple le calcul de formules moléculaires, la comparaison à des bibliothèques spectrales...). Le Tableau 1 donne un aperçu des fonctionnalités de ces logiciels et les formats de données compatibles (informations disponibles sur les sites internet ou articles des fournisseurs). En effet, chaque logiciel génère des données dans un format spécifique aux fournisseurs. Cependant, dans certains cas, le logiciel fournisseur propose également des formats différents qui peuvent être compatibles avec d'autres logiciels ou plateformes de traitement de données.

Tableau 1 : Description des logiciels fournisseurs courants

Logiciel	Fournisseur	Fonctionnalités
Chromalynx XS	Waters	<p><u>Prétraitement</u> : Soustraction du bruit, extraction des spectres</p> <p><u>Traitement</u> : Calcul de formules moléculaires, comparaison avec librairies et attribution d'un score basé sur l'erreur en masse et la distribution isotopique, comparaison de chromatogrammes deux à deux (pics communs et spécifiques)</p> <p><u>Format et compatibilité</u> : Utilisable à partir de données brutes acquises sous MassLynx (*.raw), Waters</p>
Compound discoverer	ThermoFisher	<p><u>Prétraitement</u> : Extraction des spectres, alignement en TR, groupement des isotopes, adduits et groupement sur tous les échantillons, remplissage des valeurs manquantes, normalisation des aires et soustraction du bruit</p> <p><u>Traitement</u> : Choix automatique du meilleur spectre parmi les pics groupés puis prédiction de formule moléculaire et comparaison avec une librairie (MS1 : Chempider, MS2 : mzCloud). Assignment par similarité entre spectres acquis et spectre en librairie. Représentation d'empreintes, analyse des tendances, défaut de masse, analyses multivariées (AMV), analyse par regroupement hiérarchique, réseaux moléculaires</p> <p><u>Format et compatibilité</u> : *.mgf, *.mzML, *.mzDATA, *.csv</p>
Data Analysis	Bruker	<p><u>Prétraitement</u> : Extraction des spectres, soustraction du bruit et des blancs</p> <p><u>Traitement</u> : calcul de formules moléculaires</p> <p><u>Format et compatibilité</u> : Utilisable à partir de données brutes acquises sur d'autres appareils Bruker</p>
MassHunter	Agilent	<p><u>Prétraitement</u> : Extraction des spectres, normalisation, filtration, intensités détectées et relatives</p> <p><u>Traitement</u> : Représentation d'empreintes, ANOVA, AMV, analyse par regroupement hiérarchique, calcul de masses de Kendrick, calcul de formules moléculaires, recherche en librairies (pour comparaison de spectres : librairie interne, METLIN, NIST..., sinon : Chempider et Pubchem + fragmentation in-silico) et score basé sur la corrélation des spectres expérimentaux et spectres en librairie.</p> <p><u>Format et compatibilité</u> : Différents formats compatibles et export possible pour traitement sous R</p>
Progenesis	Waters	<p><u>Prétraitement</u> : Extraction des pics, alignement et normalisation automatique, contrôle de faux positifs</p> <p><u>Traitement</u> : Abondance relative, statistiques univariés, AMV, recherche en librairies (Chempider, Metlin, Metaboscope, LipidBlast, Elemental Composition). Classement par erreur en masse, fragmentation, distribution isotopique, TR, CCS</p>

PARTIE A : Évaluation de méthodes de traitement de données

Logiciel	Fournisseur	Fonctionnalités
		<u>Format et compatibilité</u> : Compatible avec « tous les formats courants »
SIEVE	ThermoFisher	<p><u>Prétraitement</u> : Extraction des spectres, soustraction du bruit, groupement des adduits et isotopes, alignement, normalisation, filtres. Calcul de formules moléculaires et recherche en librairies (Chempider, KEGG)</p> <p><u>Traitement</u> : Représentation d'empreintes, AMV, analyse par regroupement hiérarchique, analyse des tendances (par groupe d'échantillon)</p> <p><u>Format et compatibilité</u> : N.A.</p>
UNIFI	Waters	<p><u>Prétraitement</u> : Extraction des spectres, regroupement des isotopes, alignement, normalisation</p> <p><u>Traitement</u> : Comparaison de chromatogrammes deux à deux, calcul de formules moléculaires (comparaison avec librairies internes ou Chempider + fragmentation in-silico). Scores basés sur distribution isotopique et correspondance des spectres expérimentaux et in-silico. AMV possibles via EZInfo.</p> <p><u>Format et compatibilité</u> : *.uep et données brutes issues de MassLynx (*.raw).</p>

Les logiciels fournisseurs ne permettant pas toujours à l'utilisateur d'adapter les paramètres à ses besoins ou même de connaître les paramètres utilisés par le logiciel, les étapes réellement effectuées ou leur influence sur les données, de nombreux utilisateurs s'orientent vers des formats ouverts. Ainsi, une pratique courante consiste à exporter les données brutes acquises puis de les convertir sous un format libre (mzML ou mzXML) qui permet le stockage et l'échange de données issues de la spectrométrie de masse (Martens et al., 2011). Ce type de format inclut les données brutes acquises, avec les spectres, les chromatogrammes et les métadonnées correspondantes. Cette conversion est parfois proposée directement en option dans le logiciel d'acquisition de certains fournisseurs, mais elle peut aussi être effectuée en utilisant des plateformes de conversion dédiées, la plus connue et utilisée étant ProteoWizard (Kessner et al., 2008). Nombre d'utilisateurs développent ensuite leurs propres scripts afin de prétraiter leurs données, résultant en un grand nombre de *packages* développés à ces fins et diffusés sur des plateformes de partage pour toute la communauté. Les plus connus sont *xcms* (Smith et al., 2006), *nontarget* (Loos, 2016) et plus récemment *patRoan* (Helmus et al., 2021), développés sous R. De nombreuses interfaces graphiques ont également été développées pour permettre à des utilisateurs de traiter leurs données sans utiliser leur logiciels fournisseurs et sans nécessairement savoir programmer. Parmi les plus utilisés, on retrouve MZMine et MZMine2 (Katajamaa and Orešič, 2005; Pluskal et al., 2010), *xcms Online* (Tautenhahn et al., 2012) et *EnviMass* (Loos et al., 2018).

II.2. Méthodologie pour le traitement de données

II.2.1. Workflow général pour le pré-traitement

Les méthodes de traitement de données utilisées pour le NTS environnemental sont inspirées des méthodologies employées dans le domaine de la métabolomique (Figure 3).

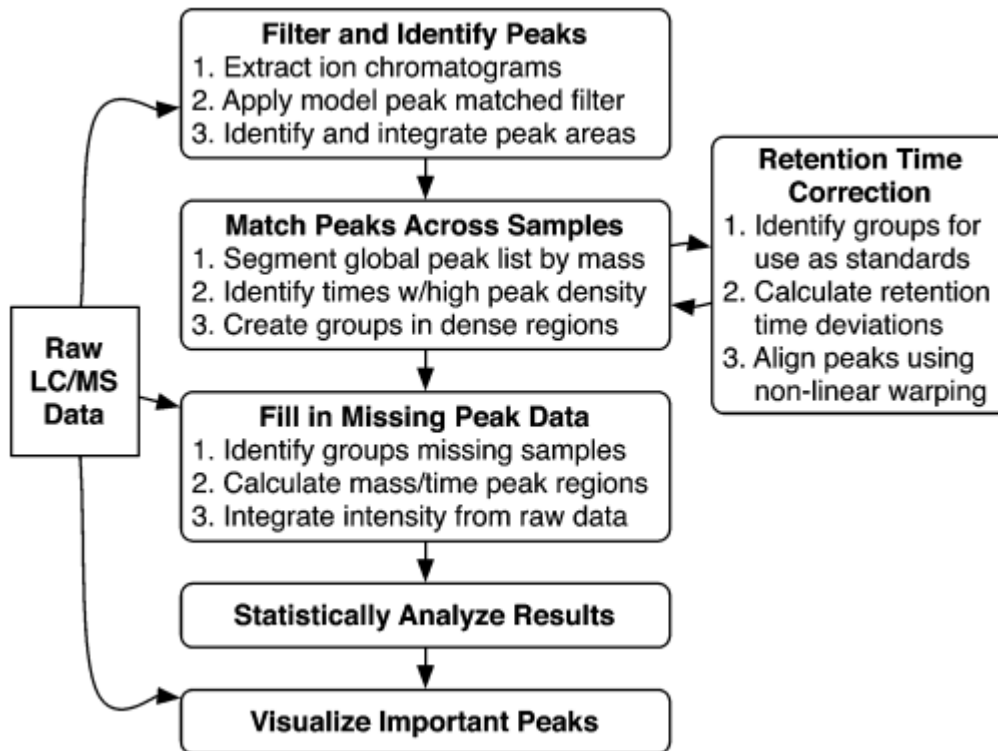


Figure 3 : Description du *workflow* général pour le prétraitement de données, selon (Smith et al., 2006)

La première étape après l'acquisition des données consiste à rendre la grande quantité de données obtenues lisible de manière simple. En effet, les données brutes consistent en une série de signaux détectés pour chaque échantillon parmi une liste d'échantillons. Il s'agit alors de faire en sorte de relever les informations pertinentes de chaque échantillon et de regrouper en un fichier unique les données de tous les échantillons afin de pouvoir comparer ces informations.

La détection de pics est réalisée en repérant des pics sur la base de leur amplitude, ou de leur correspondance avec une forme gaussienne (Bauer et al., 2011; Du et al., 2006). Cette étape est souvent suivie d'une étape de centroïdisation des données. Cette dernière consiste à transformer les pics gaussiens détectés en un unique point de ratio m/z dont l'intensité correspond à la normalisation de l'aire du pic considéré (Hohrenk et al., 2019). A ces différents

ratios m/z , détectés dans chaque échantillon, sont ensuite associés une intensité correspondante et un TR si un couplage chromatographique a été utilisé. Par la suite, pour pouvoir comparer les échantillons entre eux, il faut aligner les pics des différents échantillons entre eux en déterminant des fenêtres de m/z , de TR (et de *drift time* quand cette donnée est disponible) pour faire correspondre les signaux semblables sur tous les échantillons (Boccard et al., 2010). Les signaux n'étant pas retrouvés dans un échantillon donné se verront alors attribuer une valeur définie. Pour cela, différentes méthodes existent. Certains auteurs remplacent les valeurs manquantes par des zéros, alors que d'autres préféreront déterminer une valeur de limite de détection fictive qui peut correspondre par exemple à la plus faible valeur obtenue pour un pic donné ou à une valeur seuil d'intensité pour tous les signaux, fixée préalablement (Bader et al., 2017). Certaines études ont également distingué différents types de valeurs manquantes qui seront remplacées de différentes manières en fonction de leur type (Delaporte et al., 2019).

Dans l'étape suivante, les adduits, isotopes et fragments correspondants à une même molécule sont regroupés : c'est la « *componentization* ». Pour cela, le regroupement des isotopes appartenant à la même molécule est effectué en repérant les pics co-élués, séparés par une certaine distance en termes de m/z . En ce qui concerne les adduits, le principe est le même et se base également sur des différences de m/z en considérant tous les adduits possibles (Kuhl et al., 2012; Treutler and Neumann, 2016). Cette étape n'est pas toujours réalisée à ce stade, elle n'intervient parfois qu'à la fin du processus.

Une étape de filtration est ensuite souvent nécessaire afin de ne considérer pour la suite que les signaux se rapportant à l'échantillon. Différentes méthodes peuvent être employées en fonction des objectifs de l'étude. Trois grandes méthodes sont distinguées :

- la soustraction d'un échantillon "blanc", qui consiste à soustraire toutes les intensités des pics présents dans l'échantillon "blanc", aux pics présents dans l'échantillon (Schollée et al., 2018)
- l'exclusion du blanc, qui élimine tous les pics de l'échantillon qui ont été détectés dans le blanc, peu importe leur intensité (Li et al., 2018)
- la prise en compte des pics communs entre le blanc et l'échantillon uniquement lorsque le ratio des intensités du pic dans l'échantillon par rapport au pic du blanc est supérieur à un seuil fixé préalablement (Peter et al., 2019)

Il existe plusieurs types d'échantillons "blanc" :

- le blanc solvant (*blank*) qui est une injection du solvant dans lequel se trouve l'échantillon et qui correspond généralement à la phase mobile employée pour la séparation chromatographique
- le blanc de préparation qui est un échantillon d'eau ultrapure qui a subi toutes les étapes de traitement de l'échantillon réel (filtration, extraction...)
- le blanc échantillon (*blind*), beaucoup plus difficile à obtenir, puisqu'il s'agit d'un échantillon représentatif de la matrice et ayant également subi toutes les étapes de traitement de l'échantillon réel

Ainsi, dans le cas de solutions synthétiques dopées, un échantillon *blind* est une solution non dopée. En revanche, pour un échantillon réel, un échantillon *blind* est presque impossible à obtenir (Schollée et al., 2016). Quelques exemples d'utilisation d'échantillon *blind* pour des échantillons réels peuvent être donnés dans des cas particuliers comme l'étude de produits de transformation où l'échantillon *blind* est celui avant traitement (Park and Snyder, 2020). Un autre cas est l'étude de l'évolution de la pollution le long d'un cours d'eau où le *blind* est un échantillon prélevé juste en amont du point considéré (Ruff et al., 2015).

Une fois ces étapes réalisées, toutes les données peuvent être regroupées sous forme d'un tableau contenant un identifiant unique pour chaque pic (« marqueur ») ou groupement de pics. Ce tableau contient pour chaque marqueur les valeurs de ratio m/z , de TR, éventuellement de *drift time* ou CCS, et l'intensité de ce marqueur dans chaque échantillon. Dans la suite, ce tableau sera nommé « *marker table* ».

Une normalisation peut ensuite être appliquée sur les différents marqueurs. Ici encore, différentes méthodes peuvent être employées. Par exemple, certains auteurs normalisent les signaux par rapport à l'intensité totale de l'échantillon (Chen et al., 2009), d'autres préfèrent normaliser tous les marqueurs par rapport à un étalon unique ou un mélange d'étalons internes. Dans ce dernier cas, la normalisation peut se faire pour chaque marqueur individuellement en fonction de l'étalon qui engendra le coefficient de variation le plus faible pour ce marqueur à travers tous les échantillons de contrôle injectés (van der Kloet et al., 2009) ou en prenant pour chaque marqueur, l'étalon le plus proche en termes de TR pour effectuer la correction (Albergamo et al., 2019; Veenaas et al., 2018). L'échantillon de

contrôle, généralement nommé « *pool* » est un mélange de volume égal de tous les échantillons analysés. Cette approche est couramment utilisée dans le domaine de la métabolomique et plus généralement en NTS (Godzien et al., 2015). Les différentes utilisations de ces *pools* pour corriger les signaux sans avoir recours à des étalons internes ont été décrites et comparées dans le domaine de la lipodomique (Drotleff and Lämmerhofer, 2019), mais les observations peuvent être transposables aux autres domaines faisant appel à des approches omiques.

En général, les étapes du prétraitement sont relativement peu détaillées dans les articles relatifs au NTS. Les étapes énumérées précédemment sont parfois évoquées mais avec relativement peu d'informations concernant les paramètres utilisés, la nature des blancs, leur utilisation et la manière de corriger les données. Des efforts ont été fournis seulement récemment afin de recenser les différentes méthodes utilisées et de proposer des recommandations en fonction du type de données (Schulze et al., 2020). De nombreuses façons de prétraiter les données sont donc disponibles. Cependant, une étude a montré qu'en fonction de la manière utilisée pour traiter les données, les résultats peuvent être totalement différents. En effet, sur un même set de données, les auteurs ont utilisé différents logiciels en choisissant des paramètres comparables pour chaque étape du prétraitement. Ils ont mis en évidence seulement 10 % de recouvrement entre les quatre approches utilisées et jusqu'à 50 % de pics uniques pour l'une d'entre elles. Les sources des différences observées sont liées principalement au format d'entrée des données (en mode *centroïde* ou *profile*) et à l'étape de filtration par rapport aux réplicats qui est réalisée soit après alignement des différentes injections, soit directement en traitant les réplicats d'un même échantillon comme un groupe dès l'étape de détection des pics (Hohrenk et al., 2019).

II.2.2. Stratégies de traitement des données

Une fois une liste de marqueurs pertinents obtenue, il existe plusieurs stratégies permettant d'exploiter les données recueillies en fonction des objectifs poursuivis par l'étude. Les comparaisons globales permettent d'étudier les échantillons sans passer par une étape préalable de priorisation de marqueurs. Par exemple, la comparaison binaire permet de comparer deux échantillons ou groupes d'échantillons entre eux. Il s'agit ici de considérer les échantillons dans leur ensemble et de répertorier, généralement, le nombre et la nature des marqueurs retrouvés dans ces échantillons. Cette méthode permet d'avoir une vue

d'ensemble du comportement des différents marqueurs ; elle a notamment été utilisée pour évaluer l'efficacité de procédés de traitement en calculant des pseudos taux d'abattement basés sur des intensités relatives (Bader et al., 2017; Nürenberg et al., 2019). Les diagrammes de Venn ou d'Euler sont également des représentations très utiles qui permettent de comparer des échantillons ou groupes d'échantillons (Müller et al., 2011). Le principe reste le même puisque ces diagrammes permettent de montrer le nombre de marqueurs retrouvés dans chaque groupe d'échantillon, le nombre de marqueurs communs et le nombre de marqueurs spécifiques à chacun. Ces méthodes ont l'avantage de permettre une comparaison globale très rapide entre différents échantillons. Seulement, lorsqu'il s'agit d'identifier des molécules, elles ne suffisent plus. En effet, le temps nécessaire à l'assignation d'une identité à un marqueur étant déjà relativement long, il est impossible de tenter d'identifier tous les marqueurs détectés dans un ou plusieurs échantillons. Différentes stratégies de priorisation peuvent donc être adoptées pour sélectionner des marqueurs considérés comme pertinents pour l'étude.

Il existe ainsi des méthodes permettant de repérer des signaux en fonction de propriétés particulières. Par exemple, les représentations de Kendrick peuvent être utilisées pour identifier des familles de molécules ayant une base moléculaire commune, à laquelle un groupement moléculaire est ajouté en convertissant les masses exactes vers une nouvelle échelle qui est spécifique groupement ajouté précédemment mentionné (Hughey et al., 2001). Initialement, Kendrick avait développé ces types de graphiques pour grouper ensemble des familles d'hydrocarbures auxquelles sont ajoutés des groupements $-CH_2$ qui différencient les membres d'une même famille (Kendrick, 1963). En partant de ce principe, différentes échelles ont été explorées (Sleno, 2012), notamment la substitution d'un atome d'hydrogène par un atome de chlore dans le domaine de l'analyse environnementale (Myers et al., 2014). De la même manière, lorsque les formules moléculaires des signaux sont disponibles, les diagrammes de van Krevelen peuvent compléter ces représentations de Kendrick et permettre ainsi de distinguer les molécules en fonction de plusieurs combinaisons de ratio atomiques ce qui augmente la différenciation des compositions élémentaires (Wu et al., 2004). Enfin, dans le domaine de la métabolomique, les réseaux moléculaires sont de plus en plus utilisés (Figure 4). Ceux-ci permettent de classer des molécules en fonction de leur ressemblance structurale et donc de leurs spectres de masse. Ces réseaux permettent, par la suite, de

sélectionner et de tenter d'associer une structure à un composé inconnu, de structure et de comportement proche d'un composé déjà connu, ou par comparaison de deux réseaux moléculaires dont le fonctionnement de l'un est déjà connu (Hautbergue et al., 2019).

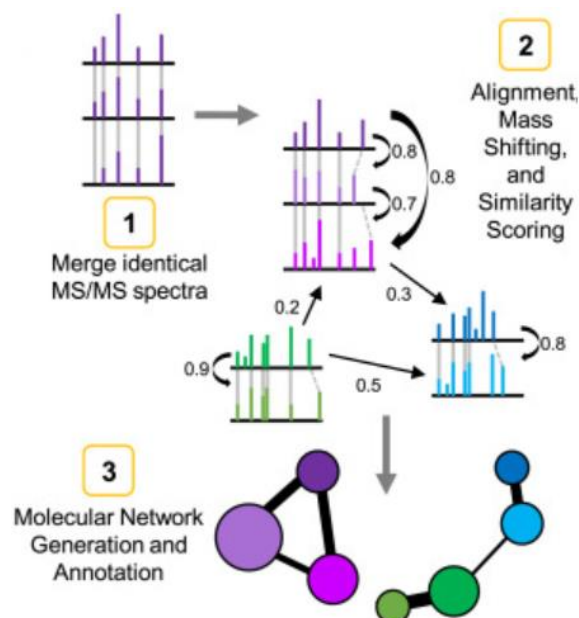


Figure 4 : Principe des réseaux moléculaires, selon (Fisher et al., 2021)

Ces trois méthodes peuvent être utilisées de façon globale, c'est-à-dire afin de comparer les empreintes de deux types d'échantillons, ou peuvent servir à prioriser des signaux intéressants pour une éventuelle identification. En effet, à partir d'une unique identification, les liens établis précédemment ou les indices concernant la composition permettront de déduire l'identité des autres signaux.

En termes de priorisation à proprement parler, certains auteurs choisissent de prioriser les marqueurs en fonction de leur distribution isotopique. En effet, certains atomes ont des distributions isotopiques très distinguables. L'intérêt de ce type de priorisation repose sur le même principe que les trois méthodes décrites précédemment puisque l'identification postérieure est facilitée par ces indices concernant certains atomes. Ainsi la présence d'atomes de soufre ou d'azote peut permettre une identification simplifiée de marqueurs (Schymanski et al., 2014b), mais la distribution isotopique la plus distinguable et, de fait, la plus fréquemment utilisée est celle du chlore (Cariou et al., 2016; Hug et al., 2014; Ruff et al., 2015). Une autre méthode consiste à sélectionner des marqueurs en fonction de tendances temporelles ou spatiales. Il s'agira ici de sélectionner des marqueurs qui ont une tendance particulière entre plusieurs échantillons. La priorisation de marqueurs dont l'intensité diminue

ou augmente de manière continue est avantageuse pour l'étude d'une cinétique de transformation (Plassmann et al., 2016, 2018). En revanche, pour une étude de déversements accidentels, les marqueurs ayant une variation très forte sur un court laps de temps seront prioritaires (Alygizakis et al., 2019). De la même manière, la variation spatiale peut être exploitée, comme lors de l'étude des micropolluants polaires dans les rejets des STEU du Rhin. Dans cette étude, les auteurs ont mis en place un système de soustraction successive de marqueurs déjà présents dans le site analysé en amont. Seuls les « nouveaux » pics à une localisation donnée sont conservés pour l'identification (Ruff et al., 2015). Une autre manière d'exploiter la variation spatiale a été utilisée pour différencier des sites de prélèvement. En utilisant la fréquence de détection d'un marqueur donné et son intensité dans chaque échantillon, un score de rareté a été assigné aux différents marqueurs. Une telle approche permet ainsi de prioriser, pour un site donné, les marqueurs qui le différencient le plus des autres sites (Krauss et al., 2019).

En parallèle, ou pour supporter les méthodes précédentes, de nombreux tests statistiques sont régulièrement employés. Parmi les tests univariés, on retrouve le test de corrélation de Spearman (Plassmann et al., 2016) ou les tests d'autocorrélation (Peters et al., 2010) qui ont été utilisés dans des cas faisant appel à une étude de la variation temporelle. La variation d'intensité seule est aussi régulièrement exploitée et permet de classer des marqueurs en différents groupes en fonction de leur variation d'intensité d'un échantillon à un autre (Bader et al., 2017; Park and Snyder, 2020). Enfin le t-test est sans doute le test univarié qui revient le plus régulièrement et sert à vérifier que les intensités d'un marqueur donné dans deux échantillons ou deux sets d'échantillons sont significativement différents l'un de l'autre (Negreira et al., 2015; Schollée et al., 2015). Parmi les tests multivariés, celui qui revient le plus souvent est l'analyse par composante principale (ACP) qui permet de distinguer de manière non supervisée des groupes d'échantillons (Samanipour et al., 2017; Sanchís et al., 2020; Schollée et al., 2018) mais aussi d'assurer un contrôle qualité sur les données acquises (Godzien et al., 2015). Différentes revues des méthodes statistiques et de leur utilité ont été réalisées dans le domaine de l'analyse de données (Ali and Bhaskar, 2016) ou de la métabolomique (Boccard et al., 2010). Celles-ci résument également le travail mené en NTS environnemental pour lequel une revue plus spécifique est aussi disponible (Schollée et al., 2016). Enfin, une dernière méthode récurrente est l'utilisation de tests toxicologiques pour

permettre de prioriser des marqueurs de dangerosité connue, aussi appelée analyse dirigée par les effets. Cette méthode passe par des tests toxicologiques sur des fractions d'un échantillon donné et une analyse chimique en parallèle. Les marqueurs se trouvant dans les fractions produisant des effets peuvent ainsi être considérés comme dangereux (Brack, 2003; Smital et al., 2013).

Les différentes méthodes présentées ci-dessus sont souvent utilisées les unes après les autres afin de réduire au maximum la liste de marqueurs et obtenir un set de données de taille raisonnable pour l'identification. Cette étape de réduction permet de sélectionner des marqueurs d'intérêt pour la problématique de l'étude menée. Les méthodes permettant la sélection de marqueurs pour une identification ultérieure sont souvent basées sur l'une des deux stratégies suivantes :

- Sélection de marqueurs liés de façon structurelle avec de nombreux autres marqueurs (comme pour la méthode de Kendrick ou celle des réseaux moléculaires). L'identification de ces marqueurs permettra de déduire l'identité des autres marqueurs auxquels ils sont liés
- Sélection de marqueurs suivant des critères (distribution isotopique, tendances ou discrimination). L'identification est réalisée avec des *a priori* plus ou moins importants, qui permettent d'éliminer des attributions aberrantes

II.3. Difficultés d'identification

II.3.1. Stratégies pour l'identification

Une fois qu'une liste de marqueurs prioritaires a été obtenue, l'étape suivante consiste à tenter de leur assigner une identité. Cela passe par plusieurs étapes de confirmation correspondant à des niveaux de confiance, les plus utilisés ayant été décrits précédemment (Schymanski et al., 2014a, 2015) (Figure 2).

Dans un premier temps, il s'agit d'assigner une formule moléculaire unique au marqueur considéré. L'étape suivante consiste à attribuer une ou plusieurs structures possibles pouvant correspondre à cette formule brute (passage du niveau 4 au niveau 3 – Figure 2). Cette étape peut nécessiter de mobiliser les connaissances de l'utilisateur concernant l'échantillon étudié, lui permettant de déterminer quelles familles ou quelles classes de molécules peuvent potentiellement être présentes dans l'échantillon. Cette étape peut également s'appuyer sur les données de fragmentation qui peuvent donner des indications quant à la structure de la molécule. Par la suite, le passage du niveau 3 vers le niveau 2 s'appuie sur des bases de données internes ou des bases disponibles en libre accès (sur les logiciels constructeurs ou en ligne). Cette étape permet de comparer les données acquises avec celles disponibles par ailleurs afin d'associer une unique structure à la formule brute. Plusieurs problèmes se posent ici. Tout d'abord, les bases de données en LC sont bien moins universelles que les bases de données en GC. Ces données sont très dépendantes des appareils grâce auxquels les données ont été acquises et, de fait, toutes, ne sont pas comparables. Enfin, il existe de nombreuses bases de données accessibles et il est impossible de toutes les consulter. Ainsi, le choix de la pertinence de l'utilisation ou non d'une librairie est crucial. Enfin, le passage du niveau 2 au niveau 1 est permis lorsque le candidat retenu correspond bien à la masse d'intérêt initiale. Cela n'est possible que lorsque les standards analytiques sont accessibles, ce qui n'est pas toujours le cas. **Ainsi, les études dont le but est d'identifier de nouvelles molécules requièrent de nombreuses étapes, d'hypothèses et de confirmations qui sont très longues et peuvent ne pas aboutir à une confirmation complète.**

II.3.2. Calcul de formules brutes

La première étape suivant la sélection d'un nombre limité de marqueurs, est de calculer une formule brute unique pouvant correspondre à ce marqueur (passage du niveau 5 au niveau 4). Cette étape repose sur la précision de la détection en masse et la haute résolution des appareils utilisés. Malgré cela, il n'est pas rare que pour une masse donnée, plusieurs formules puissent être plausibles. Une étude a permis de simuler le nombre de formules brutes assignées en fonction de la précision en masse choisie par l'opérateur et la masse détectée. Les auteurs montrent ainsi, que pour un m/z détecté à 500 avec une précision de 1 mDa, 200 formules brutes sont possibles en se restreignant à 6 atomes élémentaires (Miura et al., 2010). La majorité des auteurs se restreignent en effet aux éléments suivants : carbone (C), hydrogène (H), oxygène (O), azote (N). Certains ajoutent du phosphore (P) ou du soufre (S), et, dans le cas des études de MP en matrice aqueuses l'ajout du chlore (Cl) et du brome (Br) est assez fréquent. Pour réduire le nombre de possibilités et proposer des formules dont l'existence est chimiquement possible, un ensemble de règles, connues sous le nom de *Seven Golden Rules* (Kind and Fiehn, 2007), est souvent utilisé. Celles-ci ont été développées sur plus de 68 000 formules brutes existantes et vérifiées sur plus de 430 000 composés montrant que seulement 0,6 % d'entre elles ne respectent pas ces règles.

Deux règles élémentaires, précédemment décrites (Pellegrin, 1983), sont préalablement utilisées avant l'application des *Seven Golden Rules*. La première concerne le « *double bond equivalent* », degré d'insaturation qui permet de déterminer le nombre de liaisons possible à partir des éléments de la formule brute afin de déterminer l'existence d'une telle structure. Le degré d'insaturation est également associé à la règle de l'azote qui stipule qu'une masse nominale impaire implique un nombre d'atomes d'azote impair. Ensuite, la première règle des *Seven Golden Rules* concerne la restriction en termes de nombre d'atomes individuels dans une formule brute. Ce nombre correspond au nombre entier résultant de la division de la masse nominale considérée par la masse nominale de l'atome choisi. La seconde règle correspond aux règles chimiques fondamentales de Lewis et Senior et concernent donc la valence. La troisième règle s'appuie sur la distribution isotopique de la molécule et stipule que cette dernière doit être semblable à l'abondance isotopique naturelle des atomes entrant dans la composition de la formule. La quatrième règle s'intéresse à des ratios entre atomes permettant de déterminer l'existence ou non d'une formule moléculaire calculée. Elle est

complétée par la cinquième règle qui estime les nombres maximums d'hétéroatomes individuels différents dans une formule moléculaire en s'appuyant sur des molécules connues, et la sixième règle qui restreint le nombre total d'hétéroatomes dans la formule. Enfin, la dernière règle concerne spécifiquement l'analyse de petites molécules en GC-MS, puisqu'elle traite de la dérivation en rappelant l'importance de la soustraction du groupe utilisé pour la dérivation avant l'application des autres règles.

En parallèle, un ensemble de règles spécifiques à la matière organique naturelle (MON) a également été mis en place et repris dans les études dédiées à la MON (Koch et al., 2007). Ces règles sont globalement basées sur les mêmes principes que ceux évoqués précédemment, sauf que les éléments sont généralement restreints à C, H, N, O (avec dans certains cas, l'ajout de P et S) et les ratios entre ces différents atomes varient légèrement. Outre ces règles, la distribution isotopique joue également un rôle important dans l'assignation d'une formule brute (Thurman and Ferrer, 2010). En effet, après avoir calculé les formules moléculaires possibles, il est intéressant de prédire le spectre de masse de cette formule brute afin de le comparer au spectre acquis afin de confirmer la correspondance. Cette prédiction s'appuie sur la masse exacte et l'abondance naturelle des isotopes de chaque atome entrant dans la composition de la formule brute considérée. Généralement, seuls les isotopes dont l'abondance naturelle est supérieure à 1.10^{-6} % sont considérés. Ces deux paramètres permettent ainsi de placer les pics théoriques (par écart de masse) et d'estimer leur intensité relative respectivement (Roussis and Proulx, 2003). Les abondances naturelles des isotopes majoritaires des atomes les plus régulièrement étudiés, ainsi que leurs masses exactes, sont fournies dans le Tableau 2.

Tableau 2 : Isotopes et abondances des principaux atomes étudiés

Isotope	Masse isotopique	Abondance naturelle (%)
¹² C	12,000000	98,93
¹³ C	13,003355	1,07
¹ H	1,007825	99,98
¹⁴ N	14,003074	99,64
¹⁵ N	15,000109	0,36
¹⁶ O	15,994915	99,76
¹⁷ O	16,999132	0,04
¹⁸ O	17,999161	0,20
³¹ P	30,973762	100
³² S	31,972071	94,93
³³ S	32,971459	0,76
³⁴ S	33,967867	4,29
³⁶ S	35,967081	0,02
³⁵ Cl	34,968853	75,76
³⁷ Cl	36,965903	24,24
⁷⁹ Br	78,918337	50,69
⁸¹ Br	80,916291	49,31

Enfin, lorsque les données de fragmentation MS/MS sont disponibles, un degré de confirmation supplémentaire peut être atteint. Sur le même principe, les fragments acquis seront comparés à des fragments théoriques pour déterminer si la formule assignée est possible. Les manières d'obtenir ces derniers seront détaillés par la suite, puisque les données de MS/MS permettent généralement de confirmer l'identité d'une structure, plutôt qu'une formule moléculaire seulement.

Le Tableau 3 regroupe les algorithmes et logiciels les plus souvent retrouvés pour l'assignation de formules moléculaires à des données de HRMS, hors logiciels fournisseurs qui ont déjà été décrits dans le Tableau 1.

Tableau 3 : Description de différents logiciels permettant l'assignation de formules brutes

Nom	Langage	Système d'exploitation	Principe et format des données	Référence
Composer			<p><u>Principe</u> : Calcul de formules théoriquement possibles et rapprochement avec base de données contenant des extraits pétroliers. Les formules moléculaires se rapprochant le plus d'hydrocarbures sont considérées plus « justes ».</p> <p><u>Format entrée</u> : A partir des données brutes en profile des principaux fournisseurs. Elles seront centroidées et recalibrées directement par Composer.</p> <p><u>Format sortie</u> : Données exportables en *.xml ou en *.xlsx, avec compositions, assignement des spectres et distributions.</p>	(Sierra Analytics)
PetroOrg			<p><u>Principe</u> : N.A</p> <p><u>Format entrée</u> : Données FT-ICR : Liste de masse en *.txt à partir de données centroidées ou profile</p> <p><u>Format sortie</u> : Tableau avec différentes formules proposées et score selon l'erreur en masse, double liaisons et abondance relative isotopes en *.csv</p>	(Corilo)
WHOI code	Matlab	Windows, Linux, Mac	<p><u>Principe</u> : Etablissement de liens entre les différentes masses détectées (par addition ou soustraction des groupes fonctionnels suivants H₂, CH₂ et O). Assignation de formules moléculaires dans l'ordre croissant des masses et seulement pour la première de la série groupes fonctionnel (restreint à C, H, N, O, P, S, Na) et attribution de la formule dont l'erreur en masse est la plus faible. Les autres masses de la série sont ensuite automatiquement déterminées à partir de cette 1ère formule brute en ajoutant le groupe fonctionnel déterminé</p> <p><u>Format entrée</u> : FT-ICR, liste de masses monoisotopiques ou monoisotopiques + masses du spectre MS/MS</p> <p><u>Format sortie</u> : N.A.</p>	(Kujawinski and Behn, 2006)

PARTIE A : Évaluation de méthodes de traitement de données

Nom	Langage	Système d'exploitation	Principe et format des données	Référence
Seven golden rules	Visual Basic et C++		<p><u>Principe</u> : Calcul de toutes les formules possibles (restreint aux éléments : C, K, H, D, N, Na, S, O, P, F, Cl, Br et Si, en considérant la valence maximale de ces éléments). Toutes les formules valides (i.e., qui respectent les règles énoncées et qui correspondent en termes d'abondance isotopique et de tolérance en masse) sont comparées à une base de données interne (contenant 432 968 formules tirées de PubChem) et un score de correspondance est fourni pour classer les propositions</p> <p><u>Format entrée</u> : Masse exacte neutre, avec les abondances isotopiques relatives, normalisées à 100%</p> <p><u>Format sortie</u> : *.csv</p>	(Kind and Fiehn, 2007)
Rdisop	R (basé sur DECOMMP en C++ et disponible en Java)	Windows, Linux, Mac	<p><u>Principe</u> : Calcul de toutes les compositions possibles pour les masses entrées, puis filtration de celles dont les formules ne correspondant pas à des nombres corrects de liaisons chimiques ou de degré d'insaturation. Puis comparaison des distributions isotopiques mesurées et théoriques et classement utilisant des statistiques bayésiennes</p> <p><u>Format entrée</u> : Liste de spectres avec abondances normalisées des isotopes</p> <p><u>Format sortie</u> : Tableau avec formule calculée, score et mention valide / non valide en fonction des liaisons et nombre d'insaturation</p>	(Böcker et al., 2008)
MZMine2	Java	Windows, Linux, Mac	<p><u>Principe</u> : Calcul des formules moléculaires possibles et génération des distributions isotopiques correspondantes pour comparaison au spectre acquis. Comparaisons avec des bases de données en ligne (PubChem, KEGG, METLIN, et HMDB)</p> <p><u>Format entrée</u> : *.mzXML, *.mzDATA, *.mzML, *.NetCDF et données brutes de Thermo</p> <p><u>Format sortie</u> : N.A.</p>	(Pluskal et al., 2010)
GenformR / Molgen MS/MS	R / C++	Windows	<p><u>Principe</u> : Calcul des formules pouvant correspondre à la masse détectée en fonction des éléments choisis. Filtration des formules par correspondance isotopique et par calcul de spectres théoriques de MS/MS correspondants pour comparaison si données disponibles</p>	(Meringer et al., 2011)

PARTIE A : Évaluation de méthodes de traitement de données

Nom	Langage	Système d'exploitation	Principe et format des données	Référence
			<p><u>Format entrée</u> : Données spectrales : un pic par ligne avec valeur de m/z et intensité en ASCII, un fichier pour le MS et un second pour MS/MS</p> <p><u>Format sortie</u> : Formules moléculaires calculées avec la déviation en masse pour la MS, le score d'intensité relatif avec les données MS/MS et la correspondance entre les masses de fragments théoriques</p>	
ChemCalc	Html	En ligne	<p><u>Principe</u> : Calcul de toutes les formules théoriquement possibles à partir des éléments C, H, N, O. Pas d'option de classement lors d'assignements multiples</p> <p><u>Format d'entrée</u> : masse monoisotopique</p> <p><u>Format de sortie</u> : tableau contenant les formules brutes possibles, la masse exacte, l'erreur en masse et le degré d'insaturation, exportable en *.csv</p>	(Patiny and Borel, 2013)
HR3	C++	Windows	<p><u>Principe</u> : Appuyé sur HR2 des <i>Seven Golden Rules</i>. Augmentation du nombre d'atomes possible (C, H, N, O, P, S, F, Cl, Br, Si, I, Mg, Fe, Mo, Mn, Cu, B, Ca, Ni, Zn et Co), restreint en termes de m/z (entre 40 et 2000) et changement en termes de ratio d'éléments par rapport aux <i>Seven Golden Rules</i>. Les formules dont le degré d'insaturation est négatif sont éliminées. Changement en termes de valence pour N, P et S. Recherche de correspondance dans PubChem améliorée en termes de temps de traitement</p> <p><u>Format entrée</u> : N.A</p> <p><u>Format sortie</u> : N.A.</p>	(Lommen, 2014)

PARTIE A : Évaluation de méthodes de traitement de données

Nom	Langage	Système d'exploitation	Principe et format des données	Référence
CHOFIT	Pascal	Windows, Linux, Mac	<p><u>Principe</u> : Pour une masse donnée, assignement d'un nombre de C maximal à la masse nominale, puis remplissage du reste avec des H. Ensuite, mesure de la différence de masse entre cette formule et la masse détectée et division du résultat par la masse exacte de CH₄O₋₁. Si le résultat n'est pas un nombre entier, soustraction de la masse de C₄O₋₃ à la masse de la formule calculée et étape précédente répétée jusqu'à obtention d'un entier (X). Lorsque c'est le cas, ajout ou suppression de X fois CH₄O₋₁ à la formule qui permet d'obtenir X pour obtention de la formule brute finale.</p> <p><u>Format entrée</u> : FT-ICR, Liste de masses</p> <p><u>Format sortie</u> : N.A.</p>	(Green and Perdue, 2015)
Parallel Formula Generator	C++	Windows, Linux	<p><u>Principe</u> : Limité à C, H, N, O, P, S, F, Cl, Br, I, Si, Na, K et m/z 2000, filtration en utilisant la règle de Lewis et des restrictions en termes de ratio atomique. Le calcul de formule moléculaire s'appuie simplement sur la masse monoisotopique</p> <p><u>Format entrée</u> : masse monoisotopique</p> <p><u>Format sortie</u> : fichier *.txt contenant le nombre de chaque atome de la formule moléculaire attribuée</p>	(Zhang et al., 2016)
Formularity	Matlab	Windows	<p><u>Principe</u> : Calcul de formules brutes avec C, H, N, O, P, S, Cl et Br puis simulation des distributions isotopiques. Attribution de deux scores basés sur la présence des isotopes simulés dans les données expérimentales et les erreurs en masse associées à ces isotopes, normalisés par leur abondance relative</p> <p><u>Format entrée</u> : Fichier *.txt (tab) contenant dans les deux premières colonnes les m/z classés dans l'ordre croissant et l'intensité correspondante, et troisième colonne optionnelle avec le S/N. Format MS excel et Bruker XML acceptés</p>	(Tolić et al., 2017)

PARTIE A : Évaluation de méthodes de traitement de données

Nom	Langage	Système d'exploitation	Principe et format des données	Référence
			<p><u>Format sortie</u> : Deux fichiers listant tous les pics avec les formules correspondantes calculées, ainsi que différentes données ayant servi au calcul, les erreurs en masses observées et les scores</p>	
i-van Krevelen	Python	Windows, Linux, Mac	<p><u>Principe</u> : Calcul des formules brutes selon les <i>Seven Golden Rules</i> et limite du nombre de chaque élément défini par l'utilisateur. Seuls C, H, N, O disponibles avec adduit K et Na. Pas d'observation de la part des développeurs d'assignement multiples pour une masse donnée donc pas de système de classement des résultats</p> <p><u>Format entrée</u> : liste des masses en *.txt</p> <p><u>Format sortie</u> : Trois fichiers *.csv, le premier avec les assignements, le deuxième avec les isotopes assignés et le dernier avec les masses pour lesquelles aucun assignement n'a pu être effectué</p>	(Kew et al., 2017)
MFAssignR	R	Windows, Linux, Mac	<p><u>Principe</u> : CHOFIT adapté en ajoutant la possibilité de considérer d'autres hétéroatomes, sert de base lorsque pas d'ambiguïté d'assignation. Si ambiguïtés, application de WHOI amélioré avec ajout de groupes fonctionnels (CH₂, H₂O, O, H₂, et CH₂O) et les assignations sans ambiguïté sont ajoutées dans les formules servant de base et ainsi de suite. Les formules sont considérées sans ambiguïté quand elles respectent les <i>Seven Golden Rules</i>, des ratios O/C et H/C définis par l'utilisateur, si présence de S, vérification isotopique et vérification de la plausibilité d'appartenance à une série homologue. Si plusieurs formules possibles, assignation de la formule qui possède le plus de liens en termes de WHOI avec les formules précédemment assignées et avec le moins d'hétéroatomes (excepté O) possible</p>	(Schum, 2019)

PARTIE A : Évaluation de méthodes de traitement de données

Nom	Langage	Système d'exploitation	Principe et format des données	Référence
UltraMass Explorer	R	Windows, Linux, Mac	<p><u>Principe</u> : Comparaison avec des bibliothèques de formules brutes basées sur des masses isotopiques exactes compilées par NIST (format des bibliothèques fournies pour pouvoir en ajouter de nouvelles). Limité en m/z (100-700) et en atomes (C, H, N, O, P, S + isotope ¹³C, ³⁴S, ¹⁵N) <u>Format entrée</u> : FT-ICR, données contenant deux colonnes : la masse exacte (neutre ou ionisée) et l'intensité ou aire du pic <u>Format sortie</u> : *.csv</p>	(Leefmann et al., 2019)
SIRIUS	Java	Windows, Linux, Mac	<p><u>Principe</u> : Utilisation des données isotopiques pour la détermination de la formule brute, puis analyses sur les fragments d'un composé assigné par utilisation d'un arbre de fragmentation <u>Format entrée</u> : données centroïdées, MS et MS/MS ou MSⁿ obligatoire, en *.csv, *.ms, *.mgf, ou *.sirius <u>Format sortie</u> : Tableau avec formule moléculaire + adduit, et trois types de score : score isotopique, score de fragmentation et score total (résumé des 2 autres) pour aider aux choix lors d'assignations multiples</p>	(Dührkop et al., 2019)
TRFu et FuJHA	Matlab	Windows, Linux, Mac	<p><u>Principe</u> : TRFu : Calcul de toutes les formules possibles théoriquement à partir de la liste de masse. Lorsque plusieurs assignations pour une masse donnée, filtres successifs sur les critères suivants : nombre minimal de carbone, nombre minimal de N+S+P, puis S+P et erreur en masse minimale FuJHA : Création d'une base de formule brute qui sera comparée au set de masses données en entrée <u>Format entrée</u> : liste de masse avec intensité et ratio S/N avec des paramètres de restrictions définis en tolérance m/z, limite pour chaque éléments, isotopes, règle DBE et N <u>Format sortie</u> : N.A.</p>	(Fu et al., 2020)

II.3.3. Différents types de bases de données

Une fois les formules moléculaires uniques associées à chacun des marqueurs préalablement sélectionnés, il s'agit alors de trouver les structures et les composés correspondants. Cela passe par la confrontation à des données de référence recueillies par ailleurs. La source exacte de ces données est cruciale et permet d'attribuer un niveau différent (entre 1 et 4 selon les règles proposées par (Schymanski et al., 2014a) à l'assignation d'identité qui découlera de l'étude.

II.3.3.1. Librairies internes

Les librairies internes sont les bases de données qui fournissent les données les plus précises et qui permettent d'atteindre les plus hauts niveaux dans la classification de (Schymanski et al., 2014a). En effet, ces librairies sont construites par l'opérateur sur son propre appareil d'acquisition. Dans le meilleur des cas, elle contient toutes les informations nécessaires pour identifier un composé, en particulier si le standard correspondant a déjà été injecté dans les mêmes conditions d'analyse que celles utilisées pour le NTS. Le TR, la masse exacte de l'ion détecté ainsi que les fragments formés, et lorsqu'il est disponible le *drift time*, sont déjà connus et permettent donc d'identifier de façon certaine le composé par comparaison des données acquises en NTS aux données présentes (niveau 1). Dans un cas moins idéal, mais plus probable, seules certaines données relatives à la molécule sont disponibles. En effet, il est impossible d'avoir tous les standards des composés existants, actuellement pour deux raisons principales. La première est qu'un grand nombre de composés reste encore inconnus à ce jour, comme les produits de transformation ou de dégradation, d'où également l'intérêt du NTS comme technique d'analyse. D'autre part, en considérant seulement les standards analytiques existants, il est également peu probable de tous les posséder pour des raisons budgétaires. Dans ce cas précis, il est quand même possible d'enrichir la librairie avec des données limitées, incluant entre autres, la formule brute, la masse exacte et la structure de la molécule. Bien que ces données soient limitées, elles ont tout de même leur utilité puisqu'elles permettent d'avoir un potentiel candidat pour un marqueur donné, qu'il sera ensuite possible de confirmer ou d'infirmer par les méthodes décrites dans les paragraphes suivants. D'autre part, l'avantage d'enrichir les librairies internes avec seulement ces données est de pouvoir travailler en SS sur des composés susceptibles d'être retrouvés dans les matrices et échantillons étudiés.

II.3.3.2. Fragmentation *in-silico*

La fragmentation *in-silico* désigne une fragmentation théorique réalisée à l'aide d'algorithmes sur des structures définies. Ces informations sont utiles lorsque le standard n'est pas disponible et que peu ou pas de données de fragmentation sont disponibles par ailleurs. Elles permettent de confirmer la plausibilité de l'assignation d'une structure à un marqueur détecté. Cependant, les fragments générés restent théoriques. Ces informations doivent donc être utilisées avec précaution. A la fois lorsque le fragment généré théoriquement coïncide avec un fragment détecté, puisque ce fragment peut ne pas pouvoir être formé dans les conditions de l'analyse en question ; mais également lorsqu'aucun fragment ne correspond puisqu'une fragmentation théorique peut ne pas pouvoir prédire toutes les possibilités réelles.

Il existe plusieurs méthodes pour la génération de fragments *in-silico* (Hufsky and Böcker, 2017). La première, utilisée par exemple par le logiciel *MassFrontier* (ThermoFisher, 2018), consiste à prédire les spectres de MS/MS à partir de fragmentations connues (*i.e.*, règles de fragmentation théoriques, fragmentations retrouvées dans la littérature ou fragmentations d'un composé proche) sur des composés présents dans une base de données. Le principal inconvénient de cette méthode réside dans les règles de fragmentation utilisées. En effet, le domaine de recherche est en constante évolution, engendrant régulièrement la découverte de nouvelles règles. De plus, il reste difficile de déterminer dans quels cas, telle ou telle règle s'applique ou non. D'autres logiciels, comme *MetFrag* (Ruttkies et al., 2016), génèrent des fragments théoriques en « cassant » au fur et à mesure, toutes les liaisons possibles de la molécule considérée. Enfin, certaines techniques un peu particulières combinent l'utilisation d'arbres de fragmentation établis directement à partir du spectre MS/MS acquis afin d'établir une empreinte spectrale à l'aide de méthodes de *machine learning*, comme c'est le cas pour *CSI:FingerID* (Dührkop et al., 2015). Cette empreinte sera ensuite comparée à des empreintes spectrales présentes dans des bibliothèques pour déterminer la structure finale de l'inconnu.

II.3.3.3. Librairies “fournisseur”

Les librairies “fournisseur” sont des bases de données directement fournies sur le logiciel fournisseur de l’appareil HRMS. Elles sont variables en termes de nombre, de contenu et de précision selon les fournisseurs considérés. Elles restent cependant relativement fiables car elles correspondent généralement à des molécules qui ont été analysées sur le même type d’appareil. En revanche, dans certains cas, les informations qui sont fournies à l’utilisateur sur ces librairies sont limitées ce qui rend l’assignation d’un niveau de confiance difficile.

II.3.3.4. Bases de données en ligne

Il existe de nombreuses bases de données en ligne fournissant des informations analytiques sur de très nombreuses molécules. Elles sont constamment enrichies par la communauté, permettant ainsi un échange d’informations de plus en plus large entre les différents utilisateurs. Cependant, les informations retrouvées dépendent grandement des bases consultées. D’autre part, il est important de regarder avec précision les conditions dans lesquelles les informations données ont été recueillies, notamment le type d’appareil utilisé, les paramètres désignant une molécule n’étant pas nécessairement transférables d’un appareil à un autre. On distingue deux types de bases de données en ligne, celle qui sont désignées comme permettant l’identification, et celles qui servent à l’annotation et qui incluent donc des données spectrales (Blaženović et al., 2018) (Tableau 4).

La principale différence entre ces deux types de bases de données réside dans la présence de spectres de masses dans les libraires d’annotations. Ceux-ci permettent une comparaison des spectres acquis lors de l’analyse avec des spectres acquis par d’autres laboratoires, ils peuvent ainsi augmenter la confiance dans l’assignation d’une structure. En revanche, les bases d’identification sont beaucoup plus fournies en termes de nombre de molécules mais ne possèdent pas souvent de données d’acquisition permettant une comparaison. Elles permettent souvent d’identifier une molécule uniquement sur la masse monoisotopique et sont donc moins puissantes que les bases d’annotation.

Tableau 4 : Principales bases de données en ligne

Nom	Type de base	Nombre de composés (09/06/20)	Types d'informations	Référence
PubChem	Identification	102 716 458	Petites molécules Recherche par nom, structure, formule brute, InChI, SMILES, numéro d'identification (ex. CAS)	(Bolton et al., 2008)
ChemSpider	Identification	> 67 000 000	Petites molécules Recherche par nom, structure, formule brute, InChI, SMILES, numéro d'identification (ex. CAS)	(Pence and Williams, 2010)
KEGG	Identification	18 720	Petites molécules, biopolymers et substances chimiques d'intérêt pour les systèmes biologiques Recherche par nom ou identifiant KEGG	(Kanehisa et al., 2006)
EPA Comptox Dashboard	Identification	> 875 000	Petites molécules d'intérêt pour les sciences environnementales et la santé Recherche par nom, numéro d'enregistrement, InChI, DTXSID	
FOR-IDENT	Annotation		Petites molécules d'intérêt pour les sciences environnementales Recherche par nom, masse, formule brute, InChI, numéro d'enregistrement, SMILES, liste de m/z	
MassBank Japan	Annotation	14 382	Données EI, MS/MS et MSn Recherche par nom, masse exacte, formule brute, liste de pics, InChI	(Horai et al., 2010)
mzCloud	Annotation	17 980	Données MSn Recherche par spectre, structure, substructure, arbre de fragmentation, masse exacte, nom	
MoNA	Annotation	9 218	Données MS/MS et MSn Recherche par nom, InChI, formule brute, masse exacte, spectre	
NIST	Annotation		Données EI-MS et CID-MS/MS Recherche par nom, formule brute, numéro d'enregistrement, structure, InChI	(Linstrom and Mallard, 2020)

II.4. Partage de données et plateformes communautaires

Que ce soit pour les programmes développés pour le traitement ou le pré-traitement des données de HRMS, pour les spectres de molécules ou d'autres paramètres permettant de les identifier, voire pour les données brutes, le partage de données est essentiel afin de permettre le développement du NTS. De nombreuses plateformes de partage sont ainsi disponibles, certaines spécifiques au NTS et d'autres plus généralistes (Tableau 5).

Tableau 5 : Principales plateformes de partage de données pour le NTS

Nom	Description	Lien
Bioconductor	Partage de codes pour le traitement ou pré-traitement	https://www.bioconductor.org/
Github	Partage de codes pour le traitement ou pré-traitement	https://github.com/
Bitbucket	Partage de codes pour le traitement ou pré-traitement	https://bitbucket.org/
MassBank	Partage de données spectrales	http://www.massbank.jp/
MoNA	Partage de données spectrales	https://mona.fiehnlab.ucdavis.edu/
mzCloud	Partage de données spectrales	https://www.mzcloud.org/
Norman databases	Partage de données sur la détection de composés émergents dans l'environnement (données spectrales, occurrences, toxicologie...)	https://www.norman-network.com/nds/
Zenodo	Partage de données OpenScience	https://zenodo.org/

III. Matériels et méthodes

III.1. Présentation des outils

III.1.1. Acquisition des données

Lors de l'EIL, 6 échantillons ont été envoyés par les organisateurs à chacun des 18 laboratoires européens y participant : un blanc d'extraction (échantillonneur passif non exposé et extrait de la même manière que les échantillons), un mélange de standards analytiques (Tableau 6), 2 extraits d'eau naturelle avant entrée en station de traitement d'eau potable (S2-2 et S2-4) et 2 extraits d'eau potable prélevée en sortie d'usine de traitement (S1-2 et S1-4).

Tableau 6 : Liste des molécules présentes dans le mélange de standards analytiques de l'EIL, fournie par les organisateurs

Molécule	Numéro CAS	Formule moléculaire	[M+H] ⁺
Guanylurea	141-83-3	C ₂ H ₆ N ₄ O	103,0614
Amitrol	61-82-5	C ₂ H ₄ N ₄	85,0509
Histamine	51-45-6	C ₅ H ₉ N ₃	112,0869
Chlormequat*	999-81-5	C ₅ H ₁₃ ClN	122,0731
Methamidophos	10265-92-6	C ₂ H ₈ NO ₂ PS	142,0086
Vancomycin**	1404-90-6	C ₆₆ H ₇₅ Cl ₂ N ₉ O ₂₄	724,7224
Cefoperazone	62893-19-0	C ₂₅ H ₂₇ N ₉ O ₈ S ₂	646,1497
Trichlorfon (Dylox)	52-68-6	C ₄ H ₈ Cl ₃ O ₄ P	256,9299
Butocarboxim	34681-10-2	C ₇ H ₁₄ N ₂ O ₂ S	191,0849
Dichlorvos	62-73-7	C ₄ H ₇ Cl ₂ O ₄ P	220,9532
Tylosin	1401-69-0	C ₄₆ H ₇₇ NO ₁₇	916,5264
TCMTB	21564-17-0	C ₉ H ₆ N ₂ S ₃	238,9766
Rifaximin	80621-81-4	C ₄₃ H ₅₁ N ₃ O ₁₁	786,3596
Spinosad A (Spinosyn A)	131929-60-7	C ₄₁ H ₆₅ NO ₁₀	732,4681
Emamectin B1a	121124-29-6	C ₄₉ H ₇₅ NO ₁₃	886,5311
Avermectin B1a***	71751-41-2	C ₄₈ H ₇₂ O ₁₄	890,5260
Nigericin	28380-24-7	C ₄₀ H ₆₈ O ₁₁	725,4834
Ivermectin B1a****	70288-86-7	C ₄₈ H ₇₄ O ₁₄	892,5417

*[M]⁺ = [C₅H₁₃ClN]⁺ = 122,0731

**[M+H]²⁺ = [C₆₆H₇₅Cl₂N₉O₂₄+H]²⁺ = 724,7224

***[M+NH₄]⁺ = [C₄₈H₇₂O₁₄+NH₄]⁺ = 890,5260

****[M+NH₄]⁺ = [C₄₈H₇₄O₁₄+NH₄]⁺ = 892,5417

Les extractions ont été réalisées à l'aide d'échantillonneurs passifs avec une phase HLB, en prenant deux temps d'exposition pour chaque point, respectivement 2 jours (S2-2 et S1-2) et 4 jours (S2-4 et S1-4) (Figure 5).



Figure 5 : Stratégie d'échantillonnage de l'EIL organisé dans le réseau NORMAN (2019)

Les analyses devaient être réalisées à la fois avec une méthode analytique propre à chaque participant (en mode d'ionisation positif) et avec une méthode développée par les organisateurs sur une colonne déterminée. Les détails des protocoles chromatographiques employés sont fournis dans le Tableau 7. L'acquisition des données HRMS a été effectuée par UPLC-IMS-QToF (Vion, Waters) avec une fenêtre de 2 Hz. Après ionisation en mode positif (ESI+), les ions ont été séparés dans une cellule de mobilité ionique en *Travelling Wave Ion Mobility* utilisant de l'azote comme gaz neutre à contre-courant. Les masses ont ensuite été scannées entre 60 et 900 Da, et tous les ions détectés ont été fragmentés (MS^E) à l'aide d'une rampe de collision de 10 à 45 eV. Cet appareil fournit, pour chaque pic détecté, un triplet m/z , TR et DT unique qui le caractérise. A ce triplet sont associés l'aire correspondante détectée dans chacun des échantillons analysés et les spectres de masses caractéristiques du pic.

Tableau 7 : Protocoles de séparation chromatographique utilisés pour l'analyse des échantillons de l'EIL

	Méthode LEESU			Méthode NORMAN		
Colonne	Waters Acquity BEH C18 (100 × 2,1 mm, 1,7 µm)			Waters Acquity HSS C18 (100 × 2,1 mm, 1,7 µm)		
Débit	0,45 mL/min			0,4 mL/min		
Volume d'injection	10 µL			5 µL		
Phases mobiles	A = Eau ultra-pure (EUP) + 0,1 % acide formique (AF) B = Acétonitrile (ACN) + 0,1 % AF			A = EUP + 5 mM formate d'ammonium ajusté à pH 3 B = ACN + 0,1 % AF		
Gradient	Temps	A (%)	B (%)	Temps	A (%)	B (%)
	0,00	98	2	0,00	87	13
	1,00	98	2	0,50	87	13
	25,00	2	98	10,00	50	50
	30,00	2	98	10,75	5	95
	21,50	98	2	12,25	5	95
	34,00	98	2	12,50	87	13
				15,00	87	13

III.1.2. Logiciels de traitements : UNIFI et EZInfo

UNIFI (Waters) est le logiciel installé par défaut par le fournisseur avec l'appareil d'HRMS utilisé au cours de ce travail. Il permet l'acquisition des données en 4 dimensions (TR, DT, spectres de masses et aires des pics). Les données sont prétraitées par le logiciel en parallèle de l'acquisition (détection, alignement et annotation des pics) et parmi les options de traitement proposées par le logiciel, on retrouve :

- Identification de composés (en TS, SS) selon des critères de tolérances définis
- Quantification en TS
- Identification de composés en NTS avec recherche de données dans plusieurs types de bibliothèques (détaillées dans la partie A - II.3.3)
- Calcul de défaut de masse
- Recherche d'halogènes
- Comparaison binaire d'échantillons
- Association de produits de transformation à partir de molécules identifiées en TS ou SS
- Différentes options de filtres sur les composés identifiés (en TS ou SS)

En plus de ces options, un module complémentaire d'analyses statistiques (EZInfo) est fourni par Waters. EZInfo est un outil développé par Umetrics et permet de transférer les données directement depuis UNIFI pour réaliser des AMV. Parmi celles-ci, on retrouve les ACP, les régressions par moindres carrés discriminantes (*Partial least squares-discriminant analysis* ou PLS-DA) et régressions par moindres carrés orthogonales discriminantes (*Orthogonal partial least squares discriminant analysis* ou OPLS-DA). Différentes visualisations graphiques des scores sont disponibles (*loadings plot*, *bi-plot*, *s-plot*) ainsi que les scores des prédictions générés par le modèle.

Dans UNIFI, les données brutes sont compilées au format *.uep et les métadonnées peuvent être exportées au format *.xml. Un fichier *.uep regroupe toutes les analyses sur une séquence donnée, les échantillons individuels ne pouvant pas être exportés séparément. À la suite des différents prétraitements réalisés dans UNIFI, les données peuvent également être exportées au format *.csv (*marker table*). En ce qui concerne EZInfo, les données ne sont pas exportables. Les résultats des traitements réalisés sous EZInfo nécessitent d'être retransférés

dans UNIFI, copiés et collés manuellement dans un tableur, récupérés via des captures d'écran ou imprimés au format *.pdf.

III.1.3. Interface de programmation de UNIFI et MSConvert

MSConvert est un logiciel de la suite *Proteowizard* (Kessner et al., 2008), permettant de convertir les données brutes de HRMS vers différents formats, en particulier vers le format ouvert et universel *.mzML. Grâce à l'installation d'une interface de programmation d'application (API) par Waters (second semestre 2019), *MSConvert* a pu être utilisé avec UNIFI. L'API permet d'accéder aux données stockées dans UNIFI et aux échantillons individuels de chaque séquence, mais aussi la conversion échantillon par échantillon. Dans ce travail, *MSConvert* a été utilisé afin de convertir les données dans deux types de format :

- *.mzML afin de traiter les données avec des logiciels libres (*patRoan/xcms*)
- *.mz5 pour le partage en ligne des données sur les plateformes communautaires

Dans les deux cas, les données sont compressées en zlib en combinant les scans de mobilité ionique qui ne sont, pour le moment, pas supportés par les logiciels libres. Le format mz5, basé sur le format HDF5, permet une réduction de la taille des fichiers par rapport au format mzML, mais il n'est également pas encore supporté par les logiciels libres courants. L'algorithme *continuous wavelet transform* (CWT) (Du et al., 2006) est utilisé pour la détection des pics et la transformation des données du mode *profile* vers le mode centroïde, puis l'extraction des données spectrales en basse énergie et haute énergie est réalisée. A ce stade, les données ne sont filtrées d'aucune manière afin de garder le maximum d'informations intactes. Comparées au format uep, ces données ne contiennent plus de données de mobilité ionique (DT et CCS) et le passage vers le mode centroïde, jugé moins précis et moins fiable que le mode *profile* (Xia et al., 2011), est cependant nécessaire afin de pouvoir traiter les données par la suite avec les outils choisis.

III.1.4. R et ses différents packages

R est un langage de programmation très utilisé dans le domaine de l'analyse statistique et plusieurs packages ont été développés dans ce langage pour l'analyse de données de spectrométrie de masse. Le package *patRoan* (Helmus et al., 2021) permet ainsi de lire les données converties via *MSConvert* afin de rendre les données exploitables. Il utilise différents algorithmes déjà largement utilisés en NTS comme *xcms* (Smith et al., 2006), *OpenMS* (Sturm

et al., 2008), *SIRIUS* (Dührkop et al., 2019), *CAMERA* (Kuhl et al., 2012) ou encore *mzR* (Fischer et al., 2021). L'avantage de *patRoön* réside dans l'optimisation du temps de traitement apporté par la parallélisation et le fait que le passage d'un algorithme à un autre soit facilité par l'utilisation d'un format commun basé sur le package *data.table*. En plus des étapes de prétraitement proposées, différents traitements et types de représentation sont également codés et s'avèrent plus dédiés au NTS environnemental contrairement à la majorité des outils développés pour la métabolomique principalement. Les autres packages relatifs à chaque fonction seront décrits de manière plus détaillée dans la partie A - IV.3.2.

III.2. Présentation des bases de données

III.2.1. Construction d'une librairie interne

La création d'une librairie interne de molécules d'intérêt a été commencée au début de cette thèse et de nouvelles molécules ont été ajoutées régulièrement. Sa gestion et son utilisation sont plus simples que pour les autres librairies car les données y sont totalement contrôlées en interne : pour chaque molécule, les paramètres expérimentaux sont connus ou peuvent être retrouvés, ce qui permet l'assignation d'un niveau d'identification plus certain.

Trois types de composés sont présents dans notre librairie : les molécules qui sont disponibles en standard analytique ; des molécules d'intérêt dans le cadre de nos travaux, mais dont le standard analytique n'est pas disponible ; et des marqueurs détectés et d'intérêt pour nos études mais pour lesquels aucune identité n'a pu être associée à ce jour. Dans le second cas, seules les données correspondantes à la formule moléculaire, la masse exacte, des codes d'identification (Inchl, SMILES, éventuellement CAS) et la structure sont entrées dans la librairie. La présence de la structure dans la librairie permet au logiciel de générer une fragmentation *in silico*. La recherche en SS, s'appuie donc dans un premier temps sur la masse exacte de la molécule, et, permet de repérer la présence d'éventuels fragments expérimentaux qui correspondent aux fragments générés *in silico*. Pour ces molécules, le niveau d'identification peut, dans le meilleur des cas, atteindre le niveau 2.

Concernant les molécules disponibles en standard analytique, une procédure a été mise en place afin de les enregistrer dans la librairie. Trois niveaux de concentration sont préparés (500 µg/L, 50 µg/L et 1 µg/L) et injectés cinq fois chacun, à partir de la méthode NTS mise en place avec une colonne C18 (*méthode LEESU*, Tableau 7) dans les deux modes d'ionisation.

L'injection peut être effectuée pour un composé unique ou un mélange de composés. Les spectres (MS et MS^E) sont enregistrés dans la librairie et les moyennes sur toutes les injections des m/z , TR, DT et CCS sont réalisées et également enregistrées dans la librairie. Lorsque ces molécules sont ensuite recherchées dans des échantillons réels en SS, elles sont repérées par leur m/z , TR et DT en prenant en compte des fenêtres de tolérance définies au préalable. Les fragments observés sont également comparés aux fragments obtenus lors de l'analyse du standard avec une tolérance de 2 mDa, et les molécules peuvent donc être identifiées au niveau 1.

Concernant les marqueurs d'intérêt (niveau 5), les propriétés relatives à leur détection (m/z , TR, DT, CCS et spectres) sont répertoriées dans la librairie ce qui pourra permettre leur recherche dans d'autres échantillons. Généralement, ces marqueurs sont relevés lors d'expériences d'oxydation de composés modèles.

A ce jour, la librairie interne contient les éléments répertoriés dans le Tableau 8. La liste détaillée des composés disponibles au niveau 1 dans la librairie est présentée en Annexe 1.

Tableau 8 : Récapitulatif des composés présents dans la librairie interne (janvier 2022)

Niveau de confirmation maximal possible	Paramètres de confirmation disponibles	Type (nombre) de molécules
Niveau 1 (étalons disponibles)	<i>m/z</i> , TR, DT, CCS, fragments détectés, mode d'ionisation préférentiel	Biocides (59) Pharmaceutiques (27) Autres polluants urbains (54) Produits de transformation (14)
Niveau 2 (étalons non disponibles mais structure connue)	<i>m/z</i> , fragments <i>in silico</i> , éventuellement CCS	Biocides (19) Pharmaceutiques (145) Autres polluants urbains (90) Produits de transformation (393)
Niveau 5 (pas d'identification)	-	Produits de transformation (19)

III.2.2. Description des bases de données externes

Le logiciel UNIFI permet l'identification de substances via trois types de bases de données externes : les bases de données fournies avec le logiciel (créées par Waters), les bases de données en ligne reliées au logiciel et les bases de données externes téléchargeables par l'utilisateur et importables dans le logiciel sous certaines contraintes de mise en forme.

III.2.2.1. Librairies Waters

Par défaut, lors de l'installation de UNIFI, plusieurs librairies sont fournies. Dans notre cas, les librairies *Waters Screening Library* [UNIFI 1.8] (2 302 molécules), *Waters CCS Pesticides Library* (608 composés), *Amino Acids* (23 molécules) et *BUCT BC GF QToF library with CAS and tags 2018* (532 molécules) sont installées. Les deux premières étant plus utiles aux problématiques de cette thèse, seules celles-ci seront décrites dans la suite. Dans ces librairies, les molécules possèdent au minimum :

- Une désignation usuelle
- Un nom IUPAC
- Un code InChI et un numéro CAS
- Une formule moléculaire
- Une masse molaire moyenne et une masse monoisotopique associée
- Une structure, qui permettra au logiciel de générer des fragments *in silico*
- Un lien vers la page ChemSpider de la molécule

En plus de ces informations, certaines molécules possèdent des informations relatives à leur utilisation mais également des informations facilitant leur détection. En effet, un mode d'ionisation préférentiel avec un adduit spécifique ou des données de CCS ou de TR sont parfois précisées (1 252 molécules).

III.2.2.2. Bases de données reliées à UNIFI

L'identification dans le logiciel UNIFI peut également passer par l'utilisation de bases de données en ligne et regroupées sous ChemSpider. Cette dernière regroupe un ensemble de 276 librairies différentes contenant chacune un nombre de molécules variable, des molécules de nature très différente et avec des informations plus ou moins précises pour permettre l'identification. Les librairies les plus connues appartenant à cette base de données sont NIST, Massbank et KEGG.

Le principal inconvénient de ces librairies est l'absence d'informations précises les concernant, autant en termes de gestion que de contenu. Par exemple, le nombre de molécules que chaque librairie contient, ou encore les informations associées à chaque molécule (ex. source des données une fois les résultats d'identification affichés, détails du calcul permettant de confirmer une identification) sont inconnus. Il est donc notamment difficile d'associer un niveau de confiance au résultat d'identification s'il existe. D'autre part, les molécules ayant toutes différentes appellations possibles, aucune information n'est donnée quant à la potentielle gestion des synonymes au sein d'une même librairie. Les résultats affichés par UNIFI (Figure 6) ne précisent pas non plus de quelle librairie proviennent les molécules identifiées si plusieurs librairies sont sélectionnées simultanément. Enfin, aucune information n'est donnée quant à une potentielle mise à jour de ces librairies lors des mises à jour du logiciel par exemple. Lorsqu'une identification en NTS est réalisée grâce à ces librairies, sa fiabilité est donc très difficile à évaluer. Seul un indice de correspondance isotopique est fourni (en pourcentage), prenant en compte à la fois la présence des isotopes théoriques et leur intensité relative.

Marqueur inconnu ↓

Proposition d'identification ↓

Component Name	m/z	Elemental Composition	i-FIT Confidence (%)	Common Name	Fragment Matches	Predicted Intensity (%)	Citations
Candidate Mass 261.109438403418	261.1094	C11H12N6O2	18.81	N,N-Dimethyl-2-nitro-4-[(Z)-(4H-1,2,4...	0	0	13
Candidate Mass 261.109438403418	261.1094	C11H12N6O2	18.81	1-(4-Nitrophenyl)-5-(2-dimethylamin...	0	0	6
Candidate Mass 412.32067079429	412.3209	C27H41NO2	100.00	Cyclopamine	3	9	101
Candidate Mass 412.32067079429	412.3209	C27H41NO2	100.00	2-(3-Hydroxy-10,13-dimethyl-2,3,4,7,...	1	1	7
Candidate Mass 412.32067079429	412.3209	C27H41NO2	100.00	13b,15b-Dimethyl-2,3,4,6,6a,6b,7,7a,8,...	0	0	4
Candidate Mass 412.32067079429	412.3209	C27H41NO2	100.00	6-[1-(3-Hydroxyandrost-5-en-17-yl)et...	1	1	4

Formule moléculaire proposée et correspondance isotopique associé

Nombre de fragments correspondants et correspondance des intensités relatives

Figure 6 : Exemple de résultat d'identification d'un marqueur inconnu dans Unifi (fonction *Elucidate*)

III.2.2.3. Bases de données externes importées

Dans le cadre de cette thèse, le laboratoire a rejoint le réseau NORMAN en novembre 2019. Ce réseau regroupe plus de 70 institutions en Europe et Amérique du Nord, travaillant sur les problématiques de substances émergentes présentes dans l'environnement. Ce réseau a instauré la création de différentes bases de données collaboratives permettant de regrouper aussi bien des données d'occurrence, de détection, de toxicologie ou de spectres de molécules étudiées par les différents membres du réseau. Ces bases de données, accessibles en ligne, ont été téléchargées, puis importées progressivement dans UNIFI afin de remplacer l'utilisation de ChemSpider pour lequel les informations sont difficiles à évaluer. En effet, les bases de données NORMAN, bien qu'externes, présentent l'avantage de regrouper des molécules d'intérêt pour nos travaux. Par ailleurs, il est plus facile de gérer les données présentes dans ces bases et ainsi de voir quels paramètres ont permis l'assignation d'une molécule à un candidat. A ce jour 4 bases de données de NORMAN (*Biocides priority list*, *French monitoring list*, *Pharmaceuticals* et *Wastewater sweden suspect list*), ont été importées, regroupant 2 976 molécules avec des données de détection.

III.2.2.4. Bases de données en ligne

Lorsque les bibliothèques précédentes ne parvenaient pas à fournir des résultats, après interrogation, pour l'identification d'un inconnu ou pour tenter de confirmer une assignation, d'autres outils ont été utilisés en complément. Les spectres MS et MS^E ont été relevés dans UNIFI et les valeurs ont ensuite été transférées dans les bibliothèques suivantes, disponibles en ligne : MassBank (<https://massbank.eu/MassBank/Search>), MetFrag (<https://msbi.ipb-halle.de/MetFrag/>), PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) ou FOR-IDENT (<https://water.for-ident.org/#!home>). Les détails concernant ces bases de données sont fournis dans la partie A – III.2.2.4.

III.3. Acquisition des données

III.3.1. Contrôle qualité

Pour contrôler la qualité des analyses réalisées, plusieurs étapes ont été mises en place. La première consiste à vérifier le bon fonctionnement de l'appareil et assure la fiabilité des masses acquises. Pour cela, le système est calibré à l'aide d'une solution (20 composés organiques, Major Mix IMS/Tof calibration kit, Waters) dédiée au système Vion-IMS-QToF. A partir de cette solution, une optimisation des paramètres du détecteur et de la résolution est opérée automatiquement par le système dans les deux modes d'ionisation, de même qu'une calibration en masse et une calibration en CCS. La fiabilité des masses acquises est ensuite assurée par l'injection à intervalles réguliers d'une solution de leucine encéphaline (m/z 556,27 en ESI+ et 554,26 en ESI-) qui permet de corriger d'éventuelles déviations du système au cours d'une séquence d'analyse. Avant chaque analyse, cette solution est infusée afin de vérifier que son signal est suffisant (*i.e.*, du même ordre de grandeur que lors de l'infusion précédente). Si ce n'est pas le cas, une solution fraîche est préparée. De même, avant chaque analyse, trois échantillons blancs, ainsi qu'une solution de standards de référence injectée cinq fois et contenant 9 composés (Acétaminophène, Caféine, Sulfaguanidine, Sulfadiméthoxine, Val-Tyr-val, Verapamil, Terfenadine, Leucine encéphaline et Reserpine), sont analysés. Cette séquence permet de vérifier les performances du système en confirmant la présence de ces neuf composés et l'allure des pics chromatographiques. La moyenne quadratique de l'erreur en masse et en CCS, ainsi que la largeur moyenne des pics sont calculées pour chacun des composés sur les cinq injections et comparées à des seuils fixés : 2,00 ppm, 2,00 % et 3,0 sec respectivement. Si les valeurs calculées sont supérieures à ces

seuils, les performances du système ne correspondent pas aux attentes et une nouvelle calibration est effectuée avant de répéter ce processus.

Une fois cette première étape de vérification effectuée une séquence d'analyse débute par l'injection de trois blancs afin de vérifier la propreté de la colonne, puis cinq *pools* sont injectés. Un *pool* est un échantillon de contrôle constitué d'un volume équivalent de tous les échantillons qui seront injectés par la suite dans la séquence. Les cinq injections successives ont pour but d'équilibrer et stabiliser la colonne comme cela a été montré dans des études précédentes (Godzien et al., 2015). Par la suite, les échantillons sont injectés dans un ordre aléatoire en triplicata, et un *pool* est injecté toutes les dix injections. L'injection dans un ordre aléatoire des triplicatas et échantillons permet, d'une part, de limiter l'effet de *carry over* (contaminations par la chaîne analytique) entre les échantillons s'ils avaient été injectés dans un ordre spécifique, mais également de réduire au maximum la variabilité intra-séquence due au système analytique. L'injection régulière des *pools* permet de vérifier la stabilité du système, notamment en vérifiant qu'aucun signal n'apparaît ou disparaît au cours de la séquence.

III.3.2. Choix des critères d'identification en TS et SS

En TS ou SS, il est important de déterminer dans un premier temps, des fenêtres de tolérance en m/z , TR et DT qui permettront de savoir dans quelles mesures un pic sera considéré comme correspondant à une molécule recherchée. Dans notre cas, ces paramètres ont été choisis à partir de l'observation des variations de ces paramètres sur des molécules connues à l'aide de la méthode d'analyse tout au long d'une année. Ainsi, pour les molécules dont l'étalon a été injecté, une erreur de 5 ppm est acceptée avec une variation de TR de $\pm 0,2$ min, une erreur en CCS $< 5\%$ et la présence d'au moins un fragment enregistré.

En ce qui concerne le SS de molécules dont le standard (et donc le TR) n'était pas disponible, le critère d'erreur en masse précédemment défini a été conservé mais aucune restriction en termes de TR n'a été appliquée. En ce qui concerne les valeurs de CCS et les fragments, lorsqu'ils étaient disponibles sur des bases de données, les mêmes critères que précédemment étaient appliqués. Pour le SS, en plus des correspondances trouvées par UNIFI, une seconde étape était nécessaire pour vérifier manuellement l'aspect des pics

chromatographiques intégrés et déterminer le meilleur candidat lorsqu'il y avait plusieurs correspondances.

III.3.3. Création des *marker table* pour le NTS

III.3.3.1. Par UNIFI

En ce qui concerne le NTS, tous les pics détectés sont considérés. Il s'agit alors de déterminer dans quelle mesure un pic peut être considéré comme désignant le même composé dans des échantillons différents, par une étape d'alignement. Les paramètres d'alignement choisis pour cela (*i.e.*, tolérance en valeurs de m/z , RT et DT) sont essentiels pour assurer la fiabilité des données alignées (*marker table*). Afin d'optimiser ces paramètres, les étalons internes ont été utilisés. En effet, injectés régulièrement sur l'appareil, et tout au long de chaque séquence d'analyse, la déviation par rapport à la valeur théorique de chaque paramètre a été estimée. Ce même processus a été répété sur un an afin de détecter de potentielles variations entre les séquences ou à la suite d'une recalibration du spectromètre de masse, ou encore suite à un changement de solution mère d'étalons. Tous étalons et toutes séquences confondus, la variation maximale intra-séquence observée est de 9,2 ppm en m/z , 0,36 min en TR et de 0,18 ms pour le DT en considérant des adduits similaires lors de la séparation sur une colonne C18. La partie B de ce manuscrit présentera l'utilisation de deux colonnes supplémentaires dont les variations sont légèrement différentes. Ces dernières ont également été prises en compte pour le choix des paramètres d'alignement. Un choix a donc dû être effectué entre aligner ensemble de nombreux pics qui ne correspondent pas à la même chose, ou séparer des pics qui correspondent au même composé mais pour lesquels une trop grande variation du TR entraîne une séparation de ces marqueurs. Ici, c'est le second cas qui a été sélectionné. Tout d'abord, car, même si les marqueurs ne sont pas alignés, il est possible de les considérer comme semblables manuellement par la suite grâce aux deux autres paramètres (m/z et DT) et aux données de fragmentation. L'inverse n'est pas vrai puisqu'une fois les pics regroupés, une moyenne sur les pics regroupés pour chacun des trois paramètres (m/z , TR et DT) est effectuée pour désigner le marqueur final. Il est donc difficile, voire impossible, de retrouver les pics qui ont été alignés *a posteriori*. D'autre part, les deux autres colonnes ayant des variations de TR relativement similaires, le choix de ce paramètre d'alignement a été effectué à partir des données sur ces colonnes.

Les tolérances ainsi retenues pour l'alignement sont les suivants : 10 ppm, 0,2 min et 0,3 ms, respectivement pour le m/z , le TR et le DT. La *marker table* résultante (Figure 7) contient les informations relatives à la détection des marqueurs. Lors de l'export de ces données, des métadonnées complémentaires peuvent également être récupérées. Ces dernières sont relatives aux injections et peuvent permettre de distinguer les différents échantillons (définition des réplicats) ou fournir des informations supplémentaires sur la nature ou l'origine des échantillons (à savoir les dates de prélèvement, provenances, groupes...). Cette *marker table* est directement exportable sous cette forme en format *.csv.

Identité du marqueur
Identifiant unique

Chaque colonne
suivante représente
une injection

	PrimaryId	mz	RetentionTime	DriftTime	Comment	blanc_replicate_1	blanc_1
1	278.08097_0.035_4.44	278.0810	0.035	4.44	NA	1353.31	0.00
2	219.89776_0.039_4.02	219.8978	0.039	4.02	NA	0.00	0.00
3	219.04684_0.048_4.22	219.0468	0.048	4.22	NA	452.64	0.00
4	171.14858_0.051_3.92	171.1486	0.051	3.92	NA	418.57	0.00
5	291.11121_0.051_4.82	291.1112	0.051	4.82	NA	28654.19	0.00
6	227.12524_0.052_4.30	227.1252	0.052	4.30	NA	3635.64	0.00
7	123.05615_0.052_2.69	123.0562	0.052	2.69	NA	1504.28	0.00
8	184.88466_0.054_4.08	184.8847	0.054	4.08	NA	835.70	0.00
9	306.89221_0.055_3.98	306.8922	0.055	3.98	NA	0.00	343.37
10	307.07924_0.055_4.58	307.0792	0.055	4.58	NA	1728.06	0.00
11	229.14086_0.055_4.39	229.1409	0.055	4.39	NA	195176.97	0.00
12	228.23190_0.058_5.36	228.2319	0.058	5.36	NA	0.00	0.00

Chaque ligne correspond à un marqueur

Propriétés de détection du marqueur

Aire du marqueur désigné par la ligne dans l'injection désignée par la colonne

Figure 7 : Structure d'une *marker table*

III.3.3.2. Par *patRoom*

Lors de ce travail, une méthode complémentaire basée sur l'utilisation de logiciels libres a été employée. Pour cela, une conversion préalable des données a été effectuée (partie A – III.1.3) afin d'obtenir des fichiers au format mzML. Ces fichiers sont ensuite lus par *patRoom*.

La détection des pics est réalisée grâce à *xcms3* qui s'appuie sur l'algorithme *centwave* (Tautenhahn et al., 2008). Ces pics sont ensuite alignés, toujours à l'aide de *xcms3*, en utilisant l'algorithme *Obiwarp* avec une fenêtre de temps de rétention fixée à 12 s (soit 0,2 min, même valeur que celle choisie dans UNIFI). Enfin, la récupération des spectres de masse en basse et haute énergies est réalisée à l'aide l'algorithme *mzR* (Fischer et al., 2021a). La *marker table*

obtenue a ainsi la même structure que celle créée dans UNIFI, mais les données de MS et les données de MS^E (considérées comme de la MS²) sont également disponibles dans une autre table, contrairement aux *marker table* issues de UNIFI. En revanche, la *marker table* issue de *patRoon/xcms3* est plus petite que celle générée par UNIFI en termes de nombre de marqueurs, puisque la conversion préalable des données entraîne la perte de la séparation supplémentaire liée à la mobilité ionique. Des pics de même TR et *m/z* mais de DT différents se retrouvent donc alignés dans un même marqueur.

La récupération des spectres de masse MS et MS^E permet cependant le calcul de formule moléculaire de manière plus fiable. Ce calcul est réalisé à l'aide de *GenForm* (Meringer et al., 2011) en autorisant les atomes suivants : C, H, N, O, P, S, Cl. Les formules moléculaires découlant de ce calcul sont ensuite triées selon leur plausibilité en prenant en compte différents critères :

- Seuls les marqueurs possédant un spectre en haute énergie sont considérés pour le calcul des formules moléculaires
- Lorsque plusieurs formules moléculaires sont proposées pour un unique marqueur, la formule ayant le score combiné le plus élevé est attribuée. Le score combiné correspond au produit du score de correspondance isotopique sur la MS¹ (entre 0 et 1) et du score de plausibilité des fragments en MS² (entre 0 et 1) par rapport à la formule calculée
- Seules les formules finales ayant un score combiné supérieur à 0,5 sont conservées (*i.e.*, au minimum 75 % de correspondance pour chacun des deux scores entrant dans le calcul)

IV. Résultats

IV.1. Etude des données obtenues avec le logiciel fournisseur et optimisations

IV.1.1. Méthode employée et limites observées

Pour fournir les résultats attendus lors de cet EIL, la méthode employée pour traiter les données s'est appuyée principalement sur les logiciels fournisseurs. Les données ont été acquises et prétraitées avec UNIFI, puis exportées vers EZInfo pour un traitement par AMV. Dans un premier temps, une ACP a été effectuée (Figure 8) afin de vérifier que les triplicatas d'injection d'un même échantillon étaient bien regroupés, et de s'assurer qu'une distinction entre les deux points de prélèvement ou les deux temps d'exposition des échantillonneurs passifs était visible.

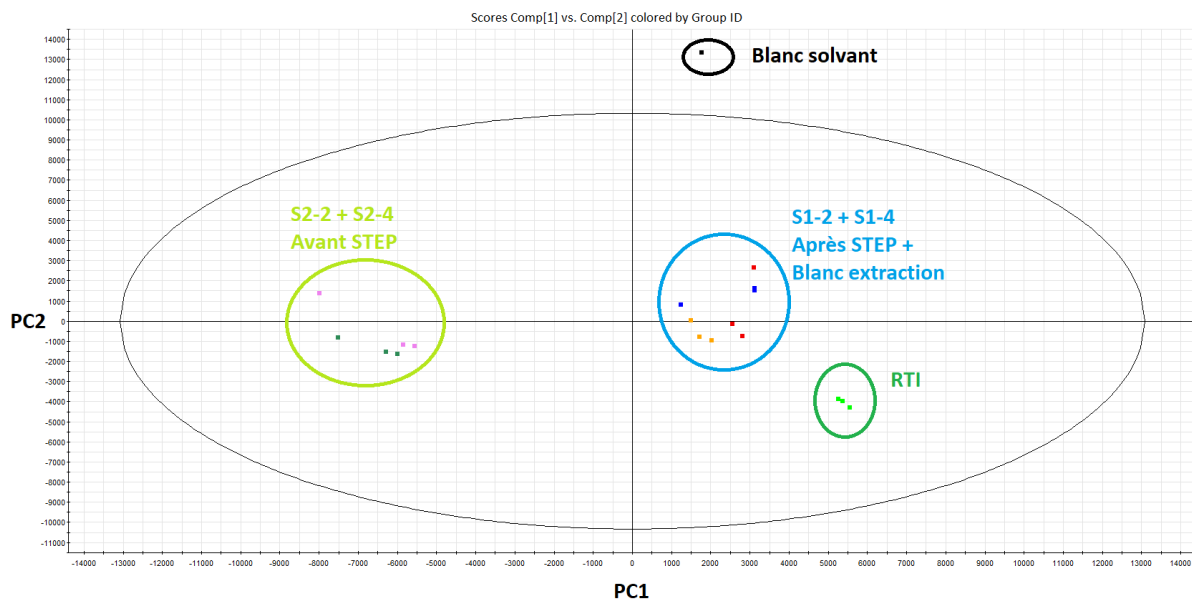


Figure 8 : Graphique des échantillons obtenu par ACP à partir des marqueurs issus de la séquence d'analyse de l'EIL (STEP = station de traitement d'eau potable, RTI = mélange de standards analytiques)

Les échantillons avant et après STEP sont bien distingués, mais peu de différences sont observées entre les deux temps d'exposition des échantillonneurs. Par ailleurs, les injections correspondant au blanc d'extraction sont retrouvées dans un cluster commun avec les échantillons après STEP, et le mélange de standards analytiques, bien que distinguable de ce cluster, se situe du même côté de l'ACP. Cette observation peut traduire le fait que les traitements réalisés en STEP sont efficaces pour l'élimination des micropolluants puisque les

empreintes des échantillons sont statistiquement proches des blancs. Par la suite, les échantillons ont été regroupés pour une comparaison deux à deux suivant le Tableau 9.

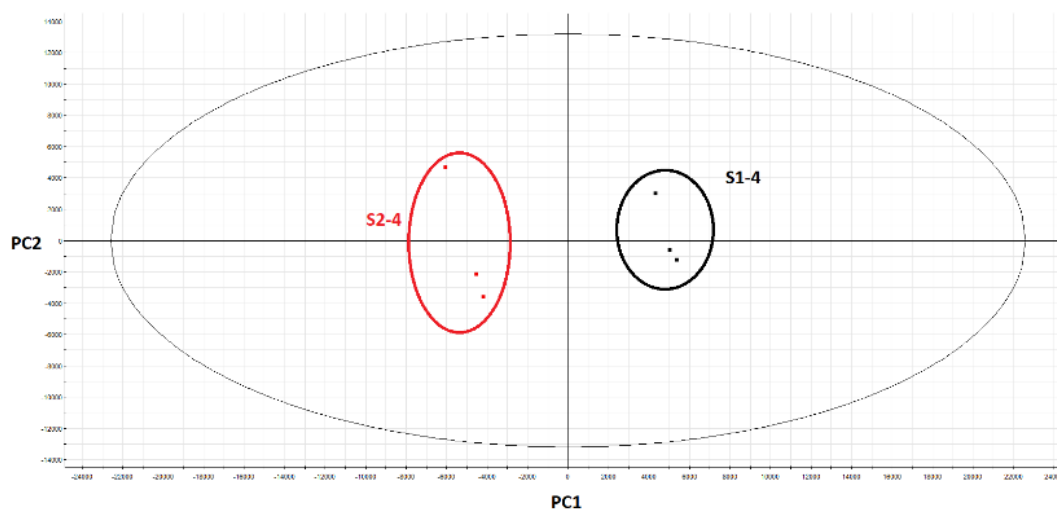
Tableau 9 : Groupes établis pour la comparaison par OPLS-DA des échantillons EIL

Echantillon 1	Echantillon 2	But de la comparaison
Avant STEP 2 jours d'exposition (S2-2)	Après STEP 2 jours d'exposition (S1-2)	Sélectionner les pics spécifiques du traitement (éliminés / formés) pour 2 jours d'exposition
Avant STEP 4 jours d'exposition (S2-4)	Après STEP 4 jours d'exposition (S1-4)	Sélectionner les pics spécifiques du traitement (éliminés / formés) pour 4 jours d'exposition
Blanc d'extraction	S2-2 + S1-2 + S2-4 + S1-4	Sélectionner les pics spécifiques au blanc d'extraction
Mélange de standards (RTI)	Blanc solvant	Sélectionner les pics spécifiques du mélange de standards

Ces regroupements ont permis de réaliser des OPLS-DA (Figure 9a) et de construire les *s-plot* (Figure 9b) correspondants afin de déterminer les marqueurs discriminants pour chaque échantillon. Par comparaison à l'ACP précédente, l'OPLS-DA est une méthode supervisée dont le but est de distinguer le plus possible des groupes définis, par rapport aux différentes observations, ici, les marqueurs analysés. Le modèle permet ainsi de créer une composante qui décrit cette différence (PC1, Figure 9a). Les observations (i.e., *marqueurs*) se verront alors attribuer un score proportionnellement à leur contribution à cette composante appelée *loading score* ($p[1]$). En traçant la corrélation de ce score avec la composante discriminante ($p(\text{corr})[1]$) en fonction du *loading score* ($p[1]$) pour toutes les observations, le *s-plot* résultant (Figure 9b) permet de sélectionner les marqueurs les plus pertinents pour expliquer la différence entre les deux groupes définis.

Les 100 marqueurs ayant le plus grand couple $p(\text{corr})[1]/p[1]$, pour chaque échantillon, ont été sélectionnés (Figure 9b). Ensuite, leur présence dans les trois réplicats d'injection et leur spécificité vis à vis des autres échantillons, ont été vérifiés manuellement. Les marqueurs ne respectant pas les deux critères précédemment énoncés ont été éliminés, et, parmi les restants, les 50 premiers marqueurs (en termes de couple $p(\text{corr})[1]/p[1]$) ont été conservés pour la suite.

a)



b)

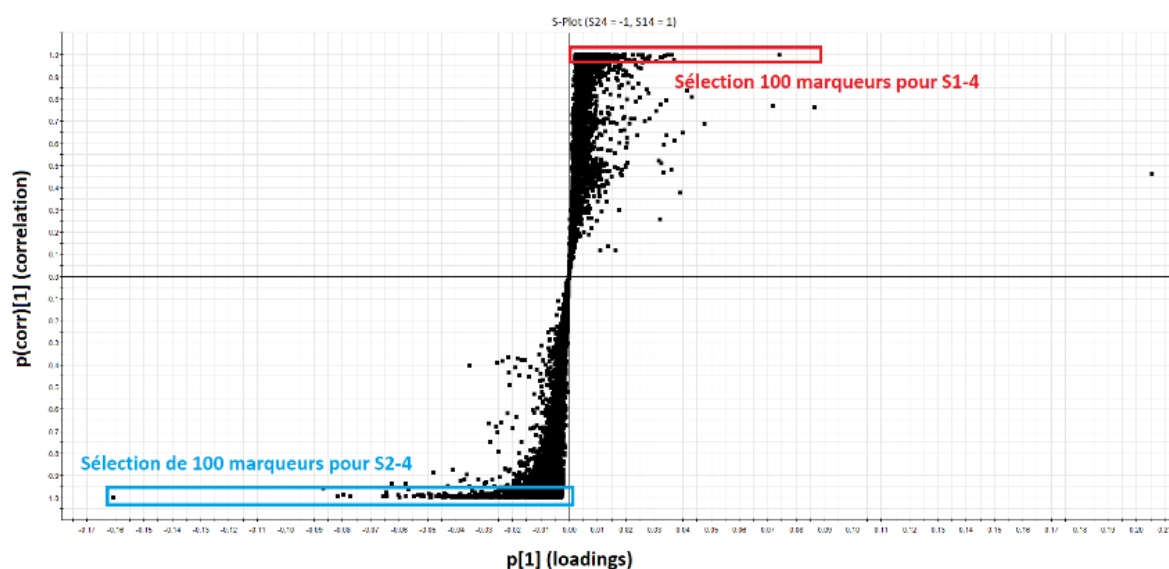


Figure 9 : Résultats de a) l'OPLS-DA et du b) s-plot pour la comparaison des marqueurs des échantillonneurs passifs (4 jours de contact)

Une fois les 50 marqueurs sélectionnés pour chaque échantillon, une tentative d'identification a été réalisée en renvoyant ces marqueurs dans le logiciel UNIFI. La fonction *Elucidate* de UNIFI permet de calculer des formules moléculaires et de rechercher une correspondance des marqueurs avec des molécules présentes dans une librairie interne ou externe. Lors d'identifications multiples pour un marqueur donné, les résultats ont été classés successivement par nombre de fragments correspondants (comparaison entre les fragments *in silico* et les fragments détectés), puis par intensité prédite des fragments identifiés par

rapport au nombre total de fragments détectés, et enfin par correspondance isotopique. Il peut être surprenant de ne prendre en compte la correspondance isotopique qu'en dernier critère de classement. Cela s'explique par le fait que lors de l'utilisation de la fonction *Elucidate*, de nombreuses formules moléculaires chimiquement improbables, voire parfois impossibles, se voient toutefois attribuer un score élevé en termes de correspondance isotopique. De même, des tests réalisés avec des étalons de référence ont parfois montré des propositions multiples de formules moléculaires, avec un score de correspondance isotopique bien plus faible pour la formule réelle que pour les autres propositions. Ces observations ayant montré le peu de fiabilité de ce critère dans ce logiciel, il a été décidé de ne le prendre en considération qu'en dernier critère d'identification.

La méthode employée pour cet essai a notamment montré ses limites lors de l'analyse de l'échantillon de standards analytiques. En effet, les autres échantillons étant totalement inconnus, il était difficile d'évaluer la qualité des résultats obtenus, à la fois en termes de sélection de marqueurs discriminants que de propositions d'identification. En utilisant la même méthode (OPLS-DA) pour distinguer les 50 marqueurs les plus discriminants entre le blanc solvant et l'échantillon de standards analytiques, aucun des 18 composés présents dans cet échantillon n'a été retrouvé parmi ces 50 marqueurs, que ce soit avec la méthode analytique imposée par les organisateurs (*méthode NORMAN*) ou avec la méthode de NTS utilisée couramment au laboratoire (*méthode LEESU*). Suite à ce constat, des optimisations ont été apportées aux étapes de traitement de données afin de permettre une meilleure identification des molécules.

IV.1.2. Optimisation de la méthode employée

IV.1.2.1. Identification des points faibles

Après les observations réalisées sur le manque de fiabilité de la méthode de traitement des données pour l'identification des molécules présentes dans le mélange de standards analytiques, une stratégie d'optimisation a été mise en place avec plusieurs objectifs. Le premier était d'identifier les étapes problématiques, c'est-à-dire, celles qui étaient responsables de l'élimination des marqueurs correspondants aux molécules connues. Le second objectif était de tenter d'optimiser les paramètres des étapes identifiées comme problématiques afin de fournir de meilleurs résultats (*i.e.*, nombre de molécules correctement identifiées).

Afin d'identifier les étapes les plus problématiques, les étalons ont tout d'abord été recherchés par SS en entrant leur structure dans la librairie interne pour recueillir les données de TR et DT correspondants à chaque étalon, ainsi que leurs spectres de masse. Ensuite, à partir de la *marker table*, les marqueurs pouvant correspondre aux étalons ont été recherchés à partir de leurs paramètres (*i.e.*, triplet m/z - TR - DT) relevés en SS. Les étapes suivantes consistent à attribuer un rang aux marqueurs par ordre décroissant de $p(\text{corr})[1]/p[1]$. Le rang correspondant aux marqueurs précédemment désignés comme pouvant correspondre aux étalons ont été relevés. Ceux-ci ont ensuite été soumis au processus d'*Elucidate*, en interrogeant les librairies ChemSpider l'une après l'autre, pour chacun des candidats pouvant correspondre aux 18 étalons. Ces différentes étapes ont été réalisées pour les deux séquences d'analyse (*méthode NORMAN* et *méthode LEESU*) afin de voir l'influence de la méthode d'analyse sur les résultats obtenus. Dans un premier temps, la comparaison entre la *méthode NORMAN* et la *méthode LEESU* montre que la seconde permet d'identifier potentiellement plus de composés que la méthode NORMAN, avec seulement 3 molécules manquantes sur 18, contre 5 sur la seconde en SS.

Ces résultats sont répertoriés dans le Tableau 10.

Tableau 10 : Comparaison de l'identification de composés d'un mélange de standards analytiques par les approches SS et NTS, avec description de l'étape limitante empêchant éventuellement l'identification

Etalon	Méthode NORMAN				Méthode LEESU			
	SS ^a	NTS ^b	Identification ^c	Elimination ^d	SS ^a	NTS ^b	Identification ^c	Elimination ^d
Guanylurea	Non	-	-	Analyse	Non	-	-	Analyse
Amitrol	Non	-	-	Analyse	Oui	1 (311)	Interne (1/1) Drugbank (FB/1) Massbank (1/1) NIST (1/6) LGC (1/3) Sigma (1/4)	Classement <i>s-plot</i>
Histamine	Non	-	-	Analyse	Non	-	-	Analyse
Chlormequat	Oui	2 (178 ; 404)	Interne (FB+/1) Drugbank (FB+/1) Massbank (FB+/1) NIST (1/1) LGC (FB+/1) Sigma (FB+/1)	Classement <i>s-plot</i> et recherche limitée à adduit H+	Oui	1 (316)	Interne (FB+/1) Drugbank (FB+/1) Massbank (FB+/1) NIST (1/1) LGC (FB+/1) Sigma (FB+/1)	Classement <i>s-plot</i> et recherche limitée à adduit H ⁺
Methamidophos	Oui	4 (344 ; 190 ; 305 ; N.C)	Interne (1/2 ; N.I; N.I) Drugbank (1/2; N.I; N.I) Massbank (FB/2; N.I; N.I) NIST (1/3; N.I; N.I) LGC (1/3; N.I; N.I) Sigma (1/5; N.I; N.I)	Classement <i>s-plot</i>	Oui	2 (593 ; 942)	Interne (1/2; 1/2) Drugbank (1/2; 1/2) Massbank (FB/2) NIST (1/2; 1/2) LGC (1/3 ;1/3) Sigma (1/3 ;1/3)	Classement <i>s-plot</i>

PARTIE A : Évaluation de méthodes de traitement de données

Etalon	Méthode NORMAN				Méthode LEESU			
	SS ^a	NTS ^b	Identification ^c	Elimination ^d	SS ^a	NTS ^b	Identification ^c	Elimination ^d
Vancomycin	Oui	3 (623 ; 712 ; N.C)	N.I ; N.I	Classement <i>s-plot</i> et mauvais calcul formule	Oui	2 (940 ; 1075)	N.I ; N.I	Classement <i>s-plot</i> et mauvais calcul formule
Cefoperazone	Oui	4 (1021 ; 1032 ; 2661 ; 5263)	N.I ; N.I ; N.I ; N.I	Classement <i>s-plot</i> et mauvais calcul formule	Oui	1 (N.C)	-	Détection non répétable
Trichlorfon (Dylox)	Oui	2 (564 ; N.C)	Interne (1/2) Drugbank (S/2) Massbank (S/2) NIST (S/2) LGC (S/2) Sigma (S/2)	Classement <i>s-plot</i>	Oui	1 (644)	Interne (1/1) Drugbank (1/1) Massbank (S/1) NIST (S/1) LGC (S/1) Sigma (S/1)	Classement <i>s-plot</i>
Butocarboxim	Non	-	-	Analyse	Oui	1 (N.C)	-	Détection non répétable
Dichlorvos	Oui	4 (44 ; 714 ; N.C ; N.C)	Interne (N.I ; 1/3) Drugbank (N.I ; 1/3) Massbank (N.I ; 1/3) NIST (N.I ; 1/3) LGC (N.I ; 1/3) Sigma (N.I ; 1/3)	Classement <i>s-plot</i> ou mauvais calcul de formule	Oui	1 (702)	1/1 dans toutes les librairies	Classement <i>s-plot</i>

PARTIE A : Évaluation de méthodes de traitement de données

Etalon	Méthode NORMAN				Méthode LEESU			
	SS ^a	NTS ^b	Identification ^c	Elimination ^d	SS ^a	NTS ^b	Identification ^c	Elimination ^d
Tylosin	Non	-	-	Processing (hors gamme <i>m/z</i> imposée)	Non	-	-	Processing (hors gamme <i>m/z</i> imposée)
TCMTB	Oui	1 (70)	N.I	Mauvais calcul formule	Oui	3 (53 ; 72 ; N.C)	Interne (N.I ; 1/1) Drugbank (N.I ; FB/1) Massbank (N.I ; FB/1) NIST (N.I ; 1/1) LGC (N.I ; 1/1) Sigma (N.I ; 1/1)	Mauvais calcul formule
Rifaximin	Oui	3 (439 ; 752 ; 1004)	N.I ; N.I ; N.I	Classement <i>s-plot</i> et mauvais calcul formule	Oui	1 (314)	Interne (1/3) Drugbank (1/3) Massbank (N.I) NIST (FB/3) LGC (1/4) Sigma (1/4)	Classement <i>s-plot</i>
Spinosad A	Oui	2 (49 ; 657)	Interne (N.I; 1/3) Drugbank (N.I; FB/3) Massbank (N.I; FB/3) NIST (N.I; FB/3) LGC (N.I; 1/3) Sigma (N.I; FB/3)	Mauvais calcul de formule ou classement <i>s-plot</i>	Oui	1 (237)	Interne (1/1) Drugbank (FB/1) Massbank (FB/1) NIST (FB/1) LGC (1/1) Sigma (FB/1)	Classement <i>s-plot</i> et absence de nom correspondant aux formules

PARTIE A : Évaluation de méthodes de traitement de données

Etalon	Méthode NORMAN				Méthode LEESU			
	SS ^a	NTS ^b	Identification ^c	Elimination ^d	SS ^a	NTS ^b	Identification ^c	Elimination ^d
Emamectin B1a	Oui	1 (1224)	Interne (1/4) Drugbank (FB/4) Massbank (FB/4) NIST (FB/4) LGC (FB/4) Sigma (FB/4)	Classement <i>s-plot</i> et absence de nom correspondant aux formules	Oui	1 (104)	N.I	Classement <i>s-plot</i> et mauvais calcul de formule
Avermectin B1a	Oui	1 (324)	N.I	Classement <i>s-plot</i> et mauvais calcul formule	Oui	1 (973)	N.I	Classement <i>s-plot</i> et mauvais calcul formule
Nigericin	Oui	1 (N.C)	-	Détection non répétable	Oui	3 (530 ; 670 ; 935)	N.I ; N.I ; N.I	Classement <i>s-plot</i> et mauvais calcul formule
Ivermectin B1a	Oui	5 (8335 ; 597 ; N.C ; N.C ; N.C)	N.I ; N.I ; N.I ; N.I ; N.I	Classement <i>s-plot</i> et mauvais calcul formule	Oui	1 (N.C)	-	Détection non répétable

^a : identifié ou non par le SS

^b : Nombre de marqueurs susceptibles de correspondre aux ID relevés en SS (rangs *s-plot* respectifs de ces marqueurs) N.C = non classé, car pas détecté dans chacun des trois réplicats d'injection

^c : Nature de l'identification pour chacune des librairies interrogées (X/nombre de propositions par librairie interrogée) pour chacun des marqueurs respectivement.

Quand X = 1 : identification correcte, X = FB : seule la formule moléculaire proposée est correcte ou seule la formule a été calculée, X = S : identification correcte par un synonyme, X = FB+ : la formule moléculaire est cohérente par rapport à l'adduit attendu, X = N.I : aucune identification correcte ou aucune identification proposée

^d : étape(s) qui ont contribué au fait que la molécule ne figure pas dans la liste finale des 50 marqueurs uniques du RTI

D'après les résultats, les plus gros points faibles de l'approche en NTS résidaient dans l'étape de sélection des marqueurs discriminants et celle de calcul des formules moléculaires, quelle que soit la méthode d'analyse employée. La *méthode LEESU* a empêché la prise en compte de légèrement plus de marqueurs que la *méthode NORMAN* à cause de l'absence de ces derniers dans un ou plusieurs réplicat(s). Cependant, la *méthode LEESU* a permis de retrouver plus de composés par une approche en SS que la *méthode NORMAN* et sera donc la seule détaillée pour la suite pour plus de lisibilité, bien que les optimisations aient été appliquées sur les deux sets de données.

IV.1.2.2. Optimisation du calcul des formules moléculaires

Afin d'optimiser le calcul des formules moléculaires, plusieurs sous-étapes ont été identifiées. Dans un premier temps, la tolérance en masse, initialement fixée à 2 mDa, a été élargie et l'influence de cette modification a été étudiée. Cette tolérance, calculée à partir de la masse mesurée, est utilisée après calcul des formules moléculaires afin de vérifier que les masses exactes des formules calculées sont comprises dans cette fenêtre avant proposition. Les atomes pris en compte pour le calcul des formules ont également été revus pour ne garder que C, H, N, O, P, S et Cl. En effet, le Fluor et le Brome ont été enlevés de la liste car ils n'apparaissent pas dans les formules des standards analytiques injectés. De plus, le Fluor ayant un seul isotope stable (^{19}F), il est retrouvé dans de très nombreuses formules moléculaires proposées puisqu'il n'a pas de massif isotopique caractéristique. A partir des formules moléculaires correspondantes aux composés d'intérêt, le nombre maximal de chaque atome autorisé pour le calcul de formule a également été réduit afin de diminuer le nombre de propositions. La recherche d'autres adduits que H^+ a été effectuée pour permettre l'identification des molécules dont des adduits différents sont attendus (ex. le chlormequat avec adduit $[\text{M}^+]$). Les différences entre les paramètres initiaux et les paramètres optimisés sont présentés dans le Tableau 11.

Tableau 11 : Optimisation des paramètres pour le calcul des formules moléculaires via UNIFI

	Paramètres initiaux	Paramètres optimisés
Éléments pris en compte	C (0-39), H (0-72), N (0-20), O (0-20), P (0-9), S (0-10), Cl (0-10), Br (0-5), F (0-16)	C (0-100), H (0-200), N (0-10), O (0-30), P (0-5), S (0-5), Cl (0-5)
Adduits recherchés	H ⁺ (automatique)	H ⁺ (manuel) et -e, NH ₄ ⁺ ou H ²⁺ pour les marqueurs spécifiques où l'adduit est précisé
Tolérance en <i>m/z</i>	2 mDa	2 mDa, puis 5 et 10 mDa
Electron state	Pair	Pair et impair pour le Chlormequat
DBE	-1,5 à 50	-1,5 à 50
Nombres d'isotopes	0 à 3	0 à 3
Autres paramètres	Règle de Senior, filtre sur le ratio C/H et C/X (avec X un hétéroatome) et filtre multi atomique	Règle de Senior, filtre sur le ratio C/H et C/X (avec X un hétéroatome) et filtre multi atomique

Le Tableau 12 regroupe les résultats obtenus pour l'identification des composés avec les deux sets de paramètres décrits. Le nombre de marqueurs correspondants est établi sur la base des données obtenues en SS : en prenant en compte tous les marqueurs détectés lors de l'analyse, ceux dont le *m/z*, le TR et le DT correspondent à ceux relevés lors de la recherche en SS ont été sélectionnés. Pour certains composés, plusieurs marqueurs pouvaient correspondre. Chaque marqueur sélectionné a ensuite été soumis à une tentative d'identification en utilisant les deux sets de paramètres décrits.

Tableau 12 : Amélioration des identifications grâce à l'optimisation de l'étape de calcul de formules moléculaires, pour l'analyse de standards analytiques avec la méthode LEESU

Composé	Identification en SS ?	Nombre de marqueurs correspondant	Paramètres initiaux		Paramètres optimisés	
			Formule moléculaire correcte ?	Identification correcte ?	Formule moléculaire correcte ?	Identification correcte ?
Amitrol	oui	1	Oui	Oui	Oui	Oui
Guanylurea	non	0	-	-	-	-
Histamine	non	0	-	-	-	-
Chlormequat	oui	1	Cohérent en recherche adduit H ⁺ mais l'adduit attendu est un M ⁺	-	Non	-
Methamidophos	oui	2	Oui	Oui	Oui	Oui
Butocarboxim	oui	1	Non	-	-	-
Dichlorvos	oui	1	Oui	Oui	Oui	Oui
TCMTB	oui	3	Oui	Oui	Oui	Oui
Trichlorfon	oui	1	Oui	Synonyme : métrifonate	Oui	Synonyme : métrifonate
Cefoperazone	oui	1	Pas de formule proposée	-	Pas de formule proposée	-
Vancomycin	oui	2	Pas de formule proposée	-	Oui	Oui
Nigericin	oui	3	Pas de formule proposée	-	Non	-
Spinosad A	oui	1	Oui	Non	Oui	Non
Rifaximin	oui	1	Oui	Non	Oui	Non
Emamectin B1a	oui	1	Non	-	Oui	Oui (à 5 mDa et 10 mDa)
Avermectin B1a	oui	1	Pas de formule proposée	-	Non	-
Ivermectin B1a	oui	1	Non	-	Avec adduit NH ₄ ⁺ à 2 mDa	Non
Tylosin	non	0	-	-	-	-

Le changement des paramètres d'identification a permis d'identifier légèrement plus de molécules du mélange que les paramètres initiaux avec la *méthode LEESU* : sur les 18 composés, 7 ont été identifiés correctement (contre 5 avec les paramètres initiaux), 3 n'ont pas été identifiés mais ont eu la bonne formule moléculaire attribuée (contre 2 avec les paramètres initiaux) et 8 n'ont pas été identifiés (contre 11 avec les paramètres initiaux). Avec la *méthode NORMAN*, le changement de paramètres n'a, en revanche, pas eu d'influence sur les résultats, avec 4 composés correctement identifiés, 3 composés avec une bonne attribution de formule moléculaire uniquement, et 11 non identifiés. Il est cependant important de noter que c'est la recherche d'adduits particuliers (et attendus) qui a permis d'augmenter le nombre d'identifications correctes. Le reste des modifications effectuées n'ont donc finalement eu que très peu d'influence sur les résultats.

IV.1.2.3. Optimisation de la sélection des marqueurs

La sélection de marqueurs pour le rendu des résultats a initialement été effectuée en combinant l'OPLS-DA au tracé du *s-plot* sur lequel les marqueurs ont été choisis en fonction de leur score de spécificité. Plusieurs problèmes ont pu être observés lors de l'utilisation de cette méthode. Tout d'abord, tous les marqueurs générés ne sont pas utilisés dans l'analyse multivariée. L'hypothèse est que les marqueurs absents n'étaient pas considérés comme significatifs par le logiciel. Cependant, aucune documentation n'a pu être trouvée sur la façon dont les marqueurs sont évalués, bien que des étapes de *trimming* et *winsorizing* soient évoquées pour d'autres logiciels d'AMV de Umetrics, supportant cette hypothèse. Généralement, environ 50 % des marqueurs générés par UNIFI sont éliminés automatiquement lorsque les données sont transférées vers EZInfo. Il est possible de les réinclure dans les tests statistiques. Pour cela, il est nécessaire de les sélectionner un par un dans un tableau qui contient à la fois les marqueurs conservés et éliminés, sans possibilité de tri automatique. Aucune tentative d'optimisation n'a été réalisée sur cette étape car il a été montré précédemment que celle-ci n'était pas une cause majeure au fait que les composés étudiés n'aient pas été identifiés.

Les deux autres problèmes sont quant à eux liés. Ils concernent la sélection des marqueurs sur le *s-plot*. En effet, il a été montré précédemment que cette étape était la plus problématique dans le processus. Tout d'abord car le classement des marqueurs sur leur score engendre l'élimination d'une grande partie des marqueurs d'intérêt, mais également que la sélection de

ces marqueurs nécessite une vérification manuelle afin de confirmer dans un premier temps leur présence réelle dans l'échantillon étudié et ensuite de leur spécificité vis-à-vis des autres échantillons. Pour tenter d'améliorer cette étape, la sélection de marqueurs d'intérêt a été opérée après établissement de l'OPLS-DA par le biais de deux autres types de représentations graphiques, le *bi-plot* et le *loadings plot*.

L'OPLS-DA a pour but de trouver les variables qui permettent discriminer le plus de groupes prédéfinis. Pour cela, l'algorithme procède par une réduction de dimensions en effectuant des projections linéaires afin de définir de nouvelles composantes qui permettent de maximiser la différence entre les moyennes des deux groupes, tout en minimisant la variabilité au sein d'un même groupe. Les sommes des carrés de ces distances par rapport aux composantes sont appelées *eigenvalue* et permettent d'obtenir un score associé à chaque observation qui servira à les placer sur l'OPLS-DA. Aux variables sont alors associées un *loading score*, représentant la contribution de chacune afin de créer ces composantes, les contributions normalisées étant appelées *eigenvector*. Le *loading plot* revient à tracer les *loading score* de chaque variable sur les deux composantes définies par l'OPLS-DA. Le *bi-plot*, permet de représenter à la fois les observations (scores, suivant leur corrélation aux composantes $t(\text{corr})$) et les variables (*loading score* suivant leur corrélation aux composantes $pc(\text{corr})$). Le *s-plot*, représente quant à lui le *loading score* suivant leur corrélation à la première composante ($p(\text{corr})$, celle qui permet de différencier les deux groupes prédéfinis) en fonction de leur *loading score* (p) sur cette composante. Les résultats des 3 types de représentation ont été comparés, en termes de spécificité des marqueurs et de sélection des marqueurs spécifiques issus du mélange de standards analytiques.

Pour le *loadings plot* la sélection des marqueurs discriminants se fait suivant la composante discriminante (PC1, axe des abscisses) (Figure 10).

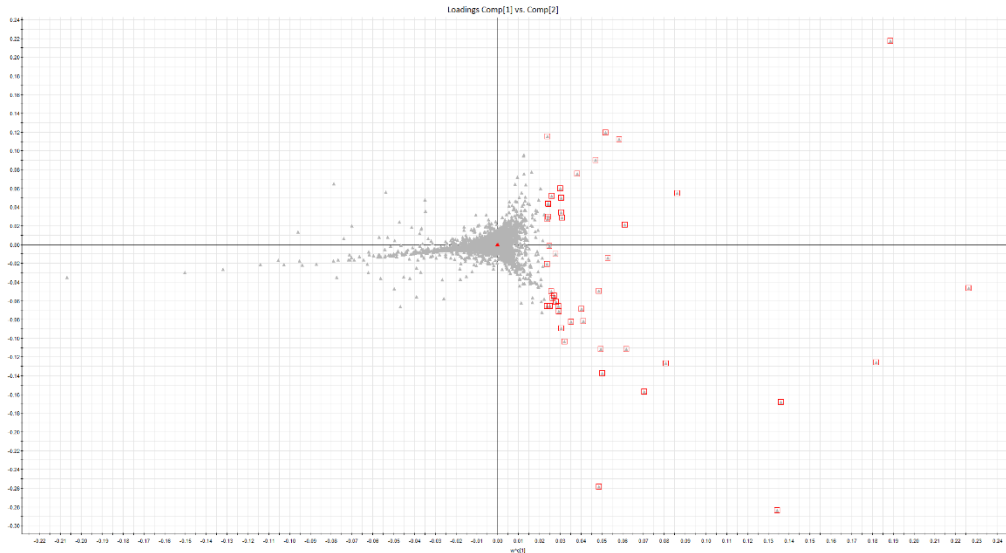


Figure 10 : Sélection des marqueurs discriminants (carrés rouges) sur le *loadings plot*

Le *bi-plot* représente sur une même figure les observations (injections) et les variables (marqueurs) suivant deux axes qui informent sur la relation qu'elles entretiennent vis-à-vis des autres (*i.e.*, corrélation, anti-corrélation, pas de corrélation...). Dans le logiciel EZInfo, le *bi-plot* permet également d'afficher un (ou plusieurs) « point représentatif » (*Observation Response*) de chaque groupe d'échantillon. Ce point représentatif représente une variable théorique modélisée par l'OPLS-DA qui discrimine le plus le groupe étudié par rapport au groupe auquel il est confronté. Pour la sélection des marqueurs d'intérêt sur cette représentation, les 100 marqueurs dont les coordonnées sont les plus proches du point représentatif de l'échantillon considéré sont sélectionnés (Figure 11).

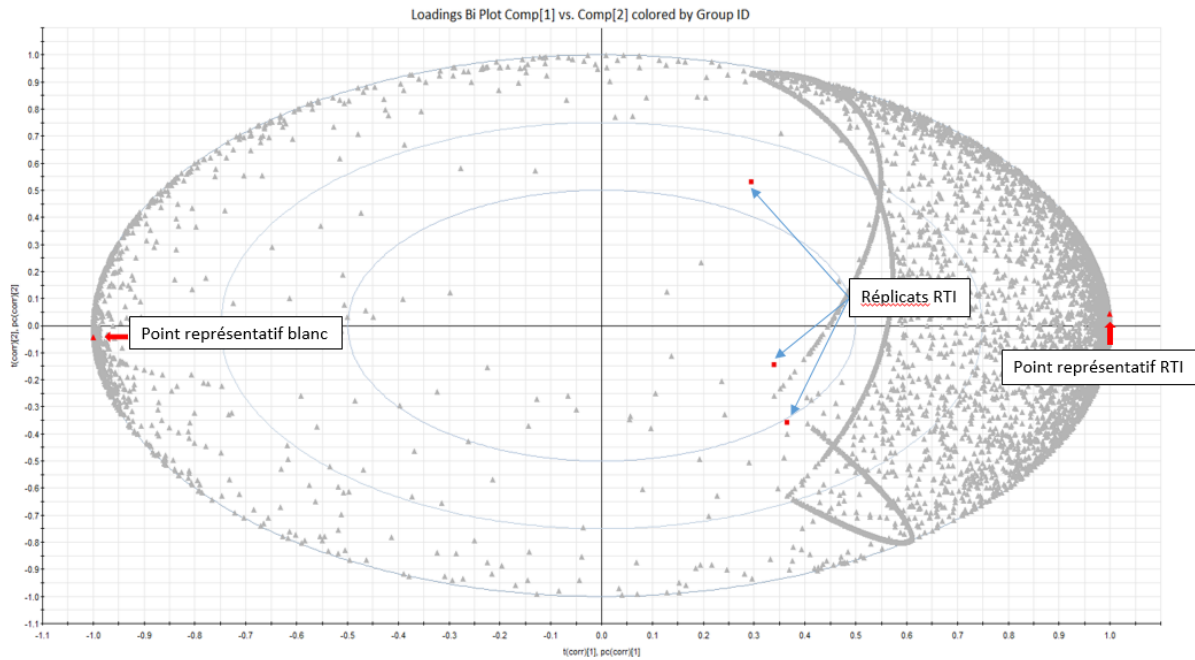


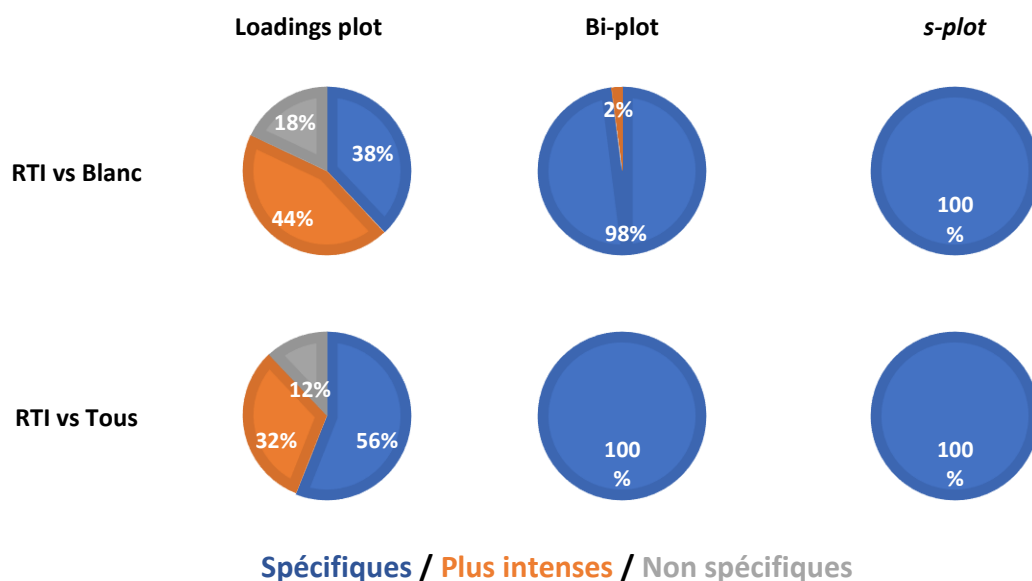
Figure 11 : Description de la sélection de marqueurs discriminants sur le *bi-plot*

Pour ces trois types de sélection, le mélange de standards a été comparé au blanc solvant (de la même manière qu'opéré dans la méthode initiale Tableau 9), puis comparé à un groupe formé de tous les autres échantillons. Les marqueurs sélectionnés ont ensuite été classés suivant 3 catégories :

- **Spécifiques** : lorsqu'ils sont présents dans les trois réplicats de l'échantillon et uniquement dans l'échantillon considéré
- **Plus intenses** : lorsqu'ils sont présents dans les trois réplicats de l'échantillon mais peuvent aussi se retrouver dans un ou plusieurs autres échantillons, mais avec une aire significativement plus faible
- **Non spécifiques** : lorsqu'ils ne sont pas présents dans les trois réplicats de l'échantillon et/ou qu'ils sont également présents dans d'autres échantillons avec une aire supérieure ou comparable à l'échantillon d'intérêt

Seuls les résultats pour l'échantillon mélange de standards analytiques seront présentés ici (Figure 12).

Figure 12 : Comparaison de la spécificité des méthodes de sélection des 100 marqueurs les plus discriminants



En termes de spécificité, le *s-plot* donne autant, voire plus de marqueurs spécifiques que le *bi-plot*, qui donne à son tour plus de marqueurs spécifiques que le *loadings plot*, ce qui montre une forte dépendance des sélections aux *loadings vectors* choisis pour l’affichage de chaque graphique (t , $p(\text{corr})$ et $pc(\text{corr})$). Il est également à noter que comparer l’échantillon de standards analytiques à tous les échantillons donne de meilleurs résultats en termes de spécificité que de le comparer seulement au blanc solvant. Cela s’explique notamment par le fait que le nombre d’échantillons utilisés ici est trop faible pour réaliser ces tests statistiques et espérer avoir des résultats très fiables comme cela a été démontré par différents auteurs. En effet, dans le cas où la comparaison se fait entre deux échantillons, seules trois observations (*i.e.*, trois répliqués d’injection) sont présentes dans chacun des groupes considérés pouvant ainsi conduire à un *over-fitting* des données (Bylesjö et al., 2006; Tapp and Kemsley, 2009). En passant à 3 observations d’un côté contre 18 observations de l’autre, les marqueurs spécifiques des 3 répliqués de standards analytiques sont mieux discriminés des autres échantillons.

Sur l’ensemble des marqueurs sélectionnés à partir de l’échantillon mélange de standards analytiques, une tentative d’identification a été effectuée pour tous les sets de données (Tableau 13). L’identification a été menée avec les paramètres optimisés décrits dans la partie A – IV.1.2.2 et en utilisant les 5 librairies employées lors de l’EIL.

Tableau 13 : Comparaison des identifications correctes pour les 3 méthodes de sélection de marqueurs. RTI = mélange de standards analytiques

	Loadings plot	Bi-plot	s-plot
RTI vs Blanc			
Composés identifiés	Dichlorvos, Spinosad A, Rifaximin, Emamectin B1a, Trichlorfon	Nigericin, Spinosad A	Dichlorvos, TCMTB
RTI vs Tous			
Composés identifiés	Chlormequat, Dichlorvos, Trichlorfon, Spinosad A, Rifaximin, Emamectin B1a, Ivermectin B1a	Nigericin, Emamectin B1a	Chlormequat, Methamidophos, Dichlorvos, TCMTB, Spinosad A, Emamectin B1a

De manière générale, comparer l'échantillon contenant les standards analytiques à l'ensemble des autres échantillons a permis d'identifier plus de composés, en accord avec la meilleure spécificité des marqueurs discriminés par cette méthode. A l'inverse, les *loadings plot* ayant montré une sélectivité moins importante (*i.e.*, moins de marqueurs *spécifiques* sélectionnés, Figure 12 [Erreur ! Source du renvoi introuvable.](#)) permettent cependant d'identifier le plus de molécules d'intérêt par rapport aux deux autres types de représentation quel que soit le type de comparaison employé.

IV.2. Comparaison avec des méthodes développées sur d'autres outils

IV.2.1. Méthode hybride : logiciel fournisseur et libre

IV.2.1.1. Sélection par des méthodes statistiques

Dans cette partie la méthode précédente a été légèrement modifiée afin de s'affranchir de l'utilisation d'EZInfo. En effet, comme mentionné précédemment, aucune information n'est disponible concernant la procédure de sélection des marqueurs utilisée par EZInfo. De plus, le logiciel EZInfo était peu performant (lenteur, impossibilité d'exportation directe des résultats).

La *marker table* créée par UNIFI a été exportée au format CSV pour être retraitée avec le logiciel R. Tout d'abord, une première filtration des données a été appliquée à la *marker table*

afin de ne prendre en compte que les marqueurs présents dans les trois réplicats d'injection. Tous les marqueurs présents dans seulement un ou deux réplicat(s) ont donc été supprimés, éliminant ainsi 40 % des marqueurs générés. Une OPLS-DA et le *s-plot* correspondant ont ensuite été générés à l'aide d'un script R développé à l'aide du package *ropls*. 100 marqueurs ont ensuite été sélectionnés de la même manière qu'à partir d'EZInfo, en fonction de leurs scores sur le *s-plot*, afin de comparer ces résultats à la sélection effectuée entièrement sous logiciel fournisseur. Des tentatives d'identification ont alors été effectuées sur ces deux sets de données en utilisant deux manières différentes. La première consiste en une identification manuelle de chaque marqueur dans UNIFI (fonction *Elucidate*). La seconde s'appuie sur l'utilisation du logiciel *GenForm* sous R pour l'obtention des formules moléculaires. Cette méthode a toutefois requis l'utilisation de UNIFI afin de récupérer les spectres de basse et haute énergies de chacun des marqueurs sélectionnés. La récupération des données spectrales est possible au cas par cas à partir de UNIFI, par copie d'un tableau des valeurs pour chaque marqueur (couples *m/z* - intensité). Ces données spectrales sont ensuite lues par *GenForm* pour le calcul des formules moléculaires. Une fois la formule moléculaire obtenue, l'identification des marqueurs a été réalisée via Metfrag et PubChem (Tableau 14).

Tableau 14 : Comparaison des outils pour l'identification après sélection statistique des marqueurs

	<i>s-plot</i> logiciel libre	<i>s-plot</i> logiciel fournisseur
<i>Elucidate</i> (UNIFI)	Chlormequat, Dichlorvos, TCMTB, Trichlorfon, Spinosad A	Chlormequat, Methamidophos, Dichlorvos, TCMTB, Spinosad A, Emamectin B1a
<i>Genform</i> (R)	Chlormequat, Dichlorvos, TCMTB, Trichlorfon, Spinosad A, Rifaximin, Emamectin B1a	Chlormequat, Dichlorvos, TCMTB, Spinosad A, Emamectin B1a

Lorsque l'identification est réalisée par le logiciel fournisseur, les 100 marqueurs sélectionnés sur le *s-plot* tracés par le logiciel fournisseur permettent d'identifier le Methamidophos en

complément des composés retrouvés dans les 100 marqueurs sélectionnés par le logiciel libre. En revanche, lorsque l'identification est réalisée par logiciel libre, le trichlorfon et la rifaximin sont retrouvés en plus dans le set de données issus du *s-plot* obtenu par logiciel libre par rapport au logiciel fournisseur. La combinaison des deux étapes réalisées sous logiciels libre permet ainsi d'identifier le plus de composés du mélange de standard (7 sur 18), en particulier, en permettant d'identifier la Rifaximin qui n'est pas retrouvé dans les autres cas.

IV.2.1.2. Sélection par des méthodes de comparaison simples

Par la suite, une seconde méthode de sélection a été testée. À l'aide d'un autre script, les marqueurs présents uniquement dans l'échantillon de standards analytiques ont été conservés par comparaison à l'ensemble des marqueurs présents dans tous les autres échantillons, soit 1 075 marqueurs sur les 59 578 initialement contenus dans la *marker table* exportée de UNIFI. Pour comparaison, le même principe a été suivi dans UNIFI avec un tri manuel des marqueurs présents uniquement dans les trois réplicats de l'échantillon de standards (vérification graphique sur chacun des 59 578 marqueurs). 989 marqueurs *spécifiques* ont été isolés de cette manière, une centaine de marqueurs ayant donc été mal catégorisés. Ceci illustre clairement l'intérêt de l'utilisation d'outils de programmation supplémentaires (avec logiciels libres) adaptés aux besoins de l'utilisateur, en comparaison des outils limités disponibles dans le logiciel fournisseur et de procédures « manuelles » (non-automatisées). De la même manière que précédemment, ces deux sets de données ont été soumis à des tentatives d'identification par les deux types de logiciels (Tableau 15).

Tableau 15 : Comparaison des outils pour l'identification après sélection des marqueurs par comparaison simple

	Sélection d'unique logiciel libre	Sélection d'unique logiciel fournisseur
<i>Elucidate (UNIFI)</i>	Chlormequat, Methamidophos, Dichlorvos, TCMTB, Trichlorfon, Spinosad A, Rifaximin	Dichlorvos, TCMTB, Trichlorfon, Rifaximin, Ivermectin B1a, Avermectin B1a
<i>Genform (R)</i>	Chlormequat, Dichlorvos, TCMTB, Trichlorfon, Spinosad A, Rifaximin, Emamectin B1a	Dichlorvos, TCMTB, Trichlorfon, Rifaximin

Ici, la sélection par logiciel libre permet toujours d'identifier plus de marqueurs par comparaison à la sélection réalisée par le logiciel fournisseur. Par comparaison des identifications de marqueurs uniquement retrouvés dans le mélange lorsque la sélection est effectuée par logiciel fournisseur, ce dernier permet d'identifier deux composés supplémentaires par rapport à l'identification par logiciel libre : l'ivermectin et l'avermectin. Ces deux composés sont attendus avec l'adduit NH_4^+ (Tableau 6) qu'il n'est pas possible d'imposer lors du processus d'identification avec Genform ce qui explique qu'ils ne sont pas identifiés ici. Dans le cas de la sélection de marqueurs sur la base de leur présence dans le mélange de standard uniquement, l'identification par logiciel libre permet donc globalement de retrouver plus d'étalons attendus.

IV.2.1.3. Comparaison des méthodes hybrides

La différence entre ces deux méthodes de sélection réside dans le fait que la première est une méthode statistique (OPLS-DA) dont le but est de déterminer les marqueurs qui permettent de discriminer le plus les deux groupes considérés. Les marqueurs discriminants peuvent donc être présents dans les deux groupes d'échantillons, mais avec des intensités significativement

différentes. La deuxième méthode est une simple comparaison deux à deux pour s'assurer de la présence ou de l'absence d'un marqueur dans un groupe d'échantillon donné. Par la suite, et, pour les deux sets de marqueurs sélectionnés, tous les marqueurs dont les m/z correspondent, à l'unité près, à un composé du mélange sont filtrés et conservés pour les étapes suivantes. Parmi les marqueurs sélectionnés (Figure 13), 9 marqueurs peuvent correspondre parmi les 100 priorités par le *s-plot* du logiciel libre ou logiciel fournisseur, 10 parmi les 989 uniques trouvés avec le logiciel fournisseur et 13 parmi les 1075 uniques sélectionnés par le logiciel libre.

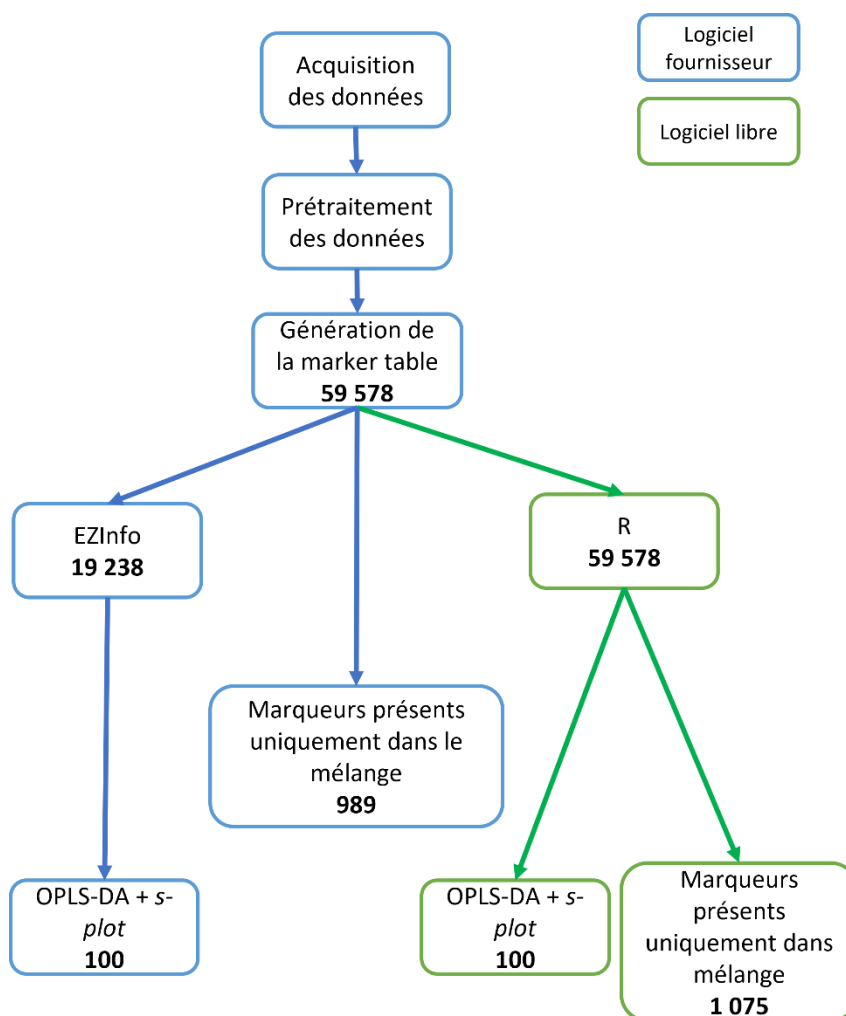


Figure 13 : Comparaison de différentes méthodes de sélection en employant les logiciels fournisseur ou libre

Lorsque l'identification est effectuée par le logiciel fournisseur, les deux méthodes de sélection réalisées par le logiciel fournisseur donnent le même nombre de composés identifiés (6 composés), alors que lors de la sélection par logiciel libre, la méthode statistique (OPLS-DA et *s-plot*) est moins efficace que la méthode qui consiste à sélectionner les marqueurs

uniquement présents dans le mélange de standard (5 vs 7 composés identifiés). Lorsque la sélection de marqueurs présents uniquement dans le mélange d'étalons est considérée, celle-ci permet d'identifier correctement 7 composés lorsqu'elle est effectuée par logiciel libre. Par comparaison, cette méthode de sélection de marqueurs uniques semble légèrement plus adaptée pour identifier les composés d'intérêt puisque le seul cas où elle s'avère moins efficace se traduit simplement par l'impossibilité d'imposer un adduit attendu. En revanche cette méthode présente l'inconvénient de ne pas fournir de classement (c'est-à-dire de priorisation statistique des marqueurs d'intérêt) et la liste de marqueurs uniquement présents dans un échantillon peut être longue. Si besoin, il serait possible d'adopter une stratégie complémentaire pour classer ces marqueurs, par exemple en fonction de leur intensité.

Il est intéressant de noter que l'identification par logiciel libre permet de retrouver presque systématiquement l'Emamectin mais pas le Metamidophos à l'inverse de l'identification réalisée par logiciel fournisseur. Cette différence peut être liée à la différence des spectres utilisés dans les deux cas. En effet, l'identification sous logiciel fournisseur est réalisée sur le marqueur (c'est-à-dire avec des spectres moyennés) alors que la récupération des spectres ne peut être effectuée que sur le spectre d'un pic détecté (c'est-à-dire dans une injection, avant alignement).

La comparaison entre une méthode réalisée entièrement sous logiciel fournisseur et une méthode utilisant à la fois le logiciel fournisseur avec une partie sous logiciel libre semble montrer que la seconde permet d'identifier plus de composés d'intérêt pour des méthodes de sélection de marqueurs comparables. Enfin, comparée à la méthode permettant d'identifier le plus de composés d'intérêt sous logiciel libre (*i.e.*, loadings plot), la méthode de sélection par les uniques et identification sous logiciel libre permet d'identifier le même nombre de composés, dont un qui diffère entre les deux méthodes sur les 7 identifiés (respectivement Ivermectin B1a par logiciels fournisseur et TCMTB pour la méthode hybride).

IV.2.2. Méthode entièrement sous logiciels libres

Etant donné les résultats obtenus dans la partie précédente, une dernière méthode a été employée pour traiter ces données, basée entièrement sur des logiciels libres (UNIFI ayant été uniquement utilisé pour l'acquisition des données). Les données brutes ont été directement converties avec MSConvert via l'interface de programmation de UNIFI, vers le format mzML et en supprimant la dimension de mobilité ionique (cf partie A – III.1.3). Le prétraitement des données, jusqu'à présent géré par UNIFI, a alors été réalisé dans R grâce au package *patRoan* (Helmus et al., 2021).

La première étape consiste à détecter les pics puis les aligner à travers les échantillons. Pour cela, deux packages sont proposés par *patRoan* : *xcms* (Smith et al., 2006) et *OpenMS* (Sturm et al., 2008). Bien que le premier semble plus couramment utilisé dans la littérature, le second est décrit par l'auteur de *patRoan* comme étant plus rapide et permettant d'obtenir un plus grand nombre de pics (Brock et al., 2020). L'inverse étant observé pour nos données (Tableau 16), la suite du traitement a donc été réalisé à partir des données après détection par *xcms*. Les paramètres pour la détection des pics et le regroupement des pics ont été choisis après avoir testé plusieurs combinaisons, pour chacun des algorithmes. Afin d'évaluer ces derniers, une première étape consistait à rechercher manuellement dans la *marker table* construite sous *patRoan*, les marqueurs pouvant correspondre aux composés du RTI en se basant sur les m/z et RT obtenus par l'étude en SS sous UNIFI. Cette étape montre que 11 marqueurs sont susceptibles de correspondre par *xcms* (sur 15 composés potentiellement identifiables, les manquants étant le Butocarboxim, la Céfoperazone, l'Emamectin et la Nigericine). Les étapes de conversion (centroïdisation, détection des pics et alignement) semblent donc conserver la majorité des marqueurs d'intérêt. Les paramètres retenus, aboutissant à la détection du plus grand nombre de composés présents dans le mélange, sont détaillés dans le Tableau 16. L'avantage de l'utilisation de *patRoan* réside dans le fait que les données spectrales sont récupérées à partir des données brutes et ensuite associées au marqueur correspondant, contrairement aux *marker table* issues de UNIFI.

Tableau 16 : Comparaison des paramètres de détection et alignement de *features* par les 3 méthodes

	Paramètres sous UNIFI	Paramètres OpenMS	Paramètres xcms
Détection des pics	Algorithme de détection 4D m/z = 10 ppm	m/z = 5 ppm	m/z = 10 ppm largeur de pic \in [5,30] sec
Alignement des pics	RT = 12 sec DT = 0,3 ms	m/z = 0,005 Da RT = 12 sec	m/z = 0,005 Da RT = 12 sec
Nombre de marqueurs résultant	59 578	1 910	11 254

Une fois l'équivalent de la *marker table* obtenue, elle est traitée sous R de la même façon que décrite dans la partie A – IV.2.1.1. Globalement, après une filtration des marqueurs suivant les réplicats, les marqueurs *uniques* du mélange ont été sélectionnés (non-présents dans les autres échantillons) et un calcul de formules moléculaires a été réalisé avec *Genform*, puis une tentative d'identifications a été menée via Metfrag et PubChem. Avec *OpenMS*, 215 marqueurs ont été classifiés dans la catégorie uniques, contre 930 avec *xcms*. Après tentative d'identification de ces marqueurs, seul le TCMTB a été correctement identifié dans la liste des marqueurs d'*OpenMS* parmi 4 marqueurs trouvés en SS, alors que pour *xcms*, le Methamidophos et le Trichlorfon ont été retrouvés en complément de celui-ci, parmi 11 composés retrouvés en SS.

De façon intéressante, l'utilisation de *Genform* pour calculer des formules moléculaires dans la partie A – IV.2.1.1 a été plus efficace en comparaison des résultats obtenus ici. La question s'est donc posée de savoir si les étapes consistant à convertir les données puis les prétraiter par des logiciels tiers avaient une influence sur le nombre de composés effectivement identifiés en fin de processus. Pour répondre à cette question, une étude plus approfondie a été menée afin d'évaluer la qualité des données obtenues après conversion.

Après l'étape de calcul de formules moléculaires par *Genform*, 4 marqueurs (Amitrol, Vancomycine, Avermectin et Ivermectine) ont été éliminés car aucune formule moléculaire calculée n'était cohérente avec la formule attendue. Par ailleurs, pour 3 autres marqueurs (Dichlorvos, Spinosad A et Rifaximin), la bonne formule moléculaire a bien été calculée et proposée par *Genform* mais avec un score combiné qui n'était pas le plus élevé parmi les différentes propositions (5e, 278e et 40e formule classée respectivement). Ceci explique que la formule correcte n'a alors pas été retenue. Ainsi, seuls 4 marqueurs s'étaient vu attribuer

la bonne formule moléculaire, parmi lesquels seul le Chlormequat a été identifié correctement après confrontation en librairie.

L'étape la plus problématique était donc le calcul de formule moléculaire, malgré les bons résultats obtenus précédemment (*i.e.*, avec les données spectrales exportés de UNIFI, partie A – IV.2.1.1). La différence majeure réside dans l'élimination de la mobilité ionique lors de la conversion des données au format mzML (fusion de tous les scans de mobilité ionique pour un même RT). La séparation supplémentaire apportée par la mobilité ionique permet d'obtenir des spectres beaucoup plus propres et donc spécifiques de la molécule détectée, débarrassés des pics de molécules co-éluées (Figure 14). La Figure 14B illustre ce phénomène : l'ion m/z 267,1698 (clairement isolé sur la Figure 14A grâce à la mobilité ionique) est masqué par d'autres ions présents au même temps de rétention.

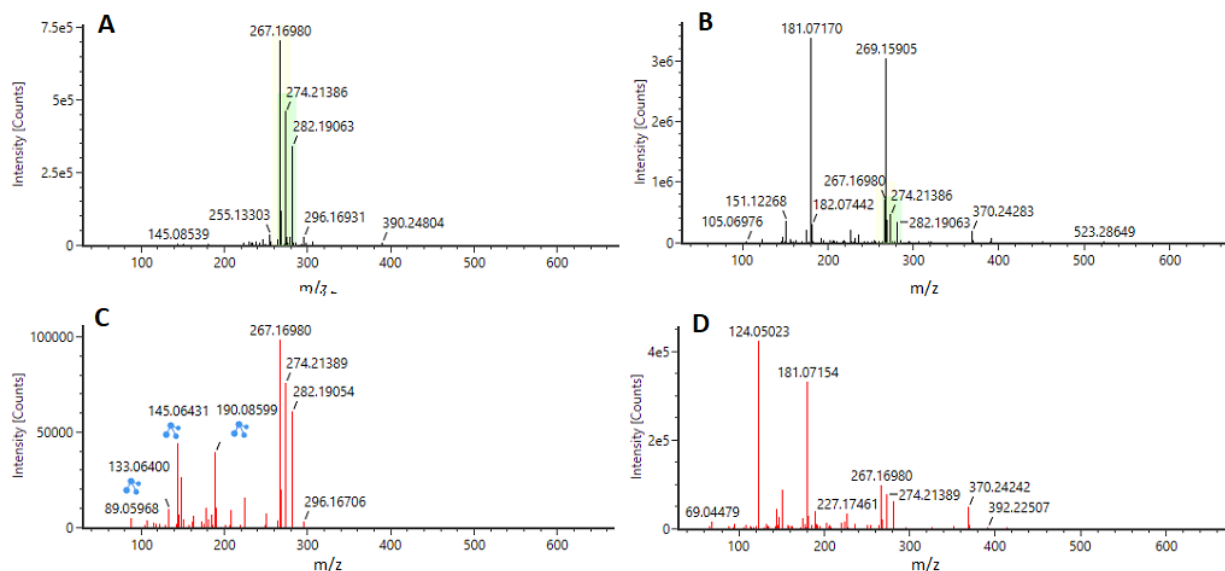


Figure 14 : Comparaison des spectres d'un même marqueur ($m/z = 267,16980$, RT = 3,0 min) A) MS avec DT, B) MS sans DT, C) MS^E avec DT, D) MS^E sans DT

Afin de vérifier cette hypothèse, une dernière étape a été implémentée. A partir des composés identifiés en SS dans UNIFI, les spectres en haute et basse énergies avec et sans IMS ont été exportés. Le calcul des formules moléculaires par *Genform* a ainsi été effectué sur les deux séries de données pour chaque marqueur et les résultats sont présentés dans le Tableau 17.

Tableau 17 : Comparaison du calcul de formule moléculaire par *Genform* à partir des spectres avec et sans mobilité ionique

Molécule	Sans IMS (exportés de UNIFI)		Avec IMS (exportés de UNIFI)		Sans IMS (via <i>patRoön</i>)
	Nombre de formule moléculaire calculées	Classement de la formule correcte parmi les propositions (NON = formule correcte pas proposée)	Nombre de formule moléculaire calculées	Classement de la formule correcte parmi les propositions (NON = formule correcte pas proposée)	Résumé de résultats
Amitrol	0	NON	5	NON	NON
Avermectin B1a	57	NON	7437	NON	NON
Butacarboxim	7	NON	7	NON	NON
Cefaporazone	1	NON	1117	590	N.A.
Chlormequat	0	NON	2	1	1
Dichlorvos	16	1	16	1	5
Emamectin B1a	9872	204	9872	1	N.A.
Ivermectin B1a	41	NON	2988	NON	NON
Methamidophos	0	NON	3	NON	1
Nigericin	799	NON	799	NON	N.A.
Rifaximin	1143	220	1143	1	40
Spinosad A	726	50	726	29	278
TCMTB	6	NON	17	1	1
Trichlorfon	16	NON	22	1	1
Vancomycin	0	NON	4575	NON	NON

Ces résultats démontrent l'intérêt de l'IMS pour l'attribution des formules moléculaires et l'identification. En effet, la propreté des spectres permet de calculer correctement la formule moléculaire pour 6 composés, contre seulement 1 sans mobilité ionique. D'autre part, il est également intéressant de remarquer une différence entre les données sans IMS issues du logiciel fournisseur et celles qui sont obtenues après conversion et détection par logiciel libre. En effet, ces dernières permettent d'obtenir 3 formules moléculaires correctes supplémentaires, suggérant que malgré la conversion des données, les spectres obtenus sous logiciels libre sont plus propres que les spectres obtenus sans IMS dans UNIFI.

IV.3. Comparaison des méthodes appliquées

IV.3.1. Avantages et inconvénients des différentes voies possibles

Dans l'étude décrite précédemment, trois voies de traitement des données ont pu être employées (Figure 15).

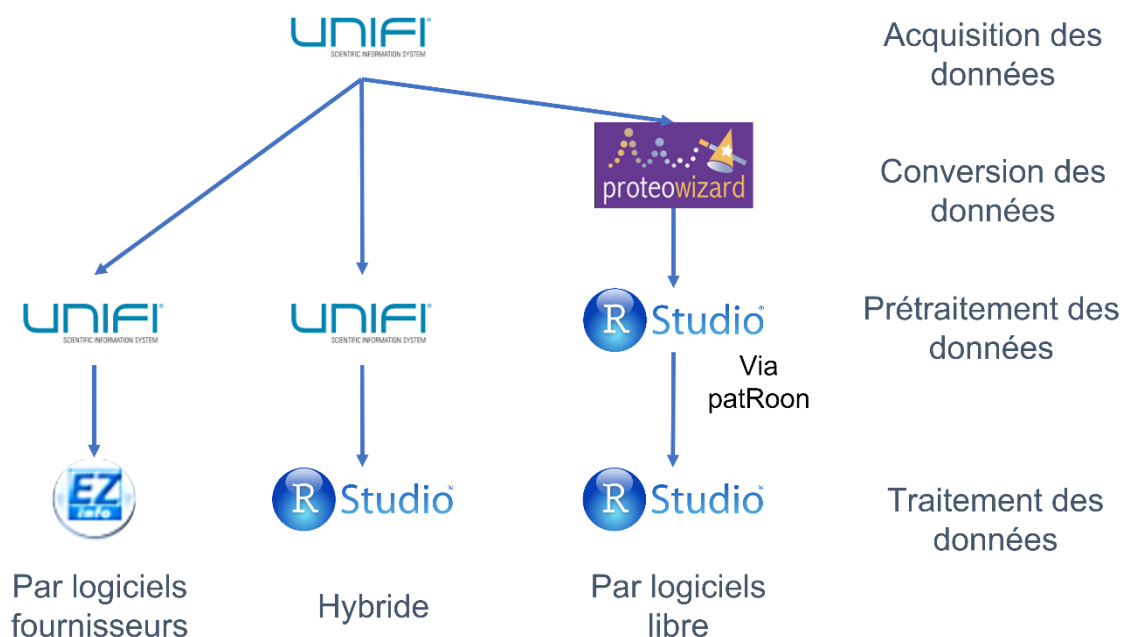


Figure 15 : Voies de traitement des données distinguées

Dans tous les cas, l'acquisition des données a été réalisée à l'aide du logiciel fournisseur. Chaque voie de traitement (par logiciels fournisseurs, hybride ou par logiciels libres) a ensuite présenté ses propres avantages et inconvénients.

IV.3.1.1. Comparaison d'ordre pratique

En ce qui concerne la voie par logiciels fournisseurs, un des plus grands inconvénients identifiés est le fait que de nombreux paramètres de traitement sont entièrement gérés par le logiciel, parfois sans aucune possibilité d'optimisation pour l'utilisateur, et sans information sur les algorithmes utilisés. D'autre part, l'utilisateur est limité aux possibilités proposées par le logiciel en termes de traitement et de visualisation, qui sont de plus souvent consommatrices en temps de traitement (notamment le traitement statistique des données de *marker table* via EZInfo). Cependant, le logiciel fournisseur est, par nature, parfaitement adapté aux données acquises ce qui permet de profiter pleinement de tous les avantages compris dans ces données, notamment concernant les données de mobilité ionique qui permettent une séparation complémentaire permettant de faciliter les tentatives d'identification.

Concernant la méthode hybride, celle-ci permet une grande variété de possibilités en termes de visualisation de données de façon très rapide. Elle se base sur les données prétraitées par le logiciel fournisseur et exportées dans la *marker table* permettant de comparer les différents échantillons analysés. Cependant, les données spectrales ne sont pas accessibles de façon aussi directe, empêchant ainsi une identification fiable en se basant uniquement sur ces données. Ainsi, cette méthode hybride est la plus avantageuse lorsqu'il s'agit d'avoir une vision globale rapide des échantillons mais reste limitée pour l'identification de marqueurs d'intérêt.

Enfin, la méthode utilisant entièrement des logiciels libres permet d'avoir un contrôle total sur les données puisque toutes les étapes de traitement et tous les paramètres associés sont choisis par l'utilisateur, et les algorithmes employés librement accessibles. De même, les données obtenues peuvent être utilisées avec les outils de traitement statistiques et de visualisation développés pour la méthode hybride, avec cette fois l'avantage de disposer des données spectrales permettant d'aller jusqu'à l'identification de marqueurs d'intérêt. Cependant, cette méthode nécessite une conversion préalable des données acquises par le logiciel fournisseur ce qui présente plusieurs inconvénients. Le premier est le temps de conversion nécessaire qui s'additionne au temps de prétraitement, habituellement directement effectué en parallèle de l'acquisition. Il est également nécessaire de disposer d'un espace de stockage non négligeable pour ces données converties. Enfin, le point négatif

majeur est la perte de la mobilité ionique lors de la conversion, ce qui en entraîne une fiabilité moindre dans l'attribution de formules moléculaires. A l'heure actuelle, la conversion en gardant la mobilité ionique est possible avec *MSCConvert* mais les logiciels libres de traitement des données ne supportent pas encore ce type de données.

De façon générale, le traitement décrit dans les parties précédentes commence par une filtration des données sur réplikat d'injections afin de pouvoir éliminer les marqueurs qui ne sont pas présents dans tous les réplikats d'un échantillon donné. Cette étape ne peut pas être mise en place de façon automatique en utilisant la voie par logiciels fournisseurs. En effet, au sein du logiciel, les injections peuvent être comparées deux à deux sur la liste des marqueurs mais il n'est pas possible de regrouper les pics communs à plusieurs injections du même échantillon.

IV.3.1.2. Comparaison en NTS

De façon globale, la méthode utilisant les logiciels fournisseurs a pu être optimisée pour détecter plus de molécules du mélange. L'utilisation d'approches de sélections de marqueurs d'intérêt et de méthodes d'identification avec des outils libres a également montré un intérêt par rapport aux outils fournisseurs seuls. Concernant la méthode utilisant entièrement les logiciels fournisseur, en exploitant le *loadings plot* résultant de la comparaison du mélange avec tous les autres échantillons à la place du *s-plot*, le nombre d'identification de composés du mélange de standards analytiques dans la limite des 50 marqueurs sélectionnés est considérablement amélioré. L'étape de calcul de formules moléculaires a également permis d'augmenter légèrement le nombre de composés identifiés, mais principalement en changeant la nature des adduits attendus. Cependant, cette information n'est normalement pas accessible pour des molécules inconnues. En ce qui concerne la méthode entièrement réalisée avec des logiciels libres, elle permet d'avoir un total contrôle des paramètres utilisés à chaque étape et de récupérer les données spectrales. Cependant, la meilleure combinaison de paramètres employée n'a permis d'identifier entièrement que 3 composés sur les 18 présents dans le mélange. Enfin, la méthode hybride a permis d'identifier le même nombre de composés que la méthode avec logiciels fournisseur uniquement. Cependant, elle présente l'avantage d'être plus rapide et plus flexible en termes de traitement de données avec comme seul désavantage de ne pas avoir accès de façon directe aux spectres de masses. Un résumé de ces résultats est fourni dans le Tableau 18.

Sur les 15 composés susceptibles d'être trouvés (d'après l'étude en SS, Tableau 10), un maximum de 7 composés a été identifié de façon correcte par les méthodes employées. Le Trichlorfon est le seul composé qui a été identifié systématiquement par chacune des méthodes employées. Les méthodes "fournisseur" et hybride ont permis d'identifier le même nombre de composés, avec pour différence une identification correcte de l'Ivermectin B1a pour la méthode fournisseur alors que la méthode hybride a permis l'identification du TCMTB.

De façon intéressante, l'erreur en masse était légèrement moins grande via le logiciel fournisseur (1,6 ppm contre 1,8 ppm respectivement). Cette différence peut s'expliquer par le fait que l'identification est réalisée sur le marqueur lorsque la méthode entièrement par logiciel fournisseur est utilisée, alors qu'elle est réalisée sur une *feature* quand la méthode hybride est employée. Dans le second cas, il s'agit d'un signal inconnu dans une injection donnée, pour laquelle les spectres MS et MS^E peuvent être affichés et récupérés individuellement. Dans le premier cas, le marqueur représente l'alignement de différentes *features* à travers toutes les injections de la séquence. Les m/z , TR, DT et spectres associés à ce marqueur résultent donc de la moyenne des propriétés des *features* qui ont été alignées. En ce qui concerne les marqueurs utilisés par le traitement de données réalisé entièrement sous R, les m/z mesurés peuvent être d'autant plus différents à cause de la conversion des données. En effet, l'étape de centroïdisation peut avoir une influence puisqu'elle transforme un signal gaussien continu en un pic discret. Par la suite, l'étape de détection des pics peut aussi être source de différence car elle est basée sur un algorithme différent de celui employé par UNIFI. Cependant, dans le cas présent, la méthode hybride et la méthode entièrement réalisée par R ont permis d'identifier deux molécules communes avec les mêmes erreurs en masse.

Tableau 18 : Comparaison des méthodes de traitement des données de l'EIL en NTS

	UNIFI + EZInfo	UNIFI + R	R
Avantages	Prétraitement réalisé en même temps que l'acquisition des données	Prétraitement réalisé en même temps que l'acquisition des données et traitement global de données rapide	Toutes les étapes de prétraitement peuvent être contrôlées par l'utilisateur
Inconvénients	Traitement des données long Beaucoup de choix gérés par les logiciels sans visibilité pour l'utilisateur	Pour l'identification, les spectres (MS et MSE) doivent être récupérés un à un manuellement	Etapes supplémentaires (conversion des données + prétraitement), longues et nécessitant une puissance de calcul non négligeable
Marqueurs générés	59 578	59 578	11 254
Marqueurs pris en compte	19 238	59 578	11 254
Marqueurs sélectionnés et méthode de sélection	100 Plus discriminants par les <i>loadings plot</i>	1 076 Les <i>spécifiques</i> du mélange de standards	930 Les <i>spécifiques</i> du mélange de standards
Composés correctement identifiés (erreur absolue en masse en ppm)	Chlormequat (2,5) Dichlorvos (0,2) Trichlorfon (0,0) Spinosad A (2,5) Rifaximin (2,1) Emamectin B1a (2,6) Ivermectin B1a (1,3)	Chlormequat (0,69) Dichlorvos (1,3) TCMTB (0,4) Trichlorfon (1,7) Spinosad A (3,8) Rifaximin (2,4) Emamectin B1a (2,6)	Methamidophos (2,5) Trichlorfon (1,7) TCMTB (0,4)

IV.3.1.3. Comparaison en SS

Sur la base de ces observations une comparaison des résultats en SS a également été effectuée. Pour cela, la méthode sous UNIFI a été comparée à la méthode optimisée avec les logiciels libres, utilisant *xcms3* pour la détection et le groupement des pics (Tableau 19). Dans cette partie en plus de l'échantillon de standards analytiques dans lequel les étalons ont été recherchés, les autres échantillons contenant six étalons d'extraction marqués isotopiquement (caféine-¹³C₃, carbamazépine-d10, cotinine-d3, diuron-d6, nicotine-d4 et simazine-d10) ont également été étudiés. Avec UNIFI, la recherche en SS a été effectuée de la

façon décrite dans la partie A – III.3.2. Dans *patRoan*, la recherche en suspect est effectuée sur la *marker table* (après le groupement des pics) en fournissant la liste des masses exactes associées à chaque composé à rechercher.

Tableau 19 : Comparaison des résultats d'identification par SS avec logiciel fournisseur (UNIFI) et logiciel libre (*patRoan*)

	Nombre de standards analytiques identifiés (sur un total de 18)	Nombre d'étalons deutérés identifiés (sur un total de 6)	Moyenne d'erreur absolue en masse
UNIFI	10	6	2 ppm
<i>patRoan</i>	11 + 3 retrouvés seulement dans 2 répliquats sur 3	5 + 1 identifié dans seulement 9 injections sur 15	3 ppm

Il est intéressant de remarquer qu'un total de 15 composés étaient potentiellement identifiable en NTS dans les données issues du logiciel libre, alors que seuls 2 ont été identifiés correctement (partie A – IV.1.2). Parmi ces derniers, 3 n'étaient pas présents dans tous les répliquats d'injection ce qui a causé leur élimination dans l'étape de filtration mise en place. L'origine de la perte d'information concernant les 11 marqueurs restants est détaillée dans la (partie A – IV.1.2). Concernant les étalons deutérés, 5 d'entre eux ont été identifiés et alignés correctement par la méthode sous logiciels libres. Seul 1 étalon deutéré (la nicotine-d4) n'a pas été détecté dans toutes les injections (manquant dans 2 répliquats du mélange de standards analytiques et de l'échantillon S1-4, et dans 1 répliquat des échantillons S2-2 et S2-4). En revanche, les étalons deutérés ont été correctement détectés dans toutes les injections lors de la recherche par UNIFI et les masses détectées étaient plus proches des valeurs théoriques par cette méthode, les masses précises perdant en précision lors de la conversion des données sous logiciels libres.

IV.3.2. Voies de développement

Afin d'approfondir les méthodes décrites précédemment, des outils complémentaires ont été développés au cours de cette thèse, sur la base de travaux précédemment décrits dans la littérature, pour pouvoir traiter d'autres types de données et fournir des informations supplémentaires. Les étapes précédemment décrites (filtration sur répliquats, soustraction de blanc, comparaison binaire, AMV) ont été regroupées dans une application *Shiny* permettant d'automatiser le traitement des *marker tables* issues de UNIFI. Dans cette application, les

différentes opérations effectuées sur les tableaux de données et les outils statistiques employés ont été optimisés en utilisant le paquet R *data.table*, permettant un gain substantiel de rapidité de traitement en comparaison du logiciel fournisseur EZInfo. Les outils suivants y ont ensuite été ajoutés.

Dans un premier temps, différents outils de visualisation ont été mis en place afin de pouvoir observer rapidement les différences entre plusieurs échantillons. Le premier consiste simplement à tracer les empreintes des différents échantillons en représentant les m/z des marqueurs en fonction de leur TR avec une aire de bulle proportionnelle à l'intensité du marqueur (nuages de points) (Figure 16).

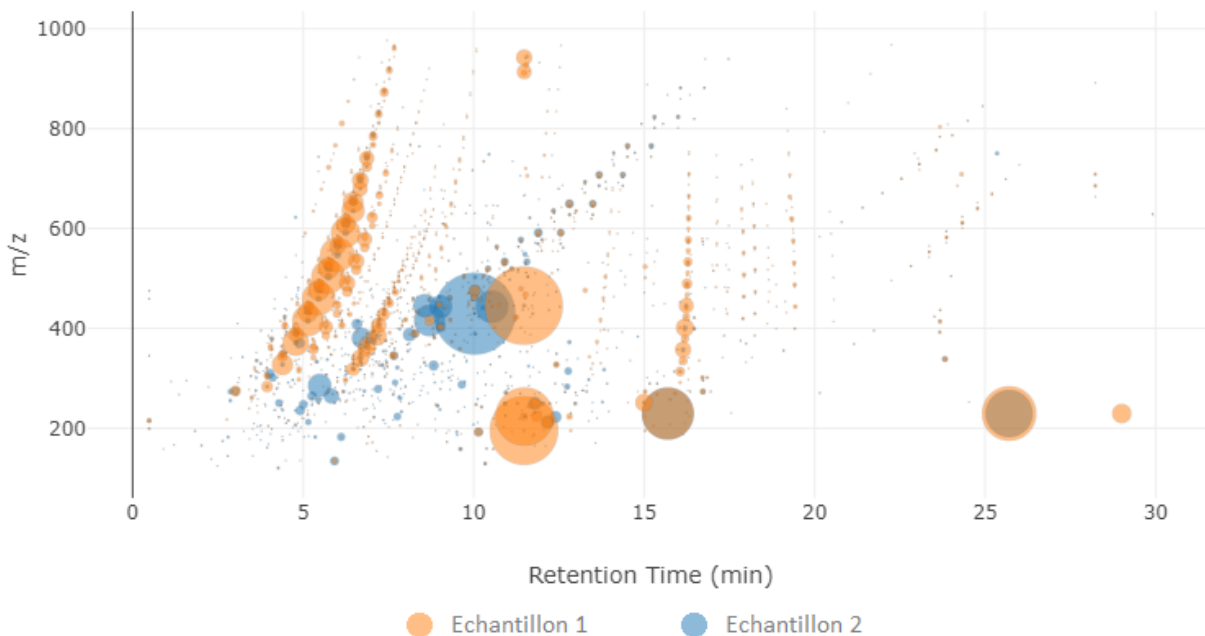


Figure 16 : Comparaison d'empreintes de deux échantillons

Différents échantillons peuvent ainsi être superposés et les différences majeures entre ces échantillons pourront être repérées rapidement. De façon plus approfondie, les différences entre échantillons ont pu être mises en valeur grâce à l'utilisation de diagrammes de Venn, de diagrammes d'Euler (Figure 17a) et de diagrammes *Upset*. Ces derniers ont plusieurs avantages par rapport à la représentation précédente. En effet, ils permettent de comparer des groupes d'échantillons et peuvent aussi mettre en avant des différences lorsqu'elles sont mineures, en visualisant rapidement les proportions de marqueurs communs et spécifiques à chaque échantillon ou groupes d'échantillons.

Cependant, ces méthodes ne permettent pas d'obtenir directement des informations sur la nature des marqueurs qui diffèrent, contrairement aux nuages de points qui sont tracés en fonction de leurs propriétés. Pour remédier à cela, une étape qui sera par la suite appelée caractérisation globale, a été ajoutée. Celle-ci consiste à calculer les moyennes en termes de m/z , TR, DT et la somme des aires des marqueurs constitutifs de différents groupes, afin d'estimer de manière globale la masse, la polarité et la taille moyenne de ces marqueurs. Ces groupes peuvent être définis comme étant des échantillons différents, ou alors résulter de la comparaison d'échantillons et ainsi représenter des groupes de marqueurs spécifiques ou communs aux échantillons. Ces moyennes permettent ainsi une comparaison rapide et globale des propriétés des marqueurs contenus dans chacun des groupes. En plus des moyennes, une distribution (Figure 17b) de ces différentes propriétés peut être tracée par échantillon afin de comptabiliser le nombre de marqueurs appartenant à une plage définie d'une propriété donnée.

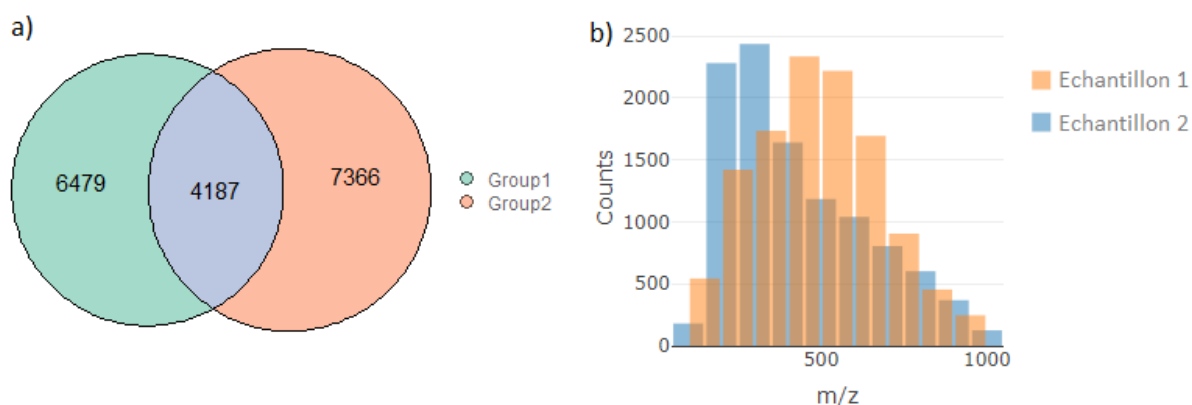


Figure 17 : Comparaison d'échantillon par a) Diagramme d'Euler et b) Distribution en m/z

Par ailleurs, une des faiblesses des identifications réalisées par les logiciels libres repose sur le fait que les données de mobilité ionique ne sont pas prises en compte, réduisant ainsi la précision des données spectrales utilisées pour le calcul des formules moléculaires. Des travaux en cours visent à récupérer ces données spectrales via l'interface de programmation de UNIFI pour les importer dans les données utilisées par les logiciels libres. L'accès à ces données permettra ainsi de récupérer les données spectrales tirant profit de la mobilité ionique tout en évitant l'étape de conversion et de prétraitement de données par logiciels libre. Une difficulté consiste à récupérer ces données et les lier aux *marker tables* exportées de UNIFI. En effet, l'interface de programmation permet d'avoir accès aux données brutes acquises pour chaque échantillon. Les spectres de masse sont donc liés à une *feature* pour un

échantillon donné, alors qu'un marqueur présent dans une *marker table* résulte de l'alignement de *features* similaires parmi tous les échantillons (pour obtenir une moyenne de *m/z*, RT et DT entre les échantillons). Pour attribuer un spectre de masse à un marqueur, il est alors nécessaire d'aligner les spectres des *features* sur les différentes injections puis de moyenniser ces spectres pour obtenir un spectre moyen correspondant au marqueur. Il est également nécessaire de choisir le meilleur format et la meilleure structure (de la même manière que sous *patRoom*) afin de regrouper ces données pour leur exploitation ultérieure. Pour cela, il est important de prendre en compte des critères de rapidité de lecture et de mise en forme, ainsi que de taille de fichier.

Enfin, afin de tirer un maximum de profit des données issues de la HRMS, deux types de représentation supplémentaires ont été choisies par rapport à leur pertinence et leur occurrence dans la littérature pour être implémentée dans l'application. Ces méthodes sont introduites ici et leur utilité sera présentée dans la suite du manuscrit. La première est le diagramme de Kendrick, développé initialement pour des données de FT-ICR-MS afin de caractériser les hydrocarbures (Kendrick, 1963). Cette méthode repose sur un changement d'échelle des masses détectées, afin de calculer des défauts de masses caractéristiques de familles de composés homologues, suivant l'équation 1.

$$\text{masse de Kendrick (M)} = \text{masse exacte(M)} \times \frac{\text{masse nominale (G)}}{\text{masse exacte (G)}} \quad \text{eq. 1}$$

Avec M : le marqueur d'intérêt

G : un groupement défini par l'opérateur. Ce groupement doit être caractéristique des séries homologues ou de mécanismes de transformation à mettre en avant

Cette représentation est intéressante puisqu'elle permet de repérer les marqueurs appartenant à la même famille homologue, c'est-à-dire, ayant le même squelette structural et différant seulement d'une ou plusieurs unités du groupement utilisé pour la conversion en masses de Kendrick (Figure 18). Ainsi, l'identification d'un seul marqueur de cette famille homologue permet d'identifier automatiquement les autres marqueurs appartenant à cette famille. Cette représentation peut également être utilisée pour mettre en avant des mécanismes de transformation simples afin de relier, par exemple, des précurseurs et leurs produits. En effet, pour des mécanismes mettant seulement en jeu des additions ou pertes de

groupements simples sans réarrangement, ces groupements pourront être utilisés pour calculer les masses de Kendrick et ainsi les précurseurs et produits de ces transformations seront alignés horizontalement de la même manière que les homologues précédemment décrits.

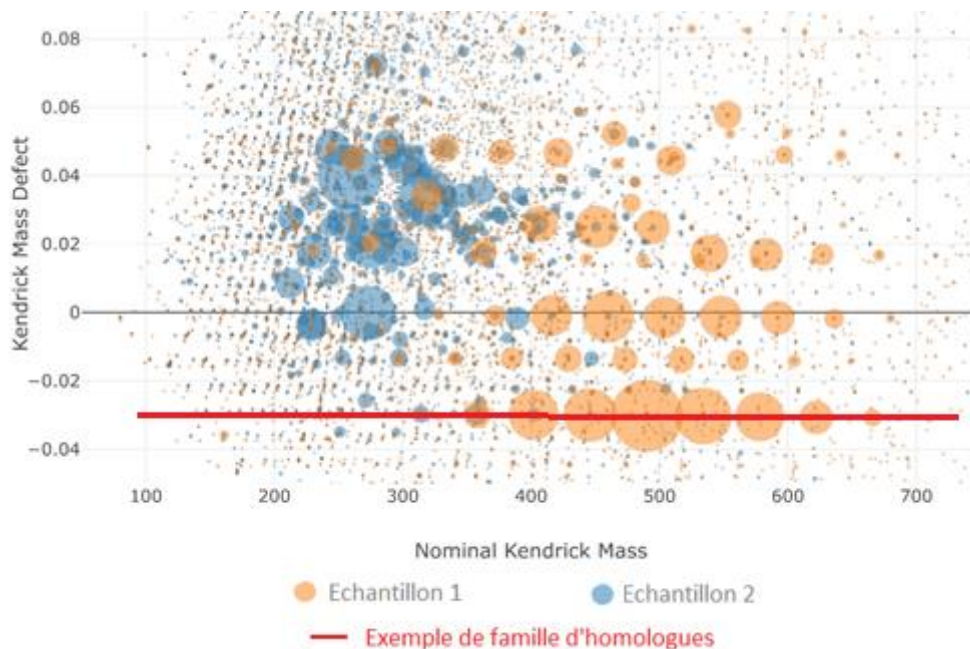


Figure 18 : Diagramme de Kendrick (base C_2H_4O). Les points alignés sur la ligne rouge appartiennent à la même famille d'homologues.

La seconde représentation consiste à utiliser les formules moléculaires calculées à partir des masses exactes afin de tracer des diagrammes de van Krevelen. Ces derniers permettent de représenter des ratios atomiques, les uns en fonction des autres et ainsi d'avoir une idée du type de composés présents dans chaque échantillon (Figure 19).

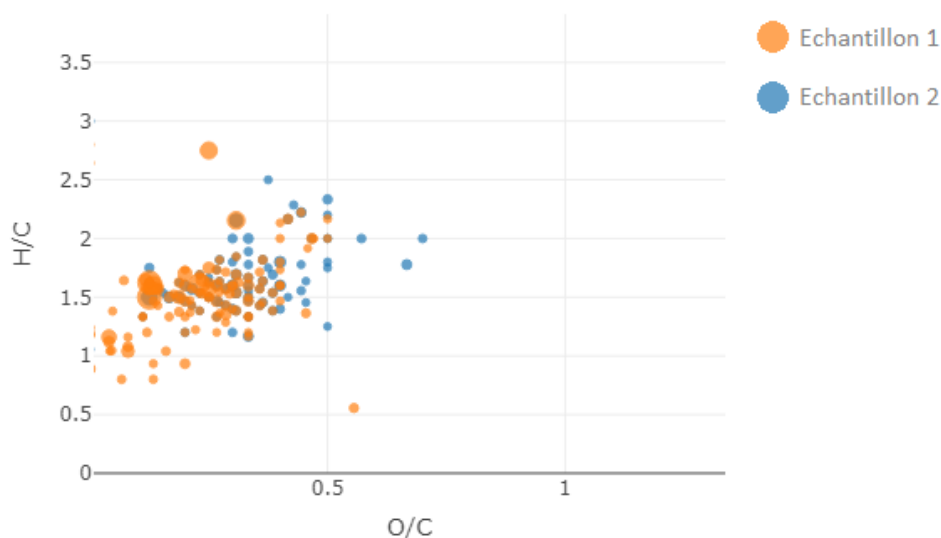


Figure 19 : Diagramme de van Krevelen

L'application développée permet à l'heure actuelle de traiter directement les *marker tables* exportées de UNIFI ou les données issues d'un prétraitement réalisé avec *patRoan*. Celles-ci peuvent également être réexportées vers un format compatible à un traitement avec *workflow4metabolomics*, une plateforme en ligne offrant une interface pour le traitement de données de spectrométrie de masse par logiciels libres (*xcms*, *Camera...*). Concernant les différentes étapes de traitement, les graphiques générés par l'application peuvent être exportés au format png ou svg, et les tables de données au format csv.

IV.4. Conclusion sur le développement de méthodes de traitement de données

Cette partie a permis le développement de différentes voies de traitement de données qui s'appuient sur plusieurs types de logiciels et approches. En particulier, l'étude a mis en évidence des avantages et inconvénients liés à chaque méthode étudiée. Les résultats ont montré qu'il n'est, à ce stade, pas possible de conclure à une méthode préférentielle de traitement de données. Le choix de la méthode de traitement dépend grandement de l'objectif fixé par l'étude.

En effet, pour réaliser des comparaisons rapides, la méthode hybride, qui s'appuie à la fois sur un prétraitement simplifié par le logiciel fournisseur et sur la liberté de traitement fournie par les logiciels libres, est plus adaptée. En revanche, cette dernière ne permet pas l'identification de marqueurs de manière directe dans son état actuel. Si le travail requiert des identifications, il est préférable d'utiliser la méthode fournisseur pour un nombre limité de marqueurs à identifier de manière poussée puisque cette méthode permet de tirer parti d'un maximum d'informations produit par l'analyse. En revanche, le logiciel ne permet pas d'effectuer un grand nombre d'attributions simultanées et les méthodes de traitement proposées ne sont parfois pas adaptées aux études réalisées. Dans ce cas, la méthode passant entièrement par logiciels libres est plus avantageuse puisqu'elle permet de contourner ces problèmes. En revanche, cette manière de procéder est plus longue puisqu'elle nécessite une conversion préalable des données et un prétraitement qui, dans les deux autres cas, est effectué en parallèle de l'acquisition. De plus, les logiciels libres ne supportent pas encore les données de

mobilité ionique et cette particularité, qui fournit une plus grande fiabilité pour l'identification, n'est donc pas mise à profit.

Afin d'aboutir à une unique méthode de traitement de données, un travail supplémentaire est en cours pour récupérer directement les données spectrales de UNIFI, en gardant donc la meilleure précision apportée par la mobilité ionique. L'objectif est de lier ces données spectrales aux *marker tables* exportées de UNIFI afin d'utiliser les méthodes d'identification disponibles dans les logiciels libres. Ce *workflow* complet alliera rapidité et fiabilité en combinant les avantages des différentes méthodes :

- Par logiciel "fournisseur" : prétraitement des données (détection des pics fiable grâce à la mobilité ionique), en simultané avec les acquisitions et fournissant des spectres précis avec l'IMS
- Hybride : grande liberté en termes de traitement de données et temps de calculs fortement réduits
- Par logiciels libres : possibilité de réaliser des attributions de formules moléculaires et des identifications sur un large nombre de marqueurs

Partie B : Comparaison de méthodes analytiques

I. Introduction et objectifs

La partie précédente s'est attachée à développer des méthodes de traitement de données permettant de conserver un maximum d'informations issues de l'analyse non-ciblée. Cette partie s'intéresse aux différentes méthodes analytiques utilisées en NTS permettant l'acquisition des données. Le NTS ayant pour objectif de fournir un maximum d'informations sans *a priori* sur les échantillons analysés, le caractère d'exhaustivité des méthodes analytiques est primordial afin de fournir un set de données représentatif de l'échantillon. Dans un premier temps, une revue bibliographique avait pour objectifs :

- De décrire les différentes méthodes développées et présentées dans la littérature pour le NTS d'échantillons aqueux environnementaux
- D'identifier s'il existe une méthode prédominante et, si c'est le cas, d'en déterminer la raison
- De mettre en avant, le cas échéant, un manque de connaissances concernant l'exhaustivité de ces méthodes et les moyens de la caractériser

Il apparaît que les méthodes d'analyse actuelles utilisées pour le NTS appliqué à des échantillons aqueux environnementaux sont diverses, mais qu'il y a cependant une forte prédominance d'analyses réalisées par LC-HRMS et employant une colonne chromatographique de type C18, après extraction sur phase solide à l'aide d'une cartouche *universelle* (généralement de type HLB) ou une cartouche multiphasique (HLB, ENV+, X-AW et X-CW) permettant d'élargir la gamme de composés retenus.

Les études décrivant la comparaison de différents protocoles analytiques en NTS sont cependant rares, et s'appuient généralement sur une approche du type SS. Une caractérisation basée sur l'ensemble des signaux détectés en NTS paraît nécessaire afin d'apporter plus de connaissances quant aux particularités de chaque méthode. Dans ce contexte, l'étude présentée dans cette partie avait pour but de diviser un même échantillon en diverses fractions ayant des propriétés physico-chimiques différentes, puis de les analyser avec des techniques différentes afin de regrouper un maximum d'informations permettant de caractériser les supports disponibles sur le marché. Pour cela, l'étude a été menée en utilisant à la fois le SS et le NTS. Afin d'exploiter ces données et apporter des informations quant aux particularités de chaque méthode, différents critères de comparaison ont été mis en place, en

particulier, en utilisant la mobilité ionique. Une partie de ces résultats fait l'objet d'un article en préparation :

Nina Huynh, Julien Le Roux, Emmanuelle Mebold, Régis Moilleron. Comparison of analytical strategies using the advantage of ion mobility for non-target screening.

II. Etat de l'art

II.1. Préparation d'échantillon

L'étape de préparation, ou prétraitement, de l'échantillon est une étape essentielle pour l'analyse. Elle consiste en une succession d'étapes ayant pour but de rendre l'échantillon analysable par la technique retenue ; elle peut avoir plusieurs objectifs : le transfert des composés d'intérêt dans une phase adaptée à la méthode d'analyse choisie, l'élimination d'interférents lors de l'analyse ou la concentration de composés présents afin d'assurer leur détection ultérieure.

Différentes techniques de préparation existent pour la préparation d'échantillons liquides comme l'extraction liquide/liquide, la microextraction en phase solide ou l'extraction sur phase solide (SPE). L'efficacité de ces méthodes d'extraction pour des polluants donnés a été comparée. Par exemple, l'extraction de la sulfonamide dans des eaux résiduaires urbaines par SPE (avec une cartouche MCX, Waters) et par microextraction en phase solide a été comparée (Balakrishnan et al., 2006). Les auteurs ont ainsi réalisé des expériences de dopages avant extraction, montrant que la seconde méthode était une méthode plus efficace, car permettant d'obtenir des valeurs expérimentales plus proches des valeurs théoriques de dopages, mais que la SPE était quant à elle une technique beaucoup plus sélective (permettant d'atteindre des concentrations plus faibles). Une autre étude comparant la SPE et l'extraction liquide-liquide pour l'extraction des 9 hydrocarbures aromatiques polycycliques réglementés a montré que la seconde était plus efficace pour les hydrocarbures de grande masse moléculaire, alors que la SPE était adaptée à toute la gamme de masse de ces hydrocarbures (Titato and Lanças, 2005). Avec le développement continu de nouvelles phases solides ayant des affinités plus ou moins grandes avec certains types de composés, une large gamme de propriétés physico-chimiques est désormais couverte par cette méthode.

En NTS, les étapes de préparation peuvent avoir une grande influence sur les résultats obtenus après analyse de l'échantillon. En effet, le NTS n'ayant pas de molécules d'intérêt prédéfinies, les étapes de préparation constituent un risque de perdre des analytes dans le processus. Dans la littérature, plusieurs techniques de préparation ont été utilisées pour des échantillons environnementaux destinés à une analyse en NTS. Nous nous attacherons ici à décrire les méthodes les plus couramment utilisées pour des échantillons aqueux.

II.1.1. Injection directe

En SS et en NTS, le but étant souvent d'analyser un maximum de composés pour lesquels peu ou pas d'informations sont disponibles quant à leur comportement, il est difficile de choisir une technique particulière de préparation d'échantillon. En effet, cette étape est parfois très spécifique aux composés à analyser et à leurs propriétés physico-chimiques. Elles sont sujettes à de nombreuses étapes d'optimisation (comme par exemple pour l'ajustement du pH, des volumes, des types de solvant...) afin d'optimiser l'analyse de ces composés en particulier. Ainsi, pour le NTS d'échantillons aqueux, certains auteurs préfèrent occulter cette étape et réaliser simplement une filtration pour enlever les matières en suspension avant d'injecter l'échantillon sur l'instrument analytique (Bader et al., 2017; Nürenberg et al., 2015; Vergeynst et al., 2014). De cette manière, l'échantillon est préservé de toute modification qui pourrait entraîner une perte d'information. Cette méthode permet également d'analyser un plus grand nombre d'échantillon puisque l'absence de pré-traitement est un gain de temps considérable, sans compter que seul un faible volume d'échantillon est nécessaire à l'analyse.

Cependant, cette technique présente également un désavantage. En effet, les résultats dépendent grandement de la sensibilité des appareils analytiques qui sont utilisés par la suite. Le plus grand risque de cette méthode réside dans le fait que seules les molécules présentes initialement en concentration suffisante soient visibles. En fonction des objectifs de l'étude ou des étapes de traitement de données suivants, cet inconvénient peut toutefois être minime. En effet, quand l'objectif est de réaliser une comparaison d'échantillons, prendre en compte les pics les plus intenses peut suffire, puisque dans certains *workflows* de traitement de données, la priorisation des pics d'intérêt est effectuée sur la base de l'intensité. Cependant, lorsqu'il s'agit d'étudier des TP qui sont potentiellement formés en de très faibles quantités, cette méthode par injection directe peut conduire à la perte de certaines informations importantes.

Pour tenter de remédier à cela en limitant au maximum la perte de certains composés, certains auteurs ont récemment choisi de concentrer l'échantillon par évaporation de l'eau, pour le récupérer dans un volume plus faible (Karki et al., 2020; Mechelke et al., 2019). Mechelke et al. (2019) ont ainsi observé des performances comparables à une SPE classique sur une liste de 590 MP organiques d'intérêt en comparant les rendements d'extraction absolus. Cependant, ils remarquent un nombre de marqueurs en NTS plus faible par cette technique par rapport à la SPE, ce qui pourrait à la fois être attribué à une contamination par le support de SPE (augmentant donc le nombre de pics détectés) ou, à l'opposé, à une perte de certains composés volatils due au chauffage à 55°C pour la méthode d'évaporation (Mechelke et al., 2019). De manière similaire, une autre étude évoque l'utilisation de la lyophilisation pour des échantillons aqueux avec reconstitution dans un solvant organique (Montes et al., 2017) pour limiter au maximum l'altération de l'échantillon. Cependant, cette évaporation totale de la phase aqueuse peut entraîner une précipitation irréversible de certains composés qui seront donc perdus.

II.1.2. Extraction sur phase solide

L'une des méthodes les plus répandues pour la préparation d'échantillons aqueux en vue de l'analyse de MP organiques est l'extraction sur phase solide (SPE). Le NTS, malgré ses spécificités, ne déroge pas à la règle (Hajeb et al., 2022). Bien que cette méthode comporte des risques de modifications de l'échantillon puisqu'il est manipulé, elle permet également de concentrer les molécules initialement présentes dans l'échantillon et rend ainsi leur analyse possible. La diversité de phases commerciales disponibles sur le marché permet de retenir désormais une large gamme de micropolluants. Certaines sont plus adaptées à des types de molécules pour lesquelles elles ont été développées. Pour autant, elles peuvent également retenir d'autres types de composés. Parmi les différentes phases existantes, plusieurs grandes familles se distinguent : les phases dérivées de la silice, les phases échangeuses d'ions, les phases en carbone graphitique et les phases polymériques fonctionnalisées (Masqué et al., 1998). Le Tableau 20 détaille le type d'interactions mises en jeu entre la phase stationnaire et les analytes, il donne un aperçu du type de phases qu'il est possible de retrouver dans chaque famille. Un exemple d'étude sur les eaux résiduaires urbaines utilisant ce type de phase pour cibler des molécules ou familles de molécules en particulier complète les informations fournies.

Tableau 20 : Présentation de différents groupes de phases stationnaire disponibles pour la SPE

Type d'interactions	Type d'interactions	Exemple de phase commerciale	Exemple d'utilisation en TS dans des eaux usées (références)
Dérivés de silice	Van der Waals	ODS C18 (Chrome Expert)	8 perturbateurs endocriniens (3 alkylphénols et 5 hormones) (Zhang et al., 2012)
Echangeuses d'ions	Electrostatiques	Oasis WCX (Waters)	11 fluoroquinolones (He and Blaney, 2015)
Carbone graphitique	Van der Waals et électrostatiques	Envi-Carb (Supelclean)	4 œstrogènes (Laganà et al., 2000)
Polymères fonctionnalisés	$\pi - \pi$	Oasis HLB (Waters)	19 antibiotiques (tétracyclines, macrolides, sulfonamides, fluoroquinolones et lincosamides) (Tylová et al., 2013)

Pour limiter la perte d'information due à la non-rétention de certaines molécules sur les phases choisies, une méthode combinant plusieurs phases commerciales a été développée afin de retenir la plus large gamme possible de composés (Kern et al., 2009). Elle s'appuie sur l'utilisation des phases suivantes :

- Une phase polymérique formée par un copolymère de divinylbenzène-vinyl pyrrolidinone (Oasis HLB, Waters), généralement considérée comme une phase universelle, permettant de retenir à la fois les composés acides, neutres et basiques
- Une phase polymérique formée d'un copolymère de polystyrène-divinylbenzène réticulés avec des groupements hydroxyles (Isolute ENV+, Biotage), dédiée à l'extraction de composés polaires en phase aqueuse
- Une phase échangeuse d'anions faibles, formée d'un polymère fonctionnalisé (Strata X-AW, Phenomenex), adaptée pour la rétention de composés acides avec un $pK_a < 5$
- Une phase échangeuse de cations faibles, formée d'un polymère fonctionnalisé (Strata X-CW, Phenomenex), adaptée pour la rétention de composés acides avec un $pK_a > 8$

L'efficacité de la combinaison de ces phases pour la rétention d'une large gamme de composés a été démontrée par l'extraction de 88 composés polaires dans des eaux souterraines, eaux de surface et eaux résiduelles urbaines. Cette liste regroupe à la fois des pesticides, des composés pharmaceutiques, des composés d'origine industrielle et leurs TP respectifs, sur

une gamme de masse comprise entre 119,0 et 747,5 g/mol, avec des pKa variant entre -4,2 et 16,0, un log Kow entre -4,17 et 5,40 et un log D entre -4,17 et 4,22 (Huntscha et al., 2012). Ayant montré une capacité à rendre quantifiable précisément 80 % des molécules de cette liste à des concentrations inférieures à 10 ng/L, cette cartouche multiphasique est désormais une méthode courante en SS et NTS par rapport à d'autres types de supports (Tableau 21). Il est cependant important de préciser qu'à ce jour, cette cartouche multiphasique n'est pas disponible commercialement, notamment à cause de la phase ENV+, peu couramment retrouvée chez les fournisseurs. Ainsi, son utilisation nécessite un assemblage manuel de la cartouche à partir des phases individuelles ce qui peut entraîner une non-reproductibilité et être source de contamination.

Tableau 21 : Exemples de méthodes de préparation d'échantillon par SPE pour l'analyse non-ciblée de matrices aqueuses. N.A. = pas spécifié

Matrice	Phase	Conditionnement*	pH	Séchage / Lavage	Élution*	Référence
Eau de surface et STEU (brute et rejet)	Oasis HLB	N.A.				(Ibáñez et al., 2008)
Eau de surface, eau souterraine et eau potable	Oasis HLB	N.A.	pH 2	Séchage N ₂ EUP acidifié à pH 2	3 × 2,5 mL ACN	(Hogenboom et al., 2009)
	Oasis MCX				2,5 mL ACN et 2 × 2,5 mL EUP + 5% NH ₄ OH	
Eau de surface STEU (brute et rejet)	Mélange de phases (HLB, ENV+, X-AW, X-CW)	MeOH et EUP	pH 6-7	N.A.	4 mL ou 6 mL MeOH/AcEt (50/50 v/v) + 2% NH ₄ OH puis 2 mL ou 3 mL MeOH/AcEt (50/50 v/v) + 1,7% AF	(Gago-Ferrero et al., 2015; Kern et al., 2009; Moschet et al., 2013; Ruff et al., 2015; Schollée et al., 2018; Schymanski et al., 2014b)
Lixiviat de décharge et eau souterraine	Isolute ENV+	6 mL n-hexane, 6 mL acetone, 6 mL MeOH, 6 mL EUP (pH 7 ou 3)	pH 7 et pH 3	Séchage N ₂	6 x 1 mL MeOH	(Müller et al., 2011)
Rejet STEU	Oasis MCX et Strata X en série	4 mL MeOH puis 4 mL of EUP + acide chlorhydrique (pH 2)	N.A.	4 mL MeOH/EUP (5/95 v/v)	2 × 4 mL of MeOH + 5% NH ₄ OH	(Nurmi et al., 2012)
STEU (brute et rejet)	Oasis HLB	N.A.	N.A.	N.A.	5 mL of MeOH	(Díaz et al., 2012)
STEU (brute et rejet)	Oasis HLB	6 mL MeOH et 6 mL EUP	N.A.	Séchage N ₂	MeOH	(Bade et al., 2015)
Eau de surface	SPE large volume mélange de phases	N.A.	N.A.	N.A.	500 mL AcEt, 500 mL MeOH (HR-X), 500 mL MeOH + 2 % NH ₄ OH (HR-	(Schymanski et al., 2015)

PARTIE B : Comparaison de méthodes analytiques

Matrice	Phase	Conditionnement*	pH	Séchage / Lavage	Élution*	Référence
	(HR-X, HR-XAW et HR-XCW)				XAW) ou 500 mL MeOH + 1 % AF (HR-XCW)	
Eau de surface	Oasis HLB, Bon-Elut PPL, Oasis MCX et Oasis MAX en série	5 mL AcEt/MeOH (50/50 v/v) et 5 mL EUP	pH 6,3 -6,7	N.A.	3 x 3 mL MeOH/AcEt (50/50 v/v) puis 3 mL MeOH/AcEt et 2 x 3 mL MeOH + 2 % AF (MAX) ou 2 x 3 mL MeOH + 5% NH ₄ OH (MCX)	(Zonja et al., 2015)
Sortie de STEU et eaux de surface	Oasis MAX et Oasis MCX en série	2 x 3 mL MeOH 2 x 3 mL EUP	N.A.	Séchage N ₂ 2 mL EUP + 5 % NH ₄ OH et 2 mL EUP + 2 % AF	6 mL MeOH/AcEt (70:30 v/v) + 2% AF 6 mL MeOH/AcEt (70/30 v/v) + 5 % NH ₄ OH	(Deeb et al., 2017)
Rejet STEU	Cunax2	8 mL MeOH, 4 mL EUP et 4 mL tampon phosphate (0,1 M à pH 6)	N.A.	5 mL tampon phosphate Séchage N ₂	10 mL CAN 10 mL ACN + 5 % NH ₄ OH	(Bergé et al., 2018)
	Chromabond HR-X	5 mL MeOH et 5 mL EUP		5 mL EUP Séchage N ₂	5 mL MeOH et 5 mL AcEt	
Infiltration eau souterraine	Oasis HLB	5 mL MeOH et 5 mL EUP	N.A.	2 mL EUP Séchage N ₂	4 x 2,5 mL MeOH	(Albergamo et al., 2019)

*ACN = acétonitrile, EUP = Eau ultra-pure, MeOH = méthanol, AcEt = acétate d'éthyle, AF = acide formique

II.1.3. Fractionnement d'échantillon

Outre la SPE, des méthodes de fractionnement peuvent aussi être utilisées en amont de l'analyse par NTS afin de réduire séquentiellement la complexité de l'échantillon. Pour réaliser ce fractionnement, la technique la plus fréquemment retrouvée est la chromatographie de préparation. Celle-ci consiste à réaliser une séparation chromatographique sur une colonne donnée et ainsi récupérer des fractions différant en termes d'affinité avec cette colonne. Pour l'étude des influents ou effluents de STEU, la colonne chromatographique la plus souvent utilisée pour le fractionnement est une colonne de type C18 (Grung et al., 2007; Hashmi et al., 2018; Purschke et al., 2020; Sonavane et al., 2018; Zwart et al., 2020). Une autre étude a préféré l'utilisation d'une colonne C8 (Reemtsma et al., 1999) tandis qu'une dernière a utilisé une séparation en série sur C18 puis sur une colonne amino-propyl (Mijangos et al., 2020). Cependant, un fractionnement par chromatographie induit une dilution de l'échantillon puisque seul un petit volume de l'échantillon est injecté sur la colonne, mélangé ensuite avec une grande proportion de phase mobile.

Cet effet de dilution est problématique pour les analyses chimiques puisque la dilution peut conduire une proportion des composés plus ou moins importante à présenter des concentrations sous leurs limites de détection respectives empêchant ainsi leur analyse. D'autre part, une grande partie des études faisant appel à cette méthode de préparation s'intéresse également à la toxicité des différentes fractions, notamment dans le cadre d'analyses dirigées par les effets. Cet effet de dilution est alors problématique pour les bioessais puisque les concentrations testées ne sont plus représentatives des concentrations réelles. Pour pallier cet effet, une étape de préconcentration, généralement réalisée par SPE est parfois nécessaire afin de concentrer l'échantillon avant de le fractionner (Sonavane et al., 2018; Zwart et al., 2020). Certaines études vont même jusqu'à reconcentrer les fractions, que ce soit par une SPE supplémentaire (Hashmi et al., 2018) ou par évaporation sous flux d'azote (Grung et al., 2007; Reemtsma et al., 1999). Cependant, peu d'études précisent le facteur de concentration final de leurs échantillons, le nombre de fractions réalisées ou les gammes de temps de rétention utilisées pour le fractionnement.

Outre le fractionnement par chromatographie d'affinité, la chromatographie d'exclusion stérique est aussi parfois utilisée afin de diviser l'échantillon en fractions de différentes gammes de masses moléculaires (Klinkow et al., 1998). De la même manière, des résines

macroporeuses et ioniques en série sont très souvent utilisées pour réaliser le fractionnement de la matière organique dissoute (MOD). Cette méthode est souvent retrouvée pour l'étude de la toxicité des échantillons car elle permet d'obtenir relativement facilement de grands facteurs de concentration (*i.e.*, > 1 000), ce qui représente un avantage pour un certain nombre de tests de toxicité qui ne sont pas assez sensibles ou qui nécessitent la reprise ou la dilution dans un autre solvant (Weller, 2012). Ce type de méthode de préparation est utilisé par différentes études de caractérisation de la MOD (Pan et al., 2016) ou bien pour l'identification de sous-produits de désinfection et de leur toxicité (Dong et al., 2017). De nombreux auteurs ont mis en évidence les possibles interactions entre MP organiques et MOD (Caupos et al., 2015; Ilani et al., 2005; Thevenot et al., 2009), ce qui montre l'intérêt de ce genre de technique de préparation dans l'étude des MP par NTS.

II.1.4. Études comparatives des méthodes de préparation

Pour des études en TS, le développement de méthode inclut très souvent la comparaison de différentes méthodes de préparation d'échantillon afin de sélectionner celle qui permet d'obtenir le résultat le plus fiable avec le coût le plus faible. Lorsque des méthodes d'extraction sont mises en place, le rendement d'extraction est un critère essentiel dans le choix final. En NTS, différentes méthodes de préparation sont également retrouvées et, étant donné la diversité des matériaux employés, il est difficile de comparer les résultats de plusieurs études de façon simple. Le caractère non-ciblé de l'analyse rend également impossible l'évaluation d'un rendement d'extraction sur la gamme large des molécules potentiellement détectées.

Une étude permettant de comparer une extraction liquide-liquide avec deux phases de SPE (HLB et C8/ENV+) pour le NTS d'eau de process issue d'une plateforme pétrolière a été réalisée (Samanipour et al., 2018). Les auteurs ont ainsi montré peu de différences entre les méthodes sur des molécules ciblées. En NTS, l'efficacité a été estimée par des calculs de rendements d'extraction sur 26 marqueurs inconnus déterminés comme étant statistiquement pertinents (F-ratio) pour la différenciation de ces trois méthodes. Les résultats montrent, cette fois, une différence d'efficacité avec la cartouche C8/ENV+ qui permet d'obtenir un rendement d'extraction relatif de 100 % pour ces 26 marqueurs alors que la cartouche HLB et l'extraction liquide-liquide permettent d'atteindre 60 % et 75 % respectivement de rendements relatifs seulement pour 3 marqueurs. Une seconde étude a comparé une SPE (phase HLB) à une

extraction par échantillonneur passif sur une durée d'une semaine (POCIS avec une phase HLB) pour des effluents de STEU (eaux brutes et rejet) et des eaux de surface (Castro et al., 2021). L'évaluation de l'efficacité de ces deux techniques a été réalisée par SS avec une liste de 3 227 MP. Sur ces derniers, 135 composés ont pu être identifiés, dont 85 % par l'échantillonnage passif contre 77 % par SPE. Les signaux obtenus par la première méthode étaient également plus intenses mais aucune conclusion n'a pu être tirée concernant les propriétés physico-chimiques des composés spécifiquement retenus par chacune des méthodes.

II.2. Analyse d'échantillon

Par rapport aux appareils de MS en basse résolution, les analyseurs en HRMS (Q-TOF, Orbitrap et FT-ICR-MS) permettent d'améliorer la résolution et la précision des masses détectées. Une étude en TS comparant la MS par trappe d'ions avec un appareil Orbitrap a montré que la HRMS permet d'abaisser les limites de détection, de réduire le bruit ainsi que les interférences et de détecter un plus grand nombre de molécules (Evans et al., 2014). Une comparaison du nombre d'articles répertoriés dans *Web of Sciences* sur les dix dernières années a été réalisée pour les trois analyseurs de HRMS cités précédemment (désignés par leur abréviation ou leur nom complet dans les titres et résumés d'articles), combinés dans un premier temps aux termes « HRMS » ou « *high-resolution mass spectrometry* » (Figure 20a), et dans un second temps à « *non-target screening* » ou « *non-target analysis* » (Figure 20b).

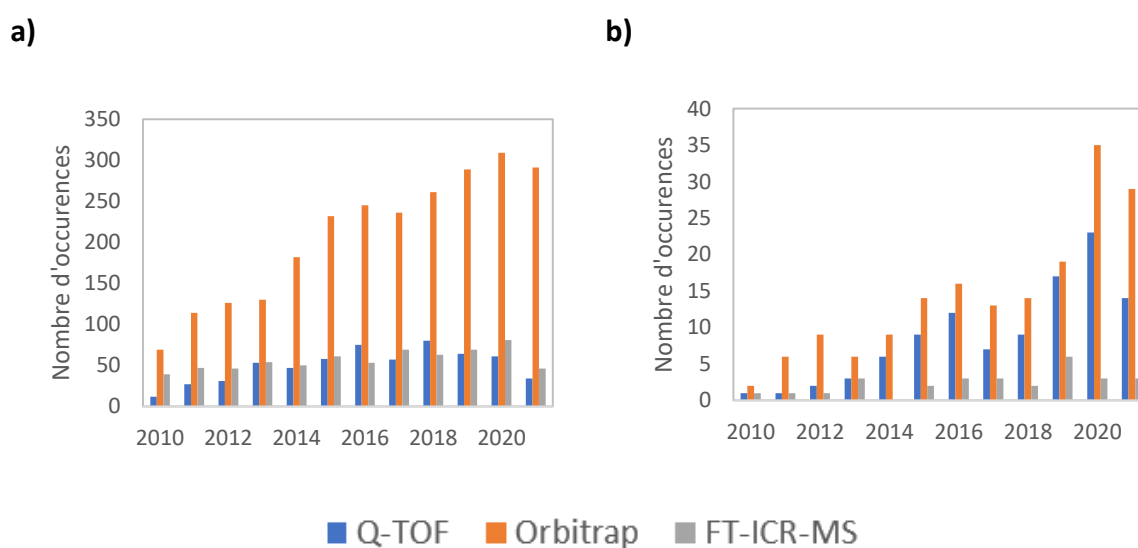


Figure 20 : Occurrence des différents analyseurs sur *Web of Sciences* a) associés à HRMS puis à b) NTS (consulté le 12/10/2021)

Cette comparaison montre dans un premier temps, l'intérêt grandissant pour ce type de technologies et ces méthodes d'analyse au fil des années. L'analyseur le plus souvent retrouvé est l'Orbitrap, tendance qui s'explique par la meilleure résolution de ce détecteur par rapport aux instruments de type Q-TOF et son prix plus abordable que celui des instruments de FT-ICR-MS. Des couplages avec une chromatographie liquide (LC) ou gazeuse (GC) en amont des détecteurs de Q-TOF ou Orbitrap sont la plupart du temps retrouvés dans la littérature, selon les propriétés des composés étudiés (Figure 21).

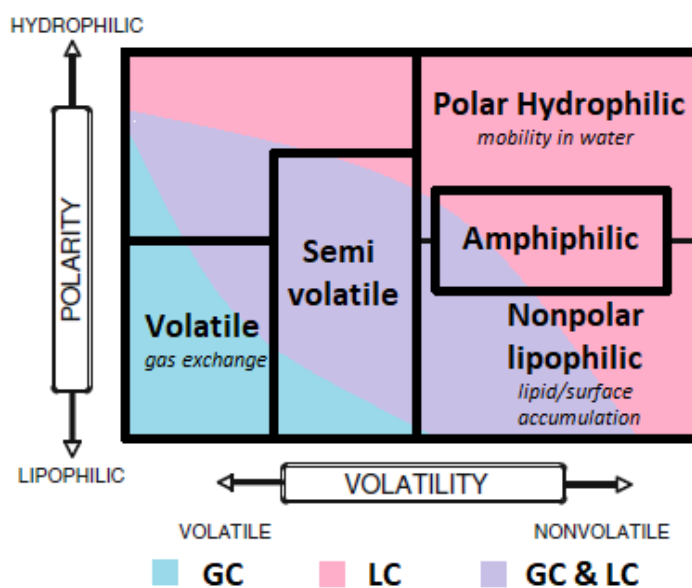


Figure 21 : Diagramme de polarité-volatilité pour le choix de la méthode de chromatographie, adapté de (Giger, 2008)

Le premier EIL mené par le réseau NORMAN a mis en évidence de nombreuses différences entre les résultats obtenus à partir d'un même extrait avec différents instruments (Schymanski et al., 2015). En effet, de plus en plus de technologies de HRMS sont développées par des fournisseurs différents, chaque appareil possédant des spécificités et des paramètres associés qui ne sont donc pas retrouvés sur d'autres instruments. Lors de cette étude où 17 laboratoires ont participé, 5 fournisseurs d'appareil HRMS étaient dénombrés, 12 modèles utilisés, et en prenant en compte le couplage avec la LC et la colonne utilisée pour la séparation chromatographique, les 17 laboratoires ont tous utilisé des méthodes analytiques différentes.

II.2.1. HRMS couplée à la chromatographie liquide

II.2.1.1. Séparation

Concernant la séparation chromatographique en phase liquide, elle est très majoritairement réalisée en phase inverse sur une colonne de type C18 (Schymanski et al., 2015) (Tableau 22). Seuls de rares auteurs ont essayé de travailler sur d'autres types de colonnes, notamment une colonne de type HILIC, plus souvent utilisée dans les études de protéomique (Boersema et al., 2008) ou de métabolomique (Cubbon et al., 2010).

Certaines études ont utilisé des types de colonnes différents pour les mêmes échantillons, mais aucune d'entre elles n'a comparé de façon fine les différences observées entre les colonnes en NTS. Par exemple, une utilisation en parallèle d'une colonne C18 et d'une colonne HILIC a montré la complémentarité de ces deux colonnes, en s'appuyant sur l'amélioration du nombre de molécules détectées en mode SS (Gago-Ferrero et al., 2015). Une seconde étude a été réalisée avec des séparations successives sur une colonne C18 suivie d'une colonne phényle-hexyle, HILIC ou penta-fluoryle-phényle, toutes trois dédiées à la séparation de composés polaires (Ouyang et al., 2015). La combinaison C18/penta-fluoryle-phényle a montré une plus grande efficacité sur l'effluent de STEU étudié, en prenant en compte le critère de séparation (*i.e.*, répartition des temps de rétention sur la deuxième colonne) qui fournit des informations supplémentaires permettant d'éliminer des attributions improbables ou de confirmer une identification. Les auteurs ont cependant montré que la combinaison C18/HILIC permettait d'obtenir une plus forte orthogonalité (*i.e.*, surface effective de séparation, calculée à partir d'un mélange standard contenant un mélange de HAP, dérivés de benzène et cétones cycliques) (Ouyang et al., 2015). D'autres auteurs ont également comparé la séparation sur différents types de colonnes (plusieurs C18, une phényl'éther et une Hypercarb) pour la mise en place d'une méthode de NTS en vue de l'évaluation de traitements réalisés en STEU (Nürnberg et al., 2015). Les critères retenus pour la comparaison des colonnes étaient la forme des pics, le nombre de plateaux théoriques pour évaluer la capacité de résolution de la colonne, le rapport signal sur bruit et le nombre de molécules ciblées détectées (80 molécules comprenant des composés pharmaceutiques, biocides, retardateurs de flammes, agents de contraste...). De façon globale, les auteurs ont montré peu de différences entre les différentes colonnes C18 et la phényl'éther. Cependant, ils ont noté que la colonne Hypercarb était peu efficace pour les composés apolaires (Nürnberg et al., 2015).

Parmi les types de LC plus rares, des travaux ont été effectués en utilisant une colonne trifonctionnelle, regroupant une séparation classique en phase inverse avec un échange cationique et anionique (Montes et al., 2017). L'évaluation de cette technique a été effectuée en SS sur un set de 3 000 MP organiques montrant que 62 % des molécules détectées sont mieux séparées par cette colonne par rapport à une phase C18 classique, notamment les acides forts (Montes et al., 2017). Une autre étude mentionne également l'utilisation de la chromatographie en fluide supercritique comparée à une séparation HILIC (Neuwald et al., 2021). L'évaluation a été effectuée en SS en prenant en compte 1 310 MPs polaires, dont 135 ont été retrouvés avec la HILIC contre 123 en fluide supercritique, dont 37 composés en commun. Les auteurs concluent notamment en montrant la complémentarité de ces deux techniques et leur efficacité à séparer les composés polaires (médiane du log D des suspects = -0,4). Une autre comparaison a été effectuée entre la chromatographie en fluide supercritique et une séparation en deux dimensions (C18 et HILIC) sur un mélange de 274 standards analytiques ayant un log D entre -7,71 et +7,74 (Bieber et al., 2017). La combinaison des deux colonnes permet de retenir 262 composés, contre 245 en fluide supercritique où il manque principalement des molécules ayant un log D < -2,5 par rapport à l'autre technique. Cependant, l'utilisation de la chromatographie par fluide supercritique permet de séparer presque autant de composés avec un gain de temps considérable en termes de séparation et les auteurs encouragent l'utilisation des deux techniques qui permettent d'obtenir plus d'informations, notamment pour l'identification ultérieure. De manière similaire, des travaux en SS (3 227 MP organiques), montrent que la chromatographie par fluide supercritique permet de détecter plus d'analytes par rapport à l'utilisation d'une colonne C18, d'une colonne trifonctionnelle ou d'une HILIC (Castro et al., 2021). Cependant, la conclusion qui ressort de cette étude est que la combinaison de plusieurs techniques permet d'élargir le spectre d'informations obtenues, puisqu'avec le fluide supercritique, seulement 67 % des composés potentiellement identifiables ont pu être observés.

Enfin, une étude dans le domaine de la métabolomique a comparé une colonne de type C18 avec 7 colonnes de type HILIC pour la recherche en TS de 764 métabolites dans un mélange de standards et une application en NTS à des échantillons de plasma humain (Wernisch and Pennathur, 2016). Les différences entre les colonnes HILIC étudiées comprenaient à la fois les phases stationnaires testées (*zwitterionic*, *mixed mode*, neutre ou par échange ionique) mais

également les diamètres de particules (entre 3,5 et 5 μm) pour une même phase stationnaire. Les auteurs ont ainsi montré que 65 % des métabolites étaient détectés après séparation sur colonne C18, et 18 à 63 % l'étaient après séparation sur les différentes HILIC. En combinant les sets de données entre la colonne C18 et les différentes HILIC, entre 75 et 85 % de ces métabolites sont retrouvés ce qui montre la complémentarité des colonnes de type C18 et HILIC.

Bien que ces études aient été réalisées à des fins de NTS, l'évaluation et la comparaison des colonnes s'appuient sur des molécules cibles ou au mieux, des suspects dont les propriétés physico-chimiques sont connues. Cependant, bien que ces molécules soient généralement choisies de façon à couvrir une large gamme de propriétés physico-chimiques, de structures ou de familles, il reste un *a priori* sur les molécules probables qui se trouvent dans l'échantillon, ce qui peut donc exercer une potentielle influence sur l'interprétation de résultats. L'obstacle majeur à ce type de comparaison avec une approche en NTS est la détermination de critères de comparaison pertinents qui puissent être applicables à l'ensemble des molécules détectées.

II.2.2.2. Analyse et détection

De manière générale, il est souvent difficile de trouver des études comparant différents appareils d'analyse puisque cela nécessite soit la disponibilité de ces appareils dans le même laboratoire ce qui, pour des appareils d'HRMS représente un coût important, soit un partenariat avec un autre laboratoire. De plus, les analyses sont souvent réalisées en utilisant le matériel disponible même s'il est montré qu'un appareil possède de nombreux avantages par rapport à un autre. En ce qui concerne la HRMS pour le NTS, l'étude qui se rapproche le plus d'une comparaison de systèmes analytiques a utilisé une UHPLC-QTOF et un LC-LTQ-Orbitrap pour le SS des mêmes échantillons d'eaux usées en entrée et sortie de STEU (Bade et al., 2015). L'étude s'est concentrée sur un nombre limité de composés pharmaceutiques et drogues, régulièrement détectés dans les eaux étudiées. De façon globale, les auteurs ont montré que les deux appareils étaient appropriés à ce type d'analyse. Peu de différences ont été observées en termes de composés identifiables. En revanche le nombre d'échantillons dans lesquels ces composés ont été effectivement détectés varie, mais sans tendance particulière vers l'un ou l'autre des appareils. Les auteurs notent cependant que l'Orbitrap a permis d'identifier légèrement plus de composés de façon automatique sans avoir recours à

une investigation manuelle des fragments (Bade et al., 2015). Le premier EIL du réseau NORMAN montre également une grande variété de systèmes analytiques avec 5 fournisseurs différents répertoriés et une majorité d'appareils de type Q-TOF par rapport aux appareils de type Orbitrap (11 contre 5). Ces instruments possèdent également des paramètres d'acquisition qui diffèrent selon l'appareil, à l'origine d'une grande diversité en termes de résultats (Schymanski et al., 2015).

Outre les systèmes analytiques utilisés, les modes d'acquisition peuvent aussi être source de différences dans les résultats obtenus. Il existe trois modes d'acquisition en NTS : le *full scan MS* simple, ou combiné à une acquisition en *Data-Dependant Analysis* (DDA) ou en *Data-Independent Analysis* (DIA) (Tableau 22). Le *full scan MS* permet de scanner tous les ratios m/z dans une gamme donnée. Une fragmentation complémentaire des ions peut être effectuée afin d'obtenir des données supplémentaires pour la détermination de structures chimiques. Pour cela deux approches peuvent être utilisées : le mode DDA qui permet de ne fragmenter qu'un nombre limité d'ions, choisis sur des critères spécifiés par l'utilisateur, ou le mode DIA qui permet théoriquement de fragmenter tous les ions détectés en MS. En mode DDA, le risque repose sur le fait que les efforts se concentrent souvent sur les ions les plus intenses qui ne sont pas forcément les plus intéressants (*e.g.*, faibles intensités attendues pour les TP). En mode DIA, le problème qui se pose est la complexité des données (présence de plusieurs fragments et précurseurs dans un même scan), ce qui rend parfois difficile l'association entre un précurseur et ses fragments.

Une étude dans le domaine de la métabolomique s'est concentrée sur la comparaison de ces trois méthodes en s'appuyant sur 5 critères de comparaisons pour l'analyse de 134 métabolites présents dans les urines. Les critères choisis sont le nombre de métabolites retrouvés, la précision en termes de quantification, la qualité des spectres de fragmentation, le nombre de métabolites ayant des données de fragmentation et l'aspect pratique de la procédure. Globalement, les auteurs montrent que chaque mode d'acquisition a ses avantages. Là où le mode DDA permet d'obtenir des spectres de fragmentation de meilleure qualité et donc s'avère être meilleur pour de l'identification par rapport au DIA, il est moins performant pour retrouver tous les métabolites présents et pour les quantifier de manière précise (Guo and Huan, 2020). De manière similaire une étude a comparé les modes DIA et DDA pour l'analyse en SS de composés pharmaceutiques et leurs TP formés après traitement

par photolyse UV. Les auteurs montrent ainsi que pour la construction de bibliothèques internes, le DDA est sans nul doute le mode d'acquisition privilégié puisqu'il permet d'établir un lien direct entre le précurseur et ses fragments générés en MS². En revanche, à des fins de SS ou NTS, les auteurs recommandent le mode DIA qui, dans cette étude particulière a permis d'obtenir des données de fragmentation pour les TP qui ne correspondaient pas aux critères déterminés (5 ions les plus intenses et liste d'inclusion avec un seuil d'intensité minimal) pour générer une fragmentation en DDA (Fagnani et al., 2022).

Le Tableau 22 récapitule les systèmes analytiques ainsi que les paramètres d'acquisition associés aux études pour lesquelles nous avons comparé les méthodes de préparation (Tableau 21). Tout comme pour les méthodes de préparation, ces données montrent la prédominance d'une méthode d'analyse, généralement caractérisée par une séparation à l'aide d'une colonne chromatographique de type C18. La spectrométrie de masse est effectuée en full-scan sur des gammes le plus souvent comprises entre 50 et 1000 m/z et est suivie d'une acquisition en mode DDA, généralement sur les ions les plus intenses, puisqu'une grande partie de ces études choisissent également de prioriser les ions par leur intensité pour une identification approfondie par la suite.

Tableau 22 : Comparaison des méthodes analytiques pour le NTS d'échantillons aqueux environnementaux

Matrice	TS / SS / NTS	Séparation	Système analytique (fournisseur)	Mode d'acquisition	Gamme m/z	Référence
Eau de surface, eau souterraine et eau potable	TS et NTS	Omnisphere C18 (100 x 2,0 mm, 3 μm) (Varian-Chrompack)	HPLC-ESI-LTQ-Orbitrap MS (ThermoFisher)	Full scan MS et DDA MS ⁿ sur les plus intenses	50 - 1300	(Hogenboom et al., 2009)
Entrée et sortie de STEU	SS et NTS	Acclaim RSLC C18 column (100 x 2,1 mm, 2,2 μm) (ThermoFisher)	UPLC (Dionex) ESI-QTOF (Bruker)	Full scan MS et DDA MS/MS	50 - 1000	(Gago-Ferrero et al., 2015)
		Acquity BEH Amide (100 x 2,1 mm, 1,7 μm) (Waters)				
Eau de surface	TS et NTS	XBridge C18 (50 x 2,1 mm, 3,5 μm) (Waters)	HPLC-ESI-LTQ-Orbitrap (ThermoFisher)	Full scan MS et DDA MS/MS	N.A.	(Kern et al., 2009)
Lixiviât de décharge et eau souterraine	NTS	Zorbax Eclipse XDB C18 (50 x 4,6 mm, 1,8 μm) (Agilent)	HPLC-ESI-QTOF (Agilent)	Full scan MS et DIA MS/MS	30 - 1050	(Müller et al., 2011)
Entrée et sortie de STEU	NTS et post-TS	Acquity C18 BEH (150 x 2,1 mm, 1,7 μm) (Waters)	UPLC-ESI-QTOF (Waters)	Full scan MS et DIA MS ^E	50 - 1000	(Díaz et al., 2012)
Sortie de STEU et eaux de surface	SS	ProntoSIL C18 (250 x 4,0 mm, 5,0 μm) (Bischoff)	HPLC-ESI-IMS-QTOF (Agilent)	Full scan MS	45 - 1700	(Deeb et al., 2017)
Eaux de surface	TS et SS	XBridge C18 (50 x 2.1 mm, 3,5 μm) (Waters)	HPLC-ESI-QExactive-Orbitrap MS(ThermoFisher)	Full scan MS et DDA MS/MS (5 plus intenses)	100 - 1000	(Moschet et al., 2013)
Eaux de surface	TS et NTS	XBridge C18 (50 x 2,1 mm, 3,5 μm) (Waters)	HPLC-ESI-LTQ-Orbitrap (ThermoFisher)	Full scan MS et DDA MS ² sur les target et les plus intenses	N.A.	(Ruff et al., 2015)
Sortie de STEU	TS et NTS	Atlantis T3 C18 (150 x 3,0 mm, 3,0 μm) (Waters)	HPLC-ESI-QExactive Orbitrap (ThermoFisher)	Full scan MS et DDA MS/MS sur les targets et les 5 plus intenses	60 - 900	(Schollée et al., 2018)
Sortie de STEU	TS, SS et NTS	XBridge	HPLC-ESI-LTQ-Orbitrap XL (ThermoFisher)	Full scan MS et DDA MS/MS sur 6 candidats	115 - 1000	(Schymanski et al., 2014b)

PARTIE B : Comparaison de méthodes analytiques

Matrice	TS / SS / NTS	Séparation	Système analytique (fournisseur)	Mode d'acquisition	Gamme m/z	Référence
		C18 column (50 x 2,1 mm, 3,5 µm) (Waters)				
Sortie de STEU	Post-target et NTS	Acquity BEH C18 (100 x 2,1 mm, 1,7 µm) (Waters)	UPLC-ESI-QTOF (Waters)	Full scan MS	100 - 500	(Nurmi et al., 2012)
Eaux de surface, Entrée et sortie de STEU	TS et SS	Acquity BEH C18 (100 x 2,1 mm, 1,7µm) (Waters)	UPLC (Waters) ESI-QTOF (Xevo)	Full scan MS et DIA MS ^E	50 - 1000	(Bade et al., 2015)
		Zorbax SB-C18 (150 x 0,5 mm, 5,0 µm) (Agilent)	HPLC-DESI-LTQ-Orbitrap XL (ThermoFisher)	Full scan MS et DDA MS/MS en CID sur les 4 plus intenses et les target	N.A.	
Eaux de surface	SS et NTS	Acquity C18 (100 x 2,1 mm, 1,7 µm) (Waters)	UPLC-ESI-QExactive Orbitrap (ThermoFisher)	Full scan MS et DDA MS/MS	200 - 1650	(Zonja et al., 2015)
Sortie de STEU	SS et NTS	Acclaim RSLC C18 (100 x 2,1 mm, 2,2 µm) (ThermoFisher)	UPLC (ThermoFisher) ESI-QTOF (Bruker)	Full scan MS et DDA MS/MS sur les 3 plus intenses et DIA en CID	80 - 1000	(Bergé et al., 2018)
Infiltration eau souterraine	NTS	Xbridge C18 (100 x 2,1 mm, 3,5 µm) (Waters)	HPLC-ESI-QExactive Orbitrap (ThermoFisher)	Full scan MS et DDA MS/MS sur les 5 plus intenses	100 - 1000 (MS) 200 - 2000 (MS/MS)	(Albergamo et al., 2019)
			UPLC (Shimadzu) ESI-QTOF (Bruker)	Full scan MS et DDA MS/MS sur les ions priorités par l'Orbitrap	50 - 1000	

II.2.2. HRMS couplée à la chromatographie gazeuse

Bien que le couplage LC-HRMS soit prédominant pour le NTS environnemental car il permet l'analyse d'un large spectre de molécules polaires et solubles dans l'eau (Hollender et al., 2017), certaines études, bien que plus rares, utilisent le couplage GC-HRMS. L'avantage des méthodes de GC par rapport à la LC réside dans le fait que l'ionisation par impact électronique est beaucoup plus reproductible que l'ESI. Par conséquent, les spectres de masse sont plus facilement transférables d'un système à l'autre, résultant en des bibliothèques spectrales bien plus fournies et plus faciles à utiliser (Vinaixa et al., 2016). De plus, en GC, les TR sont plus facilement comparables entre différentes méthodes, rendant ainsi plus facile la confirmation d'identité des composés (Milman and Zhurkovich, 2017; Moschet et al., 2016) et la création de bibliothèques indexant les TR (Babushok, 2015). Cependant, l'impact électronique est une méthode d'ionisation dite dure, qui résulte souvent en une faible intensité des ions moléculaires ce qui rend donc l'identification plus difficile (Hollender et al., 2017). En LC-HRMS, des tentatives de transferts et de prédiction de TR ont aussi émergé ces dernières années mais leur utilisation est plus rare (Aalizadeh et al., 2019, 2021; Domingo-Almenara et al., 2019). Il est aussi important de noter que l'utilisation de la GCxGC-HRMS est une technique plus répandue que la LCxLC-HRMS dans le domaine de l'analyse de matrices environnementales (Ieda et al., 2019; Lebedev et al., 2020; Lübeck et al., 2020).

Une utilisation simultanée de méthodes par LC ESI-QTOF et GC APCI-QTOF a été reportée en SS sur des eaux de surface, eaux souterraines, et rejets de STEU. L'efficacité des méthodes testées a été évaluée sur des critères de sélectivité (*i.e.*, capacité à repérer l'analyte considéré parmi les autres pics présents) et de spécificité (*i.e.*, capacité du détecteur à produire un signal pour l'analyte, distinguable du bruit). Malheureusement, aucune comparaison stricte n'a été effectuée par les auteurs entre les deux méthodes puisque ceux-ci ont choisi d'étudier des suspects spécifiques à chaque appareil. De façon générale, les auteurs ont simplement conclu que l'utilisation simultanée de ces deux méthodes permettait d'élargir la gamme de composés potentiellement analysables (Hernández et al., 2015).

II.2.3. HRMS sans couplage chromatographique

Une dernière technique à évoquer est l'infusion directe de l'échantillon sans séparation chromatographique préalable. En effet, comme démontré dans les parties précédentes, les colonnes chromatographiques apportent une séparation supplémentaire permettant de mieux distinguer les différents signaux. Cependant, cette séparation a également des désavantages comme un temps d'analyse plus long lié à la chromatographie, une pré-sélection mal maîtrisée des composés analysables selon la nature de la colonne utilisée ou encore un temps de prétraitement de données plus long dû notamment à l'intégration et l'alignement des pics chromatographiques.

La majorité des études de NTS relatant l'utilisation de l'infusion directe utilisent des appareils de type FT-ICR-MS. Généralement, ces études se concentrent sur une approche plus globale qui consiste à utiliser différents outils de visualisation permettant d'observer l'évolution de la nature de l'échantillon plutôt que d'identifier précisément les constituants de l'échantillon. Parmi celles-ci, les diagrammes de van Krevelen reviennent régulièrement avec l'utilisation de la FT-ICR-MS pour caractériser la MOD et identifier des sous-produits de désinfection (Andersson et al., 2019; Lavonen et al., 2013; Ziegler et al., 2019). L'établissement de ce type de diagrammes repose sur un pouvoir de résolution et une précision en masse supérieurs de la FT-ICR-MS par rapport à l'Orbitrap et au Q-TOF, ce qui permet une attribution plus fiable de formules moléculaires aux différents signaux observés (Krauss et al., 2010). Ces formules sont ensuite utilisées pour calculer des ratios atomiques caractéristiques de certains types ou familles de molécules (Tableau 23).

Tableau 23 : Classification établie pour catégoriser la MOD selon les ratios O/C et H/C (Minor et al., 2014; Sanchís et al., 2020)

Classification	Gamme du ratio O/C	Gamme du ratio H/C
Lipides	$0 < O/C < 0,2$	$1,5 < H/C < 2,3$
Protéines	$0,2 < O/C < 0,5$	$1,5 < H/C < 2,25$
Osamines	$0,5 < O/C < 0,7$	$1,5 < H/C < 2,25$
Glucides	$0,7 < O/C$	$1,5 < H/C < 2,4$
Hydrocarbures condensés	$0 < O/C < 0,25$	$0,5 < H/C < 1,2$
Lignines	$0,25 < O/C < 0,67$	$0,6 < H/C < 1,5$
Tanins	$0,67 < O/C < 0,95$	$0,55 < H/C < 1,3$

Cependant, il est important de noter que ces études nécessitent souvent une étape de prétraitement d'échantillon, moins cruciale que lorsqu'il y a une séparation chromatographique préalable. En effet, les échantillons contenant une grande quantité de sels inorganiques sont susceptibles d'interférer considérablement au niveau de l'étape d'ionisation. Une étape préalable de SPE permet donc, en plus de concentrer les composés organiques d'intérêt, d'éliminer ces sels.

III. Matériel et Méthodes

III.1. Matériaux utilisés

III.1.1. Etalons internes

Plusieurs étalons internes marqués isotopiquement ont été sélectionnés en fonction de leurs propriétés afin de pouvoir les utiliser par la suite lors des analyses. Les composés choisis couvrent une large gamme de polarité (Log Kow sur une gamme de 11 unités) et d'acidité (pKa sur une gamme de 8,75 unités). De même, le prix et la disponibilité commerciale de ces composés ont également été pris en compte. La liste des composés retenus et leurs propriétés sont détaillées dans le Tableau 24.

Tableau 24 : Propriétés des étalons internes sélectionnés

Molécule	Formule moléculaire	Masse monoisotopique (Da)	Mode d'analyse privilégié	Log Kow	pKa
Aténolol-d7	C ₁₄ D ₇ H ₁₅ N ₂ O ₃	273,2070	LC ESI+	0,16	9,6
Bis-(2-éthylhexyl) phtalate-d4	C ₂₄ D ₄ H ₃₄ O ₄	394,3021	GC	7,60	
Bisphénol A-d6	C ₁₅ D ₆ H ₁₀ O ₂	234,1527	LC ESI -	3,32	9,6
Caféine-d9	C ₈ D ₉ H ₁ N ₄ O ₂	203,1369	LC ESI +	-0,07	14,0
Carbamazépine-d8	C ₁₅ D ₈ H ₄ N ₂ O	244,1452	LC ESI +	2,45	13,9
Estradiol-d5	C ₁₈ D ₅ H ₁₉ O ₂	277,2090	GC	4,01	
Fluorène-d10	C ₁₃ D ₁₀	176,1410	GC	4,18	
Glyphosate- ¹³ C	C ₃ H ₈ NO ₅ P	169,0140	LC ESI -	-3,40	2,34
Naproxen-d3	C ₁₄ D ₃ H ₁₁ O ₃	233,1131	LC ESI -	3,18	4,15
Propylparabène-d4	C ₁₀ D ₄ H ₈ O ₃	184,1038	LC ESI -	3,04	8,5
Sulfaméthoxazole-d4	C ₁₀ D ₄ H ₇ N ₃ O ₃ S	257,0772	LC ESI +	0,89	1,6
Triclosan-d3	C ₁₂ D ₃ H ₄ Cl ₃ O ₂	290,9700	LC ESI -	4,76	7,9

III.1.2. Choix des méthodes et supports de fractionnement

Différentes méthodes de préparation ont été sélectionnées, avec un effort particulier qui a porté sur les méthodes de SPE puisqu'elles sont les plus couramment utilisées pour ce type d'application. Ainsi, 8 phases de SPE ayant des propriétés différentes ont été utilisées (Tableau 25), incluant en particulier, les cartouches Oasis HLB, Isolute ENV+, Strata X-AW et Strata X-

CW. Ces dernières entrent dans la composition de la cartouche multiphasique développée pour la rétention d'un large spectre de composés organiques (Kern et al., 2009) et sont régulièrement retrouvées dans les études de NTS. Des résines macroporeuses de type XAD et AGMP ont également été sélectionnées pour le fractionnement de l'échantillon, sur la base d'études destinées au fractionnement de la MOD (Croué, 2004; Leenheer, 1981). Enfin, une méthode s'appuyant sur un fractionnement en fonction de la taille a été testée en utilisant une colonne polymérique de chromatographie d'exclusion de taille.

Tableau 25 : Liste des cartouches SPE sélectionnées et leurs propriétés selon les fabricants

Phase (fournisseur)	Propriétés
SPE Strata C18 (Phenomenex)	Sélectivité pour les composés modérément hydrophobes, faible sélectivité pour les polaires
SPE Coconut Charcoal (Sigma-Aldrich)	Développé spécifiquement pour la méthode EPA 521 pour la détermination des nitrosamines dans l'eau potable
SPE DPA-6S (Sigma-Aldrich)	Sélectivité pour les composés polaires, notamment ceux contenant des groupes hydroxyles ou carboxyles
SPE Isolute ENV+ (Biotage)	Sélectivité pour les composés polaires qui ne sont pas retenus par des phases de type C8 ou C18
SPE ENVI Carb (Sigma-Aldrich)	Grande sélectivité pour les molécules organiques polaires et les composés non polaires
SPE Oasis HLB (Waters)	Utilisable pour toutes les applications, idéale pour les composés acides, basiques et neutres
SPE Strata X-AW (Phenomenex)	Sélectivité pour les composés acides avec un pKa inférieur à 5
SPE Strata X-CW (Phenomenex)	Sélectivité pour les composés ayant un pKa supérieur à 8, y compris les amines quaternaires
Résine DAX 8 (Sigma-Aldrich)	Affinité pour la matière organique hydrophobe
Résine XAD 4 (Sigma-Aldrich)	Affinité pour la matière organique transphilique
Résine AGMP 50 (Biorad)	Affinité pour la matière organique hydrophile basique
Résine AGMP 1 (Biorad)	Affinité pour la matière organique hydrophile acide
Colonne SEC polymérique (Shodex)	Seuil de coupure à 40 000 g/mol

III.1.3. Colonnes chromatographiques

Trois colonnes avec des phases stationnaires différentes ont été choisies. La première est une colonne de type C18 (Acquity UPLC BEH C18, 100 x 2,1 mm ; 1,7 μm , Waters), constituée d'une base de silice greffée avec une chaîne d'alkyle à 18 carbones, qui permet la rétention des molécules apolaires et peu polaires. La séparation s'appuie majoritairement sur des interactions hydrophobes entre les analytes et la phase stationnaire. C'est sans doute une des phases stationnaires les plus utilisées en chromatographie liquide et particulièrement pour des fins d'analyse non-ciblée de micropolluants organiques dans les eaux.

La seconde phase sélectionnée est une colonne en carbone graphite (PGC Hypercarb, 100 x 2,1 mm ; 3 μm , ThermoFisher) qui possède des propriétés relatives à la chromatographie en phase inverse (*i.e.*, rétention dépendante des interactions hydrophobes, nature des phases mobiles) et à la chromatographie en phase normale (*i.e.*, phase carbonée qui agit comme accepteur d'électrons, rétention affectée par les effets stériques). Elle permet d'augmenter les performances d'une phase inverse plus classique en augmentant l'affinité pour les composés plus polaires. Ici, les mécanismes de rétention mis en jeu sont nombreux. Tout d'abord, des interactions à courtes distances telles que des liaisons hydrogènes, des interactions dispersives et des interactions dipolaires se font entre l'éluant et les analytes. Des phénomènes de répulsions hydrophobes ont également lieu entre l'éluant et les analytes non polaires ou les groupements apolaires de certains composés. Des interactions dispersives de type London entre la surface de la phase stationnaire et l'analyte influent également sur la rétention. Enfin, les interactions induites par la charge entre l'analyte et la phase stationnaire ont lieu mais sont compensées par les interactions polaires avec l'éluant sauf dans le cas où les groupements polaires sont planaires, ce qui induit leur contact direct avec la surface de la phase stationnaire et, dans ce cas, ce mécanisme est prédominant par rapport à l'interaction polaire avec l'éluant. Ainsi, cette colonne a été utilisée par exemple pour la séparation d'isomères (Gundersen, 2001) ou encore la séparation de composés ioniques (Gu and Lim, 1990).

La dernière phase étudiée est une colonne de chromatographie d'interaction hydrophile HILIC (Acquity UPLC BEH HILIC, 100 x 2,1 mm ; 1,7 μm , Waters) qui se rapproche de la chromatographie en phase normale, constituée d'une base de silice vierge. Elle se distingue de la chromatographie en phase normale par les éluants utilisés, généralement similaires à

ceux utilisés en phase inverse. En chromatographie HILIC, différents mécanismes de rétention interagissent aussi. Tout d'abord, les analytes se partagent entre la phase stationnaire et la phase mobile, avec notamment les molécules polaires qui interagissent avec la couche polaire partiellement immobilisée de la phase stationnaire. Un mécanisme d'échange d'ions est également mis en jeu entre les groupements silanols de surface et les analytes chargés. Enfin, des liaisons hydrogènes sont également mises en jeu entre les analytes chargés positivement et les groupements silanols de surface, chargés négativement. L'utilisation de ce type de colonne est régulièrement répertoriée pour des études biologiques comme l'analyse de glycanes (Melmer et al., 2010), peptides (Yoshida, 2004) ou protéines (Periat et al., 2016).

III.2. Echantillonnage et réactifs

Les échantillons utilisés proviennent de la STEU Seine Amont (dégrillage, dessablage, décantation primaire, traitement biologique et déphosphatation tertiaire), une des six STEU du SIAAP située à Valenton.

Deux séries d'échantillons d'effluent traité (*i.e.*, rejeté en Seine) ont été prélevées dans des bouteilles en verre de 10 L, préalablement grillées pour éviter toute trace de contamination, et filtrés dans les 24 h (filtres GF/F, porosité 0,7 μm). Un premier prélèvement (6 juin 2019) a servi à la comparaison des méthodes chromatographiques et de toutes les méthodes de préparation (fractionnement sur résines, extractions SPE et chromatographie d'exclusion stérique - SEC) (parties B - IV.1 et IV.2). Un deuxième prélèvement (9 novembre 2020) a servi à la comparaison des types d'analyseurs (partie B - IV.3). Les réactifs utilisés dans cette partie sont récapitulés dans le Tableau 26.

Tableau 26 : Réactifs utilisés dans les différentes méthodes analytiques

Réactif	Qualité - Pureté	Fournisseur
Méthanol (MeOH)	ULC-MS grade	Biosolve
Acétonitrile (ACN)	ULC-MS grade	Biosolve
Acétate d'éthyle (AcEt)	Analytical grade	Merck
Dichlorométhane (DCM)	MS grade	VWR
Acétone	Analytical grade	Merck
Acide formique (AF)	ULC-MS grade	Biosolve
Formate d'ammonium	> 99 %	Merck
NH ₄ OH	35 %	Fisher Scientific
Acide sulfurique	98 %	Merck
NaOH	> 97 %	
HCl	37 %	
Acétate d'ammonium	98 %	Merck
Acide polyacrylique (1250 ; 2925 ; 7500 ; 16000 ; 28000 g/mol)	Standards	Agilent

III.3. Prétraitement des échantillons

III.3.1. Extraction sur phase solide (SPE)

Les extractions SPE ont été conduites sur 8 phases différentes dont la description et les conditions opératoires sont fournies dans le Tableau 27. Les échantillons destinés à la SPE ont été répartis dans des flacons de 1 L auxquels un mélange d'étalons internes a été ajouté et le tout a été acidifié à pH 6,5 à l'aide d'acide sulfurique (98 %).

Le conditionnement des cartouches et le chargement de l'échantillon ont été réalisés sur un appareil Autotrace (Dionex, ThermoFisher). Après séchage des cartouches pendant 30 min sous flux d'azote, celles-ci ont été éluées manuellement (Visiprep, Sigma-Aldrich) du fait de la différence de conditions d'élution pour les cartouches sélectionnées. Les solvants utilisés et les conditions opératoires ont été choisis en fonction des recommandations générales des fournisseurs (Tableau 25). Les extraits ont, par la suite, été séparés en 2 sous-extraits et évaporés à sec sous flux d'azote. L'un a été repris dans un mélange EUP/MeOH (80/20, v/v) et l'autre a été reconstitué dans un mélange EUP + 10 mM d'acétate d'ammonium + 0,1% AF / ACN (10/90, v/v), avant d'être filtrés sur un filtre seringue à 0,2 µm et transférés en vials pour analyse.

Tableau 27 : Protocoles détaillés de conditionnement et d'éluion pour les 8 cartouches SPE

Cartouche	Conditionnement	Elution
Strata C18 (500 mg)	10 mL MeOH	15 mL MeOH
	10 mL EUP	
Coconut Charcoal (2 g)	10 mL MeOH	15 mL DCM
	10 mL EUP	
DPA-6S (500 mg)	10 mL MeOH	15 mL Acétone
	10 mL EUP	
Isolute ENV+ (500 mg)	10 mL MeOH	7,5 mL MeOH
	10 mL EUP	7,5 mL acétone + 5 % NH ₄ OH
ENVI Carb (500 mg)	10 mL MeOH	15 mL MeOH + 5 % AF
	10 mL EUP	
Oasis HLB (200 mg)	10 mL MeOH	5 mL MeOH
	10 mL AcEt	5 mL AcEt
	5 mL DCM	5 mL DCM
	10 mL EUP	
Strata X-AW (500 mg)	10 mL MeOH	5 mL MeOH
	10 mL EUP	5 mL MeOH + 5 % NH ₄ OH
		5 mL MeOH + 5 % AF
Strata X-CW (500 mg)	10 mL MeOH	5 mL MeOH
	10 mL EUP	5 mL MeOH + 5 % NH ₄ OH
		5 mL MeOH + 5 % AF

III.3.2. Fractionnement sur résines

Les résines utilisées ont nécessité un lavage préalable afin d'éliminer toutes traces de contamination qui proviendraient de leur fabrication. Dans un premier temps, les résines macroporeuses (DAX 8 et XAD 4) ont été immergées dans un bain de soude (NaOH, 0,1 M) pendant une semaine et la résine ionique (AGMP 1) a été successivement immergée dans un bain d'acide chlorhydrique (HCl, 1M), puis un bain de soude (NaOH, 0,1 M) pendant une semaine. Suite à cela, des extractions successives par Soxhlet avec différents solvants ont été réalisées pendant des périodes de 24 h ou de 48 h en fonction de la résine et du solvant, selon des protocoles établis précédemment (Leenheer, 1981; Tang et al., 1981). Enfin, une fois les extractions réalisées, les résines ont été introduites dans des colonnes en verre et rincées à l'EUP jusqu'à obtention d'une valeur de carbone organique dissous (COD) en sortie de colonne inférieure à 0,5 mgC/L.

La procédure de fractionnement sur résines est schématisée en Figure 22. 30 L d'échantillon ont été chargés successivement sur les deux colonnes contenant les résines DAX 8 et XAD 4. L'échantillon, après passage sur la résine XAD 4, a ensuite été ajusté à pH 9 à l'aide de NaOH avant passage sur la colonne AGMP 1. Une fois l'échantillon passé sur chaque colonne, celles-ci ont été rincées à l'aide d'EUP acidifiée à pH 2 jusqu'à obtenir, en sortie de colonne, la même conductivité qu'en entrée de colonne. Une fois cela accompli, l'élution a été réalisée sur chaque colonne avec de l'acétate d'éthyle et les trois extraits ont été concentrés à l'évaporateur rotatif pour atteindre un facteur de concentration de 500 (égal à celui des extraits de SPE).

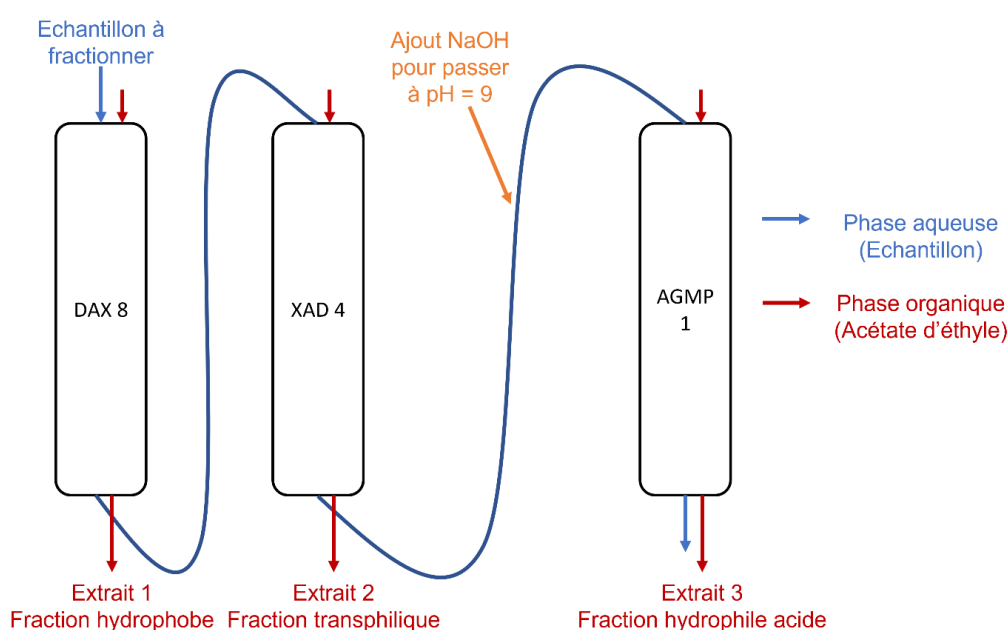


Figure 22 : Procédure de fractionnement successif sur résines

III.3.3. Fractionnement par chromatographie d'exclusion stérique (SEC)

Une chromatographie haute performance (HPLC, Shimadzu), équipée d'un détecteur UV et couplée à un collecteur de fractions, a été utilisée afin de fractionner l'échantillon en fonction de la taille des molécules, par chromatographie d'exclusion stérique (SEC). La colonne utilisée pour la séparation est une Asahipak GF-310 HQ (7,5 x 300 mm, 5 μ m, Shodex) ayant un seuil de coupure à 40 000 g/mol. L'éluant est un tampon aqueux à 100 mM d'acétate d'ammonium à pH 9. L'échantillon est injecté à un débit de 1 mL/min pendant 30 min avec une colonne maintenue à 40°C. Dans un premier temps, un étalonnage en taille de la colonne a été réalisé à l'aide d'acide polyacrylique de différentes masses (1250 ; 2925 ; 7500 ; 16000 ; 28000 Da), de l'acétone (58 Da) et du bleu de thymol (466 Da).

Les échantillons ont été préalablement concentrés (évaporateur rotatif) 15 fois (maximum atteignable avant précipitation). Après étude de la réponse de cet échantillon à une longueur d'onde de 254 nm (test sur une injection d'1 mL), des gammes de temps de rétention ont été déterminées pour effectuer la collecte (Tableau 28).

Tableau 28 : Gamme de TR utilisées pour la collection de fractions après séparation par HPLC-SEC

Fraction	Plage de collection (TR en min)
Fraction 0	0,33 – 2,00
Fraction 1	6,00 – 9,00
Fraction 2	9,10 – 12,00
Fraction 3	12,10 – 15,00
Fraction 4	15,10 – 18,00

L'échantillon (1 mL concentré 15 fois) a ainsi été injecté et les fractions précédemment définies ont été récupérées par le collecteur de fraction, puis filtrées sur un filtre seringue à 0,2 µm et introduites dans des vials pour analyse en HRMS.

III.4. Procédures analytiques

III.4.1. Analyses par HRMS

Les analyses HRMS ont été effectuées par UPLC ESI-IMS-QTOF (Vion, IMS-QTOF, Waters). La séparation a été effectuée sur trois colonnes différentes selon les conditions opératoires décrites dans le Tableau 29. Pour la séparation sur colonne C18 et PGC, les extraits repris dans le mélange EUP/MeOH (80/20, v/v) ont été utilisés alors que pour la colonne HILIC, les extraits repris dans le mélange EUP/acétate d'ammonium + AF/ACN (10/90, v/v) ont été injectés. A noter que l'injection directe d'échantillon aqueux n'a pas été effectuée avec la colonne HILIC pour des raisons d'incompatibilité avec la colonne.

Tableau 29 : Conditions chromatographiques des 3 colonnes pour les analyses HRMS

Colonne	Acquity UPLC BEH C18 (Waters)			Hypercarb (PGC) (ThermoFisher)			Acquity UPLC BEH HILIC (Waters)		
Débit	0,45 mL/min			0,3 mL/min			0,45 mL/min		
Température	Echantillon 10°C, colonne 40°C			Echantillon 10°C, colonne 60°C			Echantillon 10°C, colonne 40°C		
Gradient A = EUP + 0,1 % AF B = ACN + 0,1 % AF C = EUP + 10 mM HCOONH ₄ à pH 3	T (min)	% A	% B	T (min)	% A	% B	T (min)	% C	% B
	0	98	2	0	98	2	0	5	95
	1	98	2	1	98	2	1	5	95
	25	2	98	25	2	98	16	45	55
	30	2	98	30	2	98	18	60	40
	31,5	98	2	31,5	98	2	23	60	40
	34	98	2	34	98	2	24,5	5	95
	-	-	-	-	-	-	34	5	95

L'analyse a été effectuée en mode HDMS^E (DIA) après ionisation par une source ESI opérant dans les modes positif et négatif, et les masses ont été scannées de 100 à 1000 Da avec une fenêtre de 0,2 sec. L'énergie de collision a été fixée à 6,0 eV et pour la fragmentation, une rampe de 20,0 eV à 56,0 eV a été utilisée. Les paramètres pour les deux modes d'ionisation sont donnés dans le Tableau 30.

Tableau 30 : Paramètres d'ionisation pour les analyses en HRMS

	Positif	Négatif
Tension du capillaire	0,80 kV	2,50 kV
Température de la source	120°C	
Température de désolvation	500°C	
Débit du gaz au cône	50 L/h	
Débit du gaz de désolvation	1000 L/h	

Pour une séquence d'analyse donnée, chaque échantillon a été injecté en triplicat dans un ordre aléatoire (généralisé sous Excel grâce à la fonction *alea()*). Un échantillon de contrôle, constitué d'un mélange d'un même volume de tous les échantillons, a également été injecté en début de séquence pour équilibrer la colonne, puis, régulièrement au cours de la séquence pour vérifier la stabilité de l'instrument. En ce qui concerne le traitement de données, celui-ci a été effectué suivant les méthodes détaillées dans la partie A.

III.4.2. Analyses par UHRMS

Les analyses par UHRMS ont été réalisées, en collaboration avec le laboratoire COBRA (UMR 6014, Rouen), sur un FT-ICR (SolariX XR FTMS, Bruker Daltonics) équipé d'un aimant 12 T et d'une source ESI. Les extraits précédemment repris dans un mélange EUP/MeOH (80/20) ont été dilués au centième en gardant les mêmes proportions de solvants, avant d'être infusés directement sur l'appareil à un débit de 120 $\mu\text{L}/\text{h}$. La source ESI a été opérée en modes positif et négatif et les paramètres utilisés correspondants sont répertoriés dans le Tableau 31.

Tableau 31 : Paramètres d'acquisition en FT-ICR-MS

	Positif	Négatif
Tension du capillaire	4,5 kV	
Gaz de nébulisation	1 Bar	
Température de désolvatation	180°C	200°C
Temps d'accumulation	0,2 sec	
Nombre de scans	250	
Gamme de masse	m/z 100 - 1000	
Étalonnage externe	Sodium trifluoroacétate	
Étalonnage interne	Ions : m/z 118,08652 ; 120,05562 ; 140,06820 ; 173,07843 ; 204,14414 ; 217,10464 ; 233,02430 ; 261,13086 ; 274,21426 ; 305,15707 ; 319,18798 ; 319,22437 ; 322,04812 ; 349,18329 ; 370,17950 ; 437,23572 ; 481,26193 ; 525,28815 ; 569,31436 ; 613,34058 ; 748,48417	Ions : m/z : 112,98559 ; 118,04107 ; 129,09210 ; 157,12340 ; 183,09648 ; 211,17035 ; 239,20165 ; 255,23295 ; 293,17583 ; 325,11150 ; 329,13945 ; 339,12717 ; 353,14282 ; 367,15847 ; 427,22518 ; 431,98233 ; 441,20445 ; 601,97898

Les données ont été prétraitées et traitées avec *Data Analysis* pour générer les listes de masses et attribuer les formules moléculaires. Dans un premier temps, la détection des pics a été effectuée entre m/z 100 et 1000, sans spécifier de seuil minimal d'intensité. Par la suite, les différentes listes de masses obtenues ont été étalonnées grâce aux ions indiqués dans le Tableau 31 afin d'obtenir des listes comparables. Les blancs correspondants à chaque échantillon ont ensuite été soustraits et une attribution séquentielle de formule moléculaire a été effectuée. Le choix des paramètres de calculs de formules s'est appuyé sur les

paramètres optimisés de la méthode définie sous UNIFI (partie A – IV.1.2.2) avec quelques modifications qui se sont avérées nécessaires afin de limiter les faux positifs. Ces paramètres sont les suivants :

- Aucune limite haute en termes de C et de H n'a été fixée
- En termes d'hétéroatomes, une limite de 10 a été imposée pour N, 30 pour O, 4 pour Cl et 1 pour le S
- Les adduits Na⁺ et K⁺ ont été autorisés en ESI+
- D'après les étalonnages réalisés en ESI+ et ESI-, une déviation maximale de 0,2 ppm et 0,5 ppm, respectivement, ont été autorisés

Une fois les listes de formules moléculaires obtenues, elles ont été exportées au format csv pour générer des diagrammes de van Krevelen. Les différentes listes de masses de chaque extrait ont également été exportées au format csv et alignées (à l'aide du logiciel R) afin de construire l'équivalent d'une *marker table* et traiter les données suivant les méthodes décrites dans la partie A.

IV. Résultats

IV.1. Comparaison des méthodes chromatographiques

Dans cette partie nous nous intéresserons à la comparaison des 3 colonnes chromatographiques. Les résultats présentés s'appuieront sur les données obtenues avec le seul mode d'ionisation (ESI+) pour une meilleure lisibilité. Les marqueurs comparés ont été considérés indépendamment de la méthode de préparation employée afin d'être le plus exhaustif possible.

Les aspects pratiques liés à l'utilisation des 3 colonnes diffèrent. Alors que les mêmes phases mobiles et les mêmes gradients ont été utilisés pour les séparations réalisées sur C18 et sur PGC, la séparation sur HILIC a nécessité certaines adaptations. En effet, la HILIC n'est pas utilisée en phase inverse contrairement aux deux autres colonnes, mais se rapproche plus d'une séparation en phase normale. Le gradient doit donc commencer avec une plus forte proportion de solvant organique et l'augmentation de la proportion de phase aqueuse dans le mélange permettra de décrocher les composés les plus retenus sur la colonne. Par ailleurs, il est recommandé de ne pas dépasser 60 % de contenu aqueux dans la phase mobile afin d'éviter de décrocher la couche aqueuse permettant les interactions spécifiques à la colonne et de travailler avec des phases mobiles tamponnées afin d'éviter des pics asymétriques. D'autre part la phase de rééquilibrage de la HILIC en fin de gradient a également été augmentée par rapport aux autres colonnes afin d'assurer un retour à une pression stable avant le lancement de l'injection suivante. Par ailleurs, alors que les phases C18 et HILIC peuvent être trouvées chez différents fournisseurs, la PGC est seulement disponible chez ThermoFisher. Les connectiques étaient différentes d'un fournisseur à un autre, un adaptateur a dû être utilisé afin de fixer la colonne sur le système UPLC Waters et d'éviter les fuites.

La stabilité de ces colonnes a été comparée. Pour cela, une étude de la déviation des TR a été menée sur l'ensemble des analyses réalisées sur l'année 2019 en prenant en compte les 4 étalons deutérés visibles en ESI+ quelle que soit la méthode de préparation (Figure 23).

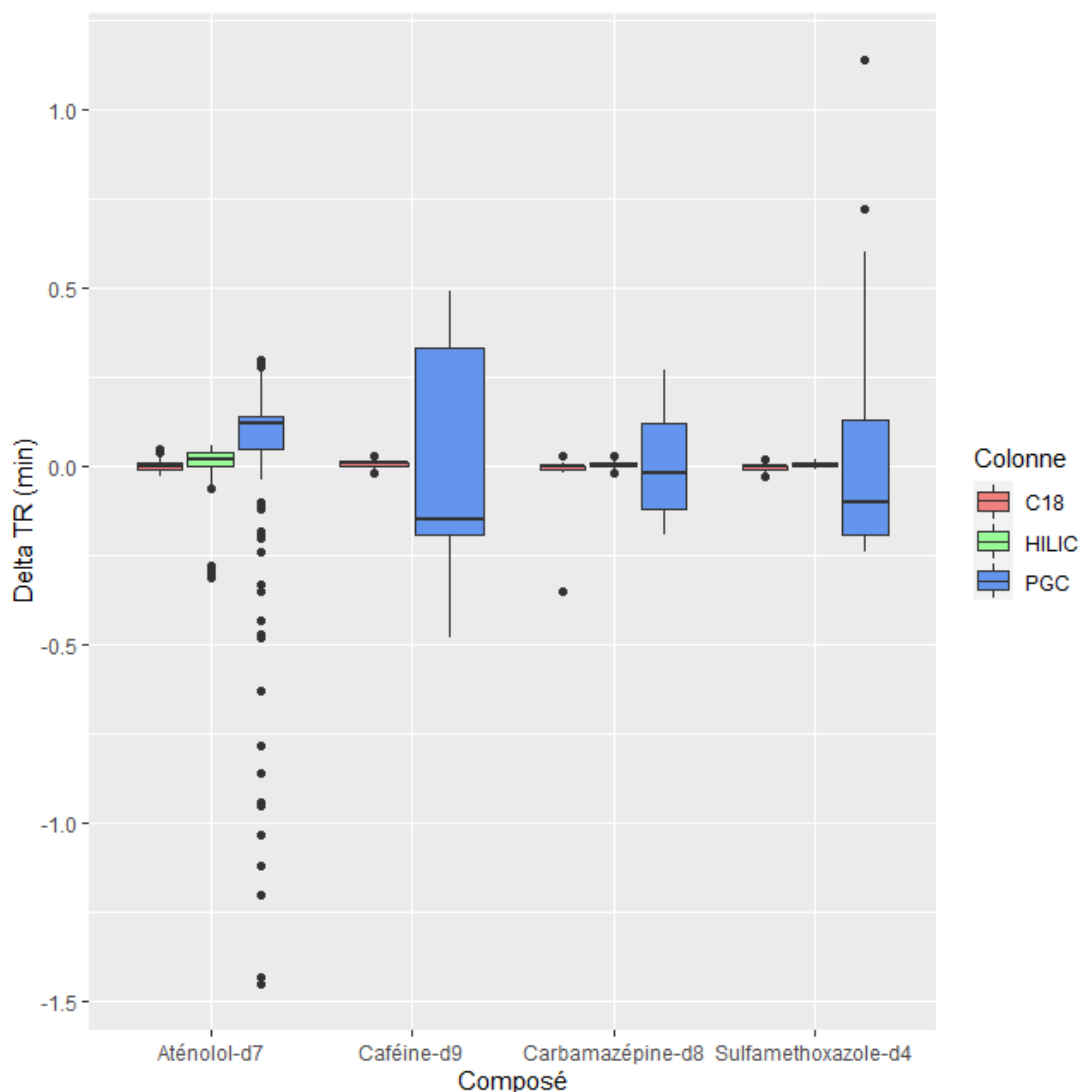


Figure 23 : Déviation des TR (par rapport à la moyenne) inter et intra-séquence des quatre étalons deutérés visibles en ESI+ pour les 3 colonnes chromatographiques) (n > 300 dans des matrices de complexité différentes)

Les variations de TR observées sur les colonnes C18 et HILIC sont globalement comparables quel que soit le composé considéré. Avec 75 % des valeurs ayant une déviation inférieure à 0,05 min, seules quelques valeurs aberrantes sont observées avec une déviation supérieure à 0,2 min pour l'aténolol-d7 sur la HILIC et pour la carbamazépine-d8 pour la C18. En revanche, la PGC montre des variations de TR beaucoup plus importantes, avec 75 % des valeurs qui montrent une variation comprise entre -0,5 et 0,5 min. De plus, l'aténolol-d7 présente beaucoup des valeurs de variation négatives, atteignant jusqu'à -1,5 min. Ces observations témoignent d'une moins bonne stabilité de la PGC comparée aux deux autres colonnes. Ces observations ont aussi été observées lors de la filtration des marqueurs entre répliqués. En effet, dans le cas de la C18 et de la HILIC, environ 50 % des marqueurs ont été éliminés par

cette étape alors que 75 % l'ont été pour la PGC. Les marqueurs étant alignés à la fois sur le TR, le m/z et le DT, les variations des deux autres paramètres ont également été étudiées, montrant peu de différences entre les colonnes. La PGC est donc la moins stable des 3 colonnes, pouvant engendrer de grandes variations en termes de TR pour certains composés. Différents temps de rééquilibrage de colonne en fin de gradient ont été testés afin de s'assurer que cette observation n'était pas liée à un mauvais équilibrage. Cependant, l'augmentation de la durée de cette étape n'a pas permis d'améliorer les observations.

Dans un second temps, les colonnes ont été comparées en termes d'informations qu'elles pouvaient fournir pour les différents échantillons analysés. Pour cela, les distributions des valeurs de m/z et DT des marqueurs séparés sur chaque colonne pour une séquence donnée ont été tracées (Figure 24).

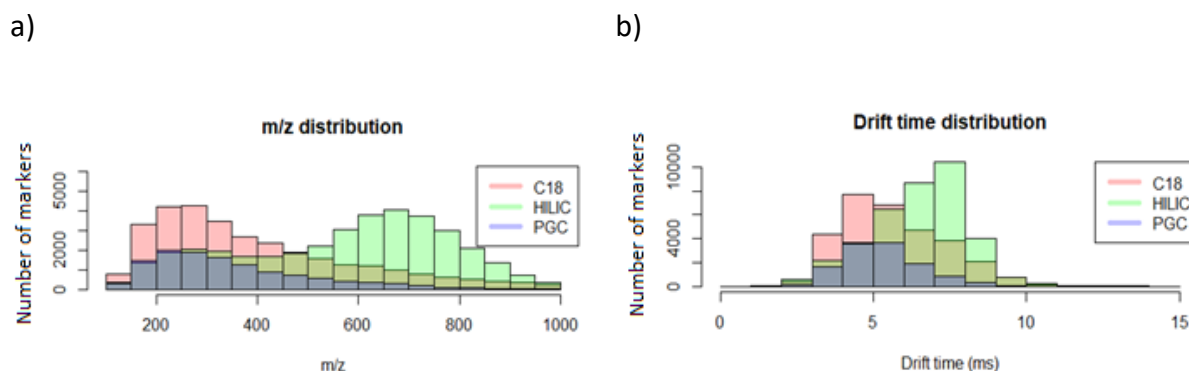


Figure 24 : Distribution des valeurs de (a) m/z et (b) DT des marqueurs séparés sur les 3 colonnes

Pour les mêmes échantillons, il est intéressant de noter que la HILIC permet de séparer et analyser des molécules de m/z globalement plus élevés que les deux autres colonnes. Cette observation est cohérente avec la distribution en DT qui montre la rétention de molécules avec des surfaces de collision plus élevées pour la HILIC. Les molécules retenues sur C18 et PGC semblent, quant à elles, être réparties de manière similaire sur les gammes de m/z et DT. En ce qui concerne les TR, les distributions n'ont pas été étudiées puisque les mécanismes de rétention mis en jeu sur les 3 colonnes sont différents, engendrant des temps de rétention difficilement comparables. En revanche, il est intéressant de noter que les TR sont répartis de manière plus uniforme tout au long du gradient sur la PGC (Figure 25a), témoignant d'une meilleure séparation, alors que sur la HILIC, la très grande majorité des marqueurs sont élués avant 10 min, sur un gradient de 18 min (Figure 25b).

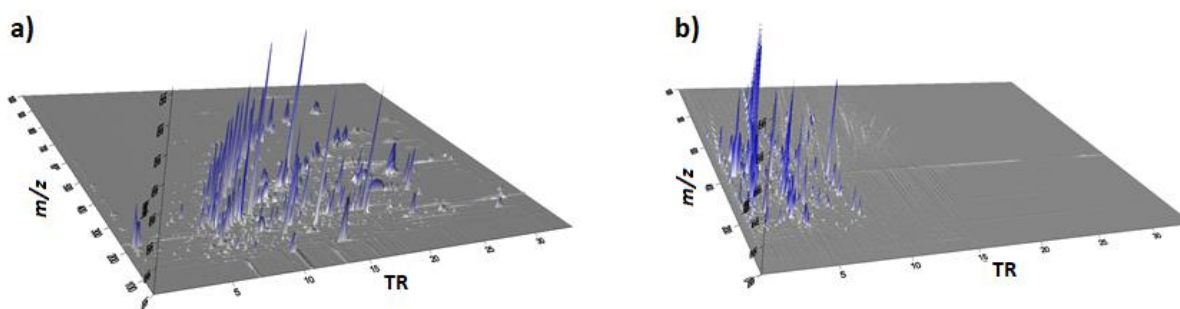


Figure 25 : Chromatogrammes d'un même extrait séparé sur a) PGC et b) HILIC

Les marqueurs retenus sur les colonnes C18 et PGC couvrant les mêmes gammes en termes de m/z et DT, la comparaison des marqueurs a été approfondie afin de savoir si ces derniers étaient effectivement similaires ou s'ils couvraient juste les mêmes gammes de masse.

Pour cela, les marqueurs ont été alignés selon leurs rapports m/z (avec une tolérance de 20 ppm) (Figure 26a) et leurs DT (avec une tolérance de 0,5 ms) (Figure 26b). Ces fenêtres de tolérance ont été déterminées à partir d'une étude inter et intra-séquence prenant en compte la variabilité des rapports m/z et des DT des différents étalons injectés sur une période d'un an. Sur l'ensemble des séquences (3 types de colonnes), la variabilité intra-séquence était inférieure à 9,6 ppm et 0,18 ms, et la variabilité inter-séquence était inférieure à 16 ppm et 0,45 ms, pour le rapport m/z et le DT respectivement. Cependant, à la suite de ces deux étapes d'alignement, certains marqueurs d'une colonne étaient alignés avec plusieurs marqueurs d'une seconde colonne. En effet, certains marqueurs pouvaient être compris dans les mêmes fenêtres de m/z et DT, mais différer seulement par leur TR, se retrouvant alors alignés avec le même marqueur de référence (*i.e.*, de première colonne considérée) (Annexe 2). Ainsi, afin de compter le nombre de marqueurs *communs* entre colonnes, une étape supplémentaire a été implémentée afin de ne compter qu'un seul alignement lorsqu'un marqueur d'une colonne était aligné avec plusieurs marqueurs d'une autre colonne (Figure 26c). Les différents alignements ont été effectués dans les deux sens (*i.e.*, en changeant la première colonne de référence sur laquelle les fenêtres de tolérance sont déterminées) et les nombres de marqueurs *communs* présentés (Figure 26) résultent de la moyenne de ces deux sens d'alignement.

Le premier alignement basé sur le rapport m/z a d'abord montré qu'environ 12 000 marqueurs présentaient des rapports m/z identiques (avec la tolérance décrite précédemment) dans les trois colonnes (Figure 26a). L'utilisation de l'étape supplémentaire d'alignement selon les

valeurs de DT a ensuite montré que la moitié de ces marqueurs ne correspondaient pas entre eux et ne pouvaient donc pas être considérés comme de "vrais *communs*" entre les trois colonnes. Ces marqueurs correspondaient donc soit à des marqueurs *spécifiques* de chaque colonne, soit à des marqueurs *communs* à seulement deux colonnes, d'où l'augmentation du nombre de marqueurs dans ces catégories (Figure 26b). Il est intéressant de noter que seul le nombre de marqueurs *communs* entre la colonne C18 et la colonne HILIC a diminué au cours du processus, avec, une fois encore, seulement la moitié des marqueurs alignés à l'aide du *m/z* correspondant également en DT. D'autre part, le nombre total de marqueurs entre la Figure 26a et la Figure 26b ne correspondait pas (46 414 contre 64 420 marqueurs respectivement).

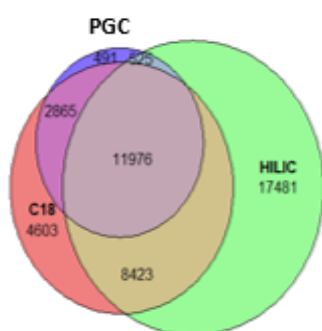
L'alignement par le DT a en effet entraîné une diminution d'environ 50 % du nombre total de marqueurs entre les 3 colonnes qui se retrouvent donc soit *communs* à seulement deux colonnes, soit *spécifiques* à une colonne en fonction des cas. De la même manière, les marqueurs initialement *communs* entre deux colonnes en Figure 26a peuvent également s'avérer *spécifiques* à une colonne car le DT des marqueurs alignés ne correspondaient pas. L'alignement par DT entraîne donc une augmentation du nombre de marqueurs *spécifiques* à chacune des 3 colonnes (+ 130 %, + 105 % et + 64 % pour la colonne C18, PGC et HILIC, respectivement) mais concernant les marqueurs *communs* à deux colonnes, deux cas sont observés :

- Il y a plus de marqueurs *communs* aux trois colonnes qui ne sont en réalité *communs* qu'à deux colonnes, par rapport au nombre de marqueurs *communs* à deux colonnes qui sont en fait *spécifiques*, suite à la prise en compte du DT. Cela conduit donc à une augmentation du nombre de marqueurs *communs* à deux colonnes dans la Figure 26b. C'est le cas pour les marqueurs *communs* entre la colonne C18 et PGC (+ 145%) et entre la PGC et la HILIC (+ 1200 %)
- Il y a moins de marqueurs *communs* aux trois colonnes qui ne sont en réalité *communs* qu'à deux colonnes, par rapport au nombre de marqueurs *communs* à deux colonnes qui sont en fait *spécifiques*, suite à la prise en compte du DT. Cela conduit donc à une diminution du nombre de marqueurs *communs* à deux colonnes dans la Figure 26b. C'est le cas pour les marqueurs *communs* entre la colonne C18 et la HILIC (- 54%)

La prise en compte du DT a donc entraîné un changement de catégorie avec :

- 10 171 marqueurs qui ont contribué à une diminution (- 50 % pour les marqueurs *communs* à 3 colonnes et - 54 % pour les marqueurs *communs* entre la colonne C18 et HILIC)
- 28 177 marqueurs qui engendrent une augmentation (augmentation du nombre de marqueurs *spécifiques* à chacune des trois colonnes et augmentation du nombre de marqueurs *communs* à la PGC et HILIC ou la PGC et colonne C18)

Soit une différence de 18 006 marqueurs observée entre la Figure 26a et la Figure 26b. Cette différence est notamment liée à l'augmentation considérable du nombre de marqueurs *communs* entre la PGC et l'une des deux autres colonnes. Il a été remarqué qu'entre 2 300 et 8 700 marqueurs sur une colonne pouvaient être alignés avec plusieurs marqueurs sur une deuxième colonne, en fonction des combinaisons considérées, d'où la nécessité d'ajouter une étape de filtration complémentaire (Figure 26c).

a) Alignement m/z 

b) Alignement DT



c) Filtration complémentaire

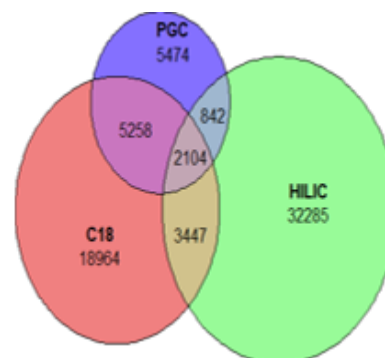


Figure 26 : Comparaison des marqueurs *communs* entre les 3 colonnes en fonction des différentes étapes d'alignement

Seule une faible proportion de marqueurs apparaît finalement commune aux trois colonnes. Tout d'abord, les colonnes C18 et PGC couvraient effectivement les mêmes gammes de m/z et DT, cependant, les marqueurs communs à ces deux colonnes ne correspondaient qu'à 50 % du nombre total de marqueurs pour la PGC, et encore moins pour la C18. En ce qui concerne la HILIC, il s'agit de la colonne qui a retenu le plus de marqueurs différents (*i.e.*, *spécifiques*, non retrouvés dans les autres colonnes). Ceci est en accord avec les différentes distributions de m/z et DT observées pour cette colonne (Figure 24). Cependant, il est important de rappeler que les phases mobiles employées pour la séparation en HILIC contiennent un tampon ammonium pouvant conduire à la formation d'adduits différents (en particulier NH_4^+), ce qui pourrait également expliquer, en partie, la différence observée. En revanche, la

différence de masse moyenne observée entre la HILIC et les deux autres colonnes était significativement plus élevée que la masse de l'ion NH_4^+ , ce qui signifie que cette seule hypothèse n'est probablement pas suffisante pour expliquer les différences observées. La HILIC permet donc effectivement la rétention de molécules ayant des masses globalement plus élevées.

Cette approche, bien que rapide, ne permet d'avoir qu'un premier aperçu approximatif des différences entre colonnes et de leurs marqueurs *communs*. En effet, en comparant les marqueurs simplement sur la base de leur m/z et leur DT, le fait que les phases mobiles puissent induire des adduits différents peut entraîner une sous-estimation du nombre de composés communs à plusieurs colonnes. A l'inverse, le fait que des marqueurs possèdent un m/z et un DT similaires après séparation sur deux colonnes différentes, n'est pas suffisant pour assurer sans ambiguïté que ceux-ci proviennent de la même molécule. Ceci est notamment confirmé par l'alignement de plusieurs marqueurs à partir d'un même marqueur. Même en ne considérant qu'une seule colonne, des marqueurs de TR différents (voire même de spectres de masse différents) peuvent posséder des valeurs de m/z et DT identiques et donc être alignés. Les TR ne pouvant pas être alignés entre ces différentes colonnes puisque les mécanismes de rétention mis en jeu sont différents, cette propriété ne pouvait pas être utilisée pour l'alignement. En ce qui concerne les spectres de masse, le problème résidait dans le fait que leur extraction ne pouvait pas être effectuée de manière automatique et qu'ils ne pouvaient donc pas être utilisés pour une étape supplémentaire d'alignement. La méthode d'alignement utilisée a toutefois permis de donner une première idée de la proportion de marqueurs *communs* ou non, et de mettre en évidence l'apport de la mobilité ionique pour discerner des marqueurs *spécifiques* de certaines colonnes grâce à leur DT.

Une comparaison rapide a également été effectuée entre les deux modes d'ionisation. Il est ainsi intéressant de remarquer que le nombre de marqueurs observé en ESI+ est toujours supérieur au nombre de marqueurs observé en ESI-. Cette observation est cohérente avec les données retrouvées dans la littérature pour ce type d'échantillon puisque certains auteurs choisissent de se limiter à l'analyse dans le mode ESI+ (Köppe et al., 2020; Peter et al., 2019; Schollée et al., 2021; Schulze et al., 2021).

IV.2. Comparaison des méthodes de préparation

IV.2.1. Comparaison des techniques de préparation

Lors de ce travail, **quatre méthodes différentes de préparation d'échantillons ont été comparées à partir d'un même échantillon de rejet de STEU**. Une SPE, méthode classique pour la préparation d'échantillon aqueux, a été réalisée sur 8 phases différentes. Le même échantillon a aussi été fractionné successivement sur trois résines macroporeuses ou séparé par HPLC-SEC. Enfin, l'échantillon aqueux ayant simplement subi une filtration a également été injecté directement pour comparaison. Différents critères ont été pris en compte afin d'évaluer l'efficacité de ces méthodes. Tout d'abord, en termes de critères pratiques : le volume d'échantillon ou de solvant nécessaire, le temps nécessaire pour mettre en place la méthode (*e.g.*, conditionnement, lavage, étalonnage...) et le temps effectif pour réaliser l'extraction ont été comparés, puis la diversité des données obtenue ainsi que leur interprétation a été étudiée.

Sur le plan pratique, la SPE semble être avantageuse à tout point de vue. En effet, malgré l'utilisation de 8 phases différentes, la consommation à la fois en termes de volume d'échantillon (8 L) qu'en volume de solvant (environ 400 mL) nécessaires ainsi que les temps de préparation et de manipulation sont raisonnables. La méthode de fractionnement sur résine est, quant à elle, celle qui possède le plus d'inconvénients concernant l'expérimentation. En effet, les résines nécessitent un long processus de lavage et de conditionnement avant de pouvoir être utilisées (de l'ordre de 3 semaines). Bien que ces étapes ne soient nécessaires que lors de la première utilisation, entre deux échantillons différents, un rinçage doit être réalisé afin d'éviter toute contamination croisée et les résines doivent également être conservées dans un solvant organique lorsqu'elles ne sont pas utilisées de façon prolongée. D'autre part, l'extraction nécessite un large volume d'échantillon et induit donc un temps de manipulation plus long. En ce qui concerne le fractionnement sur SEC, cette méthode ne requiert pas un temps de préparation ou de manipulation très long, et le volume d'échantillon nécessaire est également raisonnable. Cependant, le principal désavantage réside dans la dilution de l'échantillon dans la phase mobile, compensée par une étape préalable de concentration (15 fois, sans précipitation) par évaporateur rotatif. Enfin, l'injection directe est très avantageuse aussi bien en termes de volume d'échantillon

nécessaire que de temps de préparation et manipulation. En revanche, l'échantillon est aussi relativement peu concentré.

En ce qui concerne la richesse des données, outre le fait que les fractions soient peu concentrées pour la méthode par SEC, l'analyse par HRMS de celles-ci n'a pas permis de les différencier significativement les unes des autres en termes de gamme de masse, ce qui était pourtant attendu de cette méthode de préparation. De la même manière que les fractions de SEC, comparativement aux autres méthodes de préparation, peu de données (en termes de nombre et aire des marqueurs détectés) ont pu être tirées de l'injection directe. Cependant, cette dernière possède l'avantage, par rapport au fractionnement par SEC, de garantir l'intégrité de la phase dissoute de l'échantillon. Ainsi, cette méthode est intéressante puisqu'elle permet de distinguer quelle partie de l'échantillon peut être perdue par les autres méthodes.

Une comparaison a été effectuée entre les marqueurs obtenus après analyse des échantillons obtenus sur les 3 méthodes de fractionnement sur résines, de SPE sur 8 cartouches et d'injection directe. Le fractionnement par HPLC-SEC n'a pas été pris en compte pour des raisons pratiques. En effet, cette méthode a fourni très peu de marqueurs et ceux-ci étaient de très faible intensité. Les inclure aurait nécessité un temps de retraitement supplémentaire qui aurait permis, certes, d'apporter quelques informations complémentaires (*i.e.*, différences par rapport aux autres techniques). Elles se seraient avérées difficiles à interpréter en raison de l'absence de la discrimination en taille des fractions de la très faible intensité des marqueurs, qui engendre des spectres de masse peu caractéristiques, évoqués précédemment. En ce qui concerne la SPE et les résines, les marqueurs ont été considérés indépendamment des phases étudiées et donc tous regroupés dans une même *marker table*. Par exemple pour la SPE, même si un marqueur n'est présent que dans 1 extrait sur les 8 étudiés, il a été comptabilisé. Les marqueurs *communs* et *spécifiques* à chaque méthode de préparation sont présentés en Figure 27.

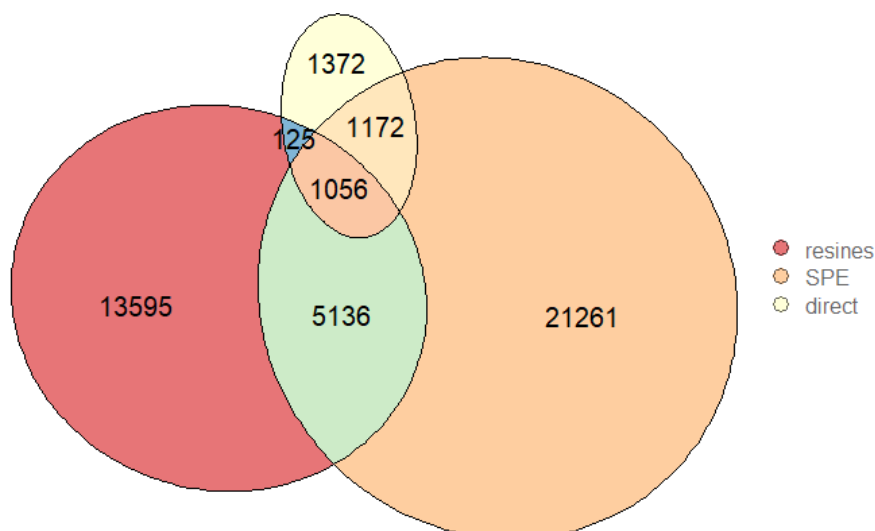


Figure 27 : Nombre de marqueurs issus des différentes méthodes de préparation (analyse sur colonne C18 en mode ESI+)

Comme attendu, le set de données obtenu par injection directe contient beaucoup moins de marqueurs que les deux autres approches. Cependant, il est intéressant de noter que plus d'un tiers des marqueurs obtenus par cette méthode lui sont *spécifiques*. En d'autres termes, malgré la combinaison de onze phases différentes, ces marqueurs, pourtant assez concentrés pour être détectés sans étape de concentration, n'ont pas pu être retenus par les différents supports utilisés. En ce qui concerne le fractionnement par résine, il a généré légèrement moins de marqueurs que la SPE. Avec 31 % de marqueurs *spécifiques*, cette technique apporte des informations complémentaires aux autres méthodes, la résine XAD 4 permettant de retenir le plus grand nombre de marqueurs. Cependant, cette méthode est à la fois consommatrice en termes de temps, d'échantillon et de matériel nécessaires. En comparaison, la combinaison des huit phases SPE permet d'obtenir 49 % de marqueurs *spécifiques* sans présenter autant d'inconvénients de manipulation.

Par la suite, une caractérisation globale des marqueurs obtenus par chaque méthode a été effectuée afin d'avoir une idée des informations récupérées par celles-ci (Tableau 32). Globalement, les composés retenus par le fractionnement sur résines présentent des masses moyennes plus élevées que l'extraction sur phase solide. En ce qui concerne le TR moyen, celui obtenu pour le fractionnement sur résine est également plus grand que pour les deux autres méthodes, et, la considération des marqueurs *spécifiques* à chaque méthode accentue d'autant plus cette différence. Ces observations semblent montrer que la méthode basée sur les résines permet de retenir des composés globalement plus gros et moins polaires. En

particulier, elle se distingue des deux autres méthodes principalement par la polarité puisque les marqueurs uniques retenus présentent cette différence. Un des objectifs de cette étude étant d'étudier plus particulièrement les composés polaires qui sont encore peu caractérisés, la méthode de fractionnement n'a pas été retenue pour la suite puisqu'elle présente beaucoup de désavantages en termes pratiques et n'apporte que peu d'information concernant les composés polaires. Ces observations ont ainsi amené à continuer l'étude en se concentrant sur la méthode de SPE, en comparant 8 phases d'extraction différentes, et l'injection directe.

Tableau 32 : Caractérisation globales de trois méthodes de préparation

	Tous les marqueurs détectés			Marqueurs <i>spécifiques</i>		
	<i>m/z</i> moyen	TR moyen (min)	DT moyen (ms)	<i>m/z</i> moyen	TR moyen (min)	DT moyen (ms)
SPE	379,67	9,37	5,68	423,87	8,01	5,84
Résines	423,10	14,67	6,10	453,27	14,32	6,26
directe	431,99	9,92	5,61	492,92	4,58	4,97

IV.2.2. Comparaison de différents supports pour l'extraction sur phase solide

Afin de comparer les différentes phases sélectionnées pour la SPE, la caractérisation globale évoquée dans la partie A – IV.3.2 a été appliquée. Brièvement, le nombre de marqueurs, l'aire totale du signal ainsi que les moyennes des valeurs de *m/z*, de DT et de TR ont été calculés pour différents groupes de marqueurs. Les différents extraits de SPE ont donc été comparés les uns aux autres et les marqueurs de chaque extrait ont été classés dans les catégories suivantes (Figure 28) :

- **Marqueurs *omniprésents*.** Cette catégorie est définie par les marqueurs qui ont été retrouvés dans tous les extraits analysés. Ils ne permettent pas de différencier les extraits entre eux, mais peuvent apporter une information sur le type de composés qui sont facilement analysables.
- **Marqueurs *communs*.** Cette catégorie est définie par des marqueurs qui sont retrouvés dans au moins un extrait différent de l'extrait considéré. Ceux-ci peuvent permettre d'établir des liens entre des cartouches.

- **Marqueurs uniques.** Cette catégorie est définie par des marqueurs qui ne sont uniquement retrouvés que dans l'extrait considéré. Ils sont donc les plus représentatifs des particularités de la phase étudiée.

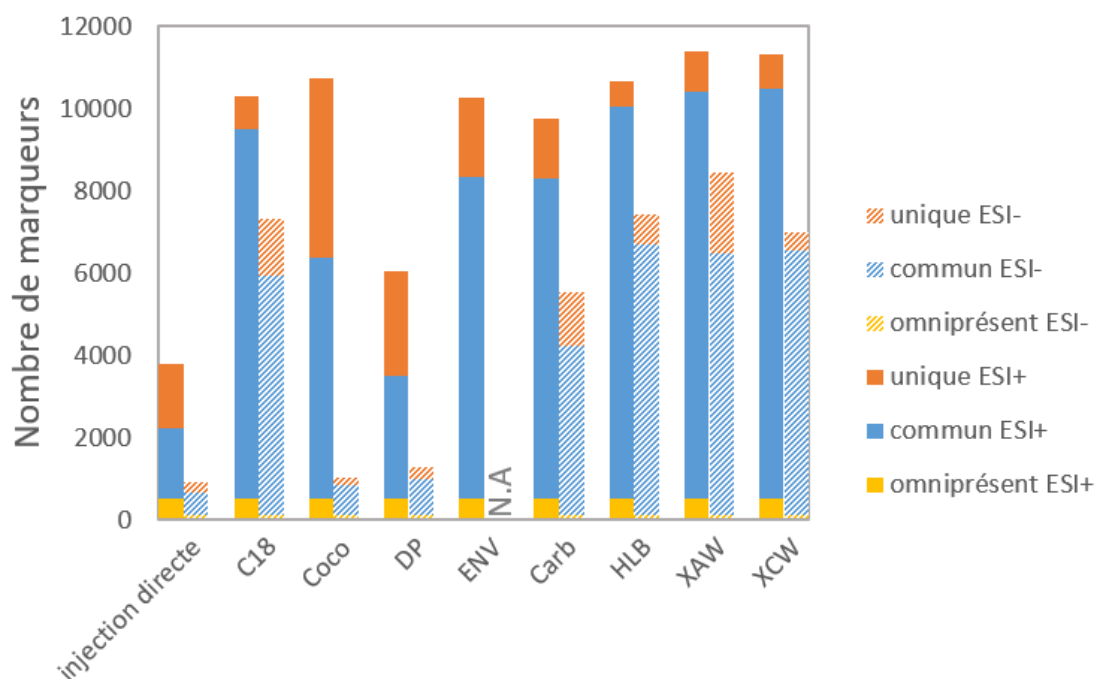


Figure 28 : Classification des marqueurs pour comparer les différents extraits d'un même échantillon, après séparation sur colonne C18

De façon cohérente avec ce qui a été observé dans la partie précédente, l'injection directe permet d'analyser moins de composés que les extraits par SPE. Au contraire, la cartouche XAW permet de retenir le plus de marqueurs dans les deux modes d'ionisation. En ESI+, près de la moitié des marqueurs des cartouches DP et Coco sont *uniques*, ce qui démontre la spécificité de ces deux phases SPE (décrites comme étant dédiées aux composés polaires, Tableau 25). Pour toutes les autres cartouches, la proportion de marqueurs *uniques* varie entre 5 et 42 %. La proportion de marqueurs *omniprésents* (c'est-à-dire que l'on retrouve toujours quel que soit l'extrait) était très faible, représentant seulement 5 à 13 % du nombre total de marqueurs selon le type d'extrait. Il est intéressant de noter que l'injection directe de l'échantillon (c'est-à-dire sans aucune étape d'extraction) présente également une grande proportion (41 %) de marqueurs *uniques*, bien que le nombre total de marqueurs soit considérablement plus faible.

Pour aller plus loin, tous les marqueurs détectés dans un extrait considéré ont été pris en compte, indépendamment de leur catégorie pour application de la caractérisation globale. Ensuite, seuls les marqueurs classés dans la catégorie *uniques* ont été considérés afin de

déterminer les propriétés des marqueurs caractéristiques de ces supports (Tableau 33). Dans un premier temps, il est intéressant de remarquer que les différences en termes de m/z entre les extraits sont plus visibles en ESI+ alors que pour les TR, celles-ci sont plus marquées en ESI- quels que soient les types de marqueurs considérés (*i.e.*, totaux ou *uniques*). **En comparant la globalité des données, il semble que les molécules analysables en ESI+ soient globalement plus petites et moins polaires que celles observées après analyse en ESI-.**

Lorsque tous les marqueurs de chaque extrait sont pris en compte, l'injection directe contient en moyenne des molécules de m/z plus élevés que les extraits, dans les deux modes d'ionisation. Par rapport aux valeurs moyennes de m/z , RT et DT sur la colonne chromatographique C18 en ESI+ (433,73 m/z , 9,45 min et 5,69 ms, respectivement), ces valeurs pour les marqueurs uniques détectés après injection directe étaient de 489,16 m/z , 4,90 min 4,98 ms, qui correspondent donc à des composés plutôt polaires et avec une structure relativement compacte (car les valeurs de DT sont faibles proportionnellement par comparaison aux valeurs m/z). Cela peut se traduire par le fait qu'une partie des molécules de m/z élevées sont relativement concentrées (car observables sans concentration) ou ont un fort potentiel d'ionisation. Cependant, celles-ci ne sont pas bien retenues ou, au contraire, trop bien retenues par les différentes phases de SPE rendant leur analyse ultérieure plus difficile. A l'inverse, cette observation peut être également liée au fait que l'étape de concentration par les phases SPE permet d'analyser plus de composés qui n'étaient pas visibles initialement par injection directe. Ceux-ci ayant potentiellement des m/z plus faibles, ils peuvent donc être responsables du décalage de la moyenne des m/z vers une valeur plus faible comparée à celle obtenue pour l'injection directe. Par ailleurs, les marqueurs uniques des différents extraits (HLB et XAW en mode ESI+ et DP en mode ESI-) présentent une moyenne de m/z bien supérieure à la moyenne des m/z de l'injection directe, montrant ainsi la capacité de ces cartouches à retenir des composés de masses élevées, qui de plus, ne sont retrouvés dans aucun autre extrait. Ainsi, la seconde hypothèse concernant l'injection directe semble plus cohérente avec ces observations puisque les cartouches semblent être capables de retenir des composés avec des valeurs élevées de m/z .

L'injection directe en ESI- montre des moyennes de TR plus faibles. Cette observation est encore plus marquée lorsque seuls les marqueurs *spécifiques* sont considérés, laissant ainsi penser que les molécules les plus polaires sont globalement moins bien retenues par les

différents supports. Contrairement à ce qui a été observé pour les valeurs de m/z , le décalage de TR vers des valeurs plus faibles ne peut pas se traduire uniquement par le fait que les cartouches retiendraient des composés de TR plus élevés. En effet, dans ce cas précis, les extraits présentaient des moyennes de TR, à la fois pour tous les marqueurs et lorsque seuls les marqueurs *uniques* sont considérés, très éloignées de celles obtenues pour l'injection directe, ce qui n'était pas le cas pour les m/z . En ESI+, l'extrait issu de la cartouche Carb a présenté ce même comportement, mais la différence était moins marquée qu'en mode ESI-. Au contraire, l'extrait issu de la cartouche Coco montre des moyennes de TR plus élevées que les autres extraits lorsque tous les marqueurs détectés sont considérés dans les deux modes d'ionisation.

Tableau 33 : Caractérisation globale du total des marqueurs détectés et des marqueurs *uniques* à chaque extrait d'un même échantillon après séparation sur colonne C18. Les valeurs extrêmes, par propriété, ont été mises en évidence pour faciliter la lecture du tableau : Vert - moyennes les plus grandes, Rouge - moyennes les plus petites

	ESI+						ESI-					
	Marqueurs détectés (total)			Marqueurs <i>uniques</i>			Marqueurs détectés (total)			Marqueurs <i>uniques</i>		
	<i>m/z</i> moyen	TR moyen (min)	DT moyen (ms)	<i>m/z</i> moyen	TR moyen (min)	DT moyen (ms)	<i>m/z</i> moyen	TR moyen (min)	DT moyen (ms)	<i>m/z</i> moyen	TR moyen (min)	DT moyen (ms)
C18	387,7	9,5	5,8	450,1	10,0	6,2	422,6	9,2	6,3	473,5	9,7	6,6
Coco	360,2	10,9	5,5	334,3	9,9	5,1	417,9	10,3	6,2	474,5	8,5	6,4
Carb	341,6	7,7	5,4	337,8	4,6	5,3	399,3	7,9	6,1	427,3	6,2	6,2
DP	367,7	10,5	5,5	371,6	7,5	5,2	457,1	7,7	5,8	614,5	3,5	6,0
ENV+	346,9	9,5	5,6	325,8	8,7	5,6	-	-	-	-	-	-
HLB	429,1	9,6	6,0	612,3	12,0	7,1	397,7	8,6	6,2	435,8	7,9	6,3
Injection directe	436,4	9,9	5,6	489,2	4,9	5,0	474,6	5,2	5,1	496,6	2,8	4,6
XAW	423,2	9,0	6,0	550,3	8,1	6,7	373,9	7,7	5,9	338,8	4,7	5,5
XCW	377,1	8,6	5,7	364,8	6,4	5,5	394,9	8,3	6,1	467,4	4,6	5,7
Moyenne globale de la série	385,5	9,5	5,7	426,2	8,0	5,8	417,3	8,1	6,0	466,0	6,0	5,9

Afin de vérifier les différentes hypothèses émises, les distributions des valeurs de m/z et de TR de l'injection directe ont été comparées à celles des extraits HLB en mode ESI+ et DP en mode ESI-. L'extrait HLB possédait la moyenne la plus élevée des valeurs de m/z pour les marqueurs *uniques* (Figure 29a) et l'extrait DP la seconde moyenne de TR la moins élevée (Figure 29b).

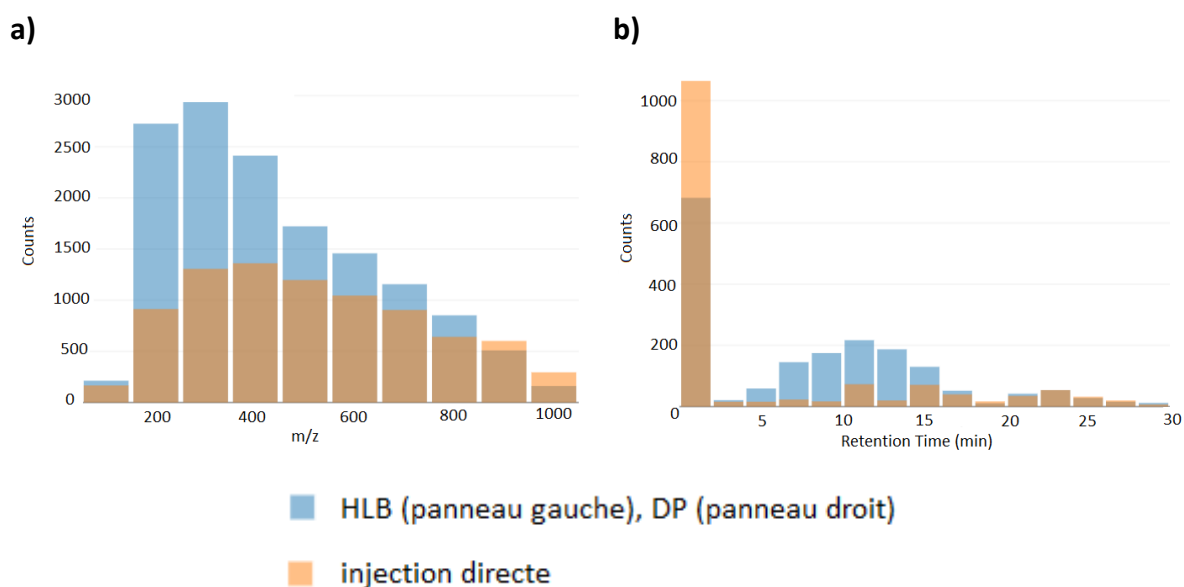


Figure 29 : Comparaison des distributions de a) m/z entre l'injection directe et l'extrait HLB en ESI+, b) TR entre l'injection directe et l'extrait DP en ESI-

Les distributions des valeurs de m/z montrent bien un nombre plus important de marqueurs ayant un m/z entre 200 et 500 pour l'extrait HLB par rapport à l'injection directe. Cette observation explique le décalage de la moyenne de m/z de cet extrait vers des valeurs plus faibles bien que ce dernier possède presque autant de marqueurs avec des valeurs de m/z entre 800 et 1000. En revanche, pour les distributions de TR, une proportion de marqueurs bien plus importante est observée dans les zones de temps de rétention faibles pour l'injection directe par rapport à l'extrait DP. Ces observations confirment ainsi les hypothèses émises précédemment montrant que les molécules analysables par injection directe ont tendance à se différencier des extraits par leur plus grande polarité plutôt que par leur masse.

Afin de vérifier si ces différences globales étaient statistiquement significatives, une ACP a été réalisée en prenant en compte tous les marqueurs détectés dans chaque extrait et dans l'injection directe, injectés sur la colonne C18 en ESI+ (Figure 30). Trois groupes principaux ont été distingués. Un premier, formé des extraits issus des cartouches ENV, Carb, XAW, XCW, C18

et HLB, un deuxième regroupant l'extrait de la cartouche DP et l'injection directe, et la cartouche Coco formant un groupe à elle seule. Sur la base des observations réalisées précédemment (Figure 28), le deuxième groupe était caractérisé par des extraits avec le nombre le plus faible de marqueurs détectés alors que le troisième groupe était associé à une plus grande proportion de marqueurs *spécifiques*. Par conséquent, la première composante principale (PC1, porteuse de 24,7 % de la variabilité du set de données) semble être liée à la spécificité des marqueurs, alors que la PC2 (responsable de 17,9 % de la variabilité) est plus influencée par le nombre total de marqueurs détectés. Les PC3 et PC4, chacune responsable de 8 % de la variabilité, permettaient de séparer les extraits ENV et Carb, respectivement, des autres extraits (Annexe 3A). Ces deux extraits sont difficilement distinguables des autres par les proportions de marqueurs dans chaque catégorie (*omniprésents, communs, uniques*) (Figure 28).

En considérant les résultats de la caractérisation globale (Tableau 33), l'ENV se distinguait des autres cartouches par une moyenne de m/z globalement plus faible alors que la Carb retenait des composés plus petits mais également plus polaires (plus faible moyenne en m/z et TR). D'autre part, l'extrait HLB, qui constitue l'extrême opposé de l'extrait Carb sur la PC4, possède des marqueurs *uniques* avec les moyennes de m/z et TR les plus élevées, confirmant ainsi que la PC4 est influencée par ces propriétés.

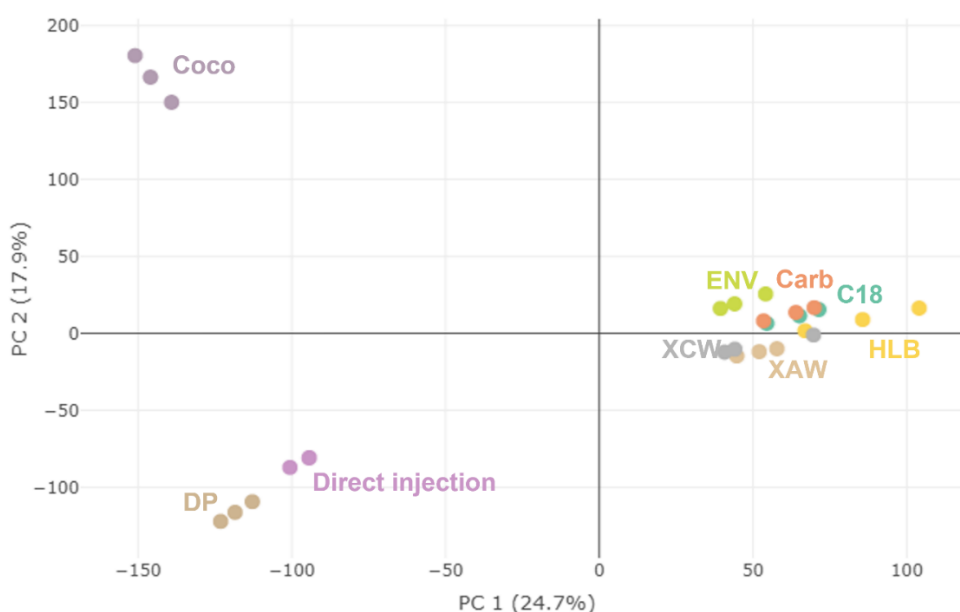


Figure 30 : ACP (Scores plot) obtenue à partir des marqueurs issus de l'analyse HRMS d'effluent de STEU après SPE sur différentes phases solides (HLB, C18, XAW, XCW, ENV, Carb, Coco, DP) ou injection directe et analyse après séparation sur colonne C18 en ESI+.

De la même manière, la classification des marqueurs en 3 catégories pour distinguer les méthodes de préparation a été également réalisée en s'appuyant sur les analyses issues des séparations sur PGC et HILIC (Figure 31). En classant les cartouches par nombre de marqueurs détectés, les extraits possédant les nombres minimum et maximum de marqueurs changent en fonction de la colonne chromatographique étudiée. Il en va de même pour la proportion de marqueurs *spécifiques* à chaque extrait. En revanche, en combinant les résultats obtenus sur les trois colonnes chromatographiques (Acquity UPLC BEH C18, Hypercarb PGC, Acquity UPLC BEH HILIC), il ressort que la XAW et la HLB permettent globalement de retenir le plus de marqueurs. La phase Coco, malgré un bon nombre de marqueurs sur la colonne C18, et la DP avec pourtant un nombre élevé sur la HILIC, regroupent globalement le moins de marqueurs en considérant les trois colonnes. Ainsi, alors que les deux premières cartouches (HLB et XAW) semblent pouvoir être utilisées en combinaison avec les trois méthodes d'analyse étudiées ici, pour les deux dernières (Coco et DP), il semble exister une méthode privilégiée d'analyse permettant de tirer un maximum d'informations issues de leur extrait. De façon intéressante, la HLB, généralement considérée comme étant une cartouche *universelle*, semble en effet remplir son rôle. La colonne chromatographique C18, également qualifiée d'*universelle*, engendre également le moins de variation entre le nombre de marqueurs détectés sur chaque extrait, par rapport aux deux autres colonnes étudiées.

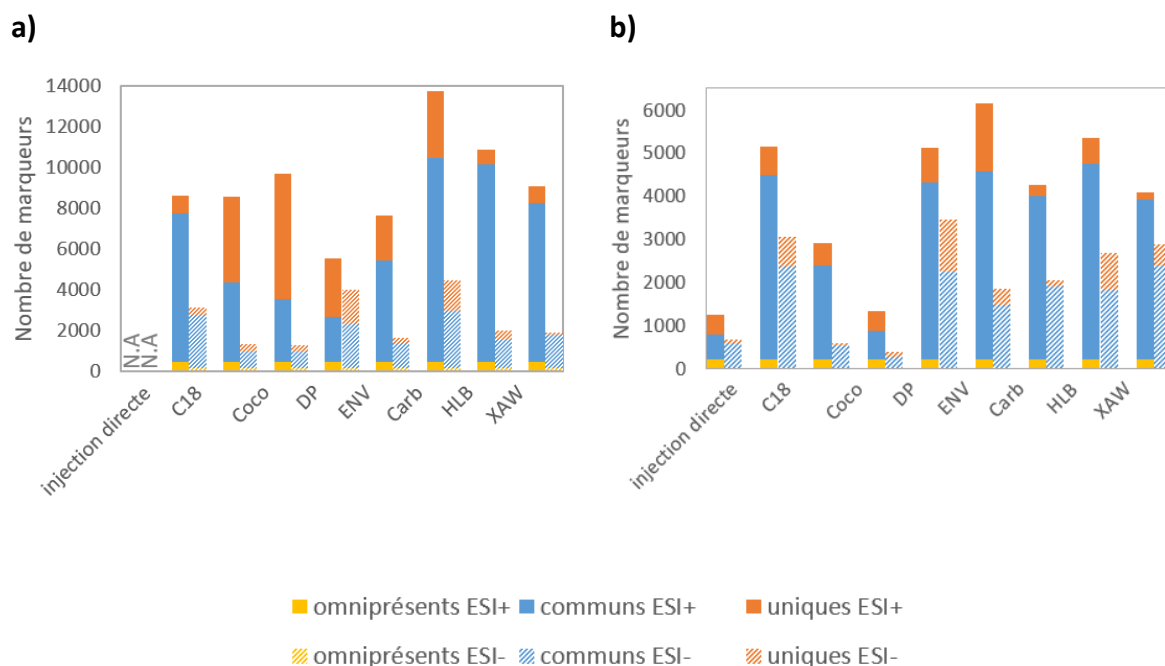


Figure 31 : Classification des marqueurs pour comparer les différents extraits d'un même échantillon après séparation sur (a) HILIC et (b) PGC

IV.3. Influence du type d'analyseurs

Une comparaison a été effectuée entre l'UPLC-IMS-QToF et un FT-ICR-MS. Cette comparaison s'appuyait principalement sur le nombre et la nature des marqueurs détectés mais aussi sur la proportion de formules moléculaires attribuées.

Comme pour l'étude des marqueurs analysés dans chaque extrait après séparation sur les trois colonnes, les marqueurs obtenus après analyse par FT-ICR ont été classés dans les trois catégories définies précédemment (*omniprésents*, *communs* ou *uniques*). En termes de nombre de marqueurs détectés, les méthodes n'ont pas été comparées car pour la FT-ICR, ces résultats étaient en partie biaisés. En effet pour certains extraits (XAW, XCW, Carb et injection directe) les analyses n'ont pas pu être répliquées (une seule injection). Ces extraits contenaient des quantités trop importantes de sels, pouvant limiter la reproductibilité des analyses et donc les comparaisons entre extraits. Ainsi, afin de comparer les quatre méthodes d'analyse (UPLC-IMS-QToF sur trois colonnes chromatographiques et FT-ICR-MS), seule la proportion de marqueurs *uniques* relevés dans chaque extrait analysé a été considérée (Figure 32).

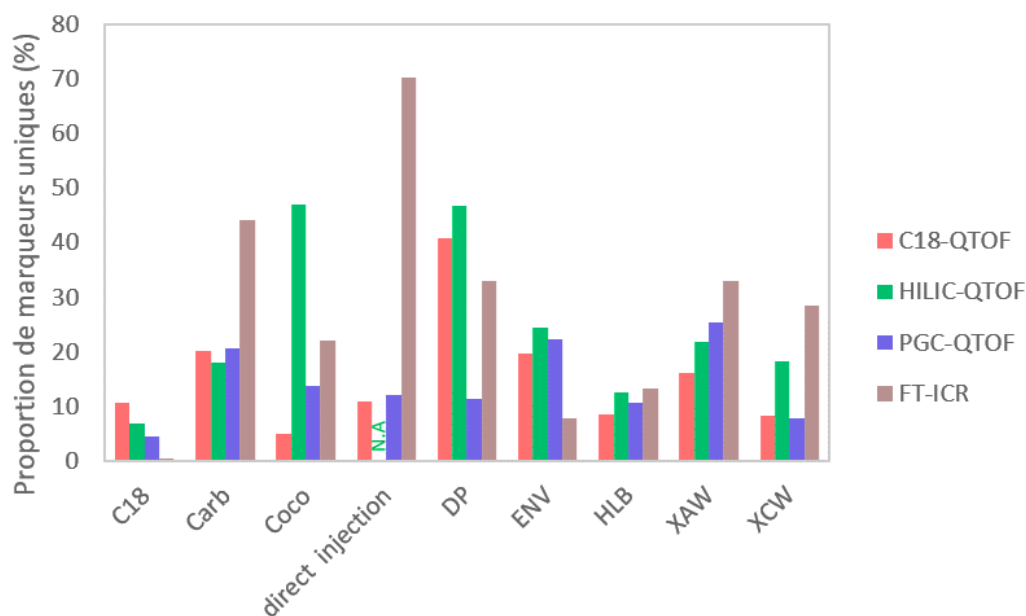


Figure 32 : Comparaison des proportions de marqueurs *uniques* entre les différents extraits pour les quatre méthodes d'analyse

Globalement il est possible de distinguer trois groupes. Le premier groupe (ENV, HLB et XAW) est défini par une variation de la proportion de marqueurs *uniques* inférieure à 35 % entre les différentes méthodes d'analyse (19 %, 11 % et 24 % en moyenne pour les phases ENV, HLB et

XAW, respectivement). Pour ces extraits, établir une proportion moyenne de marqueurs *uniques* semble avoir un sens puisque cette proportion était relativement similaire pour les quatre techniques d'analyse utilisées. De façon intéressante, les phases HLB et XAW ont également montré les plus faibles variations du nombre total de marqueurs (partie B – IV.2.2) et la phase ENV est également parfois considérée comme phase *universelle* dans la littérature (Fontanals et al., 2005; Kovalova et al., 2009).

Le deuxième groupe (DP, Carb, XCW, C18) est défini par une variation de la proportion de marqueurs *uniques* comprise entre 35 et 70 % entre les différentes méthodes d'analyse. Pour ces extraits, une des quatre techniques analytiques a produit une proportion de marqueurs *uniques* très différente des proportions obtenues par les autres méthodes. Pour les phases C18 et Carb, il s'agit de l'analyse par FT-ICR qui permet d'obtenir un nombre de marqueurs *uniques* bien plus important pour la Carb, mais moins important pour la C18. En ce qui concerne la cartouche Carb, il est utile de rappeler que les marqueurs n'ont pas pu être filtrés entre répliquats ce qui explique probablement un plus grand nombre de marqueurs retrouvés dans cet extrait et, par extension, une proportion potentiellement plus élevée de marqueurs *uniques*. Cependant, ce point ne constitue pas la seule explication puisque d'autres extraits (XAW, XCW), présentant la même situation, n'ont pas mené au même comportement. Une autre explication peut résider dans le fait que cette cartouche, d'après ses spécifications, est capable de retenir des molécules très polaires qui peuvent potentiellement être mal séparées par les méthodes chromatographiques. Dans le cas de la DP, l'analyse par PGC-QToF, a produit moins de marqueurs différents que les autres méthodes. Cette observation peut être reliée aux conformations des molécules. En effet, en considérant les deux autres colonnes (C18 et HILIC), cet extrait donne des DT moyens parmi les plus élevés dans les deux cas. La PGC étant décrite comme retenant mieux les composés ayant une conformation plus planaire, cette dernière peut être moins adaptée pour des composés retenus par la cartouche DP, qui semblent avoir des conformations plus complexes (entraînant des valeurs de DT plus grandes). Enfin, dans le cas de la XCW, il est possible de distinguer deux groupes de méthodes analytiques qui produisent des résultats proches : la C18-QToF et la PGC-QToF d'une part, la HILIC-QToF et la FT-ICR d'autre part.

Le troisième groupe (Coco et injection directe) est défini par une variation de la proportion de marqueurs *uniques* supérieure à 70 % entre les différentes méthodes d'analyse. Pour ce

dernier groupe, les comportements entre les deux types d'échantillons sont différents. En ce qui concerne l'injection directe, cette grande variation est due à l'analyse par FT-ICR qui produit une proportion de marqueurs *uniques* bien supérieure aux analyses par QToF. Cette observation s'explique de la même manière que pour l'extrait issu de la cartouche Carb dans le paragraphe précédent, avec un effet plus marqué dans le cas présent. Pour la cartouche Coco, la différence n'est pas aussi marquée que précédemment, mais les quatre méthodes d'analyse employées donnent des proportions complètement différentes les unes des autres. A ce stade et avec les informations disponibles, ce comportement n'a pas pu être expliqué. Cependant, cela met en évidence que le choix de la méthode analytique utilisée pour l'analyse de l'extrait issu de cette cartouche est d'une grande importance puisque les résultats sont très variables d'une méthode à l'autre. Lorsque la cartouche Coco est analysée par C18-QToF, elle semble n'apporter que peu d'informations complémentaires par rapport aux autres phases alors qu'en HILIC-QToF, presque la moitié des informations qu'elle apporte ne sont pas obtenues par les autres extraits. La HILIC étant décrite comme adaptée à l'analyse de composés plus polaires, cette propriété peut potentiellement être reliée aux molécules retenues par la cartouche Coco.

Afin de caractériser plus précisément la nature des marqueurs retenus, les méthodes d'analyse ont également été comparées en prenant en compte les rapports m/z détectés. Les moyennes de ces derniers, pour tous les marqueurs présents dans un extrait donné, ainsi que les moyennes de m/z des marqueurs *uniques* ont été calculées (Tableau 34). Globalement les m/z moyens obtenus par FT-ICR sont les plus faibles, suivis de l'analyse par PGC-QToF puis C18-QToF et enfin la HILIC-QToF qui permet d'obtenir les moyennes de m/z les plus élevées. En termes d'extrait, l'extrait ENV permet de retenir des composés globalement plus petits, peu importe la méthode d'analyse employée. Concernant les m/z les plus élevés obtenus, ils sont plus dépendants de la méthode d'analyse considérée. Cependant, si seuls les marqueurs *uniques* sont considérés, l'extrait de C18 présente systématiquement des valeurs de m/z moyens parmi les 3 plus élevées. En revanche, lorsque tous les marqueurs sont pris en compte, la phase DP et l'injection directe donnent des m/z moyens parmi les plus élevés pour les méthodes chromatographiques couplées au Q-ToF. En revanche, en FT-ICR, l'injection directe fait partie de extraits qui fournissent les m/z moyens les moins élevés. Pour cette méthode d'analyse, les extraits ayant les m/z moyens les plus élevés sont la XCW et la C18. De la même

manière que lors de la comparaison de la proportion de marqueurs *uniques* détectés par chaque méthode, la variation des moyennes de m/z à travers les méthodes pour chaque extrait a aussi été comparée. Que ce soit pour tous les marqueurs détectés ou lorsque seuls les marqueurs *uniques* sont considérés, les phases XCW et Carb montrent les coefficients de variations entre méthodes analytiques les plus faibles. Au contraire, la phase Coco et l'injection directe donnent les variations les plus fortes. De façon intéressante, ces dernières ont également donné le plus de différence en termes de proportions de marqueurs *uniques* et la phase Coco a aussi donné le plus de différences lorsque le nombre de marqueurs totaux était comparé. En fonction de la méthode employée pour analyser l'extrait issu de cette cartouche, les résultats obtenus peuvent donc être très différents.

Tableau 34 : Comparaison des moyennes de *m/z* totaux et *uniques* obtenus sur chaque cartouche par les différentes méthodes d'analyse en ESI+.

Les valeurs maximale et minimale par méthode d'analyse sont surlignées en vert et rouge respectivement afin de faciliter la lecture.

	Total					Uniques				
	C18-QToF <i>m/z</i> moyen	HILIC- QToF <i>m/z</i> moyen	PGC-QToF <i>m/z</i> moyen	FT-ICR <i>m/z</i> moyen	Coefficient de variation entre méthodes (%)	C18-QToF <i>m/z</i> moyen	HILIC- QToF <i>m/z</i> moyen	PGC- QToF <i>m/z</i> moyen	FT-ICR <i>m/z</i> moyen	Coefficient de variation entre méthodes (%)
C18	422,7	447,1	354,4	332,1	14	467,6	527,9	441,8	370,6	14
Carb	385,5	331,4	332,7	326,4	8	417,0	346,3	359,4	340,5	10
Coco	499,6	549,2	335,5	303,0	29	507,3	598,5	238,8	290,2	42
direct injection	493,0	-	490,9	227,8	38	465,2	-	540,3	233,6	39
DP	442,6	508,9	468,9	288,9	23	452,6	526,0	456,0	322,0	19
ENV	370,8	322,2	302,7	224,6	20	351,7	300,1	283,3	219,8	19
HLB	411,6	459,8	326,7	322,8	18	460,5	517,0	318,6	366,0	22
XAW	418,2	454,9	322,0	326,9	17	508,2	609,5	290,7	373,9	32
XCW	402,6	403,5	341,5	327,6	11	451,2	387,1	349,1	377,4	11
Moyenne	427,4	434,6	363,9	297,8	-	453,5	476,5	346,2	321,6	-

Par la suite, des tentatives d'attributions de formules moléculaires ont été effectuées sur les différentes listes et sont récapitulées dans le Tableau 35. En ce qui concerne les données issues du QToF, les formules moléculaires ont été calculées à l'aide de *Genform*, après conversion des données brutes par *MSConvert* de la même manière que décrit en partie A – III.1.3. Les données issues du FT-ICR-MS ont, quant à elles, été traitées par *Data Analysis* (Bruker), qui a également été utilisé pour le calcul des formules. Le nombre et la nature des atomes autorisés pour les formules étaient les mêmes, cependant pour les données de FT-ICR-MS, la tolérance en masse a été réduite (0,2 ppm contre 5 ppm pour le QToF) pour correspondre au type de données traitées.

Tableau 35 : Comparaison des proportions de formules moléculaires calculées après analyse par différentes méthodes en ESI+

Méthode d'analyse	Nombre de marqueurs	Proportion d'attribution (%)
UPLC(C18)-IMS-QToF	35 340	1
UPLC(PGC)-IMS-QToF	12 208	11
UPLC(HILIC)-IMS-QToF	40 687	1
FT-ICR-MS	48 455	73

De façon peu surprenante, l'analyse par FT-ICR-MS permet de calculer un nombre considérablement plus grand de formules moléculaires comparée aux analyses par UPLC-IMS-QToF. Cela provient de la très haute résolution de la FT-ICR-MS, qui permet d'obtenir à la fois des masses plus précises, et de séparer des pics de masses très proches, engendrant également une signature isotopique plus précise. Par comparaison, les analyses par QToF ont été réalisées à une résolution moyenne de $R_{\text{QToF}} = 30\,000$ alors que les analyses en FT-ICR ont pu être effectuées à $R_{\text{FT-ICR}} > 1\,000\,000$. Par ailleurs, la méthode de calcul de formules moléculaires utilisée à partir des données de QToF, basée sur des logiciels libres (partie A – IV.2), ne permet pas, à ce stade, de tirer parti de l'avantage donné par l'IMS en termes de « propreté » des données spectrales (c'est-à-dire de séparation des spectres de molécules co-éluées). Il sera intéressant d'étudier si la prise en compte de données spectrales plus spécifiques de chaque marqueur, grâce à l'IMS, permettra d'obtenir une plus grande proportion d'attributions de formules moléculaires.

IV.4. Caractérisation de l'espace chimique couvert par ces méthodes

Les précédentes comparaisons globales ayant démontré de grandes différences entre les méthodes, aussi bien d'analyse que de préparation d'échantillon, une caractérisation plus spécifique du type de molécules pouvant être analysées a été menée avec une approche par SS sur une liste de 376 molécules communément retrouvées dans les rejets urbains. Des informations telles que la formule moléculaire, la structure, le log Kow, le pKa et la CCS de ces molécules ont été enregistrées dans la librairie lorsqu'elles étaient disponibles dans des bases de données en ligne. De cette manière, des informations concernant la nature et le type de molécules analysables ont pu être récupérées afin d'établir des gammes de propriétés couvertes par les méthodes.

IV.4.1. Comparaison en SS des colonnes

En termes de colonnes chromatographiques, 36 % des molécules de la liste ont été retrouvées en combinant les trois colonnes en ESI+. Celle qui permet de retenir le plus de suspect est la PGC avec 25 %, suivie de la C18 avec 22 % et enfin la HILIC avec seulement 7 % des molécules retrouvées. Cette observation est en accord avec une étude récente qui a comparé quatre modes de chromatographie pour la recherche de suspect dans des effluents de STEU montrant que la HILIC était également moins efficace pour retenir ce type de composés (Castro et al., 2021). La Figure 33 montre les différences concernant les proportions de suspects retrouvés sur chaque colonne.

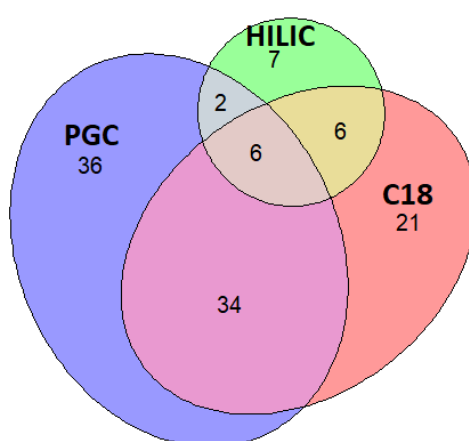


Figure 33 : Diagrammes d'Euler donnant le nombre de composés retrouvés dans les trois colonnes en SS

Il est tout d'abord intéressant de remarquer que les composés chlorés sont majoritairement retrouvés sur la colonne C18, alors que les molécules retrouvées sur HILIC sont globalement composées d'un nombre plus important d'atomes d'oxygène.

Concernant les propriétés des molécules retenues, les tendances observées en NTS par les caractérisations globales ont pu être confirmées. En effet, les molécules retenues sur la HILIC possèdent des masses molaires plus élevées que celles retenues sur les deux autres colonnes, ce qui est d'autant plus vrai pour les composés qui sont spécifiquement retenus sur cette colonne. En revanche, la caractérisation globale en NTS des colonnes C18 et PGC n'avait pas permis de les distinguer. Ici, en considérant des molécules connues, la PGC semble retenir des composés de masses molaires moins élevées que la C18. Un parallèle a pu être établi entre les moyennes de masse molaires et les moyennes de CCS, la colonne PGC retenant également les molécules possédant des valeurs de CCS plus petites.

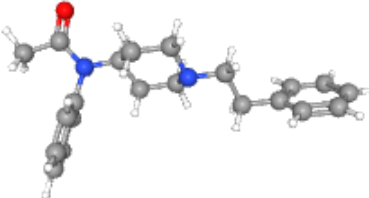
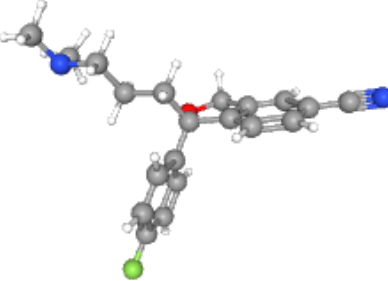
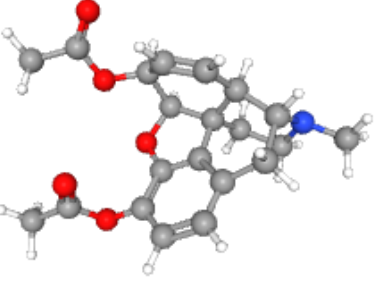
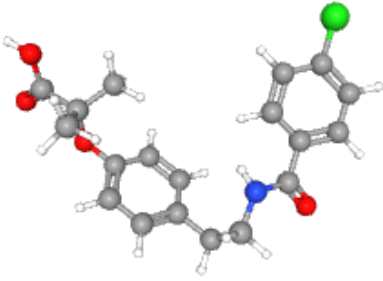
Il est intéressant de remarquer que les interactions mises en jeu pour la rétention sur PGC reposent en majorité sur la surface des composés en contact avec la surface graphite de la phase stationnaire. Ainsi, une meilleure rétention des molécules plus planaires, donc ayant une CCS plus faible, est attendue sur cette colonne, ce qui permet d'expliquer cette observation. Concernant les deux autres colonnes, les molécules retrouvées sur la HILIC ont en moyenne une masse molaire plus élevée mais également une CCS plus élevée que les molécules retrouvées sur la C18. Cependant, les données de CCS répertoriées en librairie en ligne sont encore relativement peu nombreuses. Ici, pour ce qui est des molécules retrouvées sur les colonnes C18 et PGC, la proportion de CCS manquantes était faible. En revanche, seule une valeur de CCS (pour la diacétylmorphine) était disponible sur les sept molécules retrouvées sur la HILIC. Ainsi, la comparaison effectuée n'est pas forcément représentative. Afin de remédier à cela, les valeurs de CCS mesurées ont été ensuite prises en compte. La recherche de ces molécules ayant été effectuée en SS et l'identité des molécules n'ayant pas été systématiquement confirmée, ces valeurs n'avaient pas été considérées initialement afin de ne pas introduire un biais dans l'interprétation des données dans le cas où le candidat ne correspondrait pas à la molécule. Pour ces raisons, une mesure de précaution supplémentaire a été introduite. Cette dernière consistait à estimer un ordre de grandeur de CCS en comparant les structures 3D des molécules d'intérêt à des molécules de masse molaire similaire dont la valeur de CCS existe dans la librairie. Cette estimation est ensuite comparée à

la valeur mesurée pour le candidat lors de l'analyse afin de voir si celle-ci est cohérente. Pour réaliser ces estimations, les masses molaires et structures des molécules spécifiques à la HILIC ont été répertoriées (partie gauche du Tableau 36). Ensuite, les molécules ayant des masses molaires proches et des valeurs de CCS disponibles ont été recherchées (partie droite du Tableau 36).

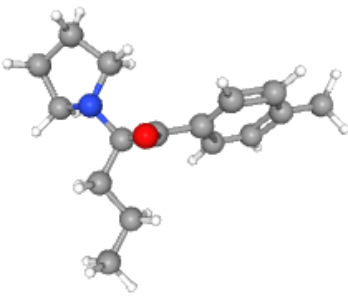
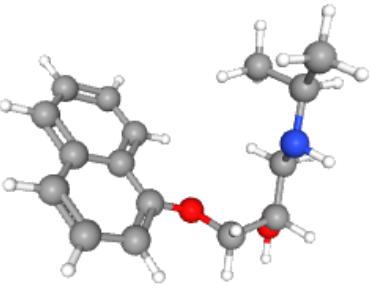
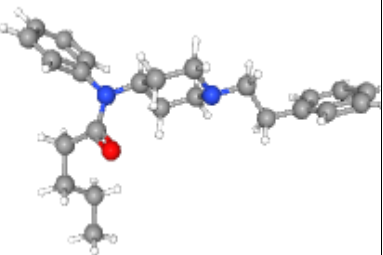
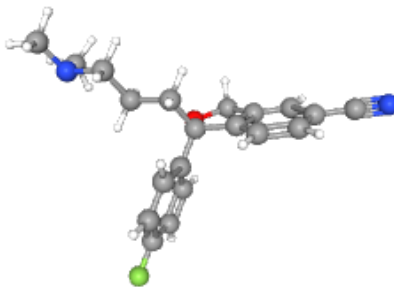
Tout d'abord, deux composés de masse molaire proche possédaient des valeurs de CCS, la diacétylmorphine avec le bezafibrate. La diacétylmorphine a une masse molaire légèrement plus élevée que le bezafibrate mais l'inverse est observé au niveau de leurs CCS respectives. Cette observation s'explique aisément en observant les structures respectives de ces composés. En effet, la diacétylmorphine possède trois cycles à 6 carbones, partageant deux à deux une liaison et partageant également chacun une liaison avec un cycle supplémentaire à 5 carbones. Cette géométrie donne à la structure une section de collision plus élevée que le bezafibrate qui ne possède que 2 cycles aromatiques séparés par des liaisons simples. En partant de ce constat, des hypothèses basées sur la comparaison des structures ont été posées afin d'attribuer un ordre de grandeur de CCS aux molécules de la partie droite du Tableau 36. En ce qui concerne l'acétyl fentanyl, celui-ci possède un cycle à 6 carbones aligné verticalement avec un groupement amide et une chaîne horizontale avec 2 cycles à 6 carbones. Il est comparé avec le citalopram qui possède également un cycle à 6 carbones compris dans un plan vertical avec une chaîne plus longue et une chaîne horizontale comprenant un autre cycle à 6 carbones. Ainsi, la section efficace du citalopram devrait être plus élevée que celle de l'acétyl fentanyl, pour lequel on peut donc émettre l'hypothèse d'une CCS < 178,90. Le même raisonnement a été appliqué à l'ensemble du Tableau 36. Les valeurs de CCS mesurées lors des analyses en mode SS ont également été répertoriées pour comparaison.

Tableau 36 : Comparaison des masses molaires, CCS et structures 3D des molécules retenues sur la HILIC et la C18 en SS.

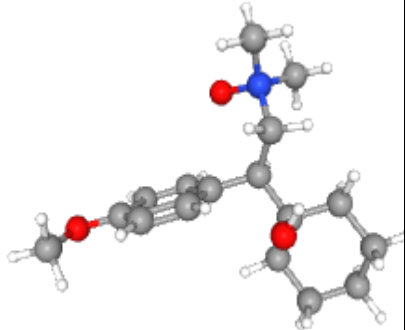
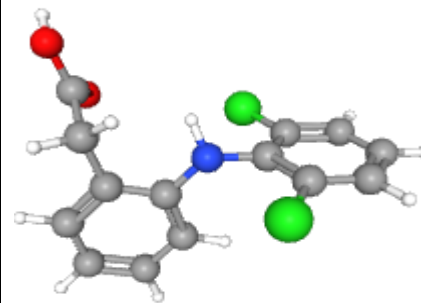
Les structures 3D ont été récupérées à partir de PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). * Valeurs estimées par comparaison des structures 3D avec la partie droite du tableau, ** Valeurs mesurées pour les candidats attribués

Molécule retenue sur HILIC	MW	CCS	Structure 3D	Molécule retenue sur C18 de masse molaire proche	MW	CCS	Structure 3D
Acétyl fentanyl	322,40	< 178,90* 175,93**		Citalopram	324,4	178,90	
Clarithromycine N-oxyde	763,47	~ 264,20* 270,60**	N.A.	Clarithromycine	747,48	264,20	N.A.
Diacétylmorphine	349,41	185,70		Bezafibrate	361,80	177,30	

PARTIE B : Comparaison de méthodes analytiques

Molécule retenue sur HILIC	MW	CCS	Structure 3D	Molécule retenue sur C18 de masse molaire proche	MW	CCS	Structure 3D
Pyrovalérone	245,36	< 162,70* 160,63**		Propranolol	259,34	162,70	
Valéryl fentanyl	364,50	~ 178,90* 189,09**		Citalopram	324,40	178,90	

PARTIE B : Comparaison de méthodes analytiques

Molécule retenue sur HILIC	MW	CCS	Structure 3D	Molécule retenue sur C18 de masse molaire proche	MW	CCS	Structure 3D
Venlafaxine N-oxyde	293,40	> 157,70* 174,61**		Diclofénac	296,10	157,70	

Il est intéressant de noter que les valeurs de CCS estimées et les valeurs effectivement mesurées (bien que non confirmées par injection de standards analytiques), sont cohérentes. En prenant en compte ces valeurs mesurées, une nouvelle moyenne de CCS plus représentative des molécules spécifiques à la HILIC a été calculée : $CCS_{\text{spécifiques_HILIC_recalculée}} = 192,76$. Cette nouvelle moyenne est légèrement supérieure à la valeur précédente (c'est-à-dire la valeur pour la diacétylmorphine, seule disponible initialement, $CCS_{\text{diacétylmorphine}} = 185,70$) et donc toujours supérieure à la moyenne obtenue pour les composés spécifiques à la C18 ($CCS_{\text{spécifiques_C18}} = 174,63$). Les tendances observées en termes de m/z et de CCS (ou DT) en NTS ont donc été retrouvées avec cette approche en SS.

En ce qui concerne les valeurs de pKa, trop peu de données étaient disponibles pour pouvoir tirer des conclusions représentatives. En revanche, pour ce qui est du log Kow, il est intéressant de remarquer que les molécules retenues spécifiquement sur la PGC ont en moyenne des log Kow plus faibles que les molécules spécifiques aux deux autres colonnes. Cette observation est cohérente avec les spécificités de la colonne, qui est décrite comme permettant de séparer des composés très polaires, ou du moins plus polaires que des phases classiques en silice réticulée, dans des conditions standards de chromatographie en phase inverse. En revanche, un log Kow moyen plus élevé pour les molécules spécifiques à la HILIC est retrouvé par rapport à la PGC, bien que toujours inférieur à la C18. Cette observation peut, une fois de plus, être expliquée par le fait que les suspects sélectionnés sont en nombre limité. Ainsi, les molécules plus hydrophiles qui ont été retrouvées en PGC ont pu être retenues par d'autres types d'interactions, ce qui explique qu'elles n'ont pas été retenues sur HILIC. Inversement, d'autres molécules d'hydrophobicité équivalente ou supérieure peuvent avoir été retenues en HILIC et non en PGC, mais ne sont simplement pas présentes dans la liste limitée des suspects employée.

IV.4.2. Comparaison en SS entre cartouches SPE

De manière similaire, à colonne chromatographique fixée, les différentes méthodes de préparation ont également été comparées en SS. Pour cela, une classification hiérarchique a été réalisée à partir des aires respectives des composés suspects détectés dans chaque extrait (Figure 34). En termes de méthodes de préparation, il a alors été possible de distinguer quatre groupes :

- Injection directe, phases DP et Coco, caractérisées par un nombre total plus faible de composés suspects retrouvés, avec des aires également plus faibles
- Phases XCW et C18, caractérisées par deux groupes de composés suspects présentant de fortes intensités (partie inférieure de la *heatmap* (Figure 34))
- Phases Carb et ENV, caractérisées par des composés suspects présentant des aires globalement plus élevées sur toute la *heatmap*
- Phases XAW et HLB, caractérisés par des aires homogènes pour tous les composés suspects retrouvés

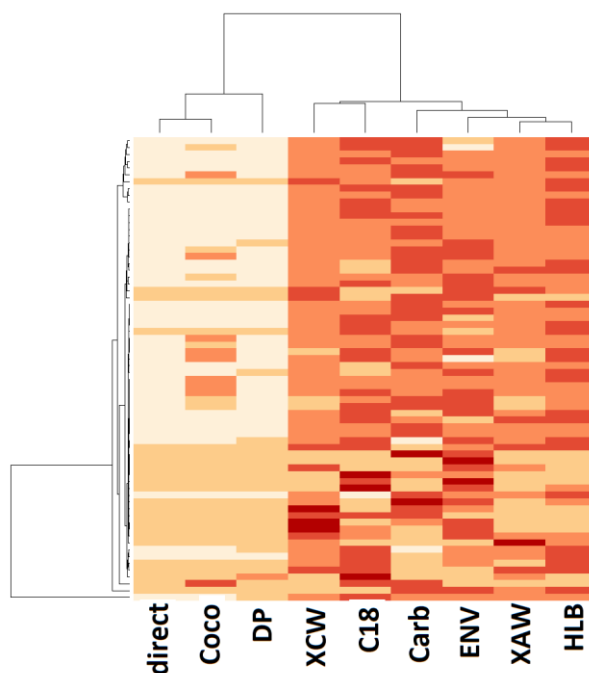


Figure 34 : Heatmap des composés suspects (dendrogramme vertical) discriminés par leurs aires relatives dans chaque extrait provenant des différentes méthodes de préparation (dendrogramme horizontal), après séparation sur colonne C18 et analyse en ESI+ (les couleurs les plus sombres indiquent des aires plus importantes)

La distribution observée ici avec la sélection de 376 composés suspects est cohérente avec celle obtenue en NTS qui considérait tous les marqueurs (Figure 28). En effet, l'injection

directe et les phases Coco et DP sont caractérisées par les plus grandes proportions de marqueurs *uniques*. Les phases XCW, C18 et HLB partagent quant à elles le plus grand nombre de marqueurs *communs*. Cependant, bien que les groupes formés par les cartouches soient facilement observables ici (Figure 34), les groupes formés par les composés (dendrogramme vertical) sont, d'une part plus difficile à repérer, et d'autre part, il est difficile de déterminer un ou plusieurs groupes verticaux de composés suspects caractéristiques d'un groupe horizontal de méthode de préparation. Ainsi, une ACP a été construite à partir de ces données en SS afin d'identifier des marqueurs discriminants qui permettraient d'expliquer cette classification (Figure 35).

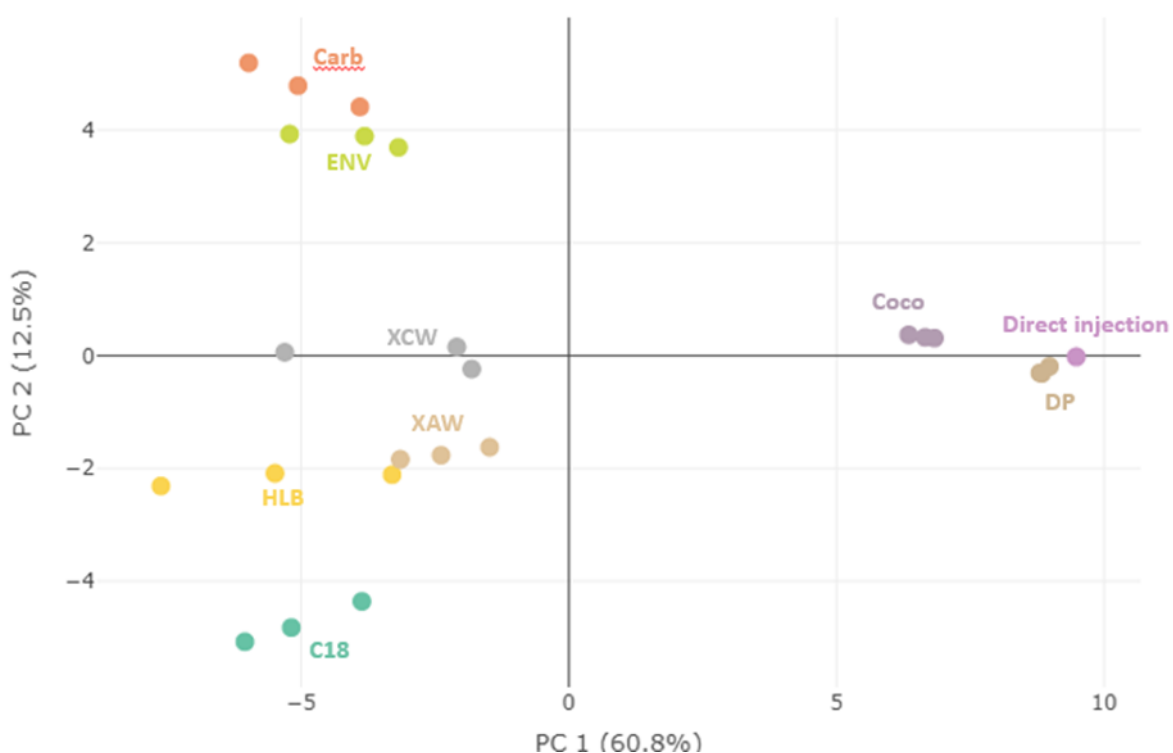


Figure 35 : ACP réalisée sur les aires des composés suspects détectés dans les extraits des différentes méthodes de préparation, après séparation sur colonne C18 et analyse en mode ESI+

Sans surprise, une plus grande variabilité est observée pour la PC1 (60,8 %) par rapport à l'ACP obtenue en mode NTS (Figure 30), puisqu'ici une très faible rétention des composés présents dans la liste de suspects est observée dans les extraits des phases Coco et DP et l'injection directe par rapport à tous les autres extraits. Cependant, malgré une variabilité plus faible sur la PC2 (12,5 %) par rapport à l'ACP en NTS, une meilleure discrimination des autres cartouches est observée. De même qu'avec la *heatmap* (Figure 34), les phases ENV et Carb forment toujours un groupe, de la même manière que les phases HLB et XAW. Cependant, les phases

C18 et XCW sont séparées sur l'ACP alors que leurs différences sont moins évidentes sur la *heatmap*. En prenant en compte les descriptions des cartouches par les fournisseurs (Tableau 25), la ENV et la Carb sont adaptées pour l'analyse de composés polaires alors que la C18 est plus adaptée pour la rétention de composés de polarité moyenne. La polarité semble donc influencer la PC2.

Afin de confirmer cette hypothèse, une étude du *loading plot* (Annexe 3B) correspondant à cette ACP a été réalisée afin de déterminer les marqueurs discriminants le plus la phase C18 des cartouches ENV et Carb. Les 3 composés suspects ayant le score le plus élevé sur la PC2 sont l'Iohexol ($\log Kow = -3,0$), l'Iomeprol ($\log Kow = -2,3$) et la 3,4-Méthylènedioxyamphétamine ($\log Kow = 1,6$) alors que les 3 composés suspects avec le score le plus faible sur la PC2 sont le Telmisartan ($\log Kow = 8,4$), la Phénytoïne ($\log Kow = 2,5$) et le diclofénac ($\log Kow = 4,4$), confirmant ainsi l'hypothèse précédente. Cependant, il convient de noter que cette observation est de moins en moins vraie à chaque suspect considéré supplémentaire. En effet, l'observation précédente se base sur les 3 composés avec les scores les plus extrêmes. Certains marqueurs obtiennent en effet des scores élevés (en valeur absolue) sur la PC2 mais des valeurs de $\log Kow$ moins extrêmes que les six cas décrits précédemment (ex. le tramadol avec un score proche de la 3,4-Méthylènedioxyamphétamine et avec un $\log Kow = 2,6$, semblable à la phénytoïne). En s'éloignant des scores les plus extrêmes, les valeurs de $\log Kow$ se chevauchent pour la plupart des marqueurs, suggérant que la polarité ne joue qu'un rôle limité pour la discrimination de ces cartouches. Ceci est en accord avec la plus faible discrimination entre les extraits C18, ENV et Carb (discriminés par les PC3 et PC4) observée en NTS (avec donc beaucoup plus de marqueurs qu'en SS, Figure 30). En SS, certains composés semblent donc entraîner une forte discrimination entre les cartouches suivant leurs valeurs extrêmes de polarité, alors que cette propriété joue un rôle plus limité dans la discrimination des marqueurs en NTS.

IV.4.3. Comparaison en SS sur l'ensemble des cartouches SPE et colonnes chromatographiques

Après avoir comparé les colonnes entre elles d'une part, indépendamment du mode d'extraction, puis les phases d'extraction d'autre part, à colonne chromatographique fixée, dans cette partie, nous nous intéresserons à l'ensemble des différences entre les modes de préparation et les techniques analytiques, et tenterons de mettre en évidence des tendances

à partir de ces données. Pour cela, les composés suspects identifiés pour chacune des colonnes chromatographiques ont été séparés en fonction de l'extrait de SPE dans lequel ils étaient retrouvés. Une *heatmap* a ainsi été établie en représentant sur l'axe vertical les 36 % de suspects identifiés et sur l'axe horizontal la combinaison cartouche SPE / colonne chromatographique qui a permis d'analyser ces molécules (Figure 36).

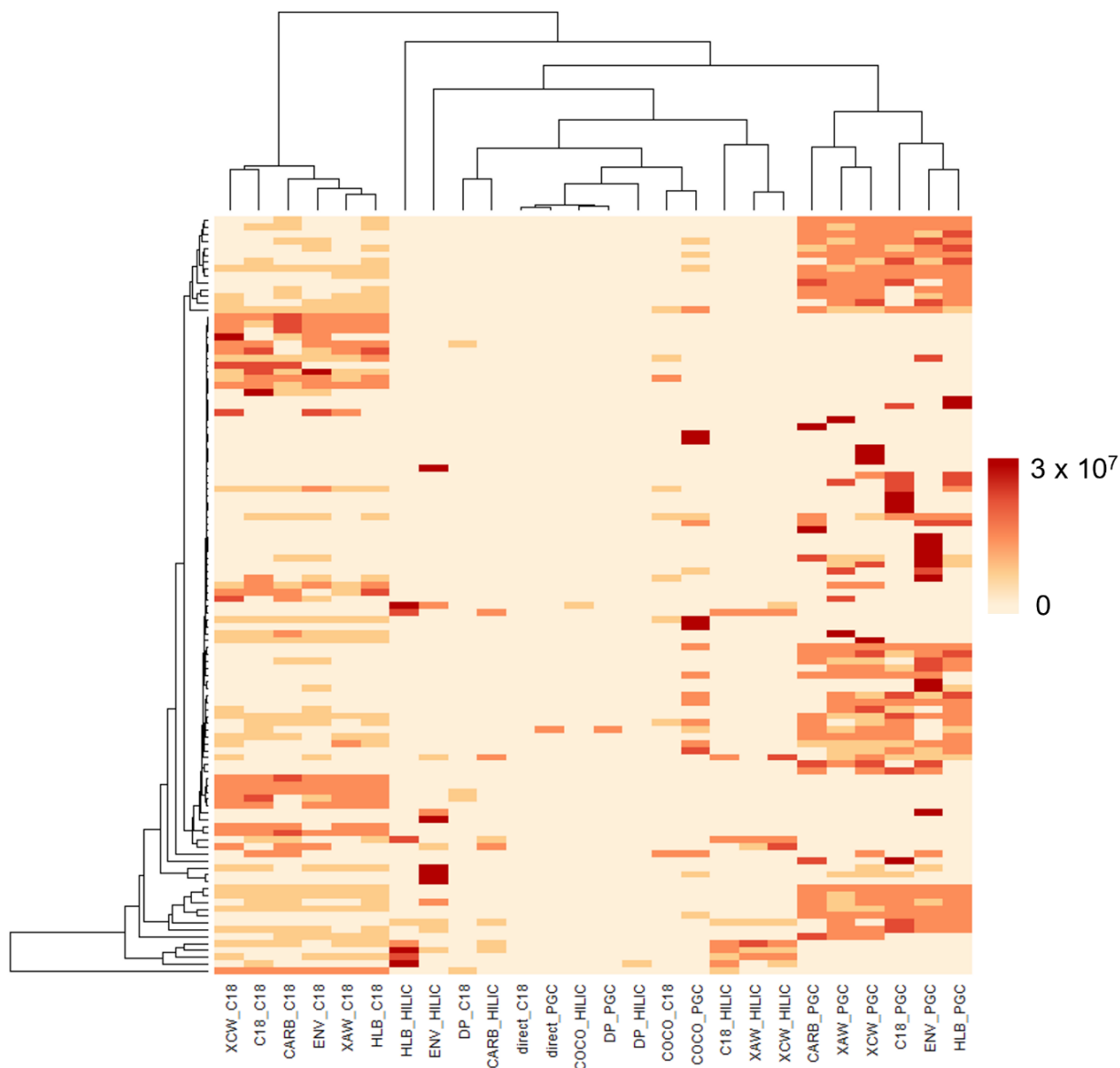


Figure 36 : *Heatmap* des composés retrouvés en SS par chacune des combinaison cartouche SPE / colonne chromatographique. L'intensité de couleur est proportionnelle à l'aire du pic analysé par UPLC-IMS-QToF

5 groupes peuvent alors être distingués sur l'axe horizontal :

- **Groupe 1 entre XCW_C18 et HLB_C18.** Ce groupe est défini par la plus grande homogénéité en aire de pics, aussi bien sur l'axe horizontal (combinaisons cartouches SPE / colonnes chromatographiques) que sur l'axe vertical (différents suspects recherchés).
- **Groupe 2 entre HLB_HILIC et CARB_HILIC.** Ce groupe est défini par un nombre limité de suspects retrouvés, qui possèdent cependant une aire élevée lorsqu'ils sont effectivement identifiés.
- **Groupe 3 entre direct_C18 et COCO_PGC.** Ce groupe est défini par l'absence de suspects retrouvés par ces combinaisons cartouche SPE / colonne chromatographique.
- **Groupe 4 entre C18_HILIC et XCW_HILIC.** Ce groupe est défini par l'absence de suspects sur la grande majorité du diagramme à l'exception de quelques-uns, principalement dans la partie inférieure.
- **Groupe 5 entre CARB_PGC et HLB_PGC.** Ce groupe est défini par des aires globalement plus élevées que les autres groupes. Il regroupe également trois groupes de suspects différents avec des aires relativement homogènes et une zone avec de fortes aires dépendant des combinaisons cartouche SPE / colonne chromatographique.

Afin de caractériser ces différents groupes, la partie verticale du diagramme a également été divisée en plusieurs zones qui permettent de distinguer ces groupes. Les propriétés des molécules de chaque zone ont ensuite été étudiées dans le but de dégager des tendances. Cependant, en considérant les propriétés compilées, les 5 groupes de combinaisons analytiques n'ont pas pu être distingués car les gammes de propriétés couvertes par les molécules de différentes zones se chevauchaient. Une autre méthode a ainsi été utilisée afin de voir s'il était possible de dégager des tendances. En reprenant ces mêmes données de SS, une ACP a été construite (Figure 37).

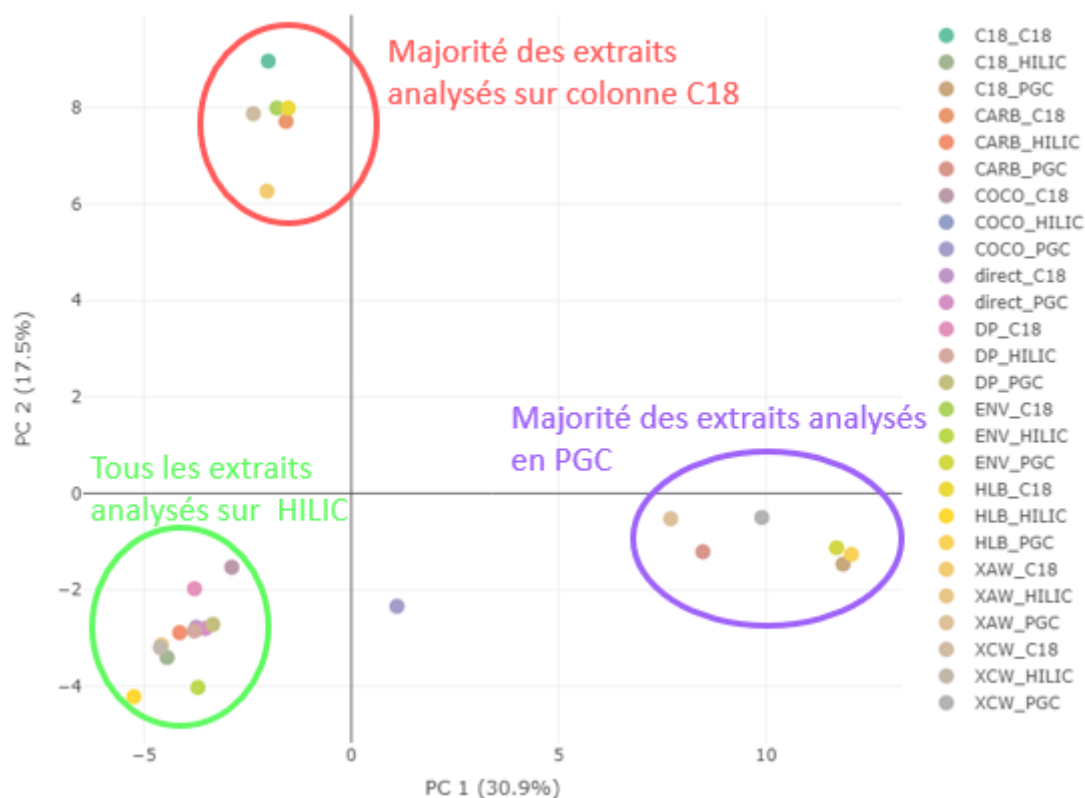


Figure 37 : ACP réalisée sur les aires des composés suspects détectés avec différentes combinaisons analytiques (cartouche SPE / colonne chromatographique)

Trois clusters sont facilement identifiables :

- **Cluster 1**, entouré en rouge, qui regroupe les extraits HLB, ENV, Carb, C18, XAW et XCW, analysés sur colonne C18
- **Cluster 2**, entouré en violet, qui regroupe les extraits HLB, ENV, Carb, C18, XAW et XCW, analysés sur PGC
- **Cluster 3**, entouré en vert, qui regroupe tous les extraits analyses sur HILIC, avec en plus l'injection directe et l'extrait DP analysés sur colonne C18 et PGC ainsi que l'extrait Coco analysé sur colonne C18

Sur la base de ces observations, les combinaisons analytiques se distinguent majoritairement par la colonne chromatographique utilisée plutôt que par la phase SPE employée (Figure 37). Cela suggère une plus grande influence du choix de la méthode analytique par rapport à la méthode de préparation. Afin d'essayer de caractériser les différences observées, le *loading plot* (Annexe 3C) correspondant a été étudié afin de repérer les composés suspects qui permettent de discriminer le plus les trois clusters définis (Tableau 37).

Tableau 37 : Composés suspects permettant de discriminer les clusters définis par l'ACP en Figure 37

	Suspects discriminant le cluster 1	Suspects discriminant le cluster 2	Suspects discriminant le cluster 3
Liste des composés	Atrazine	Aténolol	Amisulpride
	Benzyl butyl phthalate	Carbamazépine	Clarithromycine <i>N</i> -oxyde
	Bezafibrate	Climbazole	Diacétylmorphine
	Citalopram	Clopidogrel	Diphenylguanidine
	Clarithromycine	Griséofulvin	Telmisartan
	Diclofénac	Milnacipran	Triméthoprime
	Hydrocodone	Oxazepam	Venlafaxine <i>N</i> -oxyde
	Imazalil	Phénytoin	
	Irbesartan	Terbutryn	
	Losartan	Tiapride	
	Méthoxétamine	Tramadol <i>N</i> -oxyde	
	Propanolol	Venlafaxine	
	Simazine		
	Valsartan		

A partir de ces composés suspects et pour chacun des trois groupes/clusters, les valeurs minimales, maximales et moyennes de chaque propriété (Annexe 1) ont été calculées. (Figure 38).

Certaines tendances permettent de distinguer les clusters. Par exemple, le cluster 3 semblent être caractérisé par des molécules couvrant une plus large gamme de pKa par rapport aux deux autres clusters. Le cluster 2 se distingue quant à lui par des composés suspects ayant globalement des masses molaires, des CCS et des aires de surface polaires plus faibles. Ces observations sont cohérentes entre elles, puisque les CCS sont généralement corrélées avec la masse molaire. Par ailleurs, cela rejoint les observations réalisées précédemment (partie B – IV.2), à savoir que la PGC a tendance à retenir des molécules globalement plus petites. En ce qui concerne l'aire de la surface polaire topologique, celle-ci est également aisément reliée aux deux propriétés précédentes puisque si la molécule est globalement plus petite et plus compacte, son aire de surface totale l'est aussi, et il en va alors de même pour la surface polaire. Cependant, il est intéressant de noter que ces deux études en SS diffèrent des observations effectuées en NTS. En effet, en NTS, la colonne PGC a présenté des distributions de m/z et DT similaires à celles de la colonne C18. Les résultats en SS semblent quant à eux

pointer vers une gamme moins large de masses et de CCS couverte par la colonne PGC par rapport à la C18, avec une préférence pour les petites tailles. Ceci pourrait indiquer que réaliser des études en SS avec un nombre limité de composés tendrait à induire un biais vis-à-vis de ce qui est réellement observable par ces techniques analytiques. En ce qui concerne le cluster 1, il est quant à lui plus difficile à distinguer des deux autres sur la base de ces propriétés.

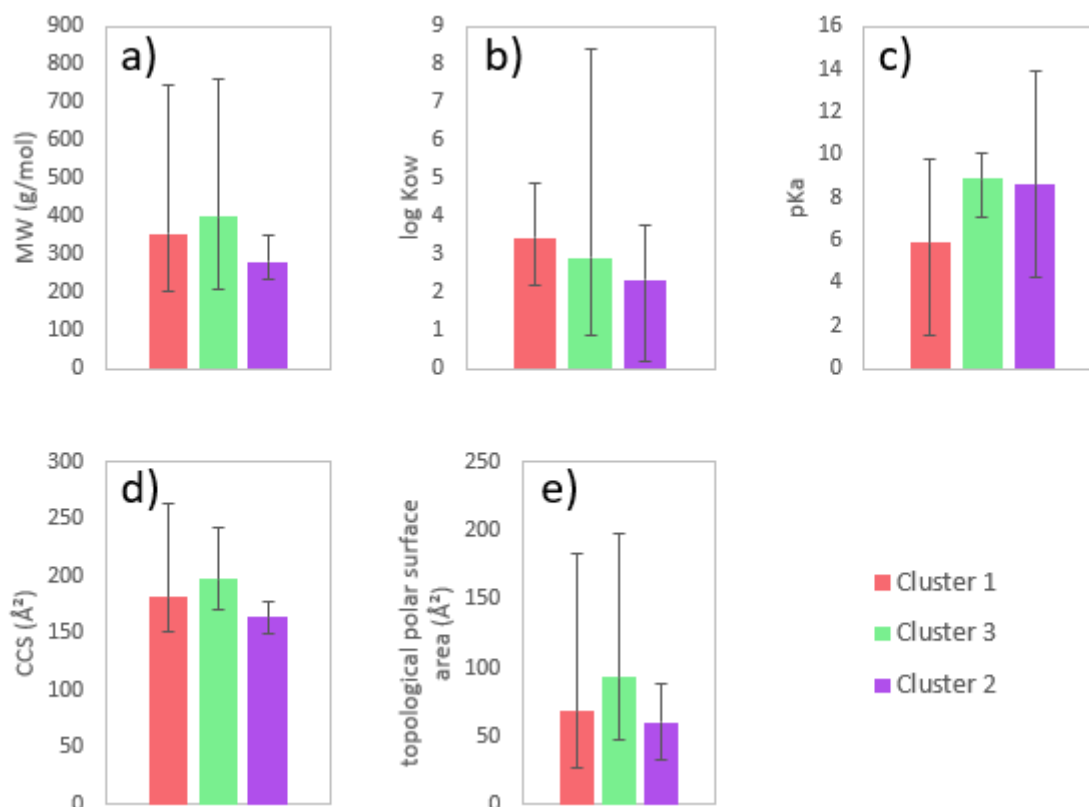


Figure 38 : Comparaison de la gamme de a) masses molaires b) log Kow, c) pKa, d) CCS et e) surface polaire topologique couverte par chaque groupe de suspects responsables de la discrimination des trois clusters définis en Figure 37. Les barres d'erreur représentent les valeurs minimales et maximales de chaque groupe.

Néanmoins, ces observations ne prennent en compte que les suspects les plus discriminants de chaque cluster. De plus, les gammes de couverture des molécules discriminantes de chaque cluster se chevauchent pour la grande majorité des propriétés étudiées (Figure 38). Il semble ainsi difficile de caractériser précisément les clusters de cette manière. Ceci pourrait être expliqué par deux facteurs. Le premier pourrait être le nombre de suspects relativement faible par rapport à d'autres études dans le domaine, pour lesquelles il a également été difficile de faire ressortir des différences significatives entre les méthodes étudiées (Bieber et al., 2017; Castro et al., 2021; Montes et al., 2017). Le second facteur concerne le choix des propriétés

étudiées, effectué sur la base des données facilement accessibles (MW, CCS, surface polaire) et sur les propriétés généralement considérées lors du développement de méthodes effectuées en TS (log Kow, pKa). L'étude de paramètres supplémentaires qui pourraient expliquer les interactions entre molécules et phases stationnaires pour la préparation ou l'analyse peut s'avérer intéressante. Des paramètres supplémentaires (disponibles sur PubChem et obtenus par modélisation avec Cactvs) tels le nombre de sites accepteurs ou donneurs de liaisons hydrogènes ou encore le nombre de liaisons capables d'effectuer des rotations, ont été considérés mais ont conduit aux mêmes conclusions. D'autres propriétés concernant l'accessibilité des différentes fonctions chimiques pour la liaison par exemple, pourraient également être étudiées pour expliquer les différences observées. Cependant, ces propriétés ne sont pas facilement accessibles et nécessitent d'associer de nouveaux outils pour leur évaluation.

IV.4.4. Comparaison en NTS entre cartouches SPE

Une tentative de comparaison des cartouches SPE, toutes colonnes chromatographiques confondues, a été effectuée en SS. Seuls deux groupes ont alors été distingués : l'un composé des cartouches Coco et DP, et de l'injection directe, qui ne permettaient d'analyser que très peu, voire pas de composés présents dans la liste de suspects, et l'autre groupe composé des six cartouches restantes qui montraient des performances relativement proches les unes des autres pour les composés étudiés. En tirant parti des données issues de l'analyse par FT-ICR-MS, une dernière comparaison de l'espace chimique de ces cartouches a pu être effectuée en NTS. Celle-ci possède l'avantage d'être moins limitante que les comparaisons précédentes effectuées en SS puisqu'elle s'appuie sur les formules moléculaires calculées, qui représentaient 73 % du set total de données. Les cartouches SPE ont ainsi pu être comparées à l'aide de diagrammes de van Krevelen afin de voir si certaines se distinguent par un rapport atomique particulier par rapport aux autres (Tableau 38).

Tableau 38 : Comparaison des ratios atomiques, pondérés par les intensités, des marqueurs retrouvés dans chaque extrait après analyse par FT-ICR-MS. Les valeurs extrêmes sont indiquées en rouge (minimum) et vert (maximum).

	ESI+				ESI-			
	H/C	O/C	N/C	Cl/C	H/C	O/C	N/C	Cl/C
C18	1,49	0,23	0,10	0,01	1,31	0,31	0,09	0,02
Carb	1,46	0,30	0,14	0,02	1,22	0,64	0,18	0,07
Coco	1,60	0,30	0,06	0,02	1,35	0,46	0,15	0,07
DP	1,53	0,27	0,15	0,02	0,97	1,13	0,41	0,20
ENV	1,98	0,17	0,12	0,00	1,30	0,38	0,09	0,01
HLB	1,51	0,23	0,09	0,01	1,28	0,20	0,07	0,02
Injection directe	1,18	0,81	0,47	0,10	0,95	0,85	0,29	0,23
XAW	1,42	0,48	0,17	0,05	1,12	1,46	0,36	0,14
XCW	1,44	0,40	0,15	0,03	1,20	1,02	0,27	0,09
Moyenne	1,51	0,35	0,16	0,03	1,18	0,71	0,21	0,09

Il est intéressant de noter qu'en mode ESI-, les ratios d'hétéroatomes sont en moyenne plus élevés qu'en mode ESI+. Cette observation est d'autant plus vraie concernant le chlore, pour lequel le ratio moyen en ESI- est trois fois plus élevé qu'en ESI+. Malgré le fait qu'une majorité de marqueurs soit détectés en ESI+ et que de nombreux auteurs se restreignent à des analyses en mode ESI+, les informations récupérées en mode ESI- n'en restent pas moins intéressantes du point de vue de la diversité des composés détectés.

En mode ESI+, l'injection directe montre les ratios d'hétéroatomes les plus élevés et l'extrait ENV répertorie les plus faibles valeurs. Ces observations permettent de confirmer que les différentes cartouches permettent de retenir des composés de nature très différente. La cartouche DP jusqu'ici très mal caractérisée en SS mais ayant montré une grande proportion de marqueurs *spécifiques* en NTS semble d'autant plus intéressante à étudier, particulièrement en ESI-, car elle semble retenir des composés de nature plus diverse (*i.e.*, variétés d'hétéroatomes), et plus particulièrement des molécules très oxygénées et azotées avec un ratio O/C = 1,3 (soit le second plus élevé après la XAW) et un ratio N/C = 0,41 (le plus élevé). En effet, une étude a estimé que la gamme classique de ratio O/C observée pour des composés organiques est comprise entre 0 et 1,2 pour 99,7 % des 45 000 formules contenues dans la librairie Wiley (Kind and Fiehn, 2007). La gamme élargie correspondante est comprise

entre 0 et 3 et permet quant à elle de caractériser correctement 99,99 % des composés. Ainsi, bien que le ratio obtenu ici semble élevé, il apparaît encore cohérent. De plus, la Coco qui est la seconde cartouche relativement mal caractérisée jusqu'à présent, au même titre que la DP, présente le ratio H/C le plus élevé en mode ESI- sans pour autant avoir les ratios d'hétéroatomes les plus faibles. Cette observation laisse penser que cette cartouche est capable de retenir des composés divers peu conjugués.

IV.5. Conclusion sur la comparaison de méthodes analytiques

Dans cette partie, différentes méthodes de préparation et d'analyse ont été utilisées afin d'analyser un même échantillon. Les fractions obtenues par chaque méthode de préparation ont été comparées avec différentes méthodes d'analyse, et celles-ci ont également été comparées entre elles.

Dans ce travail, les suspects ont été sélectionnés sur la base de travaux précédents effectués sur le type de matrice étudié. Les propriétés correspondantes ont été choisies en fonction de leur probabilité à pouvoir expliquer les différences observées entre les méthodes. Ainsi, le log Kow et le pKa ont été pris en compte puisque ce sont généralement les premiers paramètres considérés pour le développement de méthode en TS. Cependant, avec le pool de molécules sélectionnées et ces propriétés, il a été difficile de tirer des conclusions. En effet, seuls 37 % des composés suspects considérés ont effectivement été analysés par au moins une des méthodes. Par ailleurs, les molécules spécifiques à chacune des méthodes n'ont pas pu être distinguées sur la base des comparaisons de log Kow et pKa. Ceci s'explique par deux raisons différentes. La première concerne les molécules effectivement analysées qui peuvent ne pas être assez discriminantes, auquel cas, il sera nécessaire de considérer un spectre plus large de composés afin de pouvoir distinguer des différences entre les méthodes. La seconde explication concerne le nombre limité de propriétés prises en compte. En effet, ces propriétés sont les plus communes lors du développement de méthodes, mais elles ne permettent pas de caractériser entièrement les molécules. D'autres propriétés telles que la capacité des composés à former ou non des liaisons hydrogènes, le nombre de donneurs ou accepteurs d'électrons peuvent être considérées afin d'expliquer les interactions qui se produisent entre les phases de rétention et les molécules.

Néanmoins, **les observations tirées de la comparaison de plusieurs stratégies analytiques ont pu mettre en évidence que le choix de la colonne chromatographique semblait avoir une plus grande influence sur les composés analysables par rapport au choix de la méthode de préparation.** Ainsi, pour certaines méthodes de préparation, le choix de la méthode d'analyse est crucial afin de tirer le maximum d'informations possibles récupérées lors de l'extraction. D'autre part, il a été intéressant de noter que l'injection directe (*i.e.*, sans étape de préparation) permettait quant à elle de détecter un certain nombre de signaux uniques. Ainsi, bien que ce travail ait fait appel à l'utilisation de trois méthodes de préparation, incluant

onze phases d'extraction différentes, cette combinaison n'a pas suffi pour récupérer toutes les informations présentes dans l'échantillon initial. Ces signaux perdus correspondent pourtant à des molécules présentes en concentration relativement importantes initialement, puisque visibles sans étape de préconcentration. Un autre axe de développement pourrait ainsi se consacrer à développer des méthodes permettant de concentrer l'échantillon aqueux initial sans extraction et en tentant de minimiser l'altération de ce dernier.

Enfin, l'utilisation en parallèle du SS et du NTS permet également d'avoir une vision plus complète des différences qui peuvent exister entre les méthodes. En effet, le NTS est intéressant car cette technique permet d'obtenir les informations les plus représentatives de la diversité des composés pouvant être analysés par la méthode. Cependant, elle ne permet d'apporter que peu d'informations sur les propriétés de ces composés, l'interprétation des données s'en trouve donc souvent limitée. Le SS est quant à lui souvent restreint par le nombre et la nature des composés sélectionnés et peut ainsi induire un biais lors de l'interprétation des données. Cependant, une plus grande variété d'informations est obtenue par cette méthode et il est alors possible d'expliquer de manière plus approfondie les observations enregistrées. Par conséquent, la combinaison des deux stratégies permet d'obtenir un bon compromis entre représentativité et interprétation des données.

Partie C : Application des méthodes pour la comparaison d'échantillons oxydés

I. Introduction et objectifs

L'objectif de cette partie est de caractériser par NTS différents traitements d'oxydation appliqués à des effluents d'eaux usées traitées, leurs impacts sur les MP organiques et la formation de TP, en s'appuyant sur les méthodes et résultats obtenus dans les deux parties précédentes.

Après un bref état de l'art sur le traitement des micropolluants dans les eaux résiduaires urbaines et les études réalisées en NTS pour l'évaluation de tels procédés, deux cas d'application sont présentés (partie C - III) : une étude de la désinfection d'un rejet de STEU par l'acide performique, puis, l'évaluation d'un procédé d'ozonation suivi d'une adsorption sur charbon actif en poudre (CAP) pour le traitement de micropolluants dans un rejet de STEU. La partie B ayant mis en évidence de grandes différences entre les méthodes analytiques employées, une méthode unique a été choisie afin de pouvoir comparer les échantillons entre eux. L'étude de la littérature ayant montré la prédominance d'une méthode analytique basée sur une extraction sur un support multiphasique, nous avons réalisé une étude préliminaire permettant de comparer un large spectre de supports SPE, dont celui-ci, pour l'analyse d'eaux urbaines en employant le NTS. Ces résultats ont mis en avant les avantages qu'apportaient la cartouche multiphasique par rapport aux phases individuelles qui a donc été sélectionnée pour la suite. Ces travaux ont fait l'objet d'une publication disponible en Annexe 4:

Nina Huynh, Emilie Caupos, Caroline Soares Peirera, Julien Le Roux, Adèle Bressy, Régis Moilleron. *Evaluation of Sample Preparation Methods for Non-Target Screening of Organic Micropollutants in Urban Waters Using High-Resolution Mass Spectrometry*. *Molecules*. 2021; 26(23):7064. <10.3390/molecules26237064>

Brièvement, l'étude de la désinfection par l'acide performique a mis en évidence la formation de *N*-oxydes par l'emploi d'approches classiques de NTS (*i.e.*, priorisation d'un nombre limité de signaux puis identification). La seconde application a, quant à elle, permis de rechercher ces composés par des approches plus globales, basées sur les diagrammes de Kendrick développés lors de cette étude. Dans une dernière partie (partie C - IV), trois procédés d'oxydation (la chloration, l'oxydation avancée par UV/H₂O₂ et l'ozonation) ont été étudiés en appliquant à nouveau les différentes méthodes d'analyse présentées en partie B (*i.e.*, différentes cartouches SPE, injection directe, différentes colonnes chromatographiques). Le

but de cette partie était alors de mettre en évidence la multiplicité des informations recueillies en employant différentes méthodes analytiques, comme souligné en partie B, avec cette fois-ci un focus sur les produits d'oxydation. Pour cela, les échantillons soumis à oxydation par les trois procédés ont été prétraités et analysés en utilisant différentes combinaisons de protocoles analytiques, afin de déterminer les combinaisons permettant de caractériser un nombre plus important de TP formés lors de l'oxydation des eaux de STEU.

II. Etat de l'art

II.1. Traitement des eaux usées

II.1.1. Filières conventionnelles et limites des traitements face aux micropolluants

Les micropolluants organiques ont été détectés dans de nombreuses matrices environnementales comme les eaux usées (Jiang et al., 2013), les eaux de surfaces (Busch et al., 2016), les eaux souterraines ou les sols (Vodyanitskii and Yakovlev, 2016), mais également dans des organismes aquatiques (Terechovs et al., 2019). Ces contaminants entrent dans l'environnement par différentes voies, mais une grande partie transite par les STEU où ils peuvent atteindre des concentrations de l'ordre du $\mu\text{g/L}$ en entrée de station (Paijens et al., 2021) voire de la centaine de $\mu\text{g/L}$ pour certains composés les plus utilisés (Guillossou et al., 2019).

Les traitements conventionnels ont été conçus principalement pour abattre les particules solides, la macropollution carbonée, azotée et phosphorée. Pour ces paramètres, les opérateurs de STEU sont tenus par la réglementation de concevoir leur filière de traitement de sorte à atteindre des concentrations inférieures à un seuil déterminé ou *a minima*, de justifier d'un taux d'abattement minimum lorsque les concentrations en entrée sont trop élevées (Directive 91/271/CEE du Conseil, du 21 mai 1991, relative au traitement des eaux urbaines résiduaires).

Dans un premier temps, le prétraitement permet de retirer tous les gros déchets, les sables ainsi que les graisses présentes dans les eaux usées. Cette étape reposant sur des procédés mécaniques, elle n'a pas d'impact sur les MP organiques. Par la suite, le traitement primaire se base principalement sur une séparation physique par décantation afin de retirer les matières en suspension (MES). Les agents coagulants et floculants ajoutés réagissent avec les

particules en suspension et colloïdes, permettant de les agréger ce qui les rend plus propices à la décantation. De nombreuses études ont montré que cette étape du traitement était généralement peu efficace pour l'élimination des MP organiques (Dempsey and O'Melia, 1984; Snoeyink and Chen, 1985), sauf pour certains composés hydrophobes associés aux MES (Gasperi et al., 2010; Huerta-Fontela et al., 2011). Cette observation n'a cependant pas pu être confirmée par une corrélation entre le taux d'abattement observé pour les MP étudiés et leur log D (Alexander et al., 2012).

Le traitement secondaire s'opère quant à lui principalement grâce à des procédés biologiques qui permettent de décomposer la matière organique et certains MP organiques mais de manière souvent limitée. De plus, pour les MP organiques qui sont éliminés, la minéralisation complète du composé initial n'est pas toujours atteinte et des TP sont souvent formés par cette étape (Bolong et al., 2009; Luo et al., 2014). En ce qui concerne les molécules faiblement biodégradées, elles seront parfois adsorbées sur les boues (en fonction de leurs propriétés physico-chimiques) qui devront potentiellement elles-mêmes être traitées (Mailler et al., 2017), ou poursuivront leur chemin dans la filière de traitement et seront finalement rejetées dans l'environnement. De nombreuses études se sont penchées sur le taux d'abattement de MP organiques par les STEU, mais elles concernaient souvent les molécules des listes de substances prioritaires (ex. les hydrocarbures aromatiques polycycliques (Nas et al., 2020) ou les polybromodiphényléthers (Clara et al., 2012)) ou des familles de molécules particulières comme les biocides (Paijens et al., 2021), des composés pharmaceutiques (Gao et al., 2012; Guillosoy et al., 2019) ou encore des drogues (Nefau et al., 2013). Les abattements observés sont très variables suivant les molécules, même au sein d'une même famille de composés (ex. le paracétamol est abattu à 99 % dans les filières conventionnelles, alors que seul 20 % d'abattement est observé pour la carbamazépine (Guillosoy et al., 2019)).

II.1.2. Traitements avancés des micropolluants

Face à ce constat, de nouveaux traitements ont été envisagés afin de mieux traiter ce type de pollution. Parmi eux :

- les procédés d'adsorption sur charbon actif en grain (CAG) ou charbon actif en poudre (CAP) (Boehler et al., 2012)
- les procédés d'oxydation (ex. ozonation) ou d'oxydation avancée (ex. UV/H₂O₂) (Deng and Zhao, 2015; Poyatos et al., 2009)
- les procédés membranaires (Siegrist and Joss, 2012)

La Suisse, dès 2016, a commencé à implémenter de tels traitements avancés qui visent à réduire la concentration des micropolluants dans les eaux usées rejetées par les STEU (loi LEaux RS. 814.20, art 61a). Pour évaluer la qualité des rejets en termes d'abattement de micropolluants, 12 MP sont ciblés : l'amisulpride, la carbamazépine, le citalopram, la clarithromycine, le diclofénac, l'hydrochlorothiazide, le métoprolol, la venlafaxine, la benzotriazole, le candésartan, l'irbésartan et un mélange de 4-méthylbenzotriazole et 5-méthylbenzotriazole, pour lesquels les taux d'abattement doivent atteindre 80 % pour les plus grandes STEU et celles ayant été déterminées comme ayant un fort impact sur les cours d'eaux et les réserves d'eau potable. Dans ces stations, les traitements complémentaires reposent généralement sur des procédés d'adsorption sur charbon actif ou d'ozonation.

Au niveau mondial, ce genre de réglementation sur les MP en sortie de STEU est encore rare, mais de nombreuses études sont effectuées pour évaluer l'élimination de MP par différents procédés afin de réduire l'impact des rejets d'eaux usées dans l'environnement. En effet, la présence de ce type de composés dans les compartiments environnementaux peut poser divers problèmes. Notamment, des effets sur la reproduction de certaines espèces ont été répertoriés, causés par la présence de perturbateurs endocriniens dans les effluents de STEU par exemple (Vajda et al., 2008). De la même manière, la présence de nombreux composés pharmaceutiques a été associée à des dysfonctionnements chez des poissons directement exposés aux rejets, tels qu'une diminution du nombre de cellules neuronales et des problèmes au niveau du foie (Pérez et al., 2018). Par ailleurs, la présence de ces composés pharmaceutiques, et notamment des antibiotiques pose également un problème à la fois pour les organismes aquatiques qui sont exposés aux rejets, mais également d'un point de vue

thérapeutique avec le phénomène d'antibiorésistance (Rizzo et al., 2013; Väitalo et al., 2017; Zaman et al., 2017). Les MP les plus mobiles et persistants peuvent également finir par s'infiltrer dans d'autres compartiments ou dans les réserves d'eau où ils s'accumulent (Albergamo et al., 2019). Enfin, de nombreuses études ont aussi mis en évidence la capacité d'absorption des MP organiques par les plantes (Calderón-Preciado et al., 2011; Cañameras et al., 2016; Sabourin et al., 2012) qui peuvent être irriguées par des eaux impactées par ces rejets voire directement par des eaux usées traitées (réutilisation).

Pour ces différentes raisons, des études ont été menées dans le but de montrer l'amélioration de l'application des traitements complémentaires sur l'abattement de micropolluants organiques par rapport aux filières conventionnelles existantes (Tableau 39).

Tableau 39 : Exemples d'amélioration des taux d'abattelements de différents MP organiques par implémentation d'un traitement complémentaire

Nombre de MP étudiés	Filière initiale	Performances	Traitement complémentaire	Performances	Référence
220 (pharmaceutiques, TP, biocides, nitrosamines)	Boues activées	55 MP quantifiés	Ozone + filtration sur sable	11 MP persistants après ozonation sans élimination supplémentaire par le filtre à sable	(Hollender et al., 2009)
47 (pharmaceutiques)	Boues activées + filtration sur sable	41 MP détectés	UV (lampe à mercure)	13 MP avec abattement > 80 %, 14 MP entre 20 et 80 % et 14 MP < 20 %	(Kim et al., 2009)
			UV/H ₂ O ₂ (lampe à mercure)	39 MP avec un abattement > 85 % et 2 MP avec un abattement de 70 %	
6 (3 pharmaceutiques et 3 métabolites)	Boues activées	Entre 27 et 43 % d'abattement	Colonne de charbon actif en grain	Entre 88 et 98 % d'abattement	(Rúa-Gómez et al., 2012)
	Bioréacteur à membrane	Entre 20 et 37 % d'abattement	Charbon actif en poudre dans le bioréacteur à membrane	Entre 47 et 87 % d'abattement	
	Boues activées	Entre 14 et 40 % d'abattement	Ozonation + charbon actif en poudre	Entre 87 et 97 % d'abattement	
	Boues activées	Entre 21 et 56 % d'abattement	Floculation + charbon actif en poudre	Entre 79 et 88 % d'abattement	

Partie C : Application des méthodes pour la comparaison d'échantillons oxydés

Nombre de MP étudiés	Filière initiale	Performances	Traitement complémentaire	Performances	Référence	
64 (52 pharmaceutiques et 12 drogues)	Traitements biologiques suivis d'un traitement complémentaire par UV + charbon actif ou coagulation-floculation + UV	Sur les 24 MP détectés en entrée, 10 MP éliminés, 5 partiellement abattus, 4 non-abattus, 5 concentrations effluent > influent	Ozonation	Sur les 14 MP restants, seulement 2 sont persistants (valsartan et irbésartan)	(Ibáñez et al., 2013)	
572 (240 pharmaceutiques, 234 pesticides, 38 composés industriels et 60 TP)	Boues activées	Concentration totale des molécules cibles réduite de 85 %, avec un abattement moyen de 34 % mais le nombre de MP quantifiables est passé de 175 à 204	Ozonation + Réacteur en lit fluidisé	Concentration totale des molécules cibles réduite de 94 %, après ozonation, avec un abattement moyen de 67%	Abattement moyen < 5 %	(Bourgin et al., 2018)
			Ozonation + colonne en lit fixe		Abattement moyen < 5 %	
			Ozonation + charbon en grain en colonne, renouvelé tous les 3 jours		Abattement moyen 53 %	
			Ozonation + charbon en grain en colonne		Abattement moyen 19 %	
			Ozonation + filtre à sable		Abattement moyen < 5 %	
18 (10 pharmaceutiques, 3 hormones, 1 pesticide, 4 composés industriels)	Déphosphatation anaérobie + Bassins de Nitrification en anoxie / Dénitrification en aérobie	18 molécules persistants au traitement conventionnel	UV	1 composé avec abattement >99 % 4 composés avec abattement entre 20 et 53 %	(Paredes et al., 2018)	

Partie C : Application des méthodes pour la comparaison d'échantillons oxydés

Nombre de MP étudiés	Filière initiale	Performances	Traitement complémentaire	Performances	Référence
	contigus + Clarification (Biodenipho™)				
	Nitrification / Dénitrification			3 composés avec abattement entre 30 et 35 %	
48 (25 pharmaceutiques, 18 pesticides, 3 hormones, 1 TP, 1 acide perfluoré)	Biofiltration, Nitrification et dénitrification	6 MP avec abattement > 80%, 11 MP entre 20 et 80%, 1 MP <20%, 4 avec une concentration effluent > influent et 4 détectés seulement après traitement	Charbon actif en micro-grain	12 MP avec abattement > 80 %, 8 MP entre 20 et 80 %, 1 MP <20 % et 4 détectés seulement après traitement	(Guillossou et al., 2019)
47 (25 pharmaceutiques, 15 hormones, 7 pesticides)	Biofiltration	Pas de quantification effectuée en entrée de STEU, mais tous les composés étudiés semblent présents en entrée du traitement complémentaire	Ozonation	19 MP avec abattement > 80 % 8 MP entre 20 et 80 % et 20 MP sans informations	(Mathon et al., 2021)

Une première observation porte sur le fait que le nombre de MP considérés dans chacune de ces études est très variable (de la dizaine de composés à plusieurs centaines). Cette observation peut souvent être reliée aux composés d'intérêts qui ont été étudiés et à l'historique du laboratoire considéré. En ce qui concerne les traitements conventionnels, ils permettent majoritairement d'abattre au moins partiellement les composés étudiés. Dans certains cas, le traitement permet l'élimination complète, alors que pour d'autres molécules, les concentrations après traitement conventionnel sont supérieures à celles de l'influent. Dans tous les cas, le traitement complémentaire appliqué permet d'améliorer les performances de la filière. Parmi les procédés étudiés, l'ozonation semble permettre d'atteindre les meilleurs taux d'abattement finaux, d'autant plus lorsqu'elle est couplée à une adsorption complémentaire sur charbon actif.

II.1.3. Problématique des produits de transformation

Bien qu'efficaces pour éliminer les micropolluants, les traitements d'oxydation ont souvent le désavantage de seulement dégrader en partie les composés initiaux. Ainsi, ces derniers, par réaction avec l'oxydant, sont simplement transformés en un ou plusieurs autres composés (appelés produits de transformation – TP). L'élimination évoquée représente donc simplement le fait que la molécule initiale a été partiellement ou totalement transformée par la réaction, de sorte qu'elle n'est plus détectée après le traitement.

Cependant, de nombreuses études ont montré que les TP pouvaient dans certains cas être plus persistants et dangereux que les composés initiaux (Berkner and Thierbach, 2014; Escher et al., 2013). Par exemple une étude du diclofénac soumis à une photodégradation a montré la formation de 13 TP avec une dégradation du diclofénac atteignant 95 % au bout de 53 h d'irradiation. Cependant, le suivi cinétique de la toxicité de la solution a également montré que la toxicité augmentait au cours du temps, avec une multiplication par 6 de l'inhibition de la reproduction algale pour la solution à 53 h d'exposition par rapport à la solution initiale (Schmitt-Jansen et al., 2007). Il est aussi important de noter qu'en fonction des conditions dans lesquelles se déroule l'oxydation, les TP retrouvés pour une substance donnée ne sont pas les mêmes. Par exemple, une étude a été menée sur la dégradation de la caféine par l'ozone à différents pH, montrant la formation de 4 TP à pH 3 et 2 TP différents à pH 8. Cette observation a été attribuée à la différence de vitesse de décomposition de l'ozone à ces pH, amenant à différents mécanismes de décomposition de la molécule (Rosal et al., 2009).

Les TP sont ainsi de plus en plus étudiés afin d'évaluer si les traitements d'oxydation sont réellement bénéfiques. En particulier, cette problématique est récurrente dans le cadre de l'utilisation des procédés d'oxydation à des fins de désinfection dans les filières de traitement d'eau potable. Parmi les études les plus fréquentes, le potentiel de l'ozonation et des procédés à base de chlore à former des sous-produits de désinfection est souvent évalué (Gilca et al., 2020; Krasner et al., 2006; Richardson et al., 2000). En effet, de nombreux travaux ont montré que ces derniers ont un fort potentiel cancérigène ou mutagène (Cortés and Marcos, 2018; Richardson et al., 2007), qu'un grand nombre de ces TP reste à ce jour non-identifié et que ceux qui sont réglementés ne sont pas nécessairement les plus dangereux (Li et al., 2021b; Richardson and Plewa, 2020). Ces études sont souvent réalisées pour l'eau potable puisque la présence de TP peut avoir un impact direct sur la santé humaine, mais les observations effectuées sont également applicables pour les eaux usées qui contiennent généralement un nombre beaucoup plus important de composés et des concentrations plus élevées.

De façon générale, l'analyse de ces TP reste, pour le moment, compliquée puisque nombre d'entre eux ne sont pas disponibles commercialement, rendant ainsi leur analyse difficile ou imprécise, tandis qu'une grande partie des TP reste encore tout simplement inconnue. Enfin, les études de TP sont souvent réalisées dans des matrices simples et dans des conditions fixées afin de permettre une identification plus aisée de ceux-ci. Seulement, dans les eaux usées réelles, de nombreuses molécules interagissent, y compris avec la matière organique, et peuvent donc entraîner des phénomènes de compétition avec les oxydants et interférer avec les mécanismes de formation des TP.

II.2. Le NTS au service de l'évaluation des traitements d'élimination des MP organiques en STEU

Le NTS, en s'intéressant à l'ensemble des molécules présentes, permet d'apporter des informations plus complètes sur l'efficacité d'élimination des MP organiques par différents traitements. En plus de comparaisons très globales, le NTS permet également l'identification de TP formés naturellement ou après un procédé de traitement. Par exemple, des composés dont la présence n'a été observée qu'en laboratoire et pour lesquels l'analyse n'a jamais pu être effectuée dans l'environnement à cause de l'indisponibilité de standards analytiques, ont pu toutefois être analysés en mode SS (Kern et al., 2009). De la même façon, une autre étude a tenté d'identifier des centaines de molécules présentes en sortie du traitement secondaire d'une STEU. De nombreux polluants connus ont ainsi été retrouvés dans la liste résultante comme la carbamazépine, le diclofénac ou le diuron par exemple. Des molécules bien moins connues font également parties de cette liste, comme l'O-1821 ou la benzisoxazole acétamide oxime (Merel et al., 2015).

Ainsi, ce type d'analyse a également été utilisé afin d'évaluer l'influence de l'implémentation d'un traitement complémentaire en STEU. Le Tableau 40 présente les procédés étudiés ainsi que les stratégies de traitement de données mises en place pour ces études en NTS. Par comparaison avec les études en TS (Tableau 39), les études en NTS ont l'avantage de mettre en évidence la formation des TP. Par ailleurs, il est intéressant de noter la diversité des méthodes employées afin de caractériser ces TP, particulièrement en comparant leurs propriétés à celles des précurseurs (*i.e.*, aires de signaux relatifs, liens structurels, temps de rétention relatifs...). En effet, malgré une efficacité à éliminer les MP présents avant traitement complémentaire, une proportion non-négligeable des signaux détectés en fin de filière sont répertoriés comme étant des TP. Plus la caractérisation de ces marqueurs sera poussée, dans le meilleur cas allant jusqu'à une identification complète, plus il sera facile de déterminer le potentiel danger que ces TP représentent par rapport aux composés initiaux. Une évaluation plus poussée de l'efficacité du procédé est donc atteinte avec ce type d'analyse.

Tableau 40 : Comparaison des méthodes de traitement de données en NTS pour l'évaluation des procédés de traitement en STEU

Filière conventionnelle de la STEU étudiée / localisation	Traitement complémentaire appliqué	Méthodologie de traitement de données	Observations	Référence
Fossés d'oxydation / USA	UV (400, 900 and 1300 mJ/cm ²) / H ₂ O ₂ (0, 1.5 and 3 mg/L)	Diagrammes de Venn	1458 MP complètement éliminés, 939 TP spécifiques à l'UV et 2132 spécifiques à l'UV/H ₂ O ₂	(Merel et al., 2015)
Boues activées / Allemagne et Suisse	Filtration sur sable et filtration sur charbon actif en grain	Diagrammes de Venn	91 % des pics détectés en entrée de station ne sont pas retrouvés en sortie, avec la majorité éliminée après le traitement biologique 50 % de nouveaux signaux apparaissent après le traitement biologique Formation de 10 à 15 % de nouveaux signaux après traitement complémentaire	(Nürenberg et al., 2015, 2019)
Deux étapes successives de boues activées / USA	UV	Diagrammes de van Krevelen	Différence significative sur le nombre de double liaison entre les échantillons avant et après traitement complémentaire Pas de différence significative en termes de ratio atomique, mais diminution globale de l'intensité des signaux après traitement	(Maizel and Remucal, 2017)
Biofiltration, Nitrification et dénitrification / France	Adsorption sur micro-grain de charbon actif	ACP + loadings ou sélection des pics intenses	Certains composés pharmaceutiques et drogues sont réfractaires mais de nombreux analgésiques, antiviraux et pesticides sont éliminés	(Bergé et al., 2018)

Partie C : Application des méthodes pour la comparaison d'échantillons oxydés

Filière conventionnelle de la STEU étudiée / localisation	Traitement complémentaire appliqué	Méthodologie de traitement de données	Observations	Référence
Boues activées / Espagne	Filtration sur sable, UV et chloration	Diagrammes de Venn (non tracé) Diagramme de Kendrick avec CH ₂ , C ₂ H ₄ O, CF ₂ et (-H+Cl) en groupement caractéristiques	34 % de signaux en sortie de traitement complémentaires classifiés comme TP Pas d'évaluation du traitement complémentaire avec Kendrick, évaluation du traitement conventionnel montre une diminution du nombre de série homologues pour CH ₂ et C ₂ H ₄ O et apparition de nouvelles séries	(Verkh et al., 2018)
Système circulaire de boues activées / Chine	Irradiation solaire (lampe Xenon 1 700 W)	ACP pour distinguer des groupes de marqueurs avant et après oxydation puis établissement de liens de biotransformation entre précurseurs (avant) / TP (après) basé sur une différence de masse	1341 (15 %) TP, formés à partir de 2706 (31 %) précurseurs. L'ajout d'un atome d'oxygène est la transformation majoritaire observée (17 % des liens) suivis de la déalkylation (14 % des liens)	(Lian et al., 2020)
Boues activées /Suisse	Ozonation + filtration sur sable	A partir d'expériences en matrice simples sur 87 molécules modèles : identification de TP (par comparaison d'intensité, absence dans contrôle et forme des pics) pour la création d'une liste de suspects	Identification de 84 TP (6 au niveau 1, 6 niveaux 2, 71 niveaux 3 et 1 niveau 5) dont 58 % n'ont pas encore été répertoriés par d'autres études Abattement des TP meilleur par le charbon en poudre, puis le charbon en grain et faible pour la filtration sur sable	(Gulde et al., 2021a)
	Pré-ozonation + charbon actif en grain			
	Pré-ozonation + charbon actif en poudre			

Partie C : Application des méthodes pour la comparaison d'échantillons oxydés

Filière conventionnelle de la STEU étudiée / localisation	Traitement complémentaire appliqué	Méthodologie de traitement de données	Observations	Référence
Boues activées / Suisse	Ozone + charbon actif	Etablissement de tendances ou établissement de liens de biotransformation entre précurseurs / TP	80 % d'élimination après traitement biologique Plus de la moitié des marqueurs formés par l'ozonation sont éliminés par le charbon actif Dans l'effluent final : 50 % des marqueurs étaient présents dans l'influent de station + 50 % sont des TP	(Schollée et al., 2021)
Boues activées / Chine	UV (6,42 mW/cm ²) / H ₂ O ₂ , persulfate ou chlore (0,5 mM)	Diagrammes de Venn puis t-test ou Wilcoxon pour classer les marqueurs communs en 3 catégories (précurseurs, produits ou persistants). Etablissement de liens entre précurseurs et produits par une analyse des différences en masse. Diagrammes de van Krevelen	(UV + chlore) : augmentation du ratio O/C entre précurseurs et produits, avec identification de 22 produits chlorés (UV + H ₂ O ₂) et (UV + PS) forment des produits plus oxygénés que (UV + chlore) Réaction les plus communes sont l'addition de 3 atomes d'oxygène en (UV/H ₂ O ₂) et la déalkylation ou la réduction en général pour les trois procédés	(Zhang et al., 2021)

Deux grandes étapes peuvent ainsi être distinguées dans le traitement de données par NTS permettant l'évaluation des traitements utilisés en STEU, et, plus particulièrement pour la caractérisation de TP. La première consiste à appliquer des méthodes permettant de comparer les échantillons avant et après traitement afin de repérer des marqueurs discriminants. Ces derniers peuvent être spécifiques d'un groupe donné (*i.e.*, présents uniquement dans les échantillons de ce groupe), avoir subi une variation significative en termes d'intensité ou encore présenter des caractéristiques particulières (par exemple une masse, un défaut de masse, un fragment ou un massif isotopique caractéristique en fonction du cas d'étude). La seconde étape consiste quant à elle à caractériser les TP qui ont préalablement été repérés par la première étape, de façon plus ou moins poussée (jusqu'à potentiellement l'identification de la structure moléculaire précise).

II.2.1. Comparaisons basées sur les intensités des signaux

Un premier type de comparaison permettant de repérer des marqueurs discriminants s'appuie sur des méthodes statistiques. Parmi elles, l'ACP est souvent utilisée pour distinguer les échantillons avant et après le traitement complémentaire appliqué. Dans un premier temps, elle permet de voir s'il existe réellement une différence significative entre les échantillons avant et après traitement. Lorsque c'est le cas, l'étude du *loadings plot* permet ensuite de repérer les marqueurs permettant de discriminer le plus ces deux groupes d'échantillons. Ainsi, les marqueurs du groupe avant traitement correspondent à des composés très bien éliminés par le traitement, alors que les marqueurs discriminants du second groupe correspondront à des TP régulièrement retrouvés (Bergé et al., 2018; Lian et al., 2020).

De manière relativement similaire, des statistiques univariées (t-test ou Wilcoxon) sont également retrouvées. Ces tests se concentrent sur la comparaison des intensités relatives des marqueurs présents dans les deux groupes d'échantillons. Une variation significative positive sera associée à un TP, alors qu'un produit sera caractérisé par une variation significative négative et une variation non significative correspondra à un composé persistant (Zhang et al., 2021). Parmi les méthodes permettant de repérer des signaux correspondant à des TP, un autre type de méthode statistique repose sur l'utilisation de séries temporelles d'échantillons. Bien que ces méthodes n'aient pas ou peu été utilisées pour évaluer les traitements complémentaires d'eaux usées, en particulier en conditions pilote ou réelles, elles

permettent d'identifier des tendances pouvant être associées à l'apparition et la persistance des TP. Ainsi, des méthodes de corrélation sur les rangs (Plassmann et al., 2016) ou encore des relations log linéaires (Veenaas et al., 2018) ont été répertoriées.

Cependant, un des inconvénients majeurs de ce type de méthodes est qu'elles reposent sur des tests statistiques nécessitant un nombre minimum d'échantillons pour pouvoir dégager des tendances significatives, ce qui, dans certains cas n'est pas toujours possible. Ainsi, d'autres méthodes reposant sur des principes similaires sont retrouvées. Elles sont moins contraignantes en termes de nombre d'échantillons disponibles puisqu'il n'y a plus de notion de significativité statistique, mais plus le nombre d'échantillon pris en compte sera élevé, plus les marqueurs identifiés comme discriminants seront représentatifs du groupe considéré. Parmi ces méthodes, les diagrammes de Venn ou d'Euler sont les plus fréquemment retrouvés. Ils permettent de repérer les marqueurs spécifiques aux échantillons avant traitement, c'est-à-dire ceux qui sont éliminés, des marqueurs spécifiques aux échantillons après traitement, à savoir les TP (Merel et al., 2015; Nürenberg et al., 2015, 2019). Pour ce qui est des marqueurs communs, ils peuvent ensuite être distingués en sous-catégories en fonction de l'évolution de leur intensité relative (Zhang et al., 2021). Pour réaliser cette distinction, une méthode consiste aussi à calculer le ratio de l'intensité du marqueur après traitement par rapport à son intensité avant traitement. En fonction de la valeur obtenue, le marqueur est classé comme éliminé, partiellement abattu, persistant, ou formé (Bader et al., 2017). Bien que cette méthode ait été mise au point dans le cas de l'évaluation des traitements d'eau potable, sa transposition sur des eaux usées est tout à fait appropriée.

II.2.2. Comparaisons basées sur les masses détectées des signaux

Le second type de comparaison s'appuie sur la prise en compte des masses détectées liées à chaque signal. Bien qu'elle puisse également être utilisée de manière totalement indépendante, cette comparaison intervient souvent à la suite de celle qui a été décrite dans le paragraphe précédent et qui aura permis de réduire le set de données considéré aux signaux correspondants aux précurseurs et TP potentiels. La comparaison a ainsi pour but d'établir un lien entre ces signaux afin d'apporter des connaissances, autant en termes de mécanismes de transformation que de composés plus enclins à former des TP par différents procédés. Pour cela, différentes stratégies peuvent être mises en place.

La première consiste à s'appuyer sur les défauts de masse qui représentent la différence entre la masse exacte et la masse nominale d'un ion. En effet, ce défaut de masse est caractéristique de la composition élémentaire d'un composé et les mécanismes de réactions usuels (*e.g.*, oxydation, déshydratation, réduction, hydroxylation...) entraînent un changement de défaut de masse caractéristique entre le précurseur et son TP. De la même manière, lorsque des données de fragmentation sont disponibles, une stratégie similaire peut être appliquée en se basant sur le fait que les précurseurs et TP ont des structures similaires. Ainsi, leurs spectres de fragmentation devraient également être liés. Ces différentes stratégies analytiques ont récemment été décrites de manière extensive (Li et al., 2021a). Une méthode similaire consiste à s'appuyer sur les diagrammes de Kendrick, initialement développés pour caractériser différentes familles d'hydrocarbures (Kendrick, 1963). Cette méthode s'appuie sur la conversion des masses IUPAC vers une autre échelle choisie, représentative des composés ou mécanismes étudiés. Dans sa version initiale, cette échelle est basée sur le groupement CH₂, mais pour étudier les TP, celle-ci peut être adaptée aux mécanismes de réactions usuels mettant simplement en jeu l'addition, la perte ou la substitution d'un groupement caractéristique. Cette conversion permettra ainsi de tracer les nouveaux défauts de masses en fonction des nouvelles masses nominales et d'aligner horizontalement les potentiels couples précurseurs-TP. Cette méthode est régulièrement retrouvée dans les études d'oxydation de la matière organique (Maizel and Remucal, 2017; Verkh et al., 2018) et pour l'identification de TP halogénés (Jobst et al., 2013; Zhang et al., 2014). Suivant la même logique, certaines méthodes reposent sur l'utilisation de mécanismes connus afin de calculer les masses correspondantes aux TP qui pourraient être formés par ceux-ci, à partir des composés ayant été classifiés comme éliminés ou abattus par le traitement. Ces masses de TP théoriques sont ensuite recherchées parmi la liste des signaux classés comme formés afin d'établir ce lien (Gulde et al., 2021a; Schollée et al., 2015, 2021). Enfin certaines études relatent de l'utilisation des diagrammes de van Krevelen afin d'observer de potentielles modifications de la composition des eaux par le traitement. La comparaison s'appuie généralement sur les ratios atomiques et le nombre d'insaturations, permettant ainsi de donner une idée d'une potentielle modification du degré d'insaturation des molécules présentes ou de la diversité d'hétéroatomes par le traitement appliqué (Maizel and Remucal, 2017; Zhang et al., 2021).

III. Applications de méthodes à l'évaluation de procédés d'oxydation et de la formation de TP

III.1. Cas de l'acide performique

En vue de l'organisation des Jeux Olympiques de 2024 à Paris, le service public de l'assainissement parisien (SIAAP) a étudié le potentiel de l'acide performique pour désinfecter les effluents de STEU avant leur rejet en Seine afin d'améliorer la qualité de ce dernier, principalement d'un point de vue bactériologique. Cet oxydant innovant a en effet montré sa capacité à éliminer les microorganismes pathogènes présents dans les eaux de STEU. En revanche, relativement peu d'informations sont disponibles quant à son impact sur les MP présents dans l'effluent.

Dans le cadre d'une collaboration avec le SIAAP, nous avons effectué une étude préliminaire sur ces essais de désinfection afin d'évaluer le potentiel de formation de sous-produits de désinfection (*e.g.*, trihalométhanes, haloacétonitriles et N-nitrosamines) de l'acide performique, et de caractériser l'évolution globale des effluents soumis à ce traitement, à l'échelle du laboratoire et à l'échelle industrielle. Ces résultats sont publiés dans un ouvrage dédié à la désinfection par l'acide performique :

Vincent Rocher, Sam Azimi, Romain Mailler, Sabrina Rechdaoui-Guérin, Perrine Mèche, Sébastien Pichon, Angélique Goffin, Jean Bernier, Jérôme Roy, Gilles Varrault, Julien Le Roux, Nina Huynh, Maolida Nihemaiti, Thierry Pigot, Thomas Paulin, Jean-Marie Mouchel, David Du Pasquier, Laurent Paulic, Dan Angelescu, Vaizanne Huynh, Andreas Hausot, Patrizia Ragazzo, Nicoletta Chiucchini. Effectiveness of Disinfecting Wastewater Treatment Plant Discharges: Case of chemical disinfection using performic acid. IWA Publishing, 2021, 9781789062090; et plus particulièrement dans le Chapitre 4 intitulé : The fate of micropollutants and byproduct formation during the disinfection of WWTP discharge by PFA . <10.2166/9781789062106_0039>

Afin d'approfondir ce travail, une seconde étude a été consacrée à la compréhension des mécanismes de réaction de l'acide performique avec les MP et est présentée dans cette partie

III.1. Ce travail a fait l'objet d'une publication en cours de relecture :

Maolida Nihemaiti, Nina Huynh, Romain Mailler, Perrine Mèche-Ananit, Vincent Rocher, Rachid Barhdadi, Régis Moilleron, Julien Le Roux. High-resolution mass spectrometry screening of wastewater effluent for micropollutants and their transformation products during disinfection with performic acid. – Soumise à ACS ES&T Water.

III.1.1. Introduction

Les peracides tels que l'acide peracétique (PAA) et l'acide performique (PFA) sont largement utilisés en tant qu'agents oxydants en production chimique, comme désinfectants dans la chaîne alimentaire ou le domaine médical ou encore en tant que biocides dans l'industrie (Heinonen-Tanski and Miettinen, 2010; Kitis, 2004; Luukkonen and Pehkonen, 2017). Le PAA a également été utilisé dans le traitement des eaux usées dès la fin des années 80 comme alternative aux désinfectants à base de chlore (Kitis, 2004). Plus récemment, le PFA a été étudié en tant que désinfectant primaire (Gehr et al., 2009), secondaire (Karpova et al., 2013; Ragazzo et al., 2013) ou tertiaire (Luukkonen et al., 2015) des effluents d'eaux usées. Différentes villes en Europe (*e.g.*, Venise, Paris, Berlin) appliquent actuellement cet oxydant en conditions réelles pour la désinfection des effluents de STEU avant leur rejet vers l'environnement (Ragazzo et al., 2020; Rocher et al., 2021). Le PFA a démontré sa grande efficacité à inactiver les organismes pathogènes tels que *Escherichia coli* ou les entérocoques (Campo et al., 2020; Chhetri et al., 2014; Luukkonen et al., 2015; Maffettone et al., 2020). Généralement, de faibles doses (0,8 mg/L, pendant 18 min) suffisent pour réduire significativement la concentration d'*Escherichia coli* dans les effluents de STEU (Ragazzo et al., 2020).

Contrairement au PAA, le PFA est moins stable et est donc produit directement sur site avant utilisation. Cette production met en jeu une réaction réversible entre le peroxyde d'hydrogène (H₂O₂) et l'acide formique (AF)(équation 2) et peut être catalysée par des acides minéraux tels que l'acide sulfurique (Filippis et al., 2009).



Un effluent de STEU désinfecté par le PFA n'a montré aucun effet génotoxique ou mutagène à de faibles doses (0,6-1,5 mg/L) (Ragazzo et al., 2017). De manière cohérente, la quantité de composés organiques halogénés adsorbables (AOX) formés après désinfection par le PFA est bien plus faible que celle retrouvée après désinfection par le chlore ou le PAA dans des conditions similaires (Karpova et al., 2013). Par ailleurs, la désinfection par le PFA (1-3mg/L, pendant 10-30 min) d'effluents de STEU a montré un faible potentiel de formation de trihalométhanes, d'haloacétonitriles ou encore de N-nitrosamines (Ragazzo et al., 2020; Rocher et al., 2021). Cependant, peu d'études se sont concentrées sur la caractérisation chimique globale des eaux usées, en particulier sur le devenir des MP organiques soumis à un traitement par PFA.

Les traitements conventionnels appliqués en STEU sont généralement peu efficaces pour l'élimination de micropolluants organiques (Eggen et al., 2014), qui peuvent donc être transformés par les étapes de désinfection suivantes. Le TS et le SS sont des approches communément utilisées pour l'analyse de MP et de leurs TP connus (Bourgin et al., 2018; Deeb et al., 2017). Cependant, le nombre de composés étudiés par ces méthodes est souvent limité par la disponibilité d'étalons ou de spectres de références (Schymanski et al., 2014b). Récemment, le NTS a gagné en popularité pour sa capacité à caractériser de façon exhaustive les composés organiques présents dans des échantillons aqueux complexes (Gago-Ferrero et al., 2015; Schymanski et al., 2015). La haute précision en masse, le massif isotopique bien défini et les fragments obtenus lors de telles analyses permettent de tenter d'identifier des composés inconnus (Hollender et al., 2017). De plus, différentes approches statistiques peuvent être utilisées pour interpréter ces données afin de prioriser les signaux d'intérêt (Hollender et al., 2017). De telles approches (*e.g.*, ACP) ont été utilisées afin d'étudier les traitements d'eaux usées (biodégradation, ozonation) (Schollée et al., 2015, 2018) ou encore d'eau potable (procédés d'oxydation avancée) (Brunner et al., 2020), et plus particulièrement leur influence sur les MP et leur capacité à former des TP.

Le but de cette étude est d'étudier l'effet du traitement par PFA sur la composition chimique de l'effluent de STEU. Ce dernier a été soumis à différentes doses de PFA et analysé par LC-HRMS. Dans un premier temps, une analyse par NTS a été menée afin de caractériser les échantillons dans leur ensemble. Une AMV a ensuite été utilisée afin de prioriser les marqueurs discriminants du traitement. Parmi ces marqueurs, la lidocaïne *N*-oxyde a été

identifiée comme un TP formé durant le traitement. Ainsi, une étude en SS a ensuite été menée sur une sélection d'amines tertiaires et de leurs *N*-oxydes correspondantes. Un travail complémentaire en conditions contrôlées a été effectué afin de confirmer, sur des molécules cibles, que le PFA pouvait effectivement conduire à la formation de *N*-oxydes et ainsi permettre d'apporter des connaissances quant à son mécanisme d'action.

III.1.2. Matériel et méthodes

III.1.2.1. Réactifs

L'amisulpride, la clarithromycine, la lidocaïne, le tramadol, la clarithromycine *N*-oxide, l'acide sulfurique (98 %) et le peroxyde d'hydrogène (30 %) ont été achetés chez Sigma-Aldrich. Les autres *N*-oxydes correspondantes aux composés précédemment énoncés ont été obtenus chez Toronto Research Chemicals. Les solvants et additifs relatifs aux analyses ont été achetés au grade ULC-MS auprès de Biosolve.

Le PFA a été préparé suivant un protocole développé par la société Kemira (KemConnect™ DEX) et détaillé dans un précédent travail (Rocher et al., 2021). Brièvement, 10 mL d'AF (99 %) ont été acidifiés par dopage de 1 mL d'acide sulfurique (98 %). 3,5 mL d'H₂O₂ ont ensuite été ajoutés lentement à 1,5 mL d'AF acidifié et le mélange a été mis sous agitation pendant 90 min dans un bain de glace. La concentration de PFA de la solution mère a été déterminée par titrations successives. Le peroxyde d'hydrogène résiduel a tout d'abord été titré par une solution de KMnO₄ en conditions acides, puis l'ajout de KI a permis d'amorcer une réaction avec le PFA pour produire de l'iode, qui après ajout d'amidon, réagira pour former un complexe bleu avec ce dernier. La seconde étape permet ainsi la titration de l'iode par le biais du thiosulfate de sodium jusqu'à décoloration complète. De façon générale, les concentrations du PFA et de l'H₂O₂ à l'équilibre étaient de l'ordre de 10 % et 17 % en masse respectivement, ce qui est cohérent par comparaison avec des études similaires (Gehr et al., 2009; Karpova et al., 2013). Lors de l'oxydation de molécules individuelles, le PFA résiduel a été déterminé suivant une méthode colorimétrique à l'acide 2,2'-azino-bis(3-éthylbenzothiazoline-6-sulphonique) (ABTS). Brièvement, 1 mL du milieu réactionnel (PFA dilué dans l'EUP (1/10, v/v)) a été mis sous agitation avec 1 mL d'acide acétique et 1 mL d'une solution d'ABTS à 0,1 g/L pendant 20 min dans le noir. Cette solution a ensuite été analysée par spectrophotométrie UV à une longueur d'onde de 415 nm.

III.1.2.2. Traitement des rejets de STEU à l'acide performique

Les désinfections à l'échelle laboratoire et en conditions réelles ont été effectuées sur le rejet de la STEU de Seine Amont (Valenton, France). Celle-ci a une capacité de traitement en temps sec de 600 000 m³/j et les traitements effectués incluent un dégrillage et dessablage, suivis d'une décantation primaire et d'un traitement par boues activées. Ce dernier est divisé en deux voies, l'une fonctionnant par nitrification, dénitrification puis élimination physico-chimique du phosphore, tandis que l'autre procède par une déphosphatation biologique, puis nitrification et dénitrification. Les désinfections au PFA en laboratoire ont été réalisées sur des échantillons ponctuels (n=3) collectés en sortie du traitement biologique (mélange des deux voies) entre septembre et décembre 2018. Le traitement au PFA a été effectué au plus tard 2 h après échantillonnage, sur 2 L de rejet sous agitation constante pendant 10 min. La dose initiale de PFA ajoutée est comprise entre 1 et 100 mg/L. En ce qui concerne les désinfections en conditions réelles, celles-ci ont été effectuées à des doses comprises entre 0,8 et 2,5 mg/L de PFA (entre 28,3 et 74,2 ppm/min). Le PFA résiduel a été éliminé par ajout d'un excès de thiosulfate de sodium.

Ces échantillons ont été filtrés dans les 24 h sur 0,7 µm (GF/F, Whatman), acidifiés à pH 6,5 et dopés avec un mélange d'étalons internes (bisphénol A-d6, 4-n-octylphénol-d17, 4-octylphénol-diéthoxylate, et propylparaben-d4). Les échantillons ont ensuite été concentrés par SPE (Dionex AutoTrace, Thermo Scientific) à l'aide de cartouches multiphasiques assemblées manuellement (200 mg Oasis HLB, 150 mg Isolute ENV+, 100 mg Strata X-AW, 100 mg Strata X-CW). Ces cartouches SPE ont été conditionnées avec 10 mL de MeOH, suivis par 10 mL d'EUP. Après chargement des échantillons, les cartouches ont été séchées sous flux d'azote pendant 30 min et l'élution a été effectuée successivement avec 6 mL d'un mélange MeOH/AcEt (50/50, v/v) basifié avec 1,4 % d'une solution d'ammonium (35 %), puis 3 mL d'un mélange MeOH/AcEt (50/50, v/v) acidifié avec 1,7 % d'AF (98 %). L'extrait résultant a été évaporé à sec sous flux d'azote et reconstitué avec 1 mL d'un mélange EUP/MeOH (80/20, v/v). Ces échantillons finaux ont été filtrés sur 0,2 µm (PTFE) avant analyse.

III.1.2.3. Oxydation de molécules individuelles par l'acide performique

Les oxydations de molécules individuelles ont été effectuées dans des flacons ambrés, à température ambiante (22 ± 2°C). Des solutions de chaque molécule (lidocaïne, amisulpride, tramadol et clarithromycine) ont été préparées à 5 µM dans des solutions tampons

phosphatées à 50 mM (pH = 6,2-8) ou boratées (pH = 9). Les oxydations ont été amorcées par ajout d'un volume défini de la solution mère de PFA dans ces différentes solutions tampons de sorte à avoir des concentrations initiales de PFA entre 1 et 100 mg/L. Des aliquots du milieu réactionnel ont été prélevés à des intervalles de temps définis et la réaction a été arrêtée par ajout d'un excès de thiosulfate de sodium. Ces échantillons ont ensuite été directement transférés dans des vials pour analyse dans les 24 h suivantes au plus tard.

III.1.2.4. Analyse et traitement de données

Les données HRMS ont été obtenues par UPLC-IMS-QToF en utilisant une colonne C18, selon la méthode décrite en partie B – III.4.1. Les données ont été prétraitées par UNIFI et la *marker table* résultante a été exportée, puis traitées sous R selon les méthodes décrites en partie A. L'analyse en SS a été effectuée directement sous UNIFI, à partir des structures importées dans la librairie interne afin de générer une fragmentation *in silico*. Pour la lidocaïne, l'amisulpride, le tramadol et la clarithromycine, ainsi que leurs *N*-oxydes, les étalons correspondants ont été acquis et injectés a posteriori pour confirmation des identités attribuées, sur la base de leur *m/z*, TR, DT, CCS et fragments selon les tolérances définies en partie A – III.3.2.

En ce qui concerne les études d'oxydation de molécules individuelles, la lidocaïne, l'amisulpride et le tramadol, ainsi que leurs *N*-oxydes respectives, ont été quantifiées par HPLC couplée à un détecteur à barrette d'iode (HPLC DAD, SPD-M20A, Shimadzu). La séparation a été effectuée avec une colonne Ascentis Express C18 (100 x 4,6 mm, 2,7 µm, Supelco) à un débit de 1 mL/min, avec des phases mobiles composées de MeOH (A) et d'un tampon phosphate (10 mM, pH 2,3) (B). Les détails des conditions analytiques pour chaque composé sont résumés dans le Tableau 41.

Tableau 41 : Conditions pour l'analyse des composés pharmaceutiques et leurs *N*-oxydes par HPLC-DAD

Composé	Phase mobile (% de A / % de B)	Longueur d'onde (nm)
Lidocaïne	15 / 85	196
Lidocaïne <i>N</i> -oxyde	25 / 75	196
Tramadol	25 / 75	196
Tramadol <i>N</i> -oxyde	60 / 40	196
Amisulpride	25 / 75	225
Amisulpride <i>N</i> -oxyde	30 / 70	225

III.1.3. Résultats

III.1.3.1. NTS des rejets de STEU soumis à l'oxydation par PFA

Le NTS a été utilisé en première approche pour la caractérisation du rejet de STEU oxydé par le PFA. L'impact global de cette oxydation a été évalué par la classification des marqueurs selon les 5 catégories suivantes : marqueurs *éliminés*, marqueurs partiellement *abattus*, marqueurs *persistants*, marqueurs montrant une augmentation d'aire et marqueurs *formés* (Tableau 42). Ces catégories ont été déterminées en calculant la variation des aires (fc) de chaque marqueur suite à l'oxydation, selon une méthode décrite précédemment (Bader et al., 2017).

Tableau 42 : Classification des marqueurs selon leur évolution après oxydation par PFA

Catégorie*	ESI+			ESI-		
	Marqueurs (%)	m/z moyen	TR moyen (min)	Marqueurs (%)	m/z moyen	TR moyen (min)
<i>Éliminés</i> ($0,0 \leq fc < 0,2$)	27	549,34	11,02	19	441,81	9,76
<i>Partiellement abattus</i> ($0,2 \leq fc < 0,5$)	12	466,14	9,42	9	425,05	10,13
<i>Persistant</i> ($0,5 \leq fc < 2,0$)	34	339,40	7,56	46	382,40	8,43
<i>Augmentation d'aire</i> ($2,0 \leq fc < 5,0$)	1	378,27	9,26	2	383,46	9,15
<i>Formés</i> ($5,0 \leq fc$)	26	370,06	9,20	24	397,58	8,16

*fc représente le ratio entre l'aire du marqueur après oxydation et l'aire de ce même marqueur avant oxydation

Environ 50 % des marqueurs présents initialement dans l'effluent de STEU en ESI+ et ESI- ont été retrouvés après oxydation à 2 mg/L de PFA pendant 10 min. 27 % et 19 % des marqueurs en ESI+ et ESI- respectivement, ne sont plus détectables après traitement, alors que 26 % et 24 % en ESI+ et ESI- respectivement, ont été détectés seulement dans l'échantillon après oxydation. Les marqueurs éliminés étaient responsables de 25 % (8 % en ESI-) de l'aire totale du signal avant oxydation, alors que les marqueurs générés par le traitement sont porteurs de

seulement 14 % (6 % en ESI-) de l'aire totale du signal après oxydation. Par ailleurs, parmi les marqueurs présents à la fois dans les échantillons avant et après oxydation, 12 % (9 % en ESI-) ont vu leur aire diminuer d'un pourcentage compris entre 50 et 80 % à la suite du traitement. Par conséquent, l'aire totale du signal a été réduite d'un facteur 2 à 4 par l'oxydation au PFA, suggérant l'efficacité de ce dernier à éliminer, au moins partiellement, les MP en engendrant une formation limitée de TP.

Les m/z moyens et TR moyens des marqueurs *éliminés* sont globalement plus élevés que ceux des marqueurs *formés*, suggérant la formation de composés plus petits et plus polaires, à partir de marqueurs plus gros et moins polaires. Les marqueurs plus petits et plus polaires présents initialement semblent avoir été peu influencés par le PFA puisqu'ils sont majoritairement classés comme persistants. De façon générale, ces résultats suggèrent que le PFA, même à faible dose (2 mg/L, pendant 10 min), a une influence sur la composition de la matière organique présente dans les effluents de STEU.

Pour aller plus loin, les marqueurs résultant du NTS des effluents de STEU soumis à cet oxydant ($n = 12$) ont été utilisés pour réaliser une OPLS-DA en considérant des traitements entre 0,8 et 2,5 mg/L de PFA. Le *s-plot* résultant de cette régression a été généré afin de repérer les marqueurs responsables de la discrimination entre les échantillons avant et après oxydation (Tableau 37).

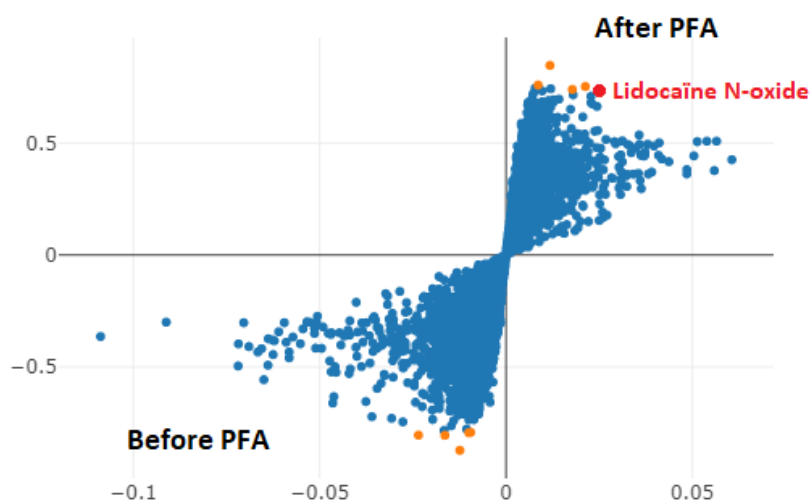


Figure 39 : S-plot correspondant à l'OPLS-DA entre les rejets de STEU ($n = 6$, injectés en triplicat) et les rejets oxydés ($n = 6$, injectés en triplicat) soumis à des doses comprises entre 0,8 et 2,5 mg/L de PFA pendant 10 min. Les marqueurs les plus discriminants sélectionnés apparaissent en orange et sont disponibles en Annexe 5A.

Les cinq marqueurs avec les plus grands couples de $p[1]/p[corr][1]$ pour chaque groupe (avant ou après oxydation) ont été sélectionnés pour une tentative d'identification en utilisant différentes bibliothèques externes (MassBank, Sigma-Aldrich, LGC Standard, NIST et Drugbank intégrées dans UNIFI ainsi que ChemSpider et ForIdent). Parmi ces marqueurs, la lidocaïne *N*-oxyde a été identifiée comme marqueur permettant de discriminer les échantillons après PFA. L'identification était basée sur la correspondance de la masse monoisotopique, du massif isotopique et des fragments avec plusieurs bibliothèques interrogées. La présence de la lidocaïne dans les échantillons non-oxydés ainsi que l'identité de sa *N*-oxyde ont par la suite été confirmées par achat, injection et correspondance avec un standard analytique. Par comparaison des échantillons, l'augmentation de l'aire de la lidocaïne *N*-oxyde peut être observée avec l'augmentation de la dose de PFA (sauf à 100 mg/L) alors que l'aire de la lidocaïne tend à diminuer (Figure 40), suggérant la transformation de cette dernière en sa *N*-oxyde à la suite du traitement par PFA.

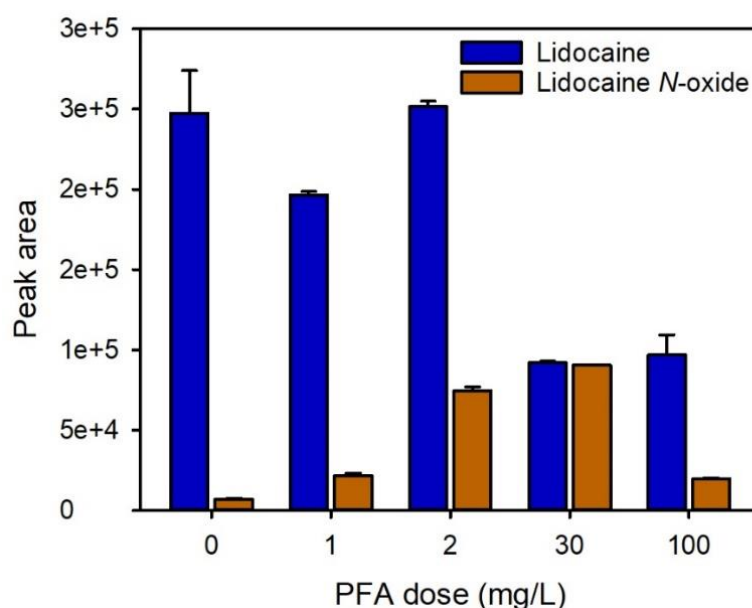


Figure 40 : Evolution des aires de la lidocaïne et de la lidocaïne *N*-oxyde dans les rejets de STEU soumis à différentes doses de PFA. Les barres d'erreur correspondent à l'écart-type entre les triplicats d'injection.

III.1.3.2. SS de N-oxydes et de leurs précurseurs dans les rejets de STEU soumis à l'oxydation au PFA

Les *N*-oxydes sont des TP majoritaires d'amines tertiaires déprotonées, retrouvées suite à des transformations biologiques ou à une oxydation par ozonation (Gulde et al., 2016; Merel et al., 2017). Des études ont montré que ces composés étaient moins biodégradables par comparaison à leurs précurseurs (Bourgin et al., 2018; Hübner et al., 2015). Afin de confirmer la réactivité du PFA avec les amines tertiaires et leur potentiel à former des *N*-oxydes-, une étude en SS a été menée sur une liste comprenant 7 MP avec une amine tertiaire et leurs *N*-oxydes respectives dans les échantillons traités au PFA (Tableau 43).

Les résultats montrent une diminution des aires de pics pour tous les précurseurs suite au traitement par PFA (2 mg/L, 10 min) alors que l'inverse est observable pour leur *N*-oxyde (sauf la clarithromycine *N*-oxyde et le tramadol *N*-oxyde), suggérant que le PFA permet la formation de ces TP.

Tableau 43 : Détection de sept amines tertiaires et de leurs N-oxydes respectives en SS dans les rejets de STEU traités par le PFA (2 mg/L, 20 min). Les aires sont moyennées sur les triplicats d'injection, normalisées par un étalon interne (propylparaben-d4)

	Formule [M+H]	m/z [M+H]	TR (min)	CCS (Å ²)	Niveau de confiance	Aire dans le rejet de STEU	Aire dans le rejet oxydé à 2 ppm de PFA
Précurseurs							
Amisulpride	C ₁₇ H ₂₈ N ₃ O ₄ S	370,1796	4,78	193,03	1	1 708 962	1 310 929
Citalopram	C ₂₀ H ₂₂ FN ₂ O	325,1710	8,49	179,57	3	285 966	237 995
Clarithromycine	C ₃₈ H ₇₀ NO ₁₃	748,4840	10,42	269,78	1	351 689	243 640
Lidocaine	C ₁₄ H ₂₃ N ₂ O	235,1804	4,81	156,31	1	12 119	8 611
Tiapride	C ₁₅ H ₂₅ N ₂ O ₄ S	329,1530	3,65	182,97	3	263 427	261 961
Tramadol	C ₁₆ H ₂₆ NO ₂	264,1954	5,73	160,75	1	331 125	273 911
Venlafaxine	C ₁₇ H ₂₈ NO ₂	278,2114	7,11	170,73	3	122 235	117 984
N-oxydes							
Amisulpride N-oxyde	C ₁₇ H ₂₈ N ₃ O ₅ S	386,1740	5,21	194,18	1	5 925	10 508
Citalopram N-oxyde	C ₂₀ H ₂₂ FN ₂ O ₂	341,1672	8,78	183,46	3	-	-
Clarithromycine N-oxyde	C ₃₈ H ₇₀ NO ₁₄	764,4794	8,35	270,56	1	181 042	131 343
Lidocaine N-oxyde	C ₁₄ H ₂₃ N ₂ O ₂	251,1750	5,78	157,33	1	10 324	15 275
Tiapride N-oxyde	C ₁₅ H ₂₅ N ₂ O ₅ S	345,1449	4,26	184,01	3	7 553	77 512
Tramadol N-oxyde	C ₁₆ H ₂₆ NO ₃	280,1905	6,08	165,54	1	18 814	14 465
Venlafaxine N-oxyde	C ₁₇ H ₂₈ NO ₃	294,2060	9,18	173,43	3	10 880	12 612

III.1.3.3. Etude des mécanismes de transformation de MP soumis au PFA

Des expériences supplémentaires ont été réalisées en solution tampon afin d'étudier les mécanismes de réaction du PFA sur une sélection de MP.

Lidocaïne. Des manipulations contrôles ont été réalisées afin d'observer l'effet de l' H_2O_2 sur la lidocaïne. Ces expériences préliminaires ont montré qu'à une dose de 33 mg/L de H_2O_2 (généralement retrouvés pour 10 mg/L de PFA) aucun effet n'était observable pour des pH compris entre 6 et 9 Annexe 5B. La lidocaïne a été progressivement dégradée selon une cinétique de premier ordre pour différentes doses de PFA appliquées (Figure 41).

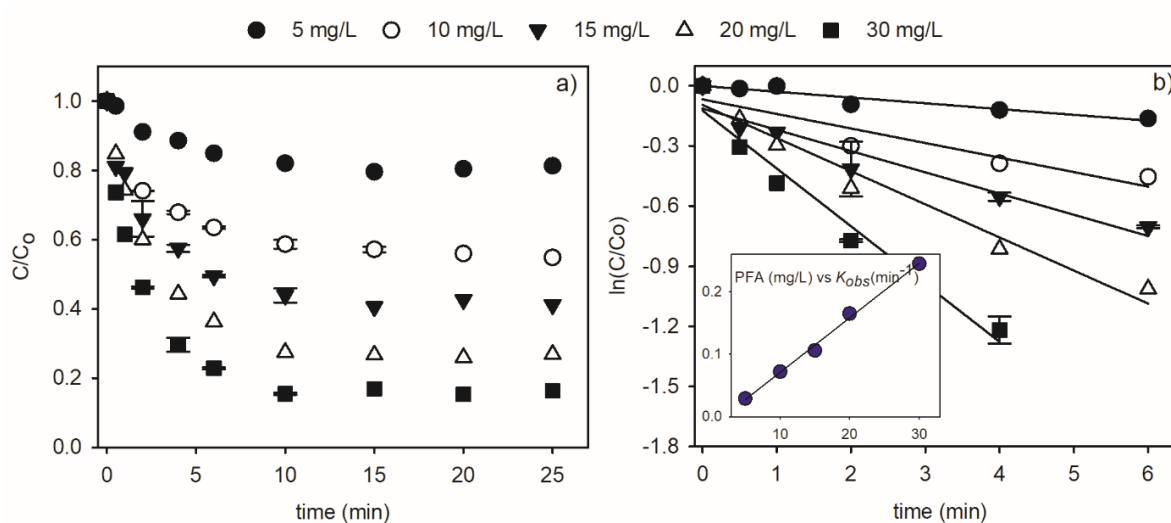


Figure 41 : Cinétique de dégradation de la lidocaïne par oxydation au PFA (concentration initiale de lidocaïne à 5 μM , PFA entre 5 et 50 mg/L dans un tampon phosphate à 50 mM pH 8).

Environ 20 % de lidocaïne a pu être éliminé en 10 min avec une dose de 5 mg/L de PFA à pH 8, alors que dans les mêmes conditions mais en multipliant la dose de PFA par 6, 80 % d'élimination a été observée. La constante de vitesse associée k_{obs} (min^{-1}), dérivée de la pente du tracé de $\ln(C/C_0)$ en fonction du temps (Figure 41b) montre une forte corrélation avec la concentration initiale de PFA (Figure 41b), suggérant que la réaction globale peut être décrite par un modèle cinétique de second ordre (équations 3) avec des constantes de vitesse apparente du second ordre de $7,54 \pm 0,90 \text{ M}^{-1} \cdot \text{s}^{-1}$ à pH 8.

$$\frac{d[\textit{lidocaine}]}{dt} = -k_{obs}[\textit{lidocaine}] \quad \text{eq. 3a}$$

$$\frac{d[\textit{lidocaine}]}{dt} = -k_{app}[\textit{lidocaine}][\textit{PFA}] \quad \text{eq. 3b}$$

La lidocaïne montre une dégradation plus rapide à pH 8 par rapport à pH 7 (i.e., $k_{app} = 2,76 \pm 0,37 \text{ M}^{-1}\cdot\text{s}^{-1}$, Annexe 5C). Cette observation peut être attribuée à une plus grande proportion de lidocaïne déprotonée au premier pH ($pK_a = 7,75$; www.chemaxon.com, consulté le 06/2021). La lidocaïne n'a pas pu être dégradée davantage au-delà de 15 min de réaction (Figure 41). Ceci s'explique par l'insuffisance de PFA résiduel en solution puisque seulement 5 % de la quantité de PFA initiale (10 mg/L) ont pu être quantifiés après 20 min de réaction à pH 8. Outre la consommation du PFA due à la réaction avec la lidocaïne, cette diminution de la concentration de l'oxydant peut aussi être attribuée à sa propre décomposition. En effet, plusieurs études ont montré que le PFA se transformait rapidement en CO_2 et H_2O en solution aqueuse (Filippis et al., 2009; Santacesaria et al., 2017; Sun et al., 2011).

La lidocaïne *N*-oxyde est, quant à elle, formée progressivement avec la dégradation de la lidocaïne (Annexe 5D). La Figure 42 montre les concentrations respectives de la lidocaïne et de sa *N*-oxyde au bout de 10 min de réaction pour différentes doses de PFA.

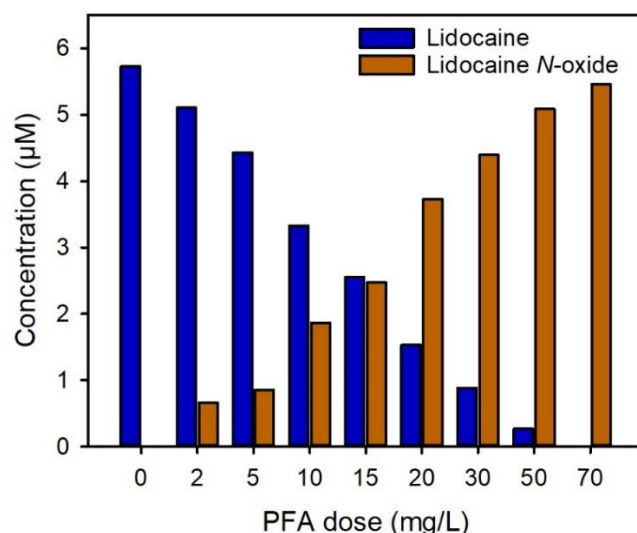


Figure 42 : Concentration de lidocaïne et de sa *N*-oxyde au bout de 10 min de réaction dans un tampon phosphate à 50 mM à pH 8

La somme des concentrations molaires des deux composés est relativement stable malgré différentes doses initiales de PFA employées, montrant que la lidocaïne est majoritairement (~95 %) transformée en sa *N*-oxyde. La présence de 0,1 M de tert-butanol n'a eu aucun impact sur la dégradation de la lidocaïne ou la formation de *N*-oxyde, indiquant que les radicaux hydroxylés ne jouent aucun rôle dans cette réaction Annexe 5E. De plus, le PFA (1-100 mg/L) semble ne pas réagir avec la lidocaïne *N*-oxyde (Annexe 5F), montrant que cette dernière est résistante à cet oxydant et peut donc être retrouvée après traitement des rejets de STEU comme TP final. Deux TP supplémentaires, observés par HRMS, ont pu être attribués à des déalkylations de la lidocaïne : $C_{12}H_{18}N_2O$ ($m/z = 206,1424$) and $C_{10}H_{13}NO$ ($m/z = 163,1003$) dont les aires augmentent avec la dose de PFA (Annexe 5G). Un mécanisme de dégradation de la lidocaïne est proposé en Figure 43.

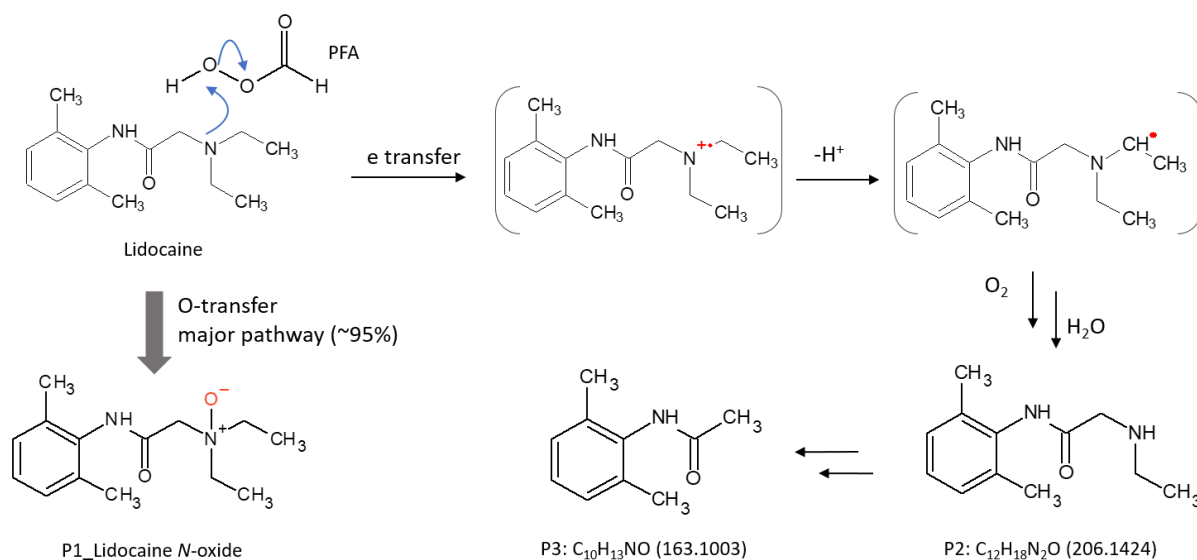


Figure 43 : Mécanisme de dégradation de la lidocaïne par action du PFA

Le mécanisme de dégradation global est similaire à l'oxydation d'amines tertiaires par l'ozone (Lange et al., 2006; von Sonntag and von Gunten, 2012). L'attaque du PFA sur le doublet non liant de l'azote permet la formation de la *N*-oxyde (P1, Figure 43), produit majoritaire formé par transfert de l'oxygène distant de la liaison peroxyde. Cette réaction de transfert est aussi observée pour d'autres types d'oxydation incluant une liaison peroxyde tels que l'acide peracétique ou le peroxymonosulfate (Kim and Huang, 2021; Nihemaiti et al., 2020; Wang et al., 2014). Le transfert direct d'électrons, réaction minoritaire, peut amorcer la formation d'une amine cationique radicalaire qui engendre un carbone radicalaire central. Ce dernier peut rapidement subir une hydrolyse permettant la formation des deux TP déalkylés (P2 et P3,

Figure 43). Une conversion presque totale de la lidocaïne vers sa *N*-oxyde n'a pas été observée à toutes les doses appliquées dans les rejets de STEU réel étudiés. La Figure 42, montre qu'à 100 mg/L de PFA et malgré la présence de lidocaïne dans cet échantillon, l'aire de sa *N*-oxyde diminuait par rapport à l'aire enregistrée pour une dose de 30 mg/L. Cette observation suggère que d'autres processus d'oxydation sont impliqués dans les matrices réelles, notamment à forte dose de PFA qui nécessiterait des études plus poussées (*e.g.*, réaction de Fenton due au dopage à forte concentration de H₂O₂ présent simultanément avec le PFA).

Amisulpride, tramadol et clarithromycine. L'amisulpride et le tramadol sont majoritairement dégradés en *N*-oxyde mais beaucoup plus lentement que la lidocaïne. Avec une concentration initiale de 5 µM à pH 8 et après 10 min de réaction avec 10 mg/L de PFA, ces composés forment moins de 0,1 µM de *N*-oxyde, contre 1,9 µM de lidocaïne *N*-oxyde dans les mêmes conditions. Cette observation s'explique par des valeurs de pKa plus élevées de ces deux composés par rapport à la lidocaïne (pKa = 8,28 pour l'amisulpride et 9,23 pour le tramadol ; www.chemaxon.com, consulté le 06/2021). De manière similaire à la lidocaïne, l'oxydation par le PFA a également conduit à la détection par HRMS de TP issus de la déalkylation des amines tertiaires pour l'amisulpride et le tramadol, tels que le C₁₅H₂₃N₃O₄S (*m/z* = 341,1414) et le C₁₅H₂₃NO₂ (*m/z* = 249,1731) respectivement (Figure 44).

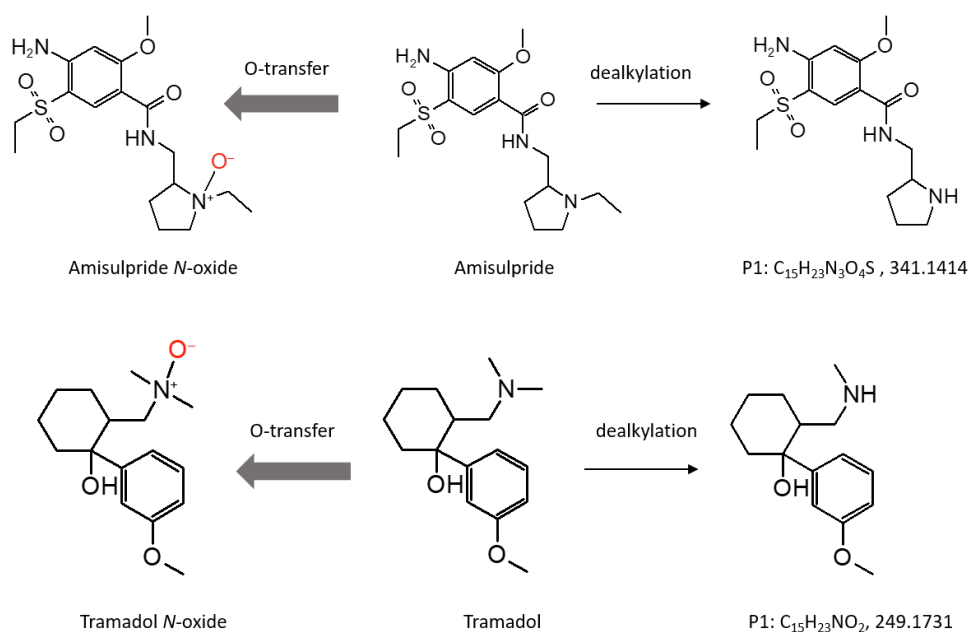


Figure 44 : Mécanismes de transformation de l'amisulpride et du tramadol par oxydation par le PFA

La clarithromycine n'a pas été quantifiée lors de cette étude. Cependant, les signaux en HRMS montrent une dégradation proportionnelle à la dose de PFA appliquée à pH 8, conduisant à la formation de sa *N*-oxyde. Deux TP supplémentaires (P1 : $m/z = 587,3667 - C_{30}H_{53}NO_{10}$ et P2 : $m/z = 605,3781 - C_{30}H_{53}NO_{11}$) ont également été détectés en HRMS, avec des aires ayant des ordres de grandeur deux fois plus faibles que celle de la *N*-oxyde. Ces TP peuvent résulter du clivage de la fonction désoxysucre de la clarithromycine et de sa *N*-oxyde Annexe 5H.

III.1.3.4. Impact environnemental

Le PFA est de plus en plus appliqué à des fins de désinfection d'effluents d'eaux usées pour son efficacité à inactiver les microorganismes et son faible potentiel à produire des sous-produits de désinfection réglementés. Cependant, de nombreuses lacunes quant à la compréhension des mécanismes de réactions restent à compléter. Les résultats obtenus par HRMS à partir de l'étude en NTS montrent que la désinfection par le PFA à faible dose (2 mg/L, pendant 10 min) permet de réduire l'aire totale des signaux en HRMS de l'effluent initial en formant un nombre limité de TP. L'évolution des marqueurs après oxydation par le PFA montre cependant le potentiel de ce dernier à former des TP plus petits et plus polaires, à partir de composés plus gros et moins polaires. Ces marqueurs ont été priorisés par analyse discriminante, permettant d'identifier la lidocaïne *N*-oxyde comme composé d'intérêt dans cette étude de l'effet du PFA sur les MP. L'étude en SS sur sept *N*-oxydes et leur précurseurs respectifs a permis de confirmer la formation de ces TP par traitement d'effluent de STEU par le PFA. L'étude en conditions contrôlées suggère que la lidocaïne est principalement convertie en sa *N*-oxyde par addition d'oxygène, avec formation mineure de deux TP issus de déalkylation. Des mécanismes similaires ont pu être observés pour l'amisulpride, le tramadol et la clarithromycine, laissant penser que le PFA réagit de façon sélective avec les amines. Des études mécanistiques complémentaires pourraient être menées sur d'autres types de MP, notamment ceux possédant des sites riches en électrons afin de compléter ces observations. Les résultats de ce travail montrent également que le PFA peut réagir avec l'azote contenu dans les biomolécules présentes dans les cellules, telles que les peptides ou les protéines, ce qui permettrait de mieux comprendre son effet sur l'inactivation des microorganismes.

III.2. Cas de l'ozonation

III.2.1. Introduction

L'ozonation est un procédé utilisé pour la production d'eau potable depuis de nombreuses années, d'abord en tant que désinfectant (von Sonntag and von Gunten, 2012), puis pour ses capacités oxydantes. Ce n'est que plus récemment que l'intérêt pour ce procédé s'est développé pour le traitement des eaux usées, notamment pour sa capacité à oxyder les MP qui sont récalcitrants aux traitements classiques appliqués en STEU, afin de réduire l'impact du rejet sur les milieux, en particulier en Suisse et en Allemagne où des réglementations spécifiques ont été émises sur les concentrations en MP dans les rejets.

Ce procédé s'appuie sur l'utilisation de la molécule d'ozone, un gaz très réactif mais également très instable. Son utilisation requiert donc une unité de production *in situ* qui permet de générer de l'ozone à partir d'oxygène. Cette réaction peut être effectuée par photolyse UV ou par le biais de décharges électriques qui engendrent une dissociation du dioxygène pour permettre une recombinaison entre dioxygène et oxygène.

L'oxydation par ozonation est régie par deux types de réactions : la première est évidemment la réaction de la molécule d'ozone avec les MP, et la seconde met en jeu les radicaux hydroxyles formés par la décomposition de l'ozone (von Gunten, 2003; Hoigné and Bader, 1975; Staehelin and Hoigne, 1985). De nombreuses études se sont penchées sur les mécanismes de réaction entre ces oxydants et les MP. Celles-ci ont mis en évidence deux principaux schémas réactionnels de l'ozone. Le premier consiste à un clivage des doubles liaisons C=C par le biais d'une cycloaddition, aussi connue sous le nom de réaction de Criegee (Criegee, 1975). Le second se traduit par une addition d'oxygène sur des sites riches en électrons (Lim et al., 2022). En ce qui concerne les radicaux hydroxyles, ces derniers sont beaucoup moins sélectifs et réagissent rapidement avec la majorité des molécules organiques (Buxton et al., 1988). L'efficacité d'une oxydation des MP organiques par ozonation peut être influencée par différents facteurs (Asghar et al., 2022). En particulier, le pH influe sur la stabilité de l'ozone en solution et détermine ainsi si les réactions liées à l'ozone ou aux radicaux hydroxyles seront majoritaires (Asghar et al., 2022; von Gunten, 2003). D'autre part, la présence d'autres espèces réactives dans le milieu pourront diminuer l'efficacité de l'ozone à éliminer les MP organiques, puisque ce dernier sera consommé par ces espèces. Parmi

celles-ci, la présence de MOD réagit principalement avec les espèces radicalaires très peu sélectives. Il est donc fréquent que pour l'ozonation de matrices chargées en MOD, la dose d'ozone appliquée soit calculée à partir de la concentration en COD de l'eau à traiter. Pour des eaux usées, cette dose se situe généralement entre 0,1 et 1,2 gO₃/gCOD (Gerrity et al., 2012; Kharel et al., 2020; Schollée et al., 2018). De plus, dans le cas des eaux usées, le traitement inclu souvent une étape de nitrification. La présence de nitrites a une grande influence sur la disponibilité de l'ozone pour oxyder les MP organiques puisque les nitrites réagissent rapidement et de façon stœchiométrique avec l'ozone pour former des nitrates (Hoigné et al., 1985).

L'étude menée dans ce paragraphe a pour objectif de montrer l'applicabilité des méthodes précédentes afin d'évaluer l'efficacité de l'ozonation pour le traitement de MP dans les rejets de STEU. En particulier, la partie C – III.1 ayant permis de mettre en évidence la formation de *N*-oxydes par oxydation d'amines tertiaires avec le PFA, ces TP connus pour l'ozone (Gulde et al., 2021b; Kharel et al., 2020; Merel et al., 2017) ont également été étudiés ici afin de confirmer leur présence dans les effluents de STEU ozonés et voir si une adsorption subséquente sur CAP permet d'éliminer ces composés qui ont montré une résistance face au PFA. Pour cela, une première approche s'appuie sur une caractérisation globale des échantillons après différentes étapes de traitements : effluent de STEU, effluent après ozonation, effluent après ozonation suivie d'une adsorption sur CAP. Cette comparaison a ensuite été complétée par une approche en SS sur les composés étudiés dans le paragraphe précédent et leurs *N*-oxydes, puis une méthode en NTS a été développée afin de pouvoir retrouver ces TP de façon plus automatisée, permettant ainsi d'identifier de potentiels nouveaux précurseurs de *N*-oxydes.

III.2.2. Matériel et méthodes

III.2.2.1. Echantillonnage

Les échantillons utilisés dans cette partie sont issus d'expérimentations réalisées dans le cadre de l'étude des bénéfices apportés par une ozonation avant adsorption sur charbon actif pour l'élimination de MP organiques dans les eaux usées. Les résultats en TS de cette étude menée sur un ensemble de 28 MP ont déjà été publiés par ailleurs (Guillossou et al., 2020). Brièvement, l'ozonation a été réalisée en continu grâce à un pilote installé sur la station de Seine Centre (Colombes, France), à une dose spécifique de 0,24 g_{O₃}/g_{DOC}. Le pilote pour l'ozonation est alimenté par l'effluent nitrifié provenant directement de l'étape de biofiltration. Les échantillons ont été prélevés en entrée du pilote, après ozonation et après une étape supplémentaire d'adsorption sur CAP. En complément des échantillons prélevés pour réaliser l'étude en TS, des sous-échantillons ont également été récupérés en parallèle pour effectuer les analyses en SS et NTS présentées dans ce travail.

III.2.2.2. Analyses

Les échantillons ont été filtrés sur 0,7 µm (GF/F, Whatmann) après prélèvement et extraits au plus tard dans les 48 h suivantes à l'aide d'une cartouche multiphasique (HLB, ENV+, XAW, XCW), selon la méthode décrite en partie C – III.1.2.4. Les extraits ont été évaporés à sec sous flux d'azote, puis reconstitués avec 1 mL d'un mélange EUP/MeOH (80/20). Les données HRMS ont été acquises par UPLC-IMS-QToF avec séparation sur colonne C18, dans les conditions définies en partie B – III.4.1.

Les données résultantes ont été traitées à l'aide de UNIFI pour l'étude en SS, en utilisant les paramètres définis en partie A – III.3.2. Quant aux données relatives au NTS, la *marker table* générée par UNIFI a été exportée pour être exploitée par l'application décrite en partie A – IV.3.2. Cependant, les tentatives d'identification ont été effectuées sous UNIFI afin de tirer parti des informations relatives à l'IMS. Les paramètres optimisés d'identification décrits en partie A – IV.1.2.2 ont été utilisés, et seules la librairie interne et les bibliothèques externes importées (décrites en partie A – III.2) ont été interrogées.

III.2.3. Résultats

III.2.3.1. Caractérisation globale

Dans un premier temps, une comparaison d'empreinte de l'échantillon aux différentes étapes de traitement a été effectuée (Figure 45).

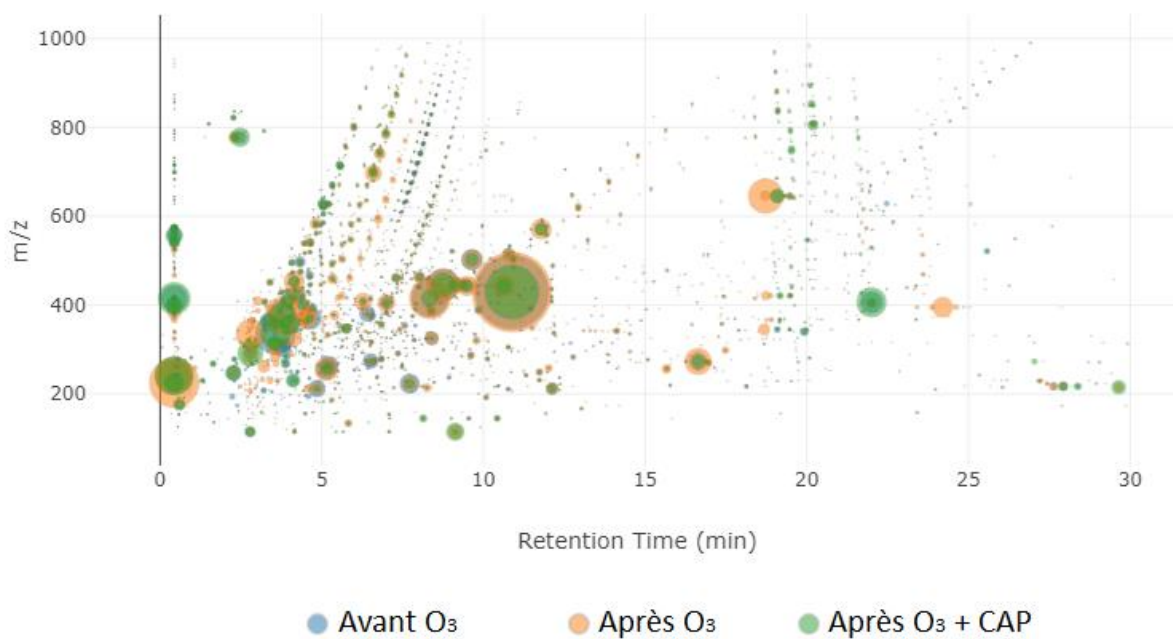


Figure 45 : Empreintes de l'échantillon à différentes étapes de traitement (seuls les marqueurs d'aires supérieures à 10 000 ont été représentés)

Par simple comparaison de ces empreintes, il est déjà possible de remarquer des différences entre les échantillons, notamment la présence de marqueurs uniquement présents dans les échantillons ozonés, à des TR supérieurs à 15 min avec des aires importantes, ou encore entre 3 et 7 min. En revanche, peu de marqueurs avec de grandes aires semblent spécifiques à l'échantillon avant ozonation. Enfin, concernant, les marqueurs avec les plus grandes aires retrouvées dans l'échantillon après ozonation et adsorption, il est intéressant de remarquer que ceux-ci sont également présents dans au moins un des deux autres échantillons avec des aires généralement plus élevées, démontrant une adsorption partielle de ces marqueurs sur le CAP. Afin d'approfondir ces premières observations, une caractérisation globale a été effectuée sur ces échantillons (Tableau 44).

Tableau 44 : Caractérisation globale des rejets de STEU soumis à ozonation et adsorption sur CAP

	Nombre de marqueurs	m/z moyen	TR moyen (min)	DT moyen (ms)	Aire totale
Avant O ₃	12 168	454,89	8,55	6,18	$7,08 \times 10^7$
Après O ₃	12 161	471,92	9,56	6,47	$9,86 \times 10^7$
Après O ₃ + 10 mg/L de CAP	11 068	487,12	9,88	6,49	$7,40 \times 10^7$

Dans cette étude, les moyennes ont été pondérées par les aires respectives des marqueurs ce qui n'a eu aucune influence (par rapport aux moyennes non pondérées) sur les tendances observées en comparant les échantillons entre eux. Ainsi, l'ozonation, dans ces conditions, permet globalement d'éliminer autant de marqueurs qu'elle n'en forme puisque le nombre total de marqueurs est quasiment identique après ozonation. L'étape d'adsorption suivante permet quant à elle de réduire ce nombre total. En revanche, concernant les propriétés moyennes de ces marqueurs, une augmentation des trois propriétés à chaque étape du traitement est observée. Ceci suggère soit une élimination de marqueurs globalement plus polaires et plus petits, ou au contraire la formation de marqueurs moins polaires et de plus grande taille.

Afin de mieux visualiser l'effet du traitement, les marqueurs *spécifiques* et *communs* à ces échantillons ont été comparés (Figure 46). Parmi les 12 168 marqueurs retrouvés dans l'échantillon avant ozonation, environ 40 % ont été complètement éliminés par le traitement puisqu'ils ne sont pas retrouvés dans les deux autres échantillons. L'étape d'ozonation est à l'origine de la formation de près de 6 600 nouveaux marqueurs, dont près du tiers ne sont pas éliminés par l'étape suivante d'adsorption sur CAP. Ces marqueurs nouvellement formés et persistants ont un m/z moyen de 547,77, un TR moyen de 11,77 min et un DT moyen de 7,00 ms. Ces marqueurs feraient donc partis des TP formés les plus gros et les moins polaires, par comparaison avec les propriétés moyennes prenant en compte tous les marqueurs de cet échantillon, ce qui est étonnant en regard d'études précédentes montrant la formation de composés globalement plus polaires par l'ozonation. La somme des aires de ces marqueurs dans l'échantillon ozoné est de 11×10^6 alors qu'elle est de 8×10^6 dans l'échantillons après CAP, montrant qu'un abattement partiel est tout de même observé sur ces marqueurs.



Figure 46 : Diagramme d'Euler pour la comparaison du nombre de marqueurs de l'échantillon à la suite du traitement par ozonation et CAP

Par ailleurs, près du tiers des marqueurs présents dans le rejet de STEU sont persistants à la combinaison de l'ozonation et de l'adsorption sur CAP. Avec un m/z moyen de 478,60, un TR moyen de 10,25 min et un DT moyen de 6,46 ms, ces marqueurs font une fois de plus partis de la moyenne haute en termes de masse et de polarité, par rapport au total des marqueurs dans cet échantillon initial. Cette observation suggère que les molécules globalement plus petites et plus polaires sont mieux éliminées par le traitement. La somme des aires de ces marqueurs dans les échantillons avant ozone, après ozone et après ozone + CAP est de 50×10^6 , 63×10^6 et 50×10^6 respectivement, montrant qu'ils sont très peu affectés par les différentes étapes.

Enfin, près de 3 000 marqueurs sont *spécifiques* de l'échantillon après ozonation et adsorption sur CAP. Ces marqueurs suggèrent soit des réactions inattendues de surface qui engendrerait la formation de ces nouveaux marqueurs, soit des interférences analytiques qui auraient masqué la présence de ces marqueurs dans les deux autres échantillons qui contiennent plus de matière organique. Ces deux groupes de marqueurs ont un m/z moyen de 482,54, un TR moyen de 10,80 min et un DT moyen de 6,29 ms et correspondent donc à des molécules légèrement moins polaires par comparaison aux autres marqueurs présents dans l'échantillon

après ozonation et adsorption. En termes d'aires, la somme des aires de ces marqueurs dans l'échantillon après ozonation et adsorption est de $14,8 \times 10^6$, soit 20 % du signal total relevé pour cet échantillon.

Ces différentes observations montrent que malgré des résultats satisfaisants lors de l'étude ciblée sur un ensemble de 28 MP (Guillossou et al., 2020), cette étude en NTS permet de nuancer le bilan. En effet, malgré l'élimination complète d'un certain nombre de marqueurs, les marqueurs qui persistent au traitement ne sont presque pas abattus par la combinaison des deux procédés étudiés. Cependant, il est rassurant de constater que les deux tiers des TP produits par l'ozonation sont correctement éliminés par adsorption sur CAP. Enfin, il serait intéressant d'étudier davantage les marqueurs spécifiques à l'échantillon après ozonation et adsorption afin de mieux comprendre leur origine.

III.2.3.2. Recherche de N-oxydes en mode suspect

Suite à l'étude menée sur l'oxydation par PFA qui a montré le potentiel de cet oxydant à réagir avec des amines tertiaires pour former des N-oxydes (partie C – III.1), une recherche en SS de ces composés a été effectuée sur les échantillons étudiés dans ce paragraphe. En effet, des études précédentes ont mis en évidence la formation de ces N-oxydes par ozonation (Gulde et al., 2021b; Kharel et al., 2020; Merel et al., 2017). Les résultats de cette recherche en SS sont présentés dans le Tableau 45.

Tableau 45 : SS de 5 amines tertiaires et de leur *N*-oxyde dans un effluent de STEU soumis à ozonation puis adsorption sur CAP. Les valeurs correspondent aux aires détectées corrigées par l'aire totale de l'échantillon considéré. ND : Non Détecté.

	Nitrifié	Ozoné	Ozoné + 10 mg/L de CAP
Amisulpride	$47,70 \times 10^{-4}$	$8,90 \times 10^{-4}$	$16,50 \times 10^{-4}$
Amisulpride <i>N</i>-oxyde	ND	$0,16 \times 10^{-4}$	ND
Lidocaïne	$6,12 \times 10^{-4}$	$2,16 \times 10^{-4}$	$2,34 \times 10^{-4}$
Lidocaïne <i>N</i>-oxyde	$0,30 \times 10^{-4}$	$3,64 \times 10^{-4}$	$3,04 \times 10^{-4}$
Tiapride	$2,45 \times 10^{-4}$	ND	$1,22 \times 10^{-4}$
Tiapride <i>N</i>-oxyde	ND	ND	ND
Tramadol	$39,80 \times 10^{-4}$	$15,70 \times 10^{-4}$	$13,80 \times 10^{-4}$
Tramadol <i>N</i>-oxyde	$0,46 \times 10^{-4}$	$1,90 \times 10^{-4}$	$1,52 \times 10^{-4}$

En comparaison avec l'étude menée avec le PFA, l'effluent étudié ici ne contenait pas autant d'amines tertiaires d'intérêt. Ainsi, le citalopram, la clarithromycine et la venlafaxine n'ont pas été détectés. A noter que les deux échantillons comparés ici ne proviennent pas de la même STEU et n'ont pas non plus été prélevés à la même période. En effet, ces composés étant des molécules pharmaceutiques, une variation saisonnière est souvent observée à la fois en entrée de STEU, mais aussi en sortie avec une efficacité de traitement réduite sur les périodes hivernales (Golovko et al., 2014; Sun et al., 2014). L'échantillon considéré dans cette partie a été prélevé au printemps alors que les échantillons du paragraphe précédent ont été prélevés en automne, ce qui pourrait expliquer les différences observées.

Quoiqu'il en soit, les *N*-oxydes respectives de l'amisulpride, de la lidocaïne et du tramadol ont pu être identifiées dans les échantillons ozonés avec des signaux augmentant significativement par rapport à l'échantillon initial, et semblent être relativement mal éliminées par l'étape complémentaire d'adsorption sur CAP. En effet, une diminution de l'ordre de 18 % en moyenne est observée pour la lidocaïne *N*-oxyde et le tramadol *N*-oxyde. Seule l'amisulpride *N*-oxyde n'a pas été détectée après adsorption, mais l'aire de ce marqueur dans l'eau ozonée était déjà faible initialement. De la même manière, la tiapride *N*-oxyde n'a pas pu être détectée, ce qui peut être expliqué par le fait que le tiapride présent dans l'eau

avant ozonation montre également un signal initial peu intense. Si conversion en sa *N*-oxyde il y a eu, il est possible que l'aire du marqueur correspondant soit trop faible pour être visible dans ces conditions. De manière surprenante, les aires de l'amisulpride dans les différents échantillons montrent une diminution de plus de 80 % entre l'échantillon nitrifié et l'échantillon ozoné, mais double entre ce dernier et l'échantillon après adsorption sur CAP.

III.2.3.3. Recherche de *N*-oxydes par approches non-ciblées

La recherche en SS a permis de mettre en évidence la présence de *N*-oxydes dans les échantillons ozonés et de leurs précurseurs dans les échantillons initiaux. Cependant, cette recherche était limitée à un nombre restreint de composés connus. Afin de tenter d'identifier de nouveaux précurseurs de *N*-oxydes dans ces échantillons, une méthode sans *a priori* a été mise en place.

Sur la base de la méthode de Kendrick (Kendrick, 1963), les masses détectées ont été converties vers une nouvelle échelle, selon l'équation 4a.

$$KM_{m/z} = m/z \times \frac{16,0000}{15,9949} \quad \text{eq. 4a}$$

$$KMn_{m/z} = \text{round}(KM_{m/z}) \quad \text{eq. 4b}$$

$$KMD_{m/z} = KMn_{m/z} - KM_{m/z} \quad \text{eq. 4c}$$

Les masses nominales de Kendrick (KMn, équation 4b) et défauts de masse de Kendrick (KMD, équation 4c) associées ont ensuite été déterminées. Selon la théorie développée par Kendrick, les marqueurs alignés horizontalement (*i.e.*, les marqueurs ayant le même défaut de masse) correspondent à des ions d'une même série d'homologues. De ce fait, ils ont une composition élémentaire identique, modulo le groupement ayant servi à calculer la nouvelle échelle, dans le cas présent l'atome d'oxygène. Une fois les nouvelles masses calculées, un lien a été établi entre les marqueurs présents dans l'échantillon nitrifié (précurseurs) et les marqueurs de l'échantillon ozoné (*N*-oxydes) possédant un même KMD et différant d'un KMn de 16 (pour l'atome d'oxygène). Une attention particulière a également été portée sur l'évolution des aires relatives de ces marqueurs entre les deux échantillons. En effet, seuls les couples dont les aires des précurseurs diminuaient et dont les aires de *N*-oxydes augmentaient d'au moins 20 % entre les deux échantillons ont été conservés (Figure 47). De la même manière, afin de faciliter les tentatives d'identification (*i.e.*, limitation du nombre de marqueurs à identifier en

ne conservant que ceux qui possédaient un spectre en haute énergie), un filtre complémentaire a été implémenté pour ne garder que les marqueurs dont les aires étaient supérieures à 5 000.

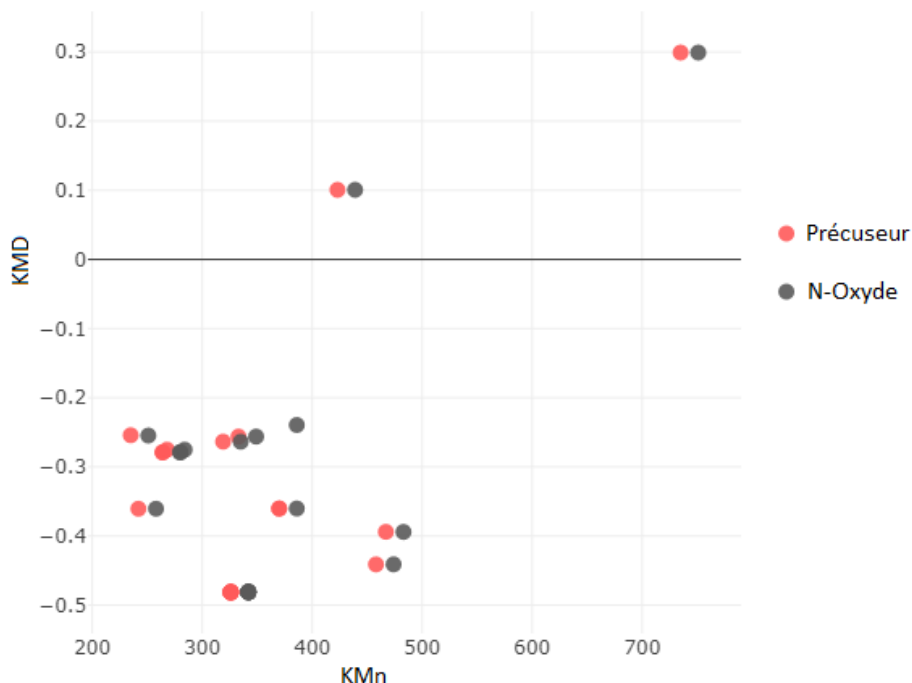


Figure 47 : Recherche non-ciblée de *N*-oxydes basée sur la méthode de Kendrick

Sur les 12 couples repérés, des tentatives d'identification ont été effectuées afin de confirmer si les marqueurs sélectionnés pouvaient effectivement correspondre à un précurseur et sa *N*-oxyde. Parmi les couples sélectionnés : 2 ont été éliminés car l'intégration effectuée automatiquement ne correspondait pas à un pic réel ; 2 ne présentaient pas de massif isotopique assez défini pour pouvoir attribuer une formule moléculaire unique ; les autres sont présentés dans le Tableau 46.

De façon encourageante, la lidocaïne et le tramadol ainsi que leur *N*-oxyde respective ont pu être retrouvées par cette méthode. Seule l'amisulpride, pourtant trouvée en SS, n'a pas été sélectionnée. Ceci s'explique par l'aire de sa *N*-oxyde, plus faible que le filtre appliqué. Ce dernier a été mis en place ici afin de limiter le nombre de couples potentiels.

Concernant les couples supplémentaires identifiés, le metoprolol et l'imiprothrin ont pu être confirmés jusqu'au niveau 3. Cependant, ces deux composés ne possèdent pas d'amines tertiaires, identifiées comme site privilégié pour la formation de *N*-oxydes. En effet, l'azote du metoprolol forme une amine secondaire alors que les deux azotes de l'imiprothrin font toutes

deux parties d'amides, qui plus est, cycliques. Dans le cas du metoprolol, le TP correspondant est cohérent par rapport à la méthodologie employée. En effet, la formule moléculaire correspond à celle du metoprolol avec un atome d'oxygène supplémentaire. Son identité a par contre été associée au metoprolol hydroxy, qui correspond à une substitution d'un hydrogène par un groupement hydroxyle plutôt qu'à la formation d'une N-oxyde. Pour l'imiprothrin, le spectre correspondant à son TP n'a pas permis d'attribuer de formule moléculaire unique au TP. Cependant, en considérant la structure de cette molécule, bien que la formation de N-oxyde soit peu probable, il est possible d'envisager la formation d'un TP hydroxylé qui serait cohérent avec la méthodologie employée.

Les autres marqueurs sélectionnés n'ont pu être confirmés que jusqu'au niveau 4, ne permettant pas de savoir si les composés sont enclins à pouvoir former une N-oxyde. Cependant, après tentatives d'identification des TP sur les 4 couples restants, seuls deux sont susceptibles d'être des couples précurseur/N-oxyde, puisque les deux autres tentatives d'identification ont conduit à l'attribution de formules moléculaires différentes par étude des spectres ($C_{21}H_{20}ClNO_4$ au lieu de $C_{16}H_{35}NO_9$ et $C_{16}H_{39}N_7O_9$ au lieu de $C_{15}H_{40}N_9O_6P$), malgré un KMD basé sur l'oxygène correspondant.

Tableau 46 : Tentative d'identification de précurseurs de *N*-oxyde par des méthodes non-ciblées

Précurseur potentiel						<i>N</i> -oxyde potentielle					
<i>m/z</i>	TR (min)	DT (ms)	Formule moléculaire	Identité	Niveau de confiance	<i>m/z</i>	TR (min)	DT (ms)	Formule moléculaire	Identité	Niveau de confiance
235,1795	4,62	4,59	C ₁₄ H ₂₂ N ₂ O	Lidocaine	1	251,1749	5,51	4,64	C ₁₄ H ₂₂ N ₂ O ₂	Lidocaine <i>N</i> -oxyde	1
264,1950	5,22	4,90	C ₁₆ H ₂₅ NO ₂	Tramadol	1	280,1899	5,83	5,02	C ₁₆ H ₂₅ NO ₃	Tramadol <i>N</i> -oxide	1
268,1899	5,51	5,23	C ₁₅ H ₂₅ NO ₃	Metoprolol	3	284,1848	3,06	5,40	C ₁₅ H ₂₅ NO ₄	Metoprolol hydroxy	3
319,1621	7,53	5,24	C ₁₇ H ₂₂ N ₂ O ₄	Imiprothrin	3	335,1572	4,75	5,33	-	-	5
326,3775	19,71	6,56	C ₂₂ H ₄₇ N	-	4	342,3720	14,85	7,07	C ₂₂ H ₄₇ NO	-	4
370,2425	2,62	5,64	C ₁₆ H ₃₅ NO ₈	-	4	386,2374	2,69	5,87	C ₂₁ H ₂₀ ClNO ₄	-	4
458,2954	3,53	6,51	C ₁₅ H ₄₀ N ₉ O ₅ P	-	4	474,2902	6,35	6,48	C ₁₆ H ₃₉ N ₇ O ₉	-	5
467,2454	4,62	6,21	C ₁₇ H ₃₄ N ₆ O ₉	-	4	483,2405	3,69	6,42	C ₁₇ H ₃₄ N ₆ O ₁₀	-	4

III.3. Conclusion

Ce paragraphe s'est intéressé à l'application de méthodes de NTS pour l'évaluation de procédés de traitement complémentaires de MP retrouvés dans les rejets de STEU. Dans deux cas d'études en conditions réelles, l'intérêt de ces méthodes a pu être mis en avant.

Ainsi, le NTS a permis d'identifier de façon inédite la lidocaïne N-oxyde comme un marqueur discriminant de désinfection des eaux usées par le PFA. Cette identification a permis d'apporter de nouvelles connaissances quant au mode d'action de cet oxydant sur les MP. D'autre part, l'étude globale a montré que le PFA semble efficace pour abattre une partie des MP persistants aux traitements conventionnels effectués en STEU. Ces nouvelles observations viennent compléter des études précédentes ayant montré que le PFA formait peu de sous-produits d'oxydation réglementés par comparaison à d'autres oxydants utilisés pour la désinfection des eaux usées dans des conditions similaires.

La seconde étude s'est portée sur un traitement par ozonation complété par une adsorption sur CAP. Les études en TS sur ces échantillons (Guillossou et al., 2020) ont montré la capacité de l'ozone à dégrader, au moins partiellement, les composés étudiés et l'adsorption sur CAP suivante permet d'obtenir des taux d'abattement totaux satisfaisants. Ce travail en SS et NTS amène globalement à des conclusions similaires concernant l'efficacité de cette combinaison de traitement pour éliminer les MP initialement présents dans le rejet de STEU. Cependant, il met également en évidence la formation de TP peu adsorbables sur CAP, en particulier les N-oxydes étudiées.

Le SS et le NTS permettent donc d'avoir une vision plus complète de l'évolution des échantillons sous l'effet d'un traitement d'oxydation. Ils peuvent donc apporter de nouvelles connaissances quant aux mécanismes de réaction mis en jeu lors de tels procédés et informer sur la formation de TP qui peuvent se retrouver dans l'environnement.

IV. Influence de la stratégie analytique sur l'évaluation des procédés d'oxydation

IV.1. Introduction

Le paragraphe précédent a permis de montrer l'applicabilité des méthodes de SS et NTS pour l'évaluation de procédés d'oxydation afin d'éliminer les MP dans les rejets de STEU. Cependant, les travaux ont été menés en fixant une méthode de préparation et une méthode d'analyse uniques. La partie B a montré la complémentarité de différentes stratégies analytiques qui, ensemble permettait d'avoir une vision plus large de la composition de l'échantillon, mais qui prises à part, pouvaient conduire à des conclusions différentes. L'objectif de cette partie était donc de déterminer dans quelle mesure l'étude de procédés d'oxydation d'eaux usées pouvait être influencée par la méthode analytique employée.

Pour cela, trois procédés d'oxydation ont été étudiés : la chloration, l'oxydation avancée par UV/H₂O₂ et l'ozonation. Ces derniers ont été choisis pour différentes raisons. La chloration a été largement étudiée, notamment dans le domaine de la production d'eau potable pour son potentiel à former des TP dont beaucoup sont déjà reportés dans la littérature et ont montré une toxicité importante. L'ozonation et l'UV/H₂O₂ sont des procédés d'oxydation avancée largement étudiés et, dans certains cas, implémentés, pour le traitement des rejets de STEU afin d'éliminer les MP.

Ces trois procédés ont été appliqués à des eaux de STEU afin de mieux comprendre les mécanismes de transformation mis en jeu lors des différentes réactions. Enfin, la comparaison de différentes stratégies analytiques permettra d'apporter des informations complémentaires et, pourrait permettre d'identifier un protocole plus adapté ou non à ce type d'étude.

IV.2. Matériel et méthodes

IV.2.1. Stratégie d'échantillonnage

Des échantillons ont été prélevés dans la station de Seine Amont (Valenton, SIAAP). Pour chaque procédé d'oxydation étudié, 18 L ont été récupérés puis filtrés sur 0,7 µm (GF/F) avant d'être séparés en deux échantillons (échantillon avant oxydation et échantillon soumis à oxydation). Les dates des prélèvements ponctuels effectués et les paramètres globaux

associés sont récapitulés dans le Tableau 47. Les deux types d'échantillons de 9 L (rejet de STEU ou rejet de STEU oxydé) ont ensuite été séparés en 8 sous-échantillons de 1 L pour extractions en parallèle sur les 8 phases de SPE décrites dans la partie B – III.3.2.

Tableau 47 : Paramètre globaux des échantillons réels

	Chloration (rejet)	UV/H ₂ O ₂ (décantée)	Ozonation (rejet)
Date	04/06/2019	11/02/2020	10/11/2020
pH	6,7	-	7,4
Ammonium (mgN/L)	0,58	38,77	-
COD (mgC/L)	7,1	-	7,5
Conductivité (mS/m)	86,6	-	1 234
DBO entière (mgO₂/L)	1,5	185,0	-
Matières en suspension (mg/L)	< 2	263,0	< 2
Nitrates (mgN/L)	16,6	2,17	-
Nitrites (mgN/L)	0,09	2,33	-
Azote Kjeldahl (mgN/L)	1,6	42,50	-
Orthophosphates (mgP/L)	0,49	5,52	-
Phosphore (mgP/L)	0,7	4,2	-
ST-DCO (mgO₂/L)	20,7	421	-
Turbidité (FNU)	1,6	> 99	-

IV.2.2. Oxydation

La chloration a été réalisée à l'aide d'une solution commerciale de NaOCl. La solution commerciale est dosée avant chaque manipulation afin de s'assurer de la concentration initiale de chlore. Pour cela, l'absorbance de la solution, diluée à 1/1000, est mesurée à $\lambda = 292$ nm. La concentration de la solution commerciale est déterminée grâce à l'équation 5.

$$[Cl_2] \left(\frac{mg}{L} \right) = \frac{A \times 71 \times 1000}{\varepsilon} \quad \text{eq. 5}$$

Avec A : l'absorbance mesurée à $\lambda = 292$ nm

ε : le coefficient d'extinction molaire de ClO_2^- , $\varepsilon = 362 \text{ M}^{-1} \cdot \text{cm}^{-1}$

La chloration a été amorcée en ajoutant un volume déterminé de la solution commerciale dans l'échantillon, en fonction de la dose de chlore souhaitée. Dans le cas présent, la dose de chlore choisie a été déterminée à partir de la valeur de COD de l'échantillon étudié, de sorte à ce que $[Cl_2] \text{ (mgCl}_2\text{/L)} = 3 \times [\text{COD}] + 15 \text{ mgC/L}$ pendant 24h. Une fois le temps de contact défini atteint, la dose de chlore résiduelle a été déterminée par une méthode iodométrique. Pour cela, 3 mL de l'échantillon pour lequel on souhaite déterminer la dose de chlore résiduel ont été prélevés. A ces 3 mL d'échantillon, 750 μL d'iodure de potassium à 1 M puis 75 μL d'acide acétique à 1,7 M ont été ajoutés. L'absorbance de ce mélange a été mesurée à $\lambda = 351$ nm et la concentration de chlore résiduel de la solution a été déterminée suivant l'équation 6.

$$[Cl_2] \text{ (mM)} = \frac{A \times (3 + 0,75 + 0,075) \times 1000}{3 \times \varepsilon \times d} \quad \text{eq. 6}$$

Avec A : l'absorbance mesurée à $\lambda = 351$ nm

ε : le coefficient d'extinction molaire, $\varepsilon = 26\,900 \text{ M}^{-1} \cdot \text{cm}^{-1}$

d : le facteur de dilution du milieu réactionnel pour faire la mesure

L'oxydation avancée par UV/ H_2O_2 a été réalisée à l'aide de 3 lampes UV émettant à $\lambda = 254$ nm (collimated beam, Figure 48) et une solution commerciale d' H_2O_2 . Dans cette étude, la réaction est amorcée par ajout de 3 mg/L d' H_2O_2 dans l'échantillon étudié, qui est ensuite placé sous les lampes UV pendant 5h.

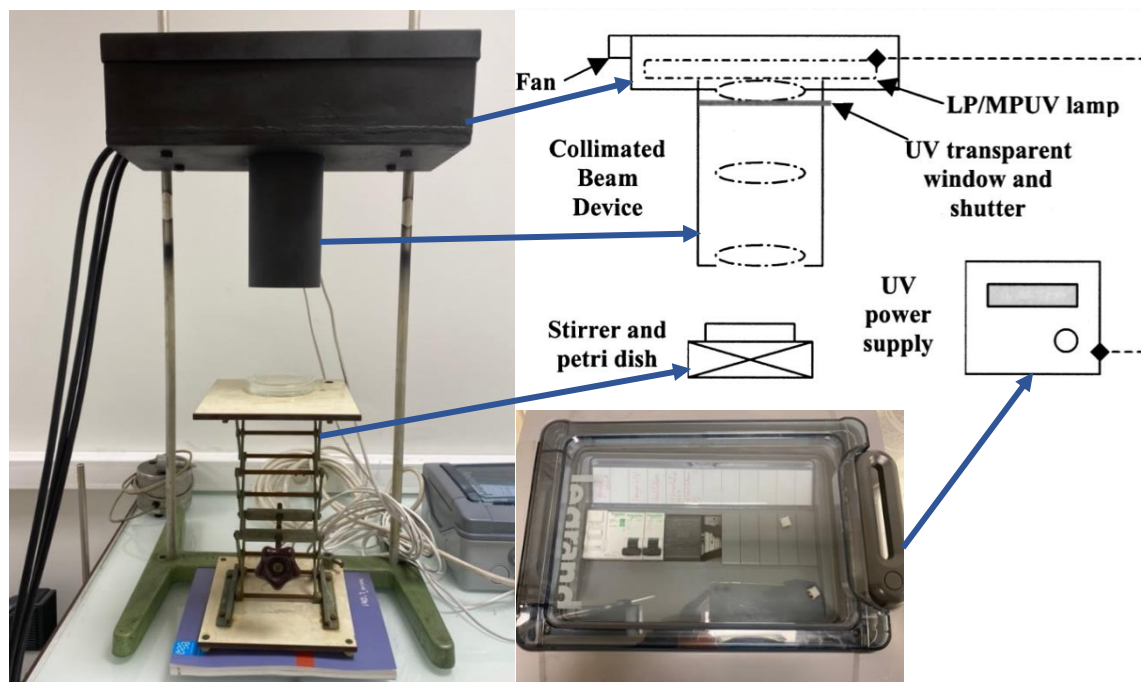


Figure 48 : Montage du pilote de photolyse UV

La lampe est allumée au minimum 30 min avant chaque manipulation pour permettre son chauffage, et des ventilateurs, situés au-dessus des lampes, tournent en permanence pendant toute la durée d'éclairage de la lampe afin d'éviter une surchauffe. L'intensité du faisceau UV est estimée régulièrement par actinométrie sur l'atrazine. Pour cela, une solution de 5 μM d'atrazine est préparée et soumise à une dégradation par photolyse UV. La cinétique de dégradation de l'atrazine est suivie pendant 30 min et la constante de dégradation (k) est estimée sur la base d'une réaction d'ordre 1. L'analyse de la concentration d'atrazine est réalisée par HPLC-UV à une absorbance de $\lambda = 220 \text{ nm}$. La séparation chromatographique est réalisée sur une colonne C18 (Ascentis Express C18, 100 x 4,6mm, 2,7 μm , Supelco), avec une phase mobile composée d'un mélange 60/40 EUP/MeOH. L'intensité de la lampe est ensuite déterminée à l'aide l'équation 7 (Canonica et al., 2008).

$$\text{Fluence (mW.cm}^{-2}\text{)} = \frac{k}{k'} \quad \text{eq. 7}$$

- k correspond à la constante de vitesse déterminée avec la cinétique de dégradation de l'atrazine
- k' correspond à une constante déterminée à l'aide plusieurs paramètres référencés par ailleurs

Pour le calcul de k', l'équation 8 est employée.

$$k'(\text{cm}^{-2}\text{mJ}^{-1}) = \frac{\Phi \times \varepsilon \times \ln(10)}{U} \quad \text{eq. 8}$$

- Φ est le rendement quantique de dégradation de l'atrazine à 254 nm. $\Phi_{254\text{nm, ATZ}} = 0,046$ mol/Einstein (Hessler et al., 1993)
- ε le coefficient d'extinction molaire de l'atrazine à 254 nm. $\varepsilon_{254\text{nm, ATZ}} = 3860$ M⁻¹cm⁻¹
- U est l'énergie portée par une mole de photon. $U_{254\text{nm}}$ est le produit de la constante de Planck par la vitesse de la lumière dans le vide et la constante d'Avogadro, le tout, divisé par la longueur d'onde de travail. $U_{254\text{nm}} = 470\,866$ J/Einstein (Bolton et al., 2015)

Au cours de ce travail, l'intensité mesurée était en moyenne de 0,3 mW/cm².

L'ozone est produit par un ozoneur, alimenté à 2 L/min par une bouteille d'oxygène de pureté supérieure à 98 %. L'ozone gazeux est produit par une décharge corona, puis diffusé dans de l'EUP préalablement refroidie (Figure 49). La dose d'ozone appliquée est calculée à partir de la valeur de COD de l'échantillon, de sorte que $[\text{O}_3]$ (mgO₃/L) = 0,5 x [COD] mgC/L.

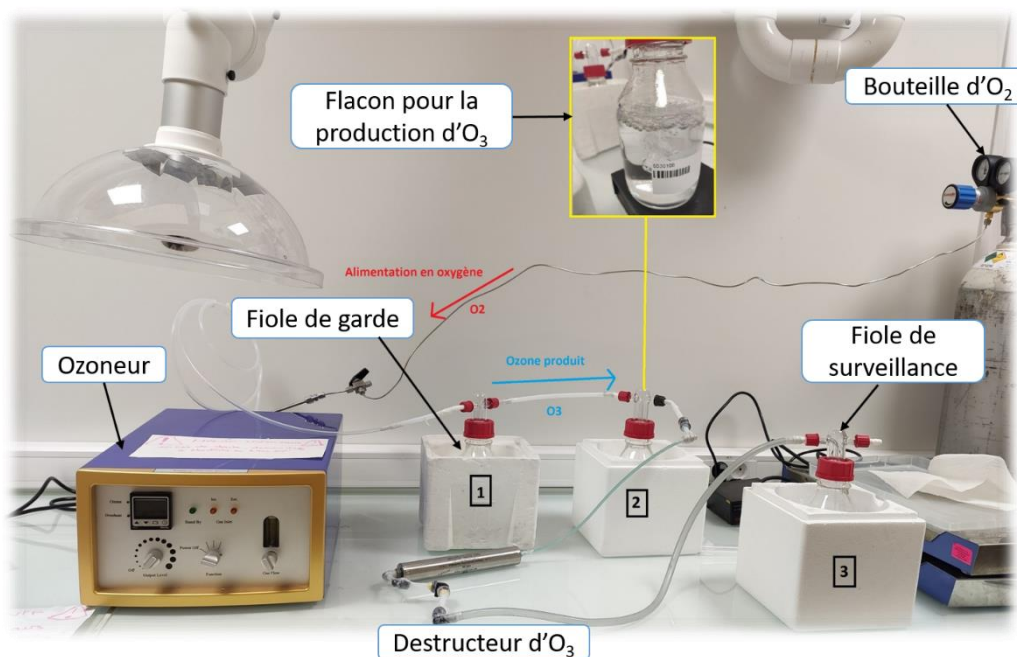


Figure 49 : Montage pour la production d'ozone en laboratoire

La fiole de garde (1) permet d'éviter tout retour de liquide dans l'ozoneur qui pourrait causer des dégâts électriques. Le flacon pour la production d'ozone (2) est placé dans un bain de glace afin de garder la solution à une température voisine de 0°C pour optimiser le transfert de l'ozone produit, de la phase gazeuse vers la phase liquide. Le destructeur d'ozone permet d'éliminer l'ozone gazeux qui n'aurait pas été transféré lors de l'étape précédente. Ce destructeur contient un filtre de Carulite, composé d'un catalyseur en oxyde de manganèse et oxyde de cuivre. La fiole de surveillance (3), remplie d'une solution d'indigo à $2,5 \times 10^{-5}$ mol/L, permet d'assurer le bon fonctionnement du destructeur d'ozone. Une décoloration de cette solution indique que l'ozone résiduel est mal éliminé et que le filtre du destructeur d'ozone doit être remplacé afin d'assurer que l'air rejeté dans la pièce ne contienne plus d'ozone. Par mesure de précaution supplémentaire, le pilote d'ozonation est placé sous un bras d'extraction pour permettre d'évacuer l'air potentiellement chargé d'ozone.

La concentration de la solution d'ozone produite est mesurée à l'aide de la méthode à l'indigo (Bader and Hoigné, 1981). Brièvement, l'ozone décolore la solution d'indigo proportionnellement à sa concentration, à une absorbance $\lambda = 600$ nm. La concentration est déterminée grâce à l'équation 9.

$$[O_3]_L = \frac{V_{tot} \times \Delta_A}{f \times b \times V_{ech}} \quad \text{eq. 9}$$

- $[O_3]_L$: concentration en ozone dissous (mg/L)
- V_{tot} : volume total (soit 10 mL)
- Δ_A : différence d'absorbance entre l'échantillon et le blanc (cm^{-1})
- b : chemin optique de la cellule
- V_{ech} : volume de l'échantillon d'eau ozonée (mL)
- f : constante de proportionnalité ($0,42 \text{ cm}^1 \text{ mg} \cdot \text{L}^{-1}$)

IV.2.3. Analyses et traitement de données

Après filtration sur filtre GF/F (porosité $0,7 \mu\text{m}$) le jour du prélèvement, les échantillons sont soumis à une extraction sur phase solide au plus tard dans les 48 h suivantes. Pour cela, ils sont divisés en 8 sous-échantillons et extraits séparément en utilisant les 8 cartouches décrites dans la partie B – III.3.1, suivant le même protocole (Tableau 27). De la même manière, l'extrait est séparé en 2 sous-extraits, évaporés à sec et repris dans 1 mL d'un mélange EUP/MeOH (80/20) ou EUP/ACN (10/90) pour analyse sur les 3 colonnes chromatographiques dans les deux modes d'ionisation. Pour la manipulation relative à l'ozonation, les échantillons ont également été analysés par FT-ICR-MS. Les protocoles analytiques correspondants sont détaillés dans la partie B – III.4.

Après acquisition des données, celles-ci ont été prétraitées par UNIFI et exportée pour exploitation sous logiciels libres (R). Une première approche a consisté à caractériser les trois procédés d'oxydation. Pour cela, plusieurs méthodes ont été appliquées de façon successive afin d'apporter des connaissances quant aux mécanismes mis en jeu lors de ces traitements d'oxydation. Dans un premier temps, les marqueurs détectés dans les échantillons ont été classés en fonction de leurs intensités relatives avant et après oxydation, par application de l'équation 10 (Bader et al., 2017).

$$Fc = \frac{\text{aire du marqueur après oxydation}}{\text{aire du marqueur avant oxydation}} \quad \text{eq. 10}$$

En fonction du paramètre Fc, les marqueurs sont classés dans les catégories suivantes :

- $Fc < 0,2$: Elimination = marqueurs qui ont été éliminés par le traitement
- $0,2 < Fc < 0,5$: Diminution = marqueurs qui ont été partiellement abattus
- $0,5 < Fc < 2$: Constant = marqueurs qui n'ont pas été affectés significativement
- $2 < Fc < 5$: Augmentation = marqueurs présents en plus grande quantité après traitement
- $Fc > 5$: Formation = marqueurs qui ont été formés par le traitement

Une fois les marqueurs *Éliminés* et *Formés* repérés, une étude de mécanismes de transformations potentiels a été menée en s'appuyant sur les méthodes de Kendrick (Kendrick, 1963), afin de tenter d'établir un lien entre les marqueurs appartenant à ces deux groupes. Cette méthode s'appuie sur les défauts de masse afin d'identifier des familles homologues et le processus est détaillé dans la partie C – III.2.3.3 (appliqué aux *N*-oxydes). Cependant, dans cette étude, aucun filtre en intensité n'a été appliqué et un nombre plus important de liens (*i.e.*, groupements moléculaires gagnés ou perdus) a été recherché (Tableau 48). Ces liens ont été sélectionnés suivant leur applicabilité à être identifiés par cette méthode (*i.e.*, perte ou gain d'un groupement d'atomes simple) et leur plausibilité vis-à-vis des procédés d'oxydation étudiés.

Tableau 48 : Etablissement de liens de transformation entre marqueurs *éliminés* et *formés*

	Groupement gagné/perdu	Ratio pour conversion dans la nouvelle échelle	Ecart de $KM_{m/z}$ (produit – précurseur)
Chloration	+Cl - H	34/33,9610	+ 34
Décarbonylation	- CO	28/27,9949	- 28
Décarboxylation	- COO	44/43,9898	- 44
Déchloration	-Cl + H	34/33,9610	- 34
Déméthylation	- CH ₂	14/14,0157	- 14
Désaturation	+ H ₂	2/2,0157	+ 2
Déshydratation	- H ₂ O	18/18,0106	- 18
2 x Hydroxylation	+ 2O	16/15,9949	+ 32
Hydratation	+ OH	17/17,0028	+ 17
Hydroxylation	+ O	16/15,9949	+ 16
Réduction	- H ₂	2/2,0157	-2

Par ailleurs, l'objectif de cette partie étant également de comparer les différentes méthodes analytiques employées dans la partie B, avec, cette fois, une attention particulière portée aux TP formés par les traitements d'oxydation, certains critères de comparaison supplémentaires ont été ajoutés. Ainsi, en plus de la comparaison des extraits suivant les catégories *omniprésent*, *commun* et *unique*, utilisés précédemment, deux catégories supplémentaires ont été ajoutées : *commun spécifique* et *unique spécifique*. La notion de marqueur *spécifique* qualifie ici les marqueurs retrouvés seulement dans l'échantillon oxydé par comparaison à l'échantillon non-oxydé. Pour cela, les extraits avant et après oxydation ont été comparés deux à deux, à cartouche SPE fixée.

L'efficacité d'une stratégie analytique donnée a ainsi été évaluée sur la richesse d'informations qu'elle pouvait apporter en termes de nombre et nature de TP formés, mais également sur la proportion de TP qui pouvait être expliquée par des mécanismes de transformation usuels, à partir des marqueurs ayant été éliminés.

IV.3. Résultats

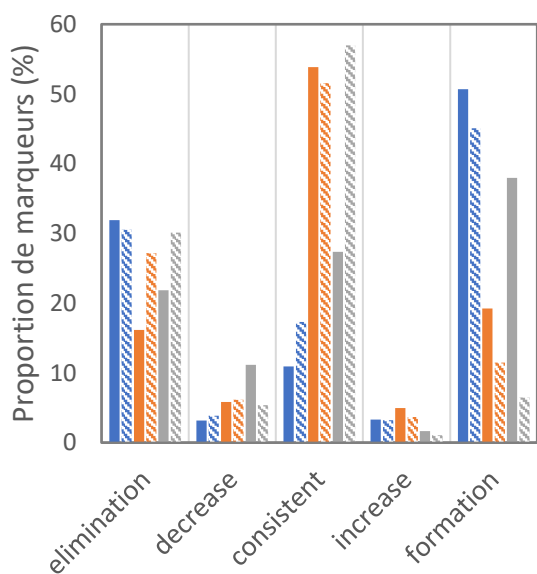
IV.3.1. Comparaison des colonnes chromatographiques

Une première approche permettant de comparer les colonnes chromatographiques vis-à-vis des différents types d'oxydation s'est appuyée sur les 5 catégories définies à partir de l'équation 10. Indépendamment de la méthode d'extraction employée, cette classification a été effectuée pour les 3 colonnes chromatographiques dans les deux modes d'ionisation (Figure 50).

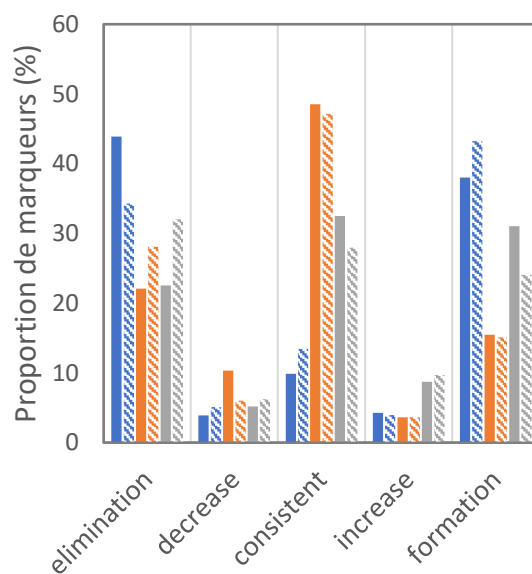
En comparant les modes d'ionisation, de légères différences sont observées en fonction du procédé d'oxydation ou de la colonne employée. La différence la plus notable entre les deux modes d'ionisation concerne l'ozonation après séparation sur colonne C18. En effet, alors qu'en ESI+, près de 40 % des marqueurs sont classifiés dans la catégorie formation, en ESI-, cette catégorie n'acompte que pour 5 %, la majorité des marqueurs étant classés comme constants. Il est cependant important de noter que l'analyse en ESI- des échantillons issus de la manipulation d'ozonation n'a donné qu'un faible nombre de marqueurs comparé à ce qui est habituellement observé. Cette observation était d'autant plus vraie en ce qui concerne la colonne C18, ce qui peut expliquer en partie cette grande différence entre les deux modes d'ionisation.

L'analyse des échantillons issus de l'oxydation avancée par UV/H₂O₂ montre également de légères différences entre les deux modes d'ionisation pour les colonnes C18 et PGC, de même que la chloration avec les colonnes PGC et HILIC. Bien que moins importantes que la différence énoncée précédemment pour l'ozonation, ces dernières sont suffisantes pour inverser les tendances entre la proportion de marqueurs *formés* et *éliminés*. Avec l'UV/H₂O₂, une plus grande proportion de marqueurs *formés* que de marqueurs *éliminés* est observée en mode ESI+, et inversement en mode ESI-. Pour la chloration, plus de marqueurs *formés* sont observables en mode ESI- comparé au mode ESI+. La colonne PGC montre particulièrement plus de différences en termes de proportions de marqueurs dans les différentes catégories entre les deux modes d'ionisation par rapport aux deux autres colonnes. Ces observations indiquent que les deux modes d'ionisation permettent d'obtenir des informations complémentaires et qu'il est donc préférable d'utiliser les deux en parallèle afin d'obtenir plus d'informations lorsque c'est possible.

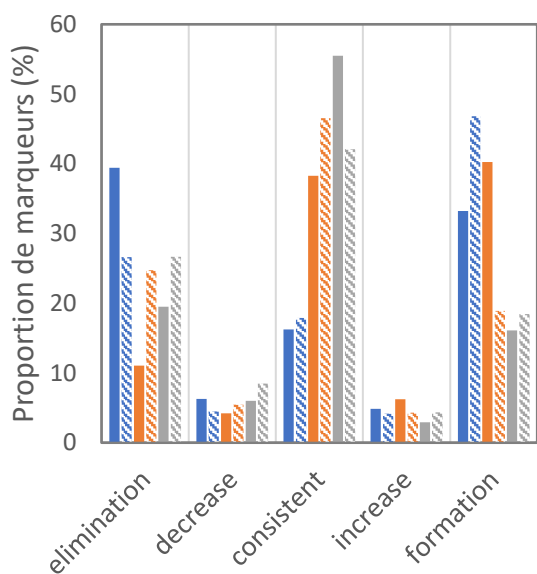
a) UPLC(C18)-IMS-QToF



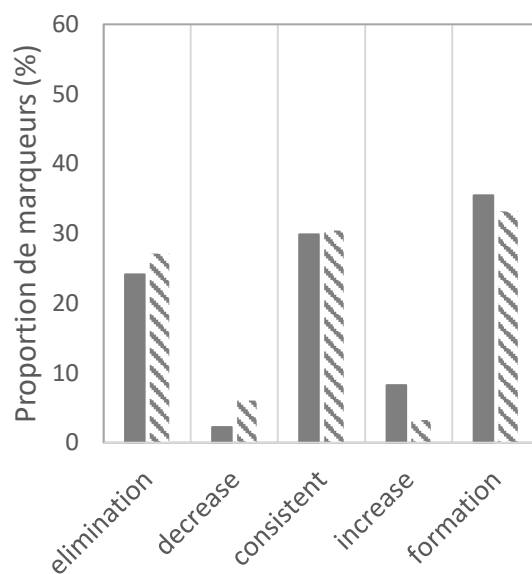
b) UPLC(HILIC)-IMS-QToF



c) UPLC(PGC)-IMS-QToF



d) FT-ICR-MS



■ chloration ESI+ ▨ chloration ESI-
 ■ UV/H2O2 ESI+ ▨ UV/H2O2 ESI-
 ■ ozonation ESI+ ▨ ozonation ESI-

Figure 50 : Classification des marqueurs suivants 5 catégories en fonction de leur comportement à la suite des 3 procédés d'oxydation et détectés par les quatre méthodes d'analyse, dans les deux modes d'ionisation (ESI+ et ESI-)

En ce qui concerne les différences entre colonnes, les tendances ne semblent pas différer d'une colonne à une autre en termes de proportion de marqueurs dans chacun des groupes lorsque les données en ESI⁻ sont considérées. En effet, bien que de légères différences soient observables, les données sur les trois colonnes montrent une plus grande proportion de marqueurs *formés* que de marqueurs *éliminés* par la chloration, alors que l'inverse est constaté pour l'UV/H₂O₂ et l'ozonation. En revanche, en considérant les analyses en ESI⁺ de plus grandes différences ont pu être observées en fonction de la méthode analytique employée, conduisant à des interprétations différentes. De manière intéressante, la séparation sur HILIC permet d'obtenir les mêmes conclusions qu'au moins l'une des deux autres colonnes pour les trois procédés d'oxydation, alors que la partie B – IV.1 a montré qu'elle semblait retenir des composés de différente nature. L'analyse de l'échantillon ozoné par FT-ICR-MS (équipé d'une source ESI) (Figure 50d) a permis une comparaison entre HRMS et UHRMS. Pour ces mêmes échantillons, les proportions obtenues se rapprochent également de ce qui a été observé pour la colonne HILIC (Figure 50b).

Par la suite, une étude des mécanismes de transformation impliqués dans chaque procédé a été menée. Pour cela, la méthode de Kendrick a été appliquée de la même manière que dans la partie C – III.2.3.3, en étendant la liste de groupements moléculaires d'intérêt (pas uniquement l'ajout d'oxygène) pour établir des liens de transformation supplémentaires (Tableau 49). La méthode a été réalisée sans filtrer les marqueurs en fonction de leur intensité, et en considérant comme précurseurs potentiels les marqueurs ayant été classés dans la catégorie *élimination*, et comme TP potentiels les marqueurs classés dans la catégorie *formation*. Le nombre de liens établis pour chaque type de transformation est exprimé en proportion du nombre de marqueurs *formés* après oxydation, pour chaque procédé et chaque méthode d'analyse (Tableau 49).

Tableau 49 : Proportion de marqueurs *formés* (%) pouvant être expliqués par des mécanismes de transformation usuels pour chaque procédé (par colonne chromatographique selon le mode d'ionisation et indépendamment du support d'extraction employé)

	Chloration			UV/H ₂ O ₂			Ozonation			
	C18 (ESI+/ESI-)	HILIC (ESI+/ESI-)	PGC (ESI+/ESI-)	C18 (ESI+/ESI-)	HILIC (ESI+/ESI-)	PGC (ESI+/ESI-)	C18 (ESI+)	HILIC (ESI+)	PGC (ESI+)	FT-ICR (ESI+)
Chloration	0,68/0,66	0,50/0,31	0,40/0,07	0,19/0,16	0,15/0,61	0,00/0,00	0,17	0,06	0,10	0,51
Décarbonylation	1,74/1,85	1,17/0,74	1,64/0,64	0,40/1,43	1,34/0,80	0,00/0,33	0,83	0,76	1,74	2,76
Décarboxylation	1,18/1,39	0,98/0,46	0,67/0,37	0,35/0,92	0,81/0,65	0,00/0,60	0,61	0,62	0,87	2,06
Déchloration	0,45/0,28	0,45/0,2	0,36/0,09	0,17/0,38	0,22/0,30	0,00/0,16	0,23	0,23	0,1	0,81
Déméthylation	2,01/1,83	1,62/0,86	2,46/1,08	0,88/3,06	2,29/1,49	0,00/1,48	1,37	1,20	2,10	3,03
Déshydratation	0,61/0,54	0,46/0,20	0,46/0,11	0,14/0,13	0,15/0,15	0,00/0,05	0,24	0,32	0,26	0,48
Déshydrogénation	1,74/1,30	1,31/0,75	1,98/0,67	0,78/1,66	1,16/1,87	0,00/0,88	1,04	0,83	1,74	2,73
Hydrogénation	2,41/2,06	1,34/1,51	2,36/1,02	0,67/1,72	1,26/1,22	2,06/1,37	1,00	1,33	1,48	3,09
Hydratation	0,65/0,72	0,52/0,27	0,46/0,10	0,09/0,22	0,15/0,19	0,00/0,11	0,23	0,29	0,15	0,42
Hydroxylation	1,73/1,70	1,46/0,72	2,46/1,02	0,20/1,91	1,46/1,87	0,00/1,37	0,84	1,01	1,64	3,23
2 x Hydroxylation	1,53/1,32	1,28/0,72	2,13/0,85	0,55/1,59	1,08/1,37	0,00/0,77	0,43	0,74	0,61	2,05
Total (ESI+/ESI-)	12,75/12,05	11,09/5,82	12,89/5,08	3,70/11,21	8,77/8,85	2,06/6,19	6,99	7,39	10,79	21,17

Dans un premier temps, en comparant les modes d'ionisation, il est intéressant de noter que les tendances sont les mêmes pour les trois colonnes chromatographiques considérées, à procédé d'oxydation fixé. En effet, pour la chloration le mode ESI+ permet d'établir plus de liens entre marqueurs *éliminés* et *formés*, alors que pour l'UV/H₂O₂ le mode ESI- permet d'établir davantage de liens.

Par comparaison des différentes méthodes d'analyse, il semble y avoir plus de différences entre les procédés d'oxydation étudiés. En effet, une plus grande proportion de liens de transformation a pu être établie avec l'analyse par séparation C18 dans le cas de la chloration, alors que la HILIC donne plus de liens avec l'UV/H₂O₂ et la PGC avec l'ozonation. Lorsque la FT-ICR a été employée (ozonation), c'est elle qui permet d'expliquer la plus grande proportion de liens entre marqueurs *éliminés* et marqueurs *formés*, à partir des mécanismes de transformation usuels. En ce qui concerne la chloration, cette observation est cohérente avec les tendances observées en partie B – IV.4.1. En effet, d'après l'étude en SS pour la comparaison des trois colonnes chromatographiques, la colonne C18 permettait de retrouver les composés chlorés qui n'étaient que très peu ou pas visibles après séparation sur les deux autres colonnes, pouvant ainsi expliquer une plus grande proportion de liens établis pour la chloration. Il avait été également constaté que la HILIC permettait de séparer des molécules ayant globalement plus d'oxygène par comparaison aux molécules retenues sur les deux autres types de colonnes (partie B – IV.4.1). Cependant, cette tendance ne semble pas être retrouvée en considérant l'hydroxylation et la double hydroxylation qui sont représentées de façon assez similaires sur la colonne C18 et la HILIC.

Globalement la déméthylation est le mécanisme le plus rencontré quel que soit le procédé d'oxydation ou le mode d'analyse, expliquant la plus grande proportion de couples précurseurs/TP, suivie par l'hydrogénation et l'hydroxylation. Dans une étude évaluant l'influence d'un traitement de rejet de STEU par ozonation, une méthode similaire a été décrite (Schollée et al., 2021). Cette dernière permettait également d'établir des liens de transformation simples, mais s'appuyait cependant sur le calcul de masses exactes théoriques de TP formés par ces mêmes transformations, à partir des *m/z* des composés catégorisés comme *éliminés*. Ces nouvelles masses théoriques ont alors été cherchées parmi les *m/z* classés comme *formés*. Les résultats de cette étude sont en accord avec les résultats que nous

avons obtenus, montrant que la déméthylation, l'hydrogénation et l'hydroxylation étaient les mécanismes permettant d'expliquer le plus grand nombre de transformation.

En termes de TP identifiés, une étude complémentaire en SS s'appuyant sur la librairie interne contenant 412 TP provenant de différentes sources a été effectuée. Les TP présents dans cette librairie sont, soit issus de la littérature (étude bibliographique de mécanismes de dégradation de micropolluants organiques d'intérêt – 178 entrées), soit identifiés à la suite d'expérience de dégradation en conditions contrôlées (cinétique de dégradation en EUP – 197 entrées), soit sont des TP connus et communément recherchés dans les eaux urbaines (37 entrées).

76 TP différents ont pu être détectés au moins une fois dans les échantillons toutes colonnes confondues, dont 43 sur la colonne C18, 13 sur la PGC et 43 sur la HILIC. En particulier, 99 % des TP détectés après séparation sur PGC l'ont été dans l'échantillon après chloration alors que presque aucun TP n'a pu être identifié sur cette colonne pour l'UV/H₂O₂ et l'ozonation. La HILIC permet quant à elle de détecter une quinzaine de TP différents dans chacune des expériences d'oxydation alors que la colonne C18 permet de détecter environ 25 TP pour la chloration et l'ozonation, mais seulement 8 TP après UV/H₂O₂. En ce qui concerne la nature des TP retrouvés, tous procédés d'oxydation confondus, la HILIC permet de détecter 23 TP qui ne sont pas retrouvés sur les deux autres colonnes, contre 6 pour la PGC et 23 sur la colonne C18. Ces tendances ne confirment que partiellement les observations faites en NTS (Figure 50), l'ozonation menant bien à des proportions plus grandes de marqueurs *formés* détectables après séparation en HILIC et C18 qu'en PGC, mais les deux autres procédés n'aboutissant pas aux mêmes distributions. Encore une fois, les résultats obtenus en SS et en NTS ne permettent pas de tirer des conclusions identiques.

IV.3.2. Comparaison des méthodes de préparation

Après la comparaison des méthodes d'analyses, une seconde étape a été consacrée à la comparaison des différentes méthodes de préparation. De la même manière que dans la partie B – IV.4.2, les marqueurs analysés dans chacun des extraits ont été classés selon différentes catégories. Deux sous-catégories (*communs-spécifiques* et *uniques-spécifiques*) ont été ajoutées aux 3 précédentes (*omniprésents*, *communs* et *uniques*) pour distinguer les marqueurs spécifiquement formés après oxydation (mais soit *uniques* à un extrait, soit en commun entre plusieurs autres extraits).

Dans un premier temps, les résultats sont comparés pour la manipulation d'ozonation afin de permettre la comparaison entre les quatre types d'analyse (Figure 51).

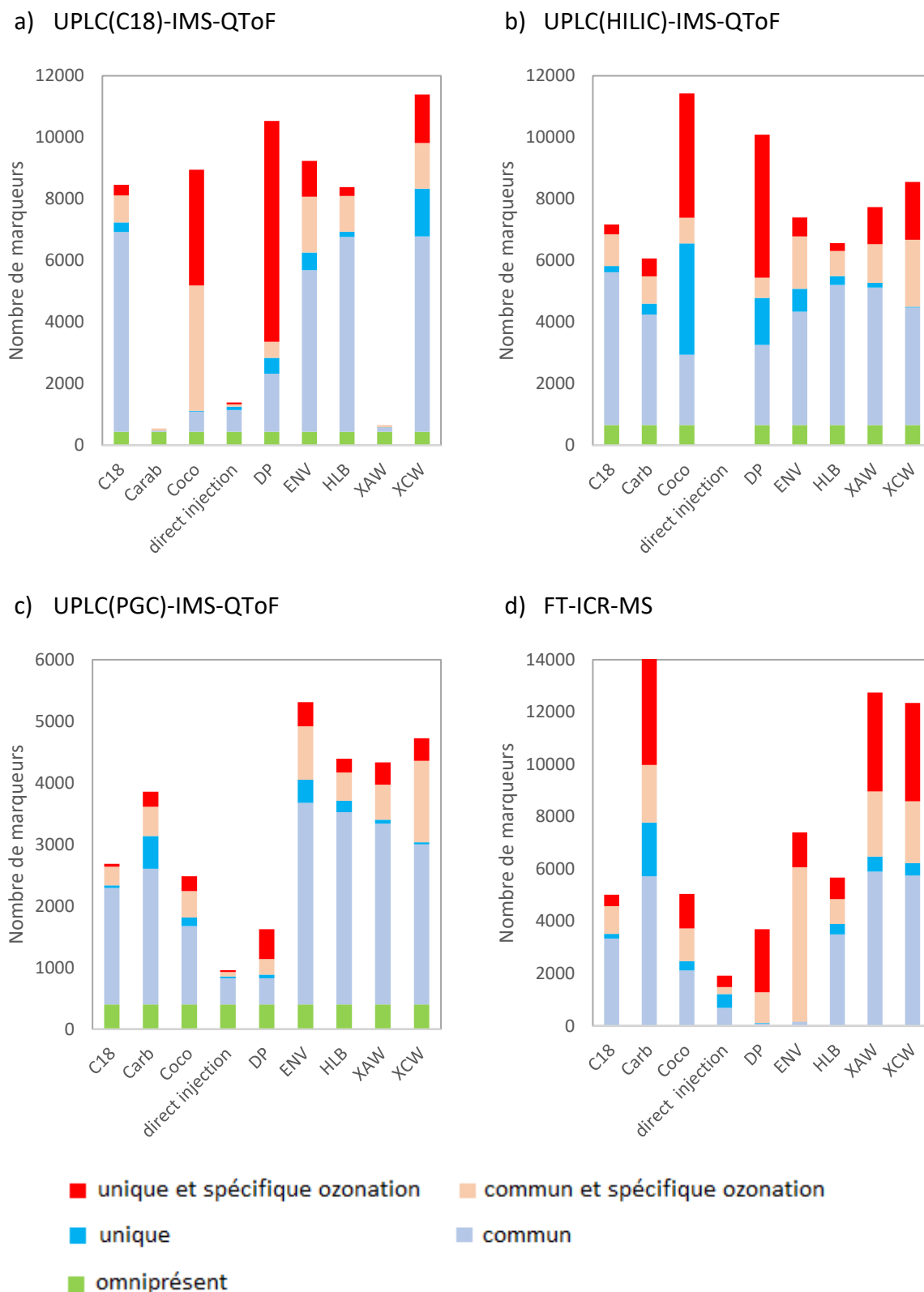


Figure 51 : Comparaison des différents types de marqueurs relevés pour différents extraits dans la manipulation d'ozonation après analyse en mode ESI+

Ainsi, il est intéressant de noter que la proportion de marqueurs *uniques* à chaque cartouche est différente, comparé aux mêmes extraits non oxydés, ce qui montre une réelle influence du traitement sur l'échantillon. Une partie des marqueurs initiaux *uniques* a donc été éliminée au profit de la formation de nouveaux marqueurs *uniques (uniques spécifiques)*. Par comparaison des quatre méthodes d'analyse, il semble que l'UPLC(PGC)-IMS-QToF permette de récupérer la proportion la plus faible de marqueurs *spécifiques* à l'ozonation ce qui est en accord avec la Figure 50, montrant une proportion de marqueurs classés dans la catégorie *Formation*, plus faible par cette technique d'analyse en comparaison avec les deux autres colonnes chromatographiques. La FT-ICR-MS donne quant à elle la plus faible proportion de marqueurs *communs* à tous les extraits, ce qui peut être expliqué par une plus grande résolution permettant de séparer des pics très proches. Ceci indiquerait que les marqueurs désignés comme *communs* par l'alignement en HRMS ne le seraient pas réellement. Concernant les deux autres méthodes, l'analyse par UPLC(C18)-IMS-QToF permet globalement d'analyser légèrement plus de marqueurs *spécifiques* de l'ozonation, mais les résultats dépendent fortement de l'extrait considéré. La cartouche DP donne par exemple 68 % de marqueurs *uniques spécifiques* de l'échantillon ozoné, alors que la cartouche HLB n'en recueille que 3 %. En prenant en compte les quatre méthodes d'analyse, les extraits Coco et DP contiennent généralement des proportions plus grandes de marqueurs *spécifiques* de l'ozonation que les autres extraits, notamment la DP dont un grand nombre de marqueurs *spécifiques* sont également classés comme étant *uniques*. De façon intéressante, ces extraits sont également ceux qui sont les moins bien caractérisés puisqu'elles retiennent assez mal les composés usuellement recherchés en SS et ne sont que très peu utilisées dans la littérature (cf partie B – IV.4.2).

La même approche a été appliquée aux deux autres procédés d'oxydation (Annexe 6). Les tendances observées avec l'ozonation sont retrouvées pour la chloration et l'UV/H₂O₂. Cependant, il est intéressant de remarquer que pour la chloration, la proportion de marqueurs *communs* et *spécifiques* à l'oxydation est plus importante mais également plus homogène entre les différents extraits par comparaison aux deux autres procédés. Cette observation suggère que les TP formés lors de la chloration sont, de manière générale, plus facilement extraits, séparés par chromatographie et détectés que les TP liés aux procédés d'oxydation avancée. Cette observation peut être soit liée au fait que les TP chlorés sont formés en

quantité plus importante (par rapport au nombre total de marqueurs) que les TP des deux autres procédés d'oxydation, ou par le fait que les TP liés à l'UV/H₂O₂ et à l'ozonation possèdent des propriétés particulières qui nécessitent donc des phases adaptées pour les extraire.

De la même manière que dans la partie C – IV.3.1, la proportion de marqueurs *formés* après oxydation et pouvant être expliqués par des mécanismes de transformation simple, à partir des marqueurs *éliminés*, a été évaluée. Pour chaque procédé d'oxydation, la méthode d'analyse permettant d'établir la plus grande proportion de liens a été sélectionnée, d'après les résultats du Tableau 49. Les résultats obtenus à partir de la chloration et analysés avec la colonne PGC sont présentés dans le Tableau 50, les deux autres procédés peuvent être retrouvés en Annexe 8.

Les plus grandes proportions de liens établis pour la chloration sont obtenues avec les extraits des cartouches Carb et ENV (7,1 et 6,9 %, respectivement). Par comparaison à la classification des marqueurs en 5 catégories (Annexe 6C), ces cartouches sont également celles qui permettent d'obtenir les nombres de marqueurs *spécifiques* de l'oxydation les plus importants. De façon cohérente, l'extrait issu de la cartouche DP et l'injection directe montrent à la fois le moins de marqueurs *spécifiques* de la chloration et également le moins de liens établis. Dans l'étude de la chloration, l'ajout d'un atome de chlore (« lien de chloration ») à des précurseurs est attendu. Par comparaison des différents extraits, la cartouche HLB permet d'obtenir une plus grande proportion de liens de chloration. Le nombre total de marqueurs classifiés comme *formés* et retenus dans cet extrait est cependant faible (1 275), le nombre de liens équivaut donc à 2 formations de TP seulement. En revanche, ce sont les extraits de Carb et ENV qui, une fois de plus, permettent d'établir le plus grand nombre de liens de chloration avec 3 et 4 liens respectivement.

En ce qui concerne la manipulation d'UV/H₂O₂, les cartouches C18 et Coco permettent d'établir le plus de liens de transformation (Annexe 8A). En revanche, en termes de nombre maximal de TP *spécifiques* retrouvés (Annexe 7A), l'injection directe suivie de l'extrait ENV donnent les meilleurs résultats. Enfin, pour l'ozonation, les extraits de Carb et HLB permettent d'établir le plus grand nombre de liens. En prenant en compte les 3 procédés, les résultats montrent donc que globalement, la cartouche Carb permet d'établir le maximum de liens de

transformations simples entre les marqueurs *éliminés* et les marqueurs *formés*, suivie de la cartouche HLB.

Tableau 50 : Proportion de marqueurs *formés* (%) pouvant être expliqués par des mécanismes de transformation usuels pour l'échantillon d'effluent d'eau usée chloré, analysé en mode ESI+ après séparation sur PGC, selon les différentes méthodes de préparation

	C18	Carb	Coco	Injection directe	DP	ENV	HLB	XAW	XCW
Chloration	0,1	0,1	0,1	0,0	0,0	0,1	0,2	0,1	0,1
Décarbonylation	0,3	0,8	0,1	0,0	0,0	0,9	0,2	0,6	0,5
Décarboxylation	0,1	0,3	0,1	0,0	0,0	0,3	0,0	0,3	0,1
Déchloration	0,0	0,2	0,1	0,0	0,0	0,1	0,2	0,1	0,1
Déméthylation	1,0	1,0	0,5	0,0	0,0	1,4	0,9	1,5	1,0
Désaturation	0,8	1,0	0,1	0,0	0,0	1,4	0,5	1,1	0,3
Déshydratation	0,2	0,1	0,0	0,0	0,0	0,0	0,2	0,2	0,0
2 x Hydroxylation	0,4	1,4	0,5	0,0	0,0	0,5	0,8	0,7	0,6
Hydratation	0,0	0,1	0,1	0,0	0,0	0,2	0,2	0,1	0,1
Oxydation	0,2	1,2	0,7	0,0	0,0	0,9	0,2	0,7	0,2
Réduction	0,6	0,9	0,4	0,7	0,0	1,1	0,9	1,4	1,0
Total	3,7	7,1	2,6	0,7	0,0	6,9	4,3	6,8	4,0

V. Conclusions sur les applications des approches non ciblées pour l'analyse des procédés d'oxydation

Cette partie a été consacrée à la mise en application des méthodes développées dans les deux parties précédentes. Bien que ces dernières aient montré qu'il n'existait pas de stratégie préférentielle, que ce soit en termes de traitement de données ou en termes de protocoles analytiques utilisés, cette partie montre qu'à méthode fixée, il est possible d'établir des comparaisons entre échantillons afin de compléter des observations effectuées en TS. En effet, dans le cas de l'évaluation de traitements d'oxydation de rejets de STEU, les approches en SS et en NTS permettent d'avoir une vision plus globale du devenir des MP et ainsi de mieux comprendre les mécanismes impliqués dans la transformation de ceux-ci.

Pour cela, une combinaison d'approches de traitement de données, employant à la fois des logiciels fournisseurs et des logiciels libres (partie A), a été utilisée afin de fournir une caractérisation globale rapide des échantillons, tout en essayant d'identifier des marqueurs spécifiques des traitements d'oxydation appliqués, en tirant partie de l'IMS. En ce qui concerne les méthodes analytiques, elles se sont appuyées principalement sur l'utilisation d'une cartouche multiphasique qui a démontré sa capacité à retenir le plus de marqueurs (Annexe 4), et de l'analyse avec séparation sur une colonne C18 qui montre une bonne stabilité et permet donc une interprétation plus poussée des résultats.

Une seconde partie s'est attachée à comparer différents procédés d'oxydation sur une même matrice et en utilisant plusieurs stratégies analytiques. La recherche de TP reste difficile dans des matrices réelles mais certaines approches (notamment l'analyse de Kendrick) permettent d'isoler des couples de marqueurs correspondant à des transformations choisies (couples précurseur/TP). D'autre part, le choix des méthodes de préparation et d'analyse peuvent avoir une très grande influence sur les données finales obtenues, ce qui confirme les observations effectuées en partie B. La comparaison de trois procédés d'oxydation a montré qu'en termes de méthodes de préparation, les cartouches SPE Coconut Charcoal (Supelclean) et DPA-6S (Discovery) ont permis de recueillir plus de signaux correspondant à des TP que d'autres cartouches plus conventionnelles. L'utilisation de ces phases dans une cartouche multiphasique pourrait donc être adaptée à l'étude de procédés d'oxydation et des TP formés,

par rapport à d'autres phases (ex. ENV) apportant moins d'informations liées aux TP. En focalisant sur certains mécanismes de transformation classiques, les phases Envi-Carb (Supelclean) et Oasis HLB (Waters) ont cependant permis d'obtenir plus d'information sur le devenir des MP suite au traitement, en particulier par ozonation.

Conclusion et Perspectives

Conclusion

Ce travail de recherche s'est intéressé aux méthodes d'analyse et de traitement de données actuellement utilisées en NTS appliquées au domaine de l'analyse environnementale. En particulier, l'objectif était de mieux comprendre l'influence du choix des méthodes employées sur les résultats obtenus.

Pour cela, une première partie a été consacrée à la comparaison de différents outils de traitement de données. Cette comparaison s'appuyait sur la capacité des logiciels à prioriser et identifier correctement les composés présents dans un mélange de standards analytiques. Le choix d'un outil dépend grandement des objectifs fixés par l'étude, mais également du type de données fourni par l'instrument. Notamment, l'utilisation du logiciel fournisseur a permis de tirer parti d'un maximum d'informations (en particulier la mobilité ionique) à partir des données acquises afin de réaliser des identifications fiables. Cependant, peu d'options de comparaison d'échantillons et de visualisation de marqueurs sont disponibles dans ce logiciel, limitant les caractérisations plus globales qui permettent de prendre en compte l'ensemble des signaux détectés. En revanche, les logiciels libres sont plus modulables en fonction des besoins des utilisateurs pour ces dernières fonctions. Cependant, les étapes d'identifications sont encore limitées. En particulier, les outils disponibles actuellement ne permettent pas encore la prise en charge des données de mobilité ionique dont l'utilité a été démontrée lors de ce travail. En effet, outre la dimension de séparation supplémentaire qu'elle apporte, permettant de détecter plus de composés différents, cette séparation permet aussi d'obtenir des spectres plus « propres » qui facilitent ainsi l'identification de substances. Par ailleurs, la mesure liée à la mobilité ionique (CCS) permet de confirmer avec d'autant plus de confiance les tentatives d'identification par comparaison aux bases de données disponibles. Il paraît donc important de pouvoir trouver un moyen d'accès plus facile à ces informations qui possèdent de nombreux avantages.

Dans un second temps, les méthodes de préparation et d'analyse de l'échantillon ont été comparées pour un même échantillon. La partie B a d'abord mis en évidence le manque de critères pour comparer des méthodes analytiques en NTS, puis, a confirmé que le choix d'une unique stratégie analytique entraîne la perte d'un grand nombre d'informations relatives à l'échantillon. Il est ainsi apparu essentiel de pouvoir caractériser au mieux différentes

méthodes afin d'apporter un maximum d'informations concernant leurs spécificités et, par conséquent, évaluer le type d'informations perdues lors de la sélection d'une stratégie unique. Ce travail a néanmoins mis en évidence le fait que le choix de la colonne analytique semble avoir plus d'influence sur le type de composés observables que la méthode de préparation employée. Parmi les trois colonnes chromatographiques étudiées lors de ce travail, la colonne C18 qui est la plus fréquemment retrouvée dans la littérature, semble être plus avantageuse d'un point de vue pratique. En effet, que ce soit en termes d'utilisation, de stabilité ou encore d'interprétation des mécanismes de rétention, cette colonne a prouvé sa supériorité par rapport aux deux autres, bien que de nombreuses molécules n'ont pas pu être retrouvées en l'utilisant en comparaison des deux autres. En ce qui concerne les méthodes de préparation, une étude approfondie des spécificités des différentes phases a été réalisée afin de mieux caractériser celles-ci. Néanmoins, les données physico-chimiques étudiées n'ont pas permis de les distinguer clairement les unes des autres, montrant qu'il existe encore de nombreuses informations manquantes. L'analyse en SS permet d'apporter plus de réponses en termes de caractérisation à partir des propriétés physico-chimiques des molécules, mais les bases de données doivent être complétées continuellement avec un grand nombre de molécules afin de permettre des comparaisons de plus en plus poussées concernant les caractéristiques des composés spécifiques aux différentes méthodes employées. Les informations obtenues en mode SS semblent en effet souvent en désaccord avec les conclusions qui peuvent être tirées à partir des données complètes acquises en mode NTS. Il est important de noter que l'injection directe (c'est-à-dire sans étape de préparation) a également été étudiée au cours de ce travail et a montré de nombreuses spécificités par comparaison à l'ensemble de différents adsorbants employés. Cette observation met en lumière la perte d'information qui se produit lors des étapes d'extraction, d'autant plus que les molécules perdues sont visibles sans étape de concentration, ce qui signifie qu'elles sont potentiellement présentes en grande concentration dans l'échantillon initial. A l'issue de ce travail, il apparaît difficile de conclure qu'une méthode unique et préférentielle puisse convenir pour le NTS d'eaux urbaines. Chaque méthode possède des particularités qui lui sont propres et qui ne peuvent pas être retrouvées par ailleurs. Cependant, il est inconcevable, d'un point de vue pratique et économique, d'utiliser cet ensemble de méthodes pour des analyses en routine. D'autant plus que ces méthodes peuvent encore être complétées par d'autres types de techniques comme la chromatographie en phase gazeuse, l'utilisation

d'autres sources d'ionisation ou d'autres phases ou méthodes de préparation. Il est en effet impossible d'avoir une vision complète de la diversité de composés contenus dans un échantillon en utilisant une unique méthode. Il s'agira alors de s'imposer une méthode donnée qui sera appliquée à tous les échantillons afin de pouvoir comparer les résultats obtenus les uns avec les autres pour des analyses en routine. Dans notre cas, une cartouche constituée d'un assemblage de quatre phases adsorbantes (HLB, ENV, XAW et XCW) a été sélectionnée puisqu'elle permet de couvrir une gamme plus large de composés analysables.

En appliquant les méthodes sélectionnées comme étant les plus avantageuses dans la partie précédente (extraction sur cartouche multiphasique et séparation sur colonne C18), deux traitements d'oxydation des MP dans les rejets d'eaux de STEU ont été étudiés en SS et en NTS. L'oxydation par le PFA montre ainsi un bon potentiel en tant que traitement complémentaire pour éliminer les MP en STEU, bien que le procédé génère un certain nombre de TP. Cette étude a notamment mis en évidence la formation de *N*-oxydes à partir d'amines tertiaires et a ainsi permis d'apporter de nouvelles connaissances quant au mode d'action de ce nouvel oxydant. L'ozone combinée à une adsorption sur CAP a également été étudiée, montrant que la seconde étape permet d'éliminer un certain nombre de TP formés au cours de l'ozonation. Cependant, un nombre non négligeable de marqueurs spécifiques a également été détecté dans l'échantillon final, suggérant l'existence de mécanismes (*i.e.*, réactions de surface) ou interférences analytique (*i.e.*, effet de matrice) qu'il serait intéressant d'étudier de façon plus approfondie. Par ailleurs, différentes stratégies analytiques ont également été appliquées à plusieurs types de procédés d'oxydation afin de mieux caractériser les méthodes d'analyses mais également les TP formés. Cette partie complémentaire a ainsi mis en avant le manque de connaissances vis-à-vis de certaines phases de SPE qui s'avèrent efficaces pour la rétention de TP. Elle a aussi permis de mettre au point des méthodes permettant de mieux comprendre les mécanismes de réaction mis en jeu lors des procédés d'oxydation, notamment en établissant des liens de transformations entre des signaux correspondant à des composés bien éliminés par le traitement et de nouveaux signaux formés lors du processus. Globalement cette partie a pu mettre en évidence l'importance des stratégies d'analyses basées sur le SS et le NTS pour l'évaluation des procédés d'oxydation qui viennent compléter les observations réalisées en TS, et parfois, les contredire. Ces éléments viennent ainsi

confirmer l'intérêt des méthodes de SS et NTS qui permettent de mieux comprendre le devenir des micropolluants organiques.

Perspectives

Ce travail a permis d'apporter de nouveaux éléments quant au processus de NTS d'échantillons aqueux environnementaux, mais a également soulevé de nouvelles interrogations.

D'un point de vue du traitement de données, il apparaît évident que de nombreux défis restent encore à surmonter. Notamment, malgré la puissance des instruments analytiques et des algorithmes de traitement de données, l'identification de composés inconnus reste un processus très long. La mobilité ionique possède un bon potentiel pour aider en ce sens mais est encore peu utilisée à son avantage par les logiciels existants. Lors de ce travail, nous avons tenté d'exploiter cette propriété, mais l'obstacle principal à cela portait sur l'accessibilité directe aux données spectrales et aux valeurs de CCS lorsque les données sont exploitées en dehors du logiciel fournisseur. Cet obstacle est en cours de résolution puisque l'installation d'une API permet désormais l'accès aux données brutes acquises. Cependant, la récupération de ces dernières nécessite des efforts supplémentaires de développement logiciel pour adapter les données récupérées aux formats utilisés par les logiciels libres. Par ailleurs, les méthodes actuelles ont encore souvent pour objectif d'identifier un nombre limité de signaux choisis à la suite d'un processus de priorisation. Les stratégies basées sur des caractérisations plus globales restent rares en comparaison, notamment à cause des informations limitées sur la nature des composés qu'elle fournit. D'autre part, dans le cadre de ce travail, certains échantillons ont été analysés à la fois par HRMS (UPLC-IMS-QToF) et UHRMS (FT-ICR-MS). Malgré une résolution et une précision en masse augmentée de cette dernière technique, les calculs de formules moléculaires ne permettent pas de faciliter considérablement l'attribution de formules dans des matrices complexes telles que les eaux de STEU. Cette observation suggère ainsi que les limitations actuelles en termes d'identification semblent plus liées aux outils de traitement de données, plutôt qu'aux instruments d'acquisition, et aux limites théoriques atteintes avec ces techniques (impossibilité d'attribuer de manière certaine une formule définitive pour certaines masses précises).

En ce qui concerne la caractérisation des méthodes d'analyse, il est nécessaire d'évaluer de manière exhaustive le type de composés qui ne peuvent pas être observés par une méthode donnée. Pour cela, le meilleur moyen est de tenter d'analyser le plus d'étalons possibles avec

une approche en SS afin de rendre compte des composés observables ou non en fonction de leurs propriétés physico-chimiques, mais une caractérisation d'échantillons complexes avec une approche en NTS peut également donner des informations complémentaires voire parfois différentes, malgré des possibilités d'interprétations plus réduites en termes de propriétés physico-chimiques. De façon évidente, la combinaison de différentes stratégies est fortement recommandée afin de regrouper le plus d'informations possibles. Ainsi, une cartouche multiphasique a été sélectionnée lors de ce travail, mais l'utilisation de cette cartouche nécessite actuellement un assemblage manuel des différentes phases qui peut ainsi introduire un biais lors de l'analyse (*i.e.*, répétabilité, contamination...). Différents fournisseurs ont été contactés afin de produire cette cartouche de façon industrielle pour éviter ces problèmes, sans succès à ce jour, malgré l'augmentation du nombre de publications décrivant l'utilisation de ce support. Par ailleurs, dans les cas de l'étude des TP d'oxydation, il peut être envisagé de modifier cette cartouche multiphasique en remplaçant une phase (ex. ENV) par une autre qui a montré une capacité à retenir de nombreux signaux spécifiques des différents traitements d'oxydation étudiés (DP ou Coco). En parallèle, il semble que l'extraction induise la perte d'un certain nombre de signaux et puisse également altérer la composition de l'échantillon. Ainsi, le développement de méthodes permettant de concentrer les échantillons sans extraction ont gagné de l'intérêt récemment, avec notamment l'emploi de l'évaporation rotative ou de la lyophilisation. En ce sens, la méthode de fractionnement par HPLC-SEC étudiée au cours de ce travail pourrait également être approfondie en concentrant davantage l'échantillon en amont du fractionnement ou en concentrant les fractions obtenues. Des investigations complémentaires devraient aussi être menées afin de voir si le changement de température ou d'état, ou encore la précipitation liée à l'évaporation totale de la phase aqueuse, aurait une plus faible influence sur l'intégrité de l'échantillon par comparaison à l'adsorption/désorption qui a lieu lors de l'extraction.

Malgré ces différentes voies d'amélioration possible, les méthodes d'analyse en SS et NTS ont prouvé leur avantage par rapport aux méthodes classiques en TS, notamment face à la problématique des TP. Avec une stratégie analytique bien définie, les caractérisations globales permettent d'apporter une vision globale et rapide de l'évolution des échantillons. Ce manuscrit a décrit l'utilisation de ces méthodes pour étudier l'évolution des échantillons suite à un procédé d'oxydation.

Ces stratégies ont également été employées pour suivre l'évolution temporelle de la composition des eaux usées, en particulier, suite à la crise sanitaire liée à l'épidémie de COVID-19. Ces travaux en cours montrent l'intérêt de mettre en lumière les informations obtenues en SS et en NTS avec des études complémentaires. En particulier dans notre cas, les données issues des paramètres physico-chimiques réglementaires mesurés en station, les caractéristiques de fluorescence, les concentrations du génome de ce virus, les données en TS sur certaines molécules ainsi que des données écotoxicologiques, permettent une meilleure interprétation des observations. L'objectif de cette étude est de déterminer si les changements d'habitudes liées aux restrictions imposées pour contenir l'épidémie pouvaient se refléter dans la qualité des eaux en entrées de STEU. De manière similaire, un suivi pérenne a également été initié avec le SIAAP dans le cadre du programme de recherche OPUR, notamment en comparant différentes fréquences de prélèvement.

Un suivi de long terme de ce type a différents objectifs :

- Permettre la détection de contaminants émergents par suivi temporel
- Acquérir une banque d'échantillons, notamment pour réaliser des études rétrospectives
- Enrichir les bases de données internes pour faciliter les identifications futures
- Suivre l'évolution des habitudes de consommation suite à différents événements (par exemple, mise en œuvre de nouvelles réglementations, prises de conscience par les consommateurs comme dans le cas des parabènes, crises sanitaires (COVID-19) ou économiques (BREXIT ou conséquences de la guerre en Ukraine...))

Il fournira ainsi des informations régulières sur la diversité de micropolluants organiques entrant en STEU, et permettra d'anticiper la mise en place de nouvelles réglementations et de suivre l'impact que celles-ci peuvent engendrer sur la qualité des eaux.

Références

- Aalizadeh, R., Nika, M.-C., and Thomaidis, N.S. (2019). Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants. *Journal of Hazardous Materials* 363, 277–285.
- Aalizadeh, R., Alygizakis, N.A., Schymanski, E.L., Krauss, M., Schulze, T., Ibáñez, M., McEachran, A.D., Chao, A., Williams, A.J., Gago-Ferrero, P., et al. (2021). Development and Application of Liquid Chromatographic Retention Time Indices in HRMS-Based Suspect and Nontarget Screening. *Anal. Chem.*
- Albergamo, V., Schollée, J.E., Schymanski, E.L., Helmus, R., Timmer, H., Hollender, J., and de Voogt, P. (2019). Nontarget Screening Reveals Time Trends of Polar Micropollutants in a Riverbank Filtration System. *Environ. Sci. Technol.*
- Alexander, J.T., Hai, F.I., and Al-aboud, T.M. (2012). Chemical coagulation-based processes for trace organic contaminant removal: Current state and future potential. *Journal of Environmental Management* 111, 195–207.
- Ali, Z., and Bhaskar, S.B. (2016). Basic statistical tools in research and data analysis. *Indian J Anaesth* 60, 662–669.
- Alygizakis, N.A., Samanipour, S., Hollender, J., Ibáñez, M., Kaserzon, S., Kokkali, V., van Leerdam, J.A., Mueller, J.F., Pijnappels, M., Reid, M.J., et al. (2018). Exploring the Potential of a Global Emerging Contaminant Early Warning Network through the Use of Retrospective Suspect Screening with High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* 52, 5135–5144.
- Alygizakis, N.A., Gago-Ferrero, P., Hollender, J., and Thomaidis, N.S. (2019). Untargeted time-pattern analysis of LC-HRMS data to detect spills and compounds with high fluctuation in influent wastewater. *Journal of Hazardous Materials* 361, 19–29.
- Andersson, A., Harir, M., Gonsior, M., Hertkorn, N., Schmitt-Kopplin, P., Kylin, H., Karlsson, S., Ashiq, M.J., Lavonen, E., Nilsson, K., et al. (2019). Waterworks-specific composition of drinking water disinfection by-products. *Environ. Sci.: Water Res. Technol.* 5, 861–872.
- Asghar, A., Lutze, H.V., Tuerk, J., and Schmidt, T.C. (2022). Influence of water matrix on the degradation of organic micropollutants by ozone based processes: A review on oxidant scavenging mechanism. *Journal of Hazardous Materials* 429, 128189.
- Babushok, V.I. (2015). Chromatographic retention indices in identification of chemical compounds. *TrAC Trends in Analytical Chemistry* 69, 98–104.
- Bade, R., Rousis, N.I., Bijlsma, L., Gracia-Lor, E., Castiglioni, S., Sancho, J.V., and Hernandez, F. (2015). Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS. *Anal Bioanal Chem* 407, 8979–8988.
- Bader, H., and Hoigné, J. (1981). Determination of ozone in water by the indigo method. *Water Research* 15, 449–456.

- Bader, T., Schulz, W., Kümmerer, K., and Winzenbacher, R. (2017). LC-HRMS Data Processing Strategy for Reliable Sample Comparison Exemplified by the Assessment of Water Treatment Processes. *Anal. Chem.* *89*, 13219–13226.
- Balakrishnan, V.K., Terry, K.A., and Toito, J. (2006). Determination of sulfonamide antibiotics in wastewater: A comparison of solid phase microextraction and solid phase extraction methods. *Journal of Chromatography A* *1131*, 1–10.
- Bauer, C., Cramer, R., and Schuchhardt, J. (2011). Evaluation of Peak-Picking Algorithms for Protein Mass Spectrometry. In *Data Mining in Proteomics: From Standards to Applications*, M. Hamacher, M. Eisenacher, and C. Stephan, eds. (Totowa, NJ: Humana Press), pp. 341–352.
- Baz-Lomba, J.A., Reid, M.J., and Thomas, K.V. (2016). Target and suspect screening of psychoactive substances in sewage-based samples by UHPLC-QTOF. *Analytica Chimica Acta* *914*, 81–90.
- Bergé, A., Buleté, A., Fildier, A., Mailler, R., Gasperi, J., Coquet, Y., Nauleau, F., Rocher, V., and Vulliet, E. (2018). Non-target strategies by HRMS to evaluate fluidized micro-grain activated carbon as a tertiary treatment of wastewater. *Chemosphere* *213*, 587–595.
- Berkner, S., and Thierbach, C. (2014). Biodegradability and transformation of human pharmaceutical active ingredients in environmentally relevant test systems. *Environ Sci Pollut Res* *21*, 9461–9467.
- Bieber, S., Greco, G., Grosse, S., and Letzel, T. (2017). RPLC-HILIC and SFC with Mass Spectrometry: Polarity-Extended Organic Molecule Screening in Environmental (Water) Samples. *Anal. Chem.* *89*, 7907–7914.
- Bijlsma, L., Bade, R., Celma, A., Mullin, L., Cleland, G., Stead, S., Hernandez, F., and Sancho, J.V. (2017). Prediction of Collision Cross-Section Values for Small Molecules: Application to Pesticide Residue Analysis. *Anal. Chem.* *89*, 6583–6589.
- Blaženović, I., Kind, T., Ji, J., and Fiehn, O. (2018). Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* *8*.
- Boccard, J., Veuthey, J.-L., and Rudaz, S. (2010). Knowledge discovery in metabolomics: An overview of MS data handling. *Journal of Separation Science* *33*, 290–304.
- Böcker, S., Lipták, Z., Martin, M., Pervukhin, A., and Sudek, H. (2008). Decomp—from interpreting Mass Spectrometry peaks to solving the Money Changing Problem. *Bioinformatics* *24*, 591–593.
- Boehler, M., Zwickenpflug, B., Hollender, J., Ternes, T., Joss, A., and Siegrist, H. (2012). Removal of micropollutants in municipal wastewater treatment plants by powder-activated carbon. *Water Science and Technology* *66*, 2115–2121.
- Boersema, P.J., Mohammed, S., and Heck, A.J.R. (2008). Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Anal Bioanal Chem* *391*, 151–159.

- Bolong, N., Ismail, A.F., Salim, M.R., and Matsuura, T. (2009). A review of the effects of emerging contaminants in wastewater and options for their removal. *Desalination* 239, 229–246.
- Bolton, E.E., Wang, Y., Thiessen, P.A., and Bryant, S.H. (2008). Chapter 12 - PubChem: Integrated Platform of Small Molecules and Biological Activities. In *Annual Reports in Computational Chemistry*, R.A. Wheeler, and D.C. Spellmeyer, eds. (Elsevier), pp. 217–241.
- Bolton, J.R., Mayor-Smith, I., and Linden, K.G. (2015). Rethinking the Concepts of Fluence (UV Dose) and Fluence Rate: The Importance of Photon-based Units – A Systemic Review. *Photochemistry and Photobiology* 91, 1252–1262.
- Bourgin, M., Beck, B., Boehler, M., Borowska, E., Fleiner, J., Salhi, E., Teichler, R., von Gunten, U., Siegrist, H., and Mc Ardell, C.S. (2018). Evaluation of a full-scale wastewater treatment plant upgraded with ozonation and biological post-treatments: Abatement of micropollutants, formation of transformation products and oxidation by-products. *Water Research* 129, 486–498.
- Brack, W. (2003). Effect-directed analysis: a promising tool for the identification of organic toxicants in complex mixtures? *Anal Bioanal Chem* 377, 397–407.
- Brock, O., Helmus, R., Kalbitz, K., and Jansen, B. (2020). Non-target screening of leaf litter-derived dissolved organic matter using liquid chromatography coupled to high-resolution mass spectrometry (LC-QTOF-MS). *European Journal of Soil Science* 71, 420–432.
- Brunner, A.M., Bertelkamp, C., Dingemans, M.M.L., Kolkman, A., Wols, B., Harmsen, D., Siegers, W., Martijn, B.J., Oorthuizen, W.A., and ter Laak, T.L. (2020). Integration of target analyses, non-target screening and effect-based monitoring to assess OMP related water quality changes in drinking water treatment. *Science of The Total Environment* 705, 135779.
- Busch, W., Schmidt, S., Kühne, R., Schulze, T., Krauss, M., and Altenburger, R. (2016). Micropollutants in European rivers: A mode of action survey to support the development of effect-based tools for water monitoring. *Environmental Toxicology and Chemistry* 35, 1887–1899.
- Buxton, G.V., Greenstock, C.L., Helman, W.P., and Ross, A.B. (1988). Critical Review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ($\cdot\text{OH}/\cdot\text{O}$ in Aqueous Solution). *Journal of Physical and Chemical Reference Data* 17, 513–886.
- Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., and Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics* 20, 341–351.
- Calderón-Preciado, D., Matamoros, V., and Bayona, J.M. (2011). Occurrence and potential crop uptake of emerging contaminants and related compounds in an agricultural irrigation network. *Science of The Total Environment* 412–413, 14–19.
- Campo, N., De Flora, C., Maffettone, R., Manoli, K., Sarathy, S., Santoro, D., Gonzalez-Olmos, R., and Auset, M. (2020). Inactivation kinetics of antibiotic resistant *Escherichia coli* in

secondary wastewater effluents by peracetic and performic acids. *Water Research* *169*, 115227.

Cañameras, N., Comas, J., and Bayona, J.M. (2016). Bioavailability and Uptake of Organic Micropollutants During Crop Irrigation with Reclaimed Wastewater: Introduction to Current Issues and Research Needs. In *Wastewater Reuse and Current Challenges*, D. Fatta-Kassinos, D.D. Dionysiou, and K. Kümmerer, eds. (Cham: Springer International Publishing), pp. 81–104.

Canonica, S., Meunier, L., and von Gunten, U. (2008). Phototransformation of selected pharmaceuticals during UV treatment of drinking water. *Water Research* *42*, 121–128.

Carballa, M., Omil, F., and Lema, J.M. (2005). Removal of cosmetic ingredients and pharmaceuticals in sewage primary treatment. *Water Research* *39*, 4790–4796.

Cariou, R., Omer, E., Léon, A., Dervilly-Pinel, G., and Le Bizec, B. (2016). Screening halogenated environmental contaminants in biota based on isotopic pattern and mass defect provided by high resolution mass spectrometry profiling. *Analytica Chimica Acta* *936*, 130–138.

Castro, V., Quintana, J.B., Carpinteiro, I., Cobas, J., Carro, N., Cela, R., and Rodil, R. (2021). Combination of different chromatographic and sampling modes for high-resolution mass spectrometric screening of organic microcontaminants in water. *Anal Bioanal Chem* *413*, 5607–5618.

Caupos, E., Touffet, A., Mazellier, P., and Croue, J.-P. (2015). Partitioning of the pesticide trifluralin between dissolved organic matter and water using automated SPME-GC/MS. *Environ Sci Pollut Res* *22*, 4201–4212.

Causon, T.J., Ivanova-Petropulos, V., Petrusheva, D., Bogeve, E., and Hann, S. (2018). Fingerprinting of traditionally produced red wines using liquid chromatography combined with drift tube ion mobility-mass spectrometry (containing tracked changes). *Analytica Chimica Acta*.

Celander, M.C. (2011). Cocktail effects on biomarker responses in fish. *Aquatic Toxicology* *105*, 72–77.

Celma, A., Sancho, J.V., Schymanski, E.L., Fabregat-Safont, D., Ibáñez, M., Goshawk, J., Barknowitz, G., Hernández, F., and Bijlsma, L. (2020). Improving Target and Suspect Screening High-Resolution Mass Spectrometry Workflows in Environmental Analysis by Ion Mobility Separation. *Environ. Sci. Technol.* *54*, 15120–15131.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* *30*, 918–920.

Chen, J., LeBoeuf, E.J., Dai, S., and Gu, B. (2003). Fluorescence spectroscopic studies of natural organic matter fractions. *Chemosphere* *50*, 639–647.

- Chen, J., Wang, W., Lv, S., Yin, P., Zhai, X., Lu, X., Zhang, F., and Xu, G. (2009). Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations. *Analytica Chimica Acta* 3–9.
- Chhetri, R.K., Thornberg, D., Berner, J., Gramstad, R., Öjstedt, U., Sharma, A.K., and Andersen, H.R. (2014). Chemical disinfection of combined sewer overflow waters using performic acid or peracetic acids. *Science of The Total Environment* 490, 1065–1072.
- Chon, K., Chon, K., and Cho, J. (2017). Characterization of size fractionated dissolved organic matter from river water and wastewater effluent using preparative high performance size exclusion chromatography. *Organic Geochemistry* 103, 105–112.
- Clara, M., Windhofer, G., Weilgony, P., Gans, O., Denner, M., Chovanec, A., and Zessner, M. (2012). Identification of relevant micropollutants in Austrian municipal wastewater and their behaviour during wastewater treatment. *Chemosphere* 87, 1265–1272.
- Coble, P.G. (1996). Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry* 51, 325–346.
- Conway, J., and Gehlenborg, N. (2019). UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets.
- Corilo, Y. PetroOrg. <https://www.research.fsu.edu/research-offices/oc/technologies/petroorg-software/>, consulté le 04/06/2020
- Cortéjade, A., Kiss, A., Cren, C., Vulliet, E., and Buleté, A. (2016). Development of an analytical method for the targeted screening and multi-residue quantification of environmental contaminants in urine by liquid chromatography coupled to high resolution mass spectrometry for evaluation of human exposures. *Talanta* 146, 694–706.
- Cortés, C., and Marcos, R. (2018). Genotoxicity of disinfection byproducts and disinfected waters: A review of recent literature. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 831, 1–12.
- Criegee, R. (1975). Mechanism of Ozonolysis. *Angewandte Chemie International Edition in English* 14, 745–752.
- Croué, J.-P. (2004). Isolation of Humic and Non-Humic NOM Fractions: Structural Characterization. *Environmental Monitoring and Assessment* 92, 193–207.
- Cubbon, S., Antonio, C., Wilson, J., and Thomas-Oates, J. (2010). Metabolomic applications of HILIC–LC–MS. *Mass Spectrometry Reviews* 29, 671–684.
- Deeb, A.A., Stephan, S., Schmitz, O.J., and Schmidt, T.C. (2017). Suspect screening of micropollutants and their transformation products in advanced wastewater treatment. *Science of The Total Environment* 601–602, 1247–1253.
- Delaporte, G., Cladière, M., and Camel, V. (2019). Missing value imputation and data cleaning in untargeted food chemical safety assessment by LC-HRMS. *Chemometrics and Intelligent Laboratory Systems* 188, 54–62.

- Dempsey, B.A., and O'Melia, C.R. (1984). Removal of naturally occurring compounds by coagulation and sedimentation. *Critical Reviews in Environmental Control* *14*, 311–331.
- Deng, Y., and Zhao, R. (2015). Advanced Oxidation Processes (AOPs) in Wastewater Treatment. *Curr Pollution Rep* *1*, 167–176.
- Díaz, R., Ibáñez, M., V. Sancho, J., and Hernández, F. (2012). Target and non-target screening strategies for organic contaminants, residues and illicit substances in food, environmental and human biological samples by UHPLC-QTOF-MS. *Analytical Methods* *4*, 196–209.
- Diaz, R., Ibáñez, M., Sancho, J.V., and Hernández, F. (2013). Qualitative validation of a liquid chromatography–quadrupole-time of flight mass spectrometry screening method for organic pollutants in waters. *Journal of Chromatography A* *1276*, 47–57.
- Domingo-Almenara, X., Montenegro-Burke, J.R., Guijas, C., Majumder, E.L.-W., Benton, H.P., and Siuzdak, G. (2019). Autonomous METLIN-Guided In-source Fragment Annotation for Untargeted Metabolomics. *Analytical Chemistry*.
- Drotleff, B., and Lämmerhofer, M. (2019). Guidelines for Selection of Internal Standard-Based Normalization Strategies in Untargeted Lipidomic Profiling by LC-HR-MS/MS. *Anal. Chem.* *91*, 9836–9843.
- Du, P., Kibbe, W.A., and Lin, S.M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* *22*, 2059–2065.
- Duan, J., Li, W., Sun, P., Lai, Q., Mulcahy, D., and Guo, S. (2013). Rapid Determination of Nine Haloacetic Acids in Wastewater Effluents Using Ultra-Performance Liquid Chromatography-Tandem Mass Spectrometry. *Analytical Letters* *46*, 569–588.
- Dührkop, K., Shen, H., Meusel, M., Rousu, J., and Böcker, S. (2015). Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U.S.A.* *112*, 12580–12585.
- Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A.V., Meusel, M., Dorrestein, P.C., Rousu, J., and Böcker, S. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods* *16*, 299–302.
- Eggen, R.I.L., Hollender, J., Joss, A., Schärer, M., and Stamm, C. (2014). Reducing the Discharge of Micropollutants in the Aquatic Environment: The Benefits of Upgrading Wastewater Treatment Plants. *Environ. Sci. Technol.* *48*, 7683–7689.
- Escher, B.I., van Daele, C., Dutt, M., Tang, J.Y.M., and Altenburger, R. (2013). Most Oxidative Stress Response In Water Samples Comes From Unknown Chemicals: The Need For Effect-Based Water Quality Trigger Values. *Environ. Sci. Technol.* *47*, 7002–7011.
- Evans, A., Bridgewater, B., Liu, Q., Mitchell, M., Robinson, R., Dai, H., Stewart, S., Dehaven, C., and Miller, L. (2014). High Resolution Mass Spectrometry Improves Data Quantity and Quality

as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. *Metabolomics* 04.

Fagnani, E., Montemurro, N., and Pérez, S. (2022). Multilayered solid phase extraction and ultra performance liquid chromatographic method for suspect screening of halogenated pharmaceuticals and photo-transformation products in freshwater - comparison between data-dependent and data-independent acquisition mass spectrometry. *Journal of Chromatography A* 1663, 462760.

Ferrer, I., Writer, J., Keen, O., Lester, Y., Padilla-Sanchez, J., and Thurman, E. (2016). LC-TOF-MS for the identification of environmental metabolites and degradation products. In *Applications of Time-of-Flight and Orbitrap Mass Spectrometry in Environmental, Food, Doping and Forensic Analysis*, (D.Barcelo), pp. 231–261.

Feuerstein, M.L., Kurulugama, R.T., Hann, S., and Causon, T. (2021). Novel acquisition strategies for metabolomics using drift tube ion mobility-quadrupole resolved all ions time-of-flight mass spectrometry (IM-QRAI-TOFMS). *Analytica Chimica Acta* 1163, 338508.

Filippis, P.D., Scarsella, M., and Verdone, N. (2009). Peroxyformic Acid Formation: A Kinetic Study. *Ind. Eng. Chem. Res.* 48, 1372–1375.

Fischer, B., Neumann, S., Gatto, L., Kou, Q., and Rainer, J. (2021). mzR: parser for netCDF, mzXML, mzData and mzML and mzIdentML files (mass spectrometry data) (Bioconductor version: Release (3.13)).

Fisher, C.M., Croley, T.R., and Knolhoff, A.M. (2021). Data Processing Strategies for Non-Targeted Analysis of Foods using Liquid Chromatography/High-Resolution Mass Spectrometry. *TrAC Trends in Analytical Chemistry* 116188.

Fontanals, N., Marcé, R.M., and Borrull, F. (2005). New hydrophilic materials for solid-phase extraction. *TrAC Trends in Analytical Chemistry* 24, 394–406.

Fontanals, N., Marcé, R.M., and Borrull, F. (2007). New materials in sorptive extraction techniques for polar compounds. *Journal of Chromatography A* 1152, 14–31.

Fu, Q.-L., Fujii, M., and Riedel, T. (2020). Development and Comparison of Formula Assignment Algorithms for Ultrahigh-resolution Mass Spectra of Natural Organic Matter. *Analytica Chimica Acta*.

Gabelica, V., Shvartsburg, A.A., Afonso, C., Barran, P., Benesch, J.L.P., Bleiholder, C., Bowers, M.T., Bilbao, A., Bush, M.F., Campbell, J.L., et al. (2019). Recommendations for reporting ion mobility Mass Spectrometry measurements. *Mass Spectrometry Reviews* 38, 291–320.

Gago-Ferrero, P., Schymanski, E.L., Bletsou, A.A., Aalizadeh, R., Hollender, J., and Thomaidis, N.S. (2015). Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environmental Science & Technology* 49, 12333–12341.

- Gao, P., Ding, Y., Li, H., and Xagorarakis, I. (2012). Occurrence of pharmaceuticals in a municipal wastewater treatment plant: Mass balance and removal processes. *Chemosphere* *88*, 17–24.
- Gasperi, J., Rocher, V., Gilbert, S., Azimi, S., and Chebbo, G. (2010). Occurrence and removal of priority pollutants by lamella clarification and biofiltration. *Water Research* *44*, 3065–3076.
- Gasperi, J., Geara, D., Lorgeoux, C., Bressy, A., Zedek, S., Rocher, V., El Samrani, A., Chebbo, G., and Moilleron, R. (2014). First assessment of triclosan, triclocarban and paraben mass loads at a very large regional scale: Case of Paris conurbation (France). *Science of The Total Environment* *493*, 854–861.
- Gehr, R., Chen, D., and Moreau, M. (2009). Performic acid (PFA): tests on an advanced primary effluent show promising disinfection performance. *Water Science and Technology* *59*, 89–96.
- Gerrity, D., Gamage, S., Jones, D., Korshin, G.V., Lee, Y., Pisarenko, A., Trenholm, R.A., von Gunten, U., Wert, E.C., and Snyder, S.A. (2012). Development of surrogate correlation models to predict trace organic contaminant oxidation and microbial inactivation during ozonation. *Water Research* *46*, 6257–6272.
- Giger, W. (2008). Hydrophilic and amphiphilic water pollutants: using advanced analytical methods for classic and emerging contaminants. *Anal Bioanal Chem* *393*, 37.
- Gilca, A.F., Teodosiu, C., Fiore, S., and Musteret, C.P. (2020). Emerging disinfection byproducts: A review on their occurrence and control in drinking water treatment processes. *Chemosphere* *259*, 127476.
- Godzien, J., Alonso-Herranz, V., Barbas, C., and Armitage, E.G. (2015). Controlling the quality of metabolomics data: new strategies to get the best out of the QC sample. *Metabolomics* *11*, 518–528.
- Golovko, O., Kumar, V., Fedorova, G., Randak, T., and Grabic, R. (2014). Removal and seasonal variability of selected analgesics/anti-inflammatory, anti-hypertensive/cardiovascular pharmaceuticals and UV filters in wastewater treatment plant. *Environ Sci Pollut Res* *21*, 7578–7585.
- González-Gaya, B., Lopez-Herguedas, N., Bilbao, D., Mijangos, L., Iker, A.M., Etxebarria, N., Irazola, M., Prieto, A., Olivares, M., and Zuloaga, O. (2021). Suspect and non-target screening: the last frontier in environmental analysis. *Anal. Methods* *13*, 1876–1904.
- Green, N.W., and Perdue, E.M. (2015). Fast Graphically Inspired Algorithm for Assignment of Molecular Formulae in Ultrahigh Resolution Mass Spectrometry. *Anal. Chem.* *87*, 5086–5094.
- Grung, M., Lichtenthaler, R., Ahel, M., Tollefsen, K.-E., Langford, K., and Thomas, K.V. (2007). Effects-directed analysis of organic toxicants in wastewater effluent from Zagreb, Croatia. *Chemosphere* *67*, 108–120.
- Gu, G., and Lim, C.K. (1990). Separation of anionic and cationic compounds of biomedical interest by high-performance liquid chromatography on porous graphitic carbon. *Journal of Chromatography A* *515*, 183–192.

- Guillossou, R., Le Roux, J., Mailler, R., Vulliet, E., Morlay, C., Nauleau, F., Gasperi, J., and Rocher, V. (2019). Organic micropollutants in a large wastewater treatment plant: What are the benefits of an advanced treatment by activated carbon adsorption in comparison to conventional treatment? *Chemosphere* *218*, 1050–1060.
- Guillossou, R., Le Roux, J., Brosillon, S., Mailler, R., Vulliet, E., Morlay, C., Nauleau, F., Rocher, V., and Gaspéri, J. (2020). Benefits of ozonation before activated carbon adsorption for the removal of organic micropollutants from wastewater effluents. *Chemosphere* *245*, 125530.
- Guillossou, R., Le Roux, J., Goffin, A., Mailler, R., Varrault, G., Vulliet, E., Morlay, C., Nauleau, F., Guérin, S., Rocher, V., et al. (2021). Fluorescence excitation/emission matrices as a tool to monitor the removal of organic micropollutants from wastewater effluents by adsorption onto activated carbon. *Water Research* *190*, 116749.
- Gulde, R., Meier, U., Schymanski, E.L., Kohler, H.-P.E., Helbling, D.E., Derrer, S., Rentsch, D., and Fenner, K. (2016). Systematic Exploration of Biotransformation Reactions of Amine-Containing Micropollutants in Activated Sludge. *Environ. Sci. Technol.* *50*, 2908–2920.
- Gulde, R., Rutsch, M., Clerc, B., Schollée, J.E., von Gunten, U., and McArdell, C.S. (2021a). Formation of transformation products during ozonation of secondary wastewater effluent and their fate in post-treatment: From laboratory- to full-scale. *Water Research* *200*, 117200.
- Gulde, R., Clerc, B., Rutsch, M., Helbing, J., Salhi, E., McArdell, C.S., and von Gunten, U. (2021b). Oxidation of 51 micropollutants during drinking water ozonation: Formation of transformation products and their fate during biological post-filtration. *Water Research* *207*, 117812.
- Gundersen, J.L. (2001). Separation of isomers of nonylphenol and select nonylphenol polyethoxylates by high-performance liquid chromatography on a graphitic carbon column. *Journal of Chromatography A* *914*, 161–166.
- von Gunten, U. (2003). Ozonation of drinking water: Part I. Oxidation kinetics and product formation. *Water Research* *37*, 1443–1467.
- Guo, J., and Huan, T. (2020). Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography–Mass Spectrometry Based Untargeted Metabolomics. *Anal. Chem.*
- Hajeb, P., Zhu, L., Bossi, R., and Vorkamp, K. (2022). Sample preparation techniques for suspect and non-target screening of emerging contaminants. *Chemosphere* *287*, 132306.
- Hashmi, M.A.K., Escher, B.I., Krauss, M., Teodorovic, I., and Brack, W. (2018). Effect-directed analysis (EDA) of Danube River water sample receiving untreated municipal wastewater from Novi Sad, Serbia. *Science of The Total Environment* *624*, 1072–1081.
- Hautbergue, T., Jamin, E.L., Costantino, R., Tadriss, S., Meneghetti, L., Tabet, J.-C., Debrauwer, L., Oswald, I.P., and Puel, O. (2019). Combination of Isotope Labeling and Molecular Networking of Tandem Mass Spectrometry Data To Reveal 69 Unknown Metabolites Produced by *Penicillium nordicum*. *Anal. Chem.*

- He, K., and Blaney, L. (2015). Systematic optimization of an SPE with HPLC-FLD method for fluoroquinolone detection in wastewater. *Journal of Hazardous Materials* 282, 96–105.
- Heinonen-Tanski, H., and Miettinen, H. (2010). Performic Acid as a Potential Disinfectant at Low Temperature. *Journal of Food Process Engineering* 33, 1159–1172.
- Helmus, R., ter Laak, T.L., van Wezel, A.P., de Voogt, P., and Schymanski, E.L. (2021). patRoom: open source software platform for environmental mass spectrometry based non-target screening. *Journal of Cheminformatics* 13, 1.
- Hernández, F., Pozo, Ó.J., Sancho, J.V., López, F.J., Marín, J.M., and Ibáñez, M. (2005). Strategies for quantification and confirmation of multi-class polar pesticides and transformation products in water by LC–MS2 using triple quadrupole and hybrid quadrupole time-of-flight analyzers. *TrAC Trends in Analytical Chemistry* 24, 596–612.
- Hessler, D.P., Gorenflo, V., and Frimmel, F.H. (1993). Degradation of Aqueous Atrazine and Metazachlor Solutions by UV and UV/H₂O₂ — Influence of pH and Herbicide Concentration. *Abbau von Atrazin und Metazachlor in wäßriger Lösung durch UV und UV/H₂O₂ — Einfluß von pH und Herbizid-Konzentration. Acta Hydrochimica et Hydrobiologica* 21, 209–214.
- Hinnenkamp, V., Klein, J., Meckelmann, S.W., Balsaa, P., Schmidt, T.C., and Schmitz, O.J. (2018). Comparison of CCS Values Determined by Traveling Wave Ion Mobility Mass Spectrometry and Drift Tube Ion Mobility Mass Spectrometry. *Anal. Chem.* 90, 12042–12050.
- Hogenboom, A.C., van Leerdam, J.A., and de Voogt, P. (2009). Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography–hybrid linear ion trap Orbitrap mass spectrometry. *Journal of Chromatography A* 1216, 510–519.
- Hohrenk, L.L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., and Schmidt, T.C. (2019). Comparison of Software Tools for Liquid Chromatography–High-Resolution Mass Spectrometry Data Processing in Nontarget Screening of Environmental Samples. *Anal. Chem.*
- Hoigné, J., and Bader, H. (1975). Ozonation of Water: Role of Hydroxyl Radicals as Oxidizing Intermediates. *Science*.
- Hoigné, J., Bader, H., Haag, W., and Staehelin, J. (1985). Rate constants of reactions of ozone with organic and inorganic compounds in water—III. Inorganic compounds and radicals. *Water Research* 19, 993–1004.
- Hollender, J., Zimmermann, S.G., Koepke, S., Krauss, M., McArdell, C.S., Ort, C., Singer, H., von Gunten, U., and Siegrist, H. (2009). Elimination of Organic Micropollutants in a Municipal Wastewater Treatment Plant Upgraded with a Full-Scale Post-Ozonation Followed by Sand Filtration. *Environ. Sci. Technol.* 43, 7862–7869.
- Hollender, J., Schymanski, E.L., Singer, H.P., and Ferguson, P.L. (2017). Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environmental Science & Technology* 51, 11505–11512.

- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry* *45*, 703–714.
- Hübner, U., von Gunten, U., and Jekel, M. (2015). Evaluation of the persistence of transformation products from ozonation of trace organic compounds – A critical review. *Water Research* *68*, 150–170.
- Huerta-Fontela, M., Galceran, M.T., and Ventura, F. (2011). Occurrence and removal of pharmaceuticals and hormones through drinking water treatment. *Water Research* *45*, 1432–1442.
- Hufsky, F., and Böcker, S. (2017). Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data. *Mass Spectrometry Reviews* *36*, 624–633.
- Hug, C., Ulrich, N., Schulze, T., Brack, W., and Krauss, M. (2014). Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environmental Pollution* *184*, 25–32.
- Hughey, C.A., Hendrickson, C.L., Rodgers, R.P., Marshall, A.G., and Qian, K. (2001). Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Analytical Chemistry* *73*, 4676–4681.
- Huntscha, S., Singer, H.P., Mc Ardell, C.S., Frank, C.E., and Hollender, J. (2012). Multiresidue analysis of 88 polar organic micropollutants in ground, surface and wastewater using online mixed-bed multilayer solid-phase extraction coupled to high performance liquid chromatography–tandem mass spectrometry. *Journal of Chromatography A* *1268*, 74–83.
- Husson, F., Josse, J., Le, S., and Mazet, J. (2020). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining.
- Huysman, S., Van Meulebroek, L., Janssens, O., Vanryckeghem, F., Van Langenhove, H., Demeestere, K., and Vanhaecke, L. (2018). Targeted quantification and untargeted screening of alkylphenols, bisphenol A and phthalates in aquatic matrices using ultra-high-performance liquid chromatography coupled to hybrid Q-Orbitrap mass spectrometry. *Analytica Chimica Acta*.
- Ibáñez, M., Sancho, J.V., Hernández, F., McMillan, D., and Rao, R. (2008). Rapid non-target screening of organic pollutants in water by ultraperformance liquid chromatography coupled to time-of-flight mass spectrometry. *TrAC Trends in Analytical Chemistry* *27*, 481–489.
- Ibáñez, M., Gracia-Lor, E., Bijlsma, L., Morales, E., Pastor, L., and Hernández, F. (2013). Removal of emerging contaminants in sewage water subjected to advanced oxidation with ozone. *Journal of Hazardous Materials* *260*, 389–398.
- Ikeda, T., Hashimoto, S., Isobe, T., Kunisue, T., and Tanabe, S. (2019). Evaluation of a data-processing method for target and non-target screening using comprehensive two-dimensional

gas chromatography coupled with high-resolution time-of-flight mass spectrometry for environmental samples. *Talanta* *194*, 461–468.

Ike, I.A., Karanfil, T., Cho, J., and Hur, J. (2019). Oxidation byproducts from the degradation of dissolved organic matter by advanced oxidation processes – A critical review. *Water Research* *114*929.

Ilani, T., Schulz, E., and Chefetz, B. (2005). Interactions of Organic Compounds with Wastewater Dissolved Organic Matter: Role of Hydrophobic Fractions. *Journal of Environmental Quality* *34*, 552–562.

Jeanneau, L., Faure, P., and Jardé, E. (2007). Influence of natural organic matter on the solid-phase extraction of organic micropollutants: Application to the water-extract from highly contaminated river sediment. *Journal of Chromatography A* *1173*, 1–9.

Jelić, A., Gros, M., Petrović, M., Ginebreda, A., and Barceló, D. (2012). Occurrence and Elimination of Pharmaceuticals During Conventional Wastewater Treatment. In *Emerging and Priority Pollutants in Rivers: Bringing Science into River Management Plans*, H. Guasch, A. Ginebreda, and A. Geiszinger, eds. (Berlin, Heidelberg: Springer), pp. 1–23.

Jiang, J.-Q., Zhou, Z., and Sharma, V.K. (2013). Occurrence, transportation, monitoring and treatment of emerging micro-pollutants in waste water — A review from global views. *Microchemical Journal* *110*, 292–300.

Jobst, K.J., Shen, L., Reiner, E.J., Taguchi, V.Y., Helm, P.A., McCrindle, R., and Backus, S. (2013). The use of mass defect plots for the identification of (novel) halogenated contaminants in the environment. *Anal Bioanal Chem* *405*, 3289–3297.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* *34*, D354–D357.

Karki, A.J., Cappelli, P., Dirks, C., Pekar, H., Hellenäs, K.-E., Rosén, J., and Westerberg, E. (2020). New efficient methodology for screening of selected organic micropollutants in raw- and drinking water from 90 Swedish water treatment plants. *Science of The Total Environment* *724*, 138069.

Karpova, T., Pekonen, P., Gramstad, R., Öjstedt, U., Laborda, S., Heinonen-Tanski, H., Chávez, A., and Jiménez, B. (2013). Performic acid for advanced wastewater disinfection. *Water Sci Technol* *68*, 2090–2096.

Kassambara, A., and Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses.

Katajamaa, M., and Orešič, M. (2005). Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* *6*, 179.

Kendrick (1963). A Mass Scale Based on CH₂ = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. | *Analytical Chemistry*. *Analytical Chemistry* *35*, 2146–2154.

- Kern, S., Fenner, K., Singer, H.P., Schwarzenbach, R.P., and Hollender, J. (2009). Identification of transformation products of organic contaminants in natural waters by computer-aided prediction and high-resolution mass spectrometry. *Environ. Sci. Technol.* *43*, 7039–7046.
- Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008). ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* *24*, 2534–2536.
- Kew, W., Blackburn, J.W.T., Clarke, D.J., and Uhrín, D. (2017). Interactive van Krevelen diagrams – Advanced visualisation of mass spectrometry data of complex mixtures. *Rapid Commun Mass Spectrom* *31*, 658–662.
- Kharel, S., Stapf, M., Miehe, U., Ekblad, M., Cimbritz, M., Falås, P., Nilsson, J., Sehlén, R., and Bester, K. (2020). Ozone dose dependent formation and removal of ozonation products of pharmaceuticals in pilot and full-scale municipal wastewater treatment plants. *Science of The Total Environment* *731*, 139064.
- Kim, J., and Huang, C.-H. (2021). Reactivity of Peracetic Acid with Organic Compounds: A Critical Review. *ACS EST Water* *1*, 15–33.
- Kim, I., Yamashita, N., and Tanaka, H. (2009). Performance of UV and UV/H₂O₂ processes for the removal of pharmaceuticals detected in secondary effluent of a sewage treatment plant in Japan. *Journal of Hazardous Materials* *166*, 1134–1140.
- Kind, T., and Fiehn, O. (2007). Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* *8*, 105.
- Kitis, M. (2004). Disinfection of wastewater with peracetic acid: a review. *Environment International* *30*, 47–55.
- Klinkow, N., Oleksy-Frenzel, J., and Jekel, M. (1998). Toxicity-directed fractionation of organic compounds in tannery wastewater with regard to their molecular weight and polarity. *Water Research* *32*, 2583–2592.
- van der Kloet, F.M., Bobeldijk, I., Verheij, E.R., and Jellema, R.H. (2009). Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. *J. Proteome Res.* *8*, 5132–5141.
- Koch, B.P., Dittmar, T., Witt, M., and Kattner, G. (2007). Fundamentals of Molecular Formula Assignment to Ultrahigh Resolution Mass Data of Natural Organic Matter. *Anal. Chem.* *79*, 1758–1763.
- Köppe, T., Jewell, K.S., Dietrich, C., Wick, A., and Ternes, T.A. (2020). Application of a non-target workflow for the identification of specific contaminants using the example of the Nidda river basin. *Water Research* 115703.
- Kovalova, L., McArdell, C.S., and Hollender, J. (2009). Challenge of high polarity and low concentrations in analysis of cytostatics and metabolites in wastewater by hydrophilic interaction chromatography/tandem mass spectrometry. *Journal of Chromatography A* *1216*, 1100–1108.

Krasner, S.W., Weinberg, H.S., Richardson, S.D., Pastor, S., Scilimenti, M.J., Onstad, G.D., and Thurston, A.D. (2006). Occurrence of a New Generation of Disinfection Byproducts. *Environmental Science and Technology* 7175–7186.

Krauss, M., Singer, H., and Hollender, J. (2010). LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Analytical and Bioanalytical Chemistry* 397, 943–951.

Krauss, M., Hug, C., Bloch, R., Schulze, T., and Brack, W. (2019). Prioritising site-specific micropollutants in surface water from LC-HRMS non-target screening data using a rarity score. *Environ Sci Eur* 31, 45.

Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T.R., and 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.

Kujawinski, E.B., and Behn, M.D. (2006). Automated Analysis of Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra of Natural Organic Matter. *Anal. Chem.* 78, 4363–4373.

Lacey, C., McMahon, G., Bones, J., Barron, L., Morrissey, A., and Tobin, J.M. (2008). An LC–MS method for the determination of pharmaceutical compounds in wastewater treatment plant influent and effluent samples. *Talanta* 75, 1089–1097.

Laganà, A., Bacaloni, A., Fago, G., and Marino, A. (2000). Trace analysis of estrogenic chemicals in sewage effluent using liquid chromatography combined with tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* 14, 401–407.

Lange, F., Cornelissen, S., Kubac, D., Sein, M.M., von Sonntag, J., Hannich, C.B., Golloch, A., Heipieper, H.J., Möder, M., and von Sonntag, C. (2006). Degradation of macrolide antibiotics by ozone: A mechanistic case study with clarithromycin. *Chemosphere* 65, 17–23.

Lardy-Fontan, S., Le Diuron, V., Drouin, C., Lalere, B., Vaslin-Reimann, S., Dauchy, X., and Rosin, C. (2017). Validation of a method to monitor the occurrence of 20 relevant pharmaceuticals and personal care products in 167 bottled waters. *Science of The Total Environment* 118–127.

Lavonen, E.E., Gonsior, M., Tranvik, L.J., Schmitt-Kopplin, P., and Köhler, S.J. (2013). Selective Chlorination of Natural Organic Matter: Identification of Previously Unknown Disinfection Byproducts. *Environmental Science and Technology* 2254–2271.

Lazzaro, S., Ogrinc, N., Lamont, L., Vecchio, G., Pappalardo, G., and Heeren, R.M.A. (2019). Ion mobility spectrometry combined with multivariate statistical analysis: revealing the effects of a drug candidate for Alzheimer's disease on A β 1-40 peptide early assembly. *Anal Bioanal Chem.*

Lebedev, A.T., Mazur, D.M., Artaev, V.B., and Tikhonov, G.Y. (2020). Better screening of non-target pollutants in complex samples using advanced chromatographic and mass spectrometric techniques. *Environ Chem Lett* 18, 1753–1760.

- Leefmann, T., Frickenhaus, S., and Koch, B.P. (2019). UltraMassExplorer - a browser-based application for the evaluation of high-resolution mass spectrometric data. *Rapid Communications in Mass Spectrometry* 33, 193–202.
- Leenheer, J.A. (1981). Comprehensive approach to preparative isolation and fractionation of dissolved organic carbon from natural waters and wastewaters. *Environ. Sci. Technol.* 15, 578–587.
- Levy, A.J., Oranzi, N.R., Ahmadireskety, A., Kemperman, R.H.J., Wei, M.S., and Yost, R.A. (2019). Recent progress in metabolomics using ion mobility-mass spectrometry. *TrAC Trends in Analytical Chemistry* 116, 274–281.
- Li, D., Liang, W., Feng, X., Ruan, T., and Jiang, G. (2021a). Recent advances in data-mining techniques for measuring transformation products by high-resolution mass spectrometry. *TrAC Trends in Analytical Chemistry* 116409.
- Li, H., Qin, Q., Shi, X., He, J., and Xu, G. (2018). Modified metabolite mapping by liquid chromatography–high resolution mass spectrometry using full scan/all ion fragmentation/neutral loss acquisition. *Journal of Chromatography A*.
- Li, J., Aziz, M.T., Granger, C.O., and Richardson, S.D. (2021b). Are Disinfection Byproducts (DBPs) Formed in My Cup of Tea? Regulated, Priority, and Unknown DBPs. *Environ. Sci. Technol.*
- Lian, L., Yan, S., Zhou, H., and Song, W. (2020). Overview of the Phototransformation of Wastewater Effluents by High-Resolution Mass Spectrometry. *Environ. Sci. Technol.*
- Lim, S., Shi, J.L., von Gunten, U., and McCurry, D.L. (2022). Ozonation of Organic Compounds in Water and Wastewater: A Critical Review. *Water Research* 118053.
- Linstrom, P., and Mallard, W.. (2020). NIST webbook of chemistry, standard reference database NIST (National Institute of Standards and Technology).
- Lommen, A. (2014). Ultrafast PubChem Searching Combined with Improved Filtering Rules for Elemental Composition Analysis. *Anal. Chem.* 86, 5463–5469.
- Loos, M. (2016). nontarget: Detecting Isotope, Adduct and Homologue Relations in LC-MS Data.
- Loos, M., Uwe Schmitt, and dutchjes (2018). Martin Loos, 2019, enviMass version 3.5 LC-HRMS trend detection workflow - R package (Zenodo).
- Lopez, B., Ollivier, P., Togola, A., Baran, N., and Ghestem, J.-P. (2015). Screening of French groundwater for regulated and emerging contaminants. *Science of The Total Environment* 562–573.
- Lübeck, J.S., Alexandrino, G.L., and Christensen, J.H. (2020). GC×GC–HRMS nontarget fingerprinting of organic micropollutants in urban freshwater sediments. *Environmental Sciences Europe* 32, 78.

- Luo, Y., Guo, W., Ngo, H.H., Nghiem, L.D., Hai, F.I., Zhang, J., Liang, S., and Wang, X.C. (2014). A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Science of The Total Environment* 473–474, 619–641.
- Luukkonen, T., and Pehkonen, S.O. (2017). Peracids in water treatment: A critical review. *Critical Reviews in Environmental Science and Technology* 47, 1–39.
- Luukkonen, T., Heyninck, T., Rämö, J., and Lassi, U. (2015). Comparison of organic peracids in wastewater treatment: Disinfection, oxidation and corrosion. *Water Research* 85, 275–285.
- Maffettone, R., Manoli, K., Santoro, D., Passalacqua, K.D., Wobus, C.E., and Sarathy, S. (2020). Performic Acid Disinfection of Municipal Secondary Effluent Wastewater: Inactivation of Murine Norovirus, Fecal Coliforms, and Enterococci. *Environ. Sci. Technol.* 54, 12761–12770.
- Magdeburg, A., Stalter, D., Schlüsener, M., Ternes, T., and Oehlmann, J. (2014). Evaluating the efficiency of advanced wastewater treatment: Target analysis of organic contaminants and (geno-)toxicity assessment tell a different story. *Water Research* 50, 35–47.
- Mailler, R., Gasperi, J., Patureau, D., Vulliet, E., Delgenes, N., Danel, A., Deshayes, S., Eudes, V., Guerin, S., Moilleron, R., et al. (2017). Fate of emerging and priority micropollutants during the sewage sludge treatment: Case study of Paris conurbation. Part 1: Contamination of the different types of sewage sludge. *Waste Management* 59, 379–393.
- Maizel, A.C., and Remucal, C.K. (2017). The effect of advanced secondary municipal wastewater treatment on the molecular composition of dissolved organic matter. *Water Research* 122, 42–52.
- Masqué, N., Marcé, R., and Borrull, F. (1998). New polymeric and other types of sorbents for solid-phase extraction of polar organic micropollutants from environmental water. *TrAC Trends in Analytical Chemistry* 17, 384–394.
- Mathon, B., Coquery, M., Liu, Z., Penru, Y., Guillon, A., Esperanza, M., Miège, C., and Choubert, J.-M. (2021). Ozonation of 47 organic micropollutants in secondary treated municipal effluents: Direct and indirect kinetic reaction rates and modelling. *Chemosphere* 262, 127969.
- Mechelke, J., Longrée, P., Singer, H., and Hollender, J. (2019). Vacuum-assisted evaporative concentration combined with LC-HRMS/MS for ultra-trace-level screening of organic micropollutants in environmental water samples. *Anal Bioanal Chem* 411, 2555–2567.
- Melmer, M., Stangler, T., Schiefermeier, M., Brunner, W., Toll, H., Rupprechter, A., Lindner, W., and Premstaller, A. (2010). HILIC analysis of fluorescence-labeled N-glycans from recombinant biopharmaceuticals. *Anal Bioanal Chem* 398, 905–914.
- Merel, S., and Zwiener, C. (2016). Accurate Mass Screening and Data Evaluation Approaches for Ozonation By-Products in Wastewater Treatment Plant Effluents. In *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening – Strategies and Workflows Volume 2*, (American Chemical Society), pp. 3–27.

- Merel, S., Anumol, T., Park, M., and Snyder, S.A. (2015). Application of surrogates, indicators, and high-resolution mass spectrometry to evaluate the efficacy of UV processes for attenuation of emerging contaminants in water. *Journal of Hazardous Materials* 282, 75–85.
- Merel, S., Lege, S., Yanez Heras, J.E., and Zwiener, C. (2017). Assessment of N-Oxide Formation during Wastewater Ozonation. *Environ. Sci. Technol.* 51, 410–417.
- Meringer, M., Reinker, S., Zhang, J., and Muller, A. (2011). MS/MS Data Improves Automated Determination of Molecular Formulas by Mass Spectrometry. *MATCH Communications in Mathematical and in Computer Chemistry* 65, 259–290.
- Mijangos, L., Krauss, M., de Miguel, L., Ziarrusta, H., Olivares, M., Zuloaga, O., Izagirre, U., Schulze, T., Brack, W., Prieto, A., et al. (2020). Application of the Sea Urchin Embryo Test in Toxicity Evaluation and Effect-Directed Analysis of Wastewater Treatment Plant Effluents. *Environ. Sci. Technol.* 54, 8890–8899.
- Milman, B.L., and Zhurkovich, I.K. (2017). The chemical space for non-target analysis. *TrAC Trends in Analytical Chemistry* 97, 179–187.
- Minor, E.C., Swenson, M.M., Mattson, B.M., and Oyler, A.R. (2014). Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environ. Sci.: Processes Impacts* 16, 2064–2079.
- Mitchell, V.G., and Diaper, C. (2006). Simulating the urban water and contaminant cycle. *Environmental Modelling & Software* 21, 129–134.
- Miura, D., Tsuji, Y., Takahashi, K., Wariishi, H., and Saito, K. (2010). A Strategy for the Determination of the Elemental Composition by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Based on Isotopic Peak Ratios. *Anal. Chem.* 82, 5887–5891.
- Montes, R., Aguirre, J., Vidal, X., Rodil, R., Cela, R., and Quintana, J.B. (2017). Screening for Polar Chemicals in Water by Trifunctional Mixed-Mode Liquid Chromatography–High Resolution Mass Spectrometry. *Environ. Sci. Technol.* 51, 6250–6259.
- Moschet, C., Piazzoli, A., Singer, H., and Hollender, J. (2013). Alleviating the Reference Standard Dilemma Using a Systematic Exact Mass Suspect Screening Approach with Liquid Chromatography–High Resolution Mass Spectrometry. *Anal. Chem.* 85, 10312–10320.
- Moschet, C., Lew, B.M., Hasenbein, S., Anumol, T., and Young, T.M. (2016). LC- and GC-QTOF-MS as Complementary Tools for a Comprehensive Micropollutant Analysis in Aquatic Systems. *Environ. Sci. Technol.*
- Müller, A., Schulz, W., Ruck, W.K.L., and Weber, W.H. (2011). A new approach to data evaluation in the non-target screening of organic trace substances in water analysis. *Chemosphere* 85, 1211–1219.
- Myers, A.L., Jobst, K.J., Mabury, S.A., and Reiner, E.J. (2014). Using mass defect plots as a discovery tool to identify novel fluoropolymer thermal decomposition products. *Journal of Mass Spectrometry* 49, 291–296.

- Nas, B., Argun, M.E., Dolu, T., Ateş, H., Yel, E., Koyuncu, S., Dinç, S., and Kara, M. (2020). Occurrence, loadings and removal of EU-priority polycyclic aromatic hydrocarbons (PAHs) in wastewater and sludge by advanced biological treatment, stabilization pond and constructed wetland. *Journal of Environmental Management* 268, 110580.
- Nefau, T., Karolak, S., Castillo, L., Boireau, V., and Levi, Y. (2013). Presence of illicit drugs and metabolites in influents and effluents of 25 sewage water treatment plants and map of drug consumption in France. *Sci. Total Environ.* 461–462, 712–722.
- Negreira, N., Regueiro, J., López de Alda, M., and Barceló, D. (2015). Degradation of the anticancer drug erlotinib during water chlorination: Non-targeted approach for the identification of transformation products. *Water Research* 85, 103–113.
- Neuwald, I., Muschket, M., Zahn, D., Berger, U., Seiwert, B., Meier, T., Kuckelkorn, J., Strobel, C., Knepper, T.P., and Reemtsma, T. (2021). Filling the knowledge gap: A suspect screening study for 1310 potentially persistent and mobile chemicals with SFC- and HILIC-HRMS in two German river systems. *Water Research* 204, 117645.
- Nihemaiti, M., Permala, R.R., and Croué, J.-P. (2020). Reactivity of unactivated peroxymonosulfate with nitrogenous compounds. *Water Research* 169, 115221.
- Nürenberg, G., Schulz, M., Kunkel, U., and Ternes, T.A. (2015). Development and validation of a generic nontarget method based on liquid chromatography – high resolution mass spectrometry analysis for the evaluation of different wastewater treatment options. *Journal of Chromatography A* 1426, 77–90.
- Nürenberg, G., Kunkel, U., Wick, A., Falås, P., Joss, A., and Ternes, T.A. (2019). Nontarget analysis: A new tool for the evaluation of wastewater processes. *Water Research* 163, 114842.
- Nurmi, J., Pellinen, J., and Rantalainen, A.-L. (2012). Critical evaluation of screening techniques for emerging environmental contaminants based on accurate mass measurements with time-of-flight mass spectrometry. *Journal of Mass Spectrometry* 47, 303–312.
- Ocampo-Duque, W., Sierra, J., Ferré-Huguet, N., Schuhmacher, M., and Domingo, J.L. (2008). Estimating the environmental impact of micro-pollutants in the low Ebro River (Spain): An approach based on screening toxicity with *Vibrio fischeri*. *Chemosphere* 72, 715–721.
- Ouyang, X., Leonards, P., Legler, J., van der Oost, R., de Boer, J., and Lamoree, M. (2015). Comprehensive two-dimensional liquid chromatography coupled to high resolution time of flight mass spectrometry for chemical characterization of sewage treatment plant effluents. *Journal of Chromatography A* 1380, 139–145.
- Paglia, G., Williams, J.P., Menikarachchi, L., Thompson, J.W., Tyldesley-Worster, R., Halldórsson, S., Rolfsson, O., Moseley, A., Grant, D., Langridge, J., et al. (2014). Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Anal. Chem.* 86, 3985–3993.

- Paijens, C., Bressy, A., Frère, B., Tedoldi, D., Mailler, R., Rocher, V., Neveu, P., and Moilleron, R. (2021). Urban pathways of biocides towards surface waters during dry and wet weathers: Assessment at the Paris conurbation scale. *Journal of Hazardous Materials* 402, 123765.
- Pan, Y., Li, H., Zhang, X., and Li, A. (2016). Characterization of natural organic matter in drinking water: Sample preparation and analytical approaches. *Trends in Environmental Analytical Chemistry* 23–30.
- Paredes, L., Omil, F., Lema, J.M., and Carballa, M. (2018). What happens with organic micropollutants during UV disinfection in WWTPs? A global perspective from laboratory to full-scale. *Journal of Hazardous Materials* 342, 670–678.
- Park, M., and Snyder, S.A. (2020). Statistical profiling for identifying transformation products in an engineered treatment process. *Chemosphere* 251, 126401.
- Parlanti, E. (2000). Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Organic Geochemistry* 17.
- Patiny, L., and Borel, A. (2013). ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. *J. Chem. Inf. Model.* 53, 1223–1228.
- Pellegrin, V. (1983). Molecular formulas of organic compounds: the nitrogen rule and degree of unsaturation. *J. Chem. Educ.* 60, 626.
- Pence, H.E., and Williams, A. (2010). ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* 87, 1123–1124.
- Pérez, M.R., Rossi, A.S., Bacchetta, C., Elorriaga, Y., Carriquiriborde, P., and Cazenave, J. (2018). In situ evaluation of the toxicological impact of a wastewater effluent on the fish *Prochilodus lineatus*: biochemical and histological assessment. *Ecological Indicators* 84, 345–353.
- Periat, A., Fekete, S., Cusumano, A., Veuthey, J.-L., Beck, A., Lauber, M., and Guillarme, D. (2016). Potential of hydrophilic interaction chromatography for the analytical characterization of protein biopharmaceuticals. *Journal of Chromatography A* 1448, 81–92.
- Peter, K.T., Wu, C., Tian, Z., and Kolodziej, E.P. (2019). Application of Nontarget High Resolution Mass Spectrometry Data to Quantitative Source Apportionment. *Environ. Sci. Technol.* 53, 12257–12268.
- Peters, S., Janssen, H.-G., and Vivó-Truyols, G. (2010). Trend analysis of time-series data: A novel method for untargeted metabolite discovery. *Analytica Chimica Acta* 663, 98–104.
- Plassmann, M.M., Tengstrand, E., Åberg, K.M., and Benskin, J.P. (2016). Non-target time trend screening: a data reduction strategy for detecting emerging contaminants in biological samples. *Anal Bioanal Chem* 408, 4203–4208.
- Plassmann, M.M., Fischer, S., and Benskin, J.P. (2018). Nontarget Time Trend Screening in Human Blood. *Environ. Sci. Technol. Lett.* 5, 335–340.

- Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* *11*, 395.
- Pluskal, T., Uehara, T., and Yanagida, M. (2012). Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching. *Anal. Chem.* *84*, 4396–4403.
- Poyatos, J.M., Muñio, M.M., Almecija, M.C., Torres, J.C., Hontoria, E., and Osorio, F. (2009). Advanced Oxidation Processes for Wastewater Treatment: State of the Art. *Water Air Soil Pollut* *205*, 187.
- Purschke, K., Zoell, C., Leonhardt, J., Weber, M., and Schmidt, T.C. (2020). Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. *Science of The Total Environment* *706*, 135835.
- Ragazzo, P., Chiucchini, N., Piccolo, V., and Ostoich, M. (2013). A new disinfection system for wastewater treatment: performic acid full-scale trial evaluations. *Water Science and Technology* *67*, 2476–2487.
- Ragazzo, P., Feretti, D., Monarca, S., Dominici, L., Ceretti, E., Viola, G., Piccolo, V., Chiucchini, N., and Villarini, M. (2017). Evaluation of cytotoxicity, genotoxicity, and apoptosis of wastewater before and after disinfection with performic acid. *Water Research* *116*, 44–52.
- Ragazzo, P., Chiucchini, N., Piccolo, V., Spadolini, M., Carrer, S., Zanon, F., and Gehr, R. (2020). Wastewater disinfection: long-term laboratory and full-scale studies on performic acid in comparison with peracetic acid and chlorine. *Water Research* *184*, 116169.
- Reemtsma, T., Fiehn, O., and Jekel, M. (1999). A modified method for the analysis of organics in industrial wastewater as directed by their toxicity to *Vibrio fischeri*. *Fresenius J Anal Chem* *363*, 771–776.
- Richard, J., Boergers, A., vom Eyser, C., Bester, K., and Tuerk, J. (2014). Toxicity of the micropollutants Bisphenol A, Ciprofloxacin, Metoprolol and Sulfamethoxazole in water samples before and after the oxidative treatment. *International Journal of Hygiene and Environmental Health* *217*, 506–514.
- Richardson, S.D., and Plewa, M.J. (2020). To regulate or not to regulate? What to do with more toxic disinfection by-products? *Journal of Environmental Chemical Engineering* *8*, 103939.
- Richardson, S.D., and Ternes, T.A. (2005). Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* *77*, 3807–3838.
- Richardson, S.D., Thruston, A.D., Caughran, T.V., Chen, P.H., Collette, T.W., Schenck, K.M., Lykins, B.W., Rav-Acha, C., and Glezer, V. (2000). Identification of New Drinking Water Disinfection by - Products from Ozone, Chlorine Dioxide, Chloramine, and Chlorine. *Water, Air, & Soil Pollution* *123*, 95–102.

- Richardson, S.D., Plewa, M.J., Wagner, E.D., Schoeny, R., and DeMarini, D.M. (2007). Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. *Mutation Research/Reviews in Mutation Research* 636, 178–242.
- Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M.C., Michael, I., and Fatta-Kassinos, D. (2013). Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. *Science of The Total Environment* 447, 345–360.
- Rocher, V., Azimi, S., Mailler, R., Rechdaoui-Guérin, S., Mèche, P., Pichon, S., Goffin, A., Bernier, J., Roy, J., Varrault, G., et al. (2021). Effectiveness of Disinfecting Wastewater Treatment Plant Discharges: Case of chemical disinfection using performic acid (IWA Publishing).
- Rodriguez-Aller, M., Gurny, R., Veuthey, J.-L., and Guillarme, D. (2013). Coupling ultra high-pressure liquid chromatography with mass spectrometry: Constraints and possible applications. *Journal of Chromatography A* 1292, 2–18.
- Rosal, R., Rodríguez, A., Perdigón-Melón, J.A., Petre, A., García-Calvo, E., Gómez, M.J., Agüera, A., and Fernández-Alba, A.R. (2009). Degradation of caffeine and identification of the transformation products generated by ozonation. *Chemosphere* 74, 825–831.
- Roussis, S.G., and Proulx, R. (2003). Reduction of Chemical Formulas from the Isotopic Peak Distributions of High-Resolution Mass Spectra. *Anal. Chem.* 75, 1470–1482.
- Rúa-Gómez, P.C., Guedez, A.A., Ania, C.O., and Püttmann, W. (2012). Upgrading of Wastewater Treatment Plants Through the Use of Unconventional Treatment Technologies: Removal of Lidocaine, Tramadol, Venlafaxine and Their Metabolites. *Water* 4, 650–669.
- Ruff, M., Mueller, M.S., Loos, M., and Singer, H.P. (2015). Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass spectrometry e Identification of unknown sources and compounds. *Water Research* 145–154.
- Sabourin, L., Duenk, P., Bonte-Gelok, S., Payne, M., Lapen, D.R., and Topp, E. (2012). Uptake of pharmaceuticals, hormones and parabens into vegetables grown in soil fertilized with municipal biosolids. *Science of The Total Environment* 431, 233–236.
- Samanipour, S., Reid, M.J., and Thomas, K.V. (2017). Statistical Variable Selection: An Alternative Prioritization Strategy during the Nontarget Analysis of LC-HR-MS Data. *Anal. Chem.* 89, 5585–5591.
- Samanipour, S., Baz-Lomba, J.A., Reid, M.J., Ciceri, E., Rowland, S., Nilsson, P., and Thomas, K.V. (2018). Assessing sample extraction efficiencies for the analysis of complex unresolved mixtures of organic pollutants: A comprehensive non-target approach. *Analytica Chimica Acta* 1025, 92–98.

- Sanchís, J., Jaén-Gil, A., Gago-Ferrero, P., Munthali, E., and Farré, M.J. (2020). Characterization of organic matter by HRMS in surface waters: Effects of chlorination on molecular fingerprints and correlation with DBP formation potential. *Water Research* 115743.
- Santacesaria, E., Russo, V., Tesser, R., Turco, R., and Di Serio, M. (2017). Kinetics of Performic Acid Synthesis and Decomposition. *Ind. Eng. Chem. Res.* 56, 12940–12952.
- Schmitt-Jansen, M., Bartels, P., Adler, N., and Altenburger, R. (2007). Phytotoxicity assessment of diclofenac and its phototransformation products. *Anal Bioanal Chem* 387, 1389–1396.
- Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., and Hollender, J. (2015). Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* 87, 12121–12129.
- Schollée, J.E., Schymanski, E.L., and Hollender, J. (2016). Statistical Approaches for LC-HRMS Data To Characterize, Prioritize, and Identify Transformation Products from Water Treatment Processes. In *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening – Strategies and Workflows Volume 1*, (American Chemical Society), pp. 45–65.
- Schollée, J.E., Bourgin, M., von Gunten, U., McArdell, C.S., and Hollender, J. (2018). Non-target screening to trace ozonation transformation products in a wastewater treatment train including different post-treatments. *Water Research* 142, 267–278.
- Schollée, J.E., Hollender, J., and McArdell, C.S. (2021). Characterization of advanced wastewater treatment with ozone and activated carbon using LC-HRMS based non-target screening with automated trend assignment. *Water Research* 117209.
- Schulz, W., Lucke, T., Balsaa, P., Hinnenkamp, V., Brüggem, S., Dünnbier, U., Liebmann, D., Fink, A., Götz, S., Geiss, S., et al. (2019). Non-Target Screening in Water Analysis- Guideline for the application of LC-ESI-HRMS for screening.
- Schulze, B., Jeon, Y., Kaserzon, S., Heffernan, A.L., Dewapriya, P., O'Brien, J., Gomez Ramos, M.J., Ghorbani Gorji, S., Mueller, J.F., Thomas, K.V., et al. (2020). An assessment of quality assurance/quality control efforts in high resolution mass spectrometry non-target workflows for analysis of environmental samples. *TrAC Trends in Analytical Chemistry* 133, 116063.
- Schulze, B., van Herwerden, D., Allan, I., Bijlsma, L., Etxebarria, N., Hansen, M., Merel, S., Vrana, B., Aalizadeh, R., Bajema, B., et al. (2021). Inter-laboratory mass spectrometry dataset based on passive sampling of drinking water for non-target analysis. *Sci Data* 8, 1–10.
- Schum, S. (2019). Molecular characterization of free tropospheric organic aerosol and the development of computational tools for molecular formula assignment. *Dissertations, Master's Theses and Master's Reports*.
- Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., and Hollender, J. (2014a). Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* 48, 2097–2098.

- Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., and Hollender, J. (2014b). Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. *Environmental Science & Technology* 48, 1811–1818.
- Schymanski, E.L., Singer, H.P., Slobodnik, J., Ipolyi, I.M., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., et al. (2015). Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Analytical and Bioanalytical Chemistry* 407, 6237–6255.
- Sgroi, M., Roccaro, P., Korshin, G.V., Greco, V., Sciuto, S., Anumol, T., Snyder, S.A., and Vagliasindi, F.G.A. (2017). Use of fluorescence EEM to monitor the removal of emerging contaminants in full scale wastewater treatment plants. *Journal of Hazardous Materials* 323, 367–376.
- Siegrist, H., and Joss, A. (2012). Review on the fate of organic micropollutants in wastewater treatment and water reuse with membranes. *Water Science and Technology* 66, 1369–1376.
- SierraAnalytics Composer. <http://massspec.com/composer/> , consulté le 04/06/2020
- Singer, H.P., Wössner, A.E., McArdell, C.S., and Fenner, K. (2016). Rapid Screening for Exposure to “Non-Target” Pharmaceuticals from Wastewater Effluents by Combining HRMS-Based Suspect Screening and Exposure Modeling. *Environ. Sci. Technol.* 50, 6698–6707.
- Sleno (2012). The use of mass defect in modern mass spectrometry - Sleno - 2012 - *Journal of Mass Spectrometry - Wiley Online Library*. *Journal of Mass Spectrometry* 47, 226–236.
- Smital, T., Terzić, S., Lončar, J., Senta, I., Žaja, R., Popović, M., Mikac, I., Tollefsen, K.-E., Thomas, K.V., and Ahel, M. (2013). Prioritisation of organic contaminants in a river basin using chemical analyses and bioassays. *Environ Sci Pollut Res* 20, 1384–1395.
- Smith, R.M. (2003). Before the injection—modern methods of sample preparation for separation techniques. *Journal of Chromatography A* 1000, 3–27.
- Smith, C.A., Want, E.J., O’Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry* 78, 779–787.
- Snoeyink, V.L., and Chen, A.S.C. (1985). Removal of organic micropollutants by coagulation and adsorption. *Science of The Total Environment* 47, 155–167.
- Sobus, J.R., Grossman, J.N., Chao, A., Singh, R., Williams, A.J., Grulke, C.M., Richard, A.M., Newton, S.R., McEachran, A.D., and Ulrich, E.M. (2019). Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance. *Anal Bioanal Chem.*
- Sonavane, M., Schollée, J.E., Hidasí, A.O., Creusot, N., Brion, F., Suter, M.J.-F., Hollender, J., and Ait-Aïssa, S. (2018). An integrative approach combining passive sampling, bioassays, and effect-directed analysis to assess the impact of wastewater effluent. *Environmental Toxicology and Chemistry* 37, 2079–2088.

von Sonntag, C., and von Gunten, U. (2012). *Chemistry of Ozone in Water and Wastewater Treatment: From Basic Principles to Applications*.

Soulier, C., Boiteux, V., Candido, P., Caupos, E., Chachignon, M., Couturier, G., Dauchy, X., Devier, M.-H., Esperanza, M., Fildier, A., et al. (2021). La spectrométrie de masse haute résolution pour la recherche de micropolluants organiques dans l'environnement. *Techniques Sciences Méthodes* 43.

Staelin, J., and Hoigne, J. (1985). Decomposition of ozone in water in the presence of organic solutes acting as promoters and inhibitors of radical chain reactions. *Environ. Sci. Technol.* 19, 1206–1213.

Sturm, M., Bertsch, A., Gröpl, C., Hildebrandt, A., Hussong, R., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., Zerck, A., Reinert, K., et al. (2008). OpenMS – An open-source software framework for mass spectrometry. *BMC Bioinformatics* 9, 163.

Sun, Q., Lv, M., Hu, A., Yang, X., and Yu, C.-P. (2014). Seasonal variation in the occurrence and removal of pharmaceuticals and personal care products in a wastewater treatment plant in Xiamen, China. *Journal of Hazardous Materials* 277, 69–75.

Sun, X., Zhao, X., Du, W., and Liu, D. (2011). Kinetics of Formic Acid-autocatalyzed Preparation of Performic Acid in Aqueous Phase. *Chinese Journal of Chemical Engineering* 19, 964–971.

Tang, J.I.S., Kawahara, F.K., and Yen, T.F. (1981). THE USE OF MACRORETICULAR RESIN FOR SEPARATION OF COAL CONVERSION PROCESS WASTEWATER. *American Chemical Society Symposium on the Chemistry of Engine Combustion Deposits* 6.

Tapp, H.S., and Kemsley, E.K. (2009). Notes on the practical utility of OPLS. *TrAC Trends in Analytical Chemistry* 28, 1322–1327.

Tautenhahn, R., Böttcher, C., and Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* 9, 504.

Tautenhahn, R., Patti, G.J., Rinehart, D., and Siuzdak, G. (2012). XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* 84, 5035–5039.

Tejada-Casado, C., Hernández-Mesa, M., Monteau, F., Lara, F.J., Olmo-Iruela, M. del, García-Campaña, A.M., Le Bizec, B., and Dervilly-Pinel, G. (2018). Collision cross section (CCS) as a complementary parameter to characterize human and veterinary drugs. *Analytica Chimica Acta* 1043, 52–63.

Terechovs, A.K.E., Ansari, A.J., McDonald, J.A., Khan, S.J., Hai, F.I., Knott, N.A., Zhou, J., and Nghiem, L.D. (2019). Occurrence and bioconcentration of micropollutants in Silver Perch (*Bidyanus bidyanus*) in a reclaimed water reservoir. *Science of The Total Environment* 650, 585–593.

ThermoFisher (2018). *Powering Confident Insights - Explore your Small-Molecule Data to its Core*.

- Thevenot, M., Dousset, S., Hertkorn, N., Schmitt-Kopplin, P., and Andreux, F. (2009). Interactions of diuron with dissolved organic matter from organic amendments. *Science of The Total Environment* 407, 4297–4302.
- Thurman, E.M., and Ferrer, I. (2010). The isotopic mass defect: a tool for limiting molecular formulas by accurate mass. *Anal Bioanal Chem* 397, 2807–2816.
- Titato, G.M., and Lanças, P.D.F.M. (2005). Comparison Between Different Extraction (LLE and SPE) and Determination (HPLC and Capillary-LC) Techniques in the Analysis of Selected PAHs in Water Samples. *Journal of Liquid Chromatography & Related Technologies* 28, 3045–3056.
- Tolić, N., Liu, Y., Liyu, A., Shen, Y., Tfaily, M.M., Kujawinski, E.B., Longnecker, K., Kuo, L.-J., Robinson, E.W., Paša-Tolić, L., et al. (2017). Formularity: Software for Automated Formula Assignment of Natural and Other Organic Matter from Ultrahigh-Resolution Mass Spectra. *Anal. Chem.* 89, 12659–12665.
- Treutler, H., and Neumann, S. (2016). Prediction, Detection, and Validation of Isotope Clusters in Mass Spectrometry Data. *Metabolites* 6.
- Tylová, T., Flieger, M., and Olšovská, J. (2013). Determination of antibiotics in influents and effluents of wastewater-treatment-plants in the Czech Republic – development and application of the SPE and a UHPLC-ToFMS method. *Anal. Methods* 5, 2110–2118.
- Ulrich, E.M., Sobus, J.R., Grulke, C.M., Richard, A.M., Newton, S.R., Strynar, M.J., Mansouri, K., and Williams, A.J. (2018). EPA’s non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. *Anal Bioanal Chem.*
- Vajda, A.M., Barber, L.B., Gray, J.L., Lopez, E.M., Woodling, J.D., and Norris, D.O. (2008). Reproductive Disruption in Fish Downstream from an Estrogenic Wastewater Effluent. *Environ. Sci. Technol.* 42, 3407–3414.
- Välitalo, P., Kruglova, A., Mikola, A., and Vahala, R. (2017). Toxicological impacts of antibiotics on aquatic micro-organisms: A mini-review. *International Journal of Hygiene and Environmental Health* 220, 558–569.
- Vasquez, M.I., Lambrianides, A., Schneider, M., Kümmerer, K., and Fatta-Kassinos, D. (2014). Environmental side effects of pharmaceutical cocktails: What we know and what we should know. *Journal of Hazardous Materials* 279, 169–189.
- Veenaas, C., Bignert, A., Liljelind, P., and Haglund, P. (2018). Nontarget Screening and Time-Trend Analysis of Sewage Sludge Contaminants via Two-Dimensional Gas Chromatography–High Resolution Mass Spectrometry. *Environ. Sci. Technol.* 52, 7813–7822.
- Vergeynst, L., Van Langenhove, H., Joos, P., and Demeestere, K. (2014). Suspect screening and target quantification of multi-class pharmaceuticals in surface water based on large-volume injection liquid chromatography and time-of-flight mass spectrometry. *Anal Bioanal Chem* 406, 2533–2547.

- Verkh, Y., Rozman, M., and Petrovic, M. (2018). A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment. *Chemosphere* *200*, 397–404.
- Vinaixa, M., Schymanski, E.L., Neumann, S., Navarro, M., Salek, R.M., and Yanes, O. (2016). Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *TrAC Trends in Analytical Chemistry* *78*, 23–35.
- Vodyanitskii, Y.N., and Yakovlev, A.S. (2016). Contamination of soils and groundwater with new organic micropollutants: A review. *Eurasian Soil Sc.* *49*, 560–569.
- Wagner, E.D., and Plewa, M.J. (2017). CHO cell cytotoxicity and genotoxicity analyses of disinfection by-products: An updated review. *Journal of Environmental Sciences* *58*, 64–76.
- Wang, Z., Bush, R.T., Sullivan, L.A., Chen, C., and Liu, J. (2014). Selective Oxidation of Arsenite by Peroxymonosulfate with High Utilization Efficiency of Oxidant. *Environ. Sci. Technol.* *48*, 3978–3985.
- Weller, M.G. (2012). A Unifying Review of Bioassay-Guided Fractionation, Effect-Directed Analysis and Related Techniques. *Sensors* *12*, 9181–9209.
- Wernisch, S., and Pennathur, S. (2016). Evaluation of coverage, retention patterns, and selectivity of seven liquid chromatographic methods for metabolomics. *Anal Bioanal Chem* *408*, 6079–6091.
- Wickramasekara, S., Hernández-Ruiz, S., Abrell, L., Arnold, R., and Chorover, J. (2012). Natural dissolved organic matter affects electrospray ionization during analysis of emerging contaminants by mass spectrometry. *Analytica Chimica Acta* *717*, 77–84.
- Wu, Z., Rodgers, R.P., and Marshall, A.G. (2004). Two- and Three-Dimensional van Krevelen Diagrams: A Graphical Analysis Complementary to the Kendrick Mass Plot for Sorting Elemental Compositions of Complex Organic Mixtures Based on Ultrahigh-Resolution Broadband Fourier Transform Ion Cyclotron Resonance Mass Measurements. *Analytical Chemistry* *76*, 2511–2516.
- Xia, Y.-Q., Lau, J., Olah, T., and Jemal, M. (2011). Targeted quantitative bioanalysis in plasma using liquid chromatography/high-resolution accurate mass spectrometry: an evaluation of global selectivity as a function of mass resolving power and extraction window, with comparison of centroid and profile modes. *Rapid Communications in Mass Spectrometry* *25*, 2863–2878.
- Yoshida, T. (2004). Peptide separation by Hydrophilic-Interaction Chromatography: a review. *Journal of Biochemical and Biophysical Methods* *60*, 265–280.
- Zaman, S.B., Hussain, M.A., Nye, R., Mehta, V., Mamun, K.T., and Hossain, N. (2017). A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* *9*, e1403.

- Zhang, B., Wang, X., Fang, Z., Wang, S., Shan, C., Wei, S., and Pan, B. (2021). Unravelling molecular transformation of dissolved effluent organic matter in UV/H₂O₂, UV/persulfate, and UV/chlorine processes based on FT-ICR-MS analysis. *Water Research* 117158.
- Zhang, H., Zhang, Y., Shi, Q., Zheng, H., and Yang, M. (2014). Characterization of Unknown Brominated Disinfection Byproducts during Chlorination Using Ultrahigh Resolution Mass Spectrometry. *Environ. Sci. Technol.* 48, 3112–3119.
- Zhang, M., Zhang, Z., Chen, C., Lu, H., and Liang, Y. (2016). Parallel formula generator based on branch-and-bound algorithm for elucidating high resolution mass spectra. *Chemometrics and Intelligent Laboratory Systems* 153, 106–109.
- Zhang, S., You, J., Sun, Z., Song, C., Ning, S., Zhao, C., and Suo, Y. (2012). A sensitive method for extraction and determination of endocrine-disrupting compounds from wastewater using 10-ethyl-acridone-2-sulfonyl chloride as pre-column labeling reagent by high-performance liquid chromatography with fluorescence detection. *Microchemical Journal* 103, 90–96.
- Zhang, T., Lu, J., Ma, J., and Qiang, Z. (2008). Fluorescence spectroscopic characterization of DOM fractions isolated from a filtered river water after ozonation and catalytic ozonation. *Chemosphere* 71, 911–921.
- Ziegler, G., Gonsior, M., Fisher, D.J., Schmitt-Kopplin, P., and Tamburri, M.N. (2019). Formation of Brominated Organic Compounds and Molecular Transformations in Dissolved Organic Matter (DOM) after Ballast Water Treatment with Sodium Dichloroisocyanurate Dihydrate (DICD). *Environ. Sci. Technol.* 53, 8006–8016.
- Zonja, B., Delgado, A., Pérez, S., and Barceló, D. (2015). LC-HRMS Suspect Screening for Detection-Based Prioritization of Iodinated Contrast Media Photodegradates in Surface Waters. *Environ. Sci. Technol.* 49, 3464–3472.
- Zwart, N., Jonker, W., Broek, R. ten, de Boer, J., Somsen, G., Kool, J., Hamers, T., Houtman, C.J., and Lamoree, M.H. (2020). Identification of mutagenic and endocrine disrupting compounds in surface water and wastewater treatment plant effluents using high-resolution effect-directed analysis. *Water Research* 168, 115204.

Annexes

Liste des annexes

Annexe 1 : Etalons analytiques entrés dans la librairie interne (identification possible au niveau 1).....	286
Annexe 2 : Réplicats après alignement de différentes colonnes chromatographiques	293
Annexe 3 : Comparaison des méthodes analytiques	294
Annexe 4 : Article Molecules, MDPI	296
Annexe 5 : Informations complémentaires sur l'étude de désinfection par le PFA	323
Annexe 6 : Comparaison des méthodes de préparation pour l'analyse de TPs ESI+.....	328
Annexe 7 : Comparaison des méthodes de préparation pour l'analyse de TPs ESI-	329
Annexe 8 : Proportion de liens de biotransformation selon la méthode de préparation	330

Annexe 1 : Etalons analytiques entrés dans la librairie interne (identification possible au niveau 1)

Nom	Formule moléculaire	Masse monoisotopique	CAS	TR mesuré (min)	CCS* mesuré (Å ²)	CCS* (Å ²) en librairie**	log Kow**	pKa**	Surface topologique polaire (Å ²)**
10-Hydroxycarbamazepine	C15H14N2O2	254,1055	29331-92-8	6,8	154,82		1,5		66,6
2-(methylthio)benzothiazole	C8H7NS2	181,0020	615-22-5	11,7	129,73	131,32	3,2		66,4
2-aminobenzothiazole	C7H6N2S	150,0252	136-95-8	3,4	126,32	128,60	1,9		67,2
2-hydroxybenzothiazole	C7H5NOS	151,0092	934-34-9	6,9	125,78	123,75	1,8		54,4
2-methylbenzothiazole	C8H7NS	149,0299	120-75-2	8,8	124,97		3,0		41,1
5-chlorobenzotriazole	C6H4ClN3	153,0094	94-97-3	6,7	126,66	131,88	1,7		41,6
5-methyl-1H-benzotriazole	C7H7N3	133,0640	136-85-6	5,9	125,22	124,64	1,4		41,6
Acephate	C4H10NO3PS	183,0119	30560-19-1	2,4	138,55	139,15	-0,8		80,7
Acetaminophen	C8H9NO2	151,0633	103-90-2	2,7	131,78	132,20	0,5	-4,4 ; 9,4	49,3
Acetamiprid	C10H11ClN4	222,0672	160430-64-8			150,24	0,8	0,7	52,3
Adenosine	C10H13N5O4	267,0968	58-61-7	1,2	158,91	155,70	-1,1	3,6 ; 12,4	140,0
Amisulpride	C17H27N3O4S	369,1722	71675-85-9	4,9	193,79	193,10	1,5	9,4	110,0
Amisulpride N-oxide	C17H27N3O5S	385,1671	71676-01-2	5,3	195,82		1,9		125,0
Ampicilline	C16H19N3O4S	349,1096	69-53-4	4,5	185,73	187,50	1,4	2,5 ; 7,3	138,0
Aspartame	C14H18N2O5	294,1216	22839-47-0	5,0	168,41	169,60	-2,7	3,2 ; 7,9	119,0
Atenolol	C14H22N2O3	266,1630	29122-68-7	3,0	159,04	159,50	0,2	9,6	84,6
Atenolol-d7	C14H15D7N2O3	273,2070	1202864-50-3	3,0	158,93		0,2		84,6
Atrazine	C8H14ClN5	215,0938	1912-24-9	9,7	149,41	150,74	2,6	1,6	62,7
Atrazine-2-OH	C8H15N5O	197,1277	2163-68-0	4,6	147,76	147,71	0,1		77,9
Atrazine-d5	C8H9D5ClN5	220,1252	163165-75-1	9,6	150,31		2,6	1,6	62,7
Atrazine-desethyl	C6H10ClN5	187,0625	6190-65-4	5,9	139,79	139,64	1,5		76,7

Annexes

Nom	Formule moléculaire	Masse monoisotopique	CAS	TR mesuré (min)	CCS* mesuré (Å ²)	CCS* (Å ²) en librairie**	log Kow**	pKa**	Surface topologique polaire (Å ²)**
Atrazine-desethyl-2-OH	C6H11N5O	169,0964	19988-24-0	1,7	137,25		-0,6		91,9
Atrazine-desethyl-desisopropyl	C3H4CIN5	145,0155	3397-62-4	5,8	123,10		-0,1		90,7
Atrazine-desethyl-desisopropyl-2-OH	C3H5N5O	127,0494	645-95-1	1,7	121,75		-1,8		106,0
Atrazine-desisopropyl	C5H8CIN5	173,0468	1007-28-9	4,4	133,26	132,85	1,1		76,7
Atrazine-desisopropyl-2-OH	C5H9N5O	155,0807	7313-54-4	0,9	131,15		-1,1		91,9
Avermectin B1a	C48H72O14	872,4922	65195-55-3	19,1	295,47	290,50	3,8		170,0
Benzotriazole	C6H5N3	119,0483	95-14-7	4,3	121,42		1,0	8,4	41,6
Benzotriazole-5-carboxylic acid	C7H5N3O2	163,0382	23814-12-2	3,4	131,33		0,5		78,9
BenzylParaben	C14H12O3	228,0786	94-18-8	11,8	201,78		3,6		46,5
Bifenthrin	C23H22CIF3O2	422,1260	439680-76-9				6,0		26,3
Boscalid	C18H12Cl2N2O	342,0327	188425-85-6	12,8	171,28	171,67	3,0		42,0
Butocarboxim	C7H14N2O2S	190,0776	34681-10-2	7,1	144,49		1,1		76,0
ButylParaben	C11H14O3	194,0943	94-26-8	11,7	150,51	150,14	3,6	8,5	46,5
BZK C12	C21H38CIN	339,2693	139-07-1	15,0	213,62				0,0
BZK C14	C23H42CIN	367,3006	5285-67-6	16,8	221,67		8,4		0,0
BZK C16	C25H46CIN	395,3319	122-18-9	18,4	228,12				0,0
BZK C18	C27H50CIN	423,3632	206752-43-4	19,8	235,34				1,0
Caffeine	C8H10N4O2	194,0804	58-08-2	4,0	138,21	140,90	-0,1	14,0	58,4
Caffeine-13C3	C5[13C]3H10N4O2	197,0904	78072-66-9	4,0	137,36				
Carbamazepine	C15H12N2O	236,0950	298-46-4	9,0	149,09	150,30	2,5	13,9	46,3
Carbamazepine-10,11-epoxide	C15H12N2O2	253,0924	36407-30-9	7,5	152,08	151,47	1,3		58,9
Carbamazepine-d10	C15H2D10N2O	246,1577	132183-78-9	8,9	149,04		2,5		46,3

Annexes

Nom	Formule moléculaire	Masse monoisotopique	CAS	TR mesuré (min)	CCS* mesuré (Å ²)	CCS* (Å ²) en librairie**	log Kow**	pKa**	Surface topologique polaire (Å ²)**
Carbendazime	C9H9N3O2	191,0695	10605-21-7	3,6	137,60	134,94	1,5	4,3	67,0
Cefoperazone	C25H27N9O8S2	645,1424	62893-19-0	6,4	235,49		-0,7		271,0
Chlormequat	C5H13ClN	122,0737	7003-89-6	0,6	123,22		0,9		0,0
Chlorothiazide	C7H6ClN3O4S2	294,9488	58-94-6	3,0	161,40	161,00	-0,2	6,9 ; 9,5	135,0
Chlorpyrifos	C9H11Cl3NO3PS	348,9263	2921-88-2	17,8	163,73	163,12	5,3		72,7
Chlortoluron	C10H13ClN2O	212,0716	15545-48-9	9,6	145,99	145,96	2,4		32,3
Clarithromycin	C38H69NO13	747,4769	81103-11-9	10,5	269,04	264,20	3,2	9,0	183,0
Clarithromycin N-oxide	C38H69NO14	763,4718	118074-07-0	10,9	273,69		3,2		198,0
Climbazole	C15H17ClN2O2	292,0979	38083-17-9	9,8	171,26		3,7		44,1
Clothianidin	C6H8ClN5O2S	249,0087	210880-92-5		151,55		0,7	11,1	123,0
CMIT	C4H4ClNOS	148,9702	26172-55-4	3,9	124,22		0,4		45,6
Cotinine-d3	C10H9[2H]3N2O	179,1138	110952-70-0	1,0	142,28		-0,3		33,2
Cybutryne	C12H21N5	235,1797	28159-98-0				3,9		88,0
Cypermethrin	C22H19Cl2NO3	415,0742	52315-07-8			200,60	6,0		59,3
DCOIT	C11H17Cl2NOS	281,0408	64359-81-5	17,2	164,94		2,8		45,6
DDAC	C22H48ClN	361,3475	7173-51-5	17,6	227,93		2,6		0,0
Deltamethrin	C22H19Br2NO3	502,9732	52918-63-5			206,32	6,2		59,3
Dichlorvos	C4H7Cl2O4P	219,9459	62-73-7	8,4	133,38	133,07	1,4		44,8
Diclofenac	C14H11Cl2NO2	295,0167	15307-86-5	13,4	156,22	157,70	4,4	4,2	49,3
Diflubenzuron	C14H9ClF2N2O2	310,0321	35367-38-5			160,98	3,9		58,2
Dimethomorph - forme 1	C21H22ClNO4	387,1237	113210-97-2	11,9	196,54	199,03	3,9		48,0
Dimethomorph - forme 2	C21H22ClNO4	387,1237	113210-98-3	11,6	196,95	201,14	3,9		48,0
Dimethyl phthalate	C10H10O4	194,0579	131-11-3	8,5	147,61		1,6		52,6
Dimethyl phthalate-CH4O	C9H6O3	162,0317		8,5	124,16				
Diphenyl phthalate	C20H14O4	318,0892	84-62-8	15,8	243,53	181,27	4,5		52,6
Diphenyl phthalate-C6H6O	C14H8O3	224,0473		15,8	144,09				

Annexes

Nom	Formule moléculaire	Masse monoisotopique	CAS	TR mesuré (min)	CCS* mesuré (Å ²)	CCS* (Å ²) en librairie**	log Kow**	pKa**	Surface topologique polaire (Å ²)**
Diphenylguanidine	C13H13N3	211,1109	102-06-7	5,1	151,20		2,4	10,1	50,4
Diuron	C9H10Cl2N2O	232,0170	330-54-1	10,2	147,68	152,30	2,7		32,3
Diuron-d6	C9H4[2H]6Cl2N2O	238,0547	1007536-67-5	10,1	148,45		2,7		32,3
Efavirenz	C14H9ClF3NO2	315,0274	154598-52-4	14,2	166,48		4,0	10,2 ; 12,5	38,3
Emamectin B1a	C49H75NO13	885,5238	119791-41-2	15,4	296,95		4,1	4,2 ; 8,7	162,0
EthylParaben	C9H10O3	166,0630	120-47-8	8,3	134,08	125,40	2,5	8,3	46,5
Fenhexamide	C14H17Cl2NO2	301,0636	126833-17-8	12,9	163,71	164,82	3,5	7,3	49,3
Fipronil	C12H4Cl2F6N4OS	435,9387	120068-37-3			180,77	4,5		104,0
Furosemide	C12H11ClN2O5S	330,0077	54-31-9	8,7	172,56	169,30	2,0	3,9	131,0
Haloperidol	C21H23ClFNO2	375,1401	52-86-8	8,9	193,59	193,70	3,2	8,7	40,5
Haloperidol-d4	C21H19ClD4FNO2	379,1652	1189986-59-1	8,9	193,53		3,2	8,7	40,5
Hexaflumuron	C16H10F6N2O3	392,0596	86479-06-3			188,22	5,7		67,4
Imazalil	C14H14Cl2N2O	296,0483	35554-44-0	9,6	165,43	165,20	3,8	6,5	27,0
Imidaclopride	C9H10ClN5O2	255,0523	138261-41-3				0,6	1,6	
IPBC	C8H12INO2	280,9913	55406-53-6	11,2	145,81		2,1		38,3
Irbesartan	C25H28N6O	428,2325	138402-11-6	9,9	201,80	200,80	4,1	4,1	87,1
IsoButylParaben	C11H14O3	194,0943	4247-02-3	11,5	191,10		3,4		46,5
Isoproturon	C12H18N2O	206,1419	34123-59-6	10,2	152,45	150,75	2,9		32,3
Ketoconazole	C26H28Cl2N4O4	530,1488	65277-42-1	9,7	214,95	214,50	4,3	4,0 ; 6,8	69,1
lidocaine	C14H22N2O	234,1732	137-58-6	4,4	157,74	157,70	2,3	8,0	32,3
lidocaine N-oxide	C14H22N2O2	250,1681	2903-45-9	5,1	170,95		2,6		47,2
L-phenylalanine	C9H11NO2	165,0790	63-91-2	2,3	138,60	135,00	-1,5	1,2	63,3
Mecoprop	C10H11ClO3	214,0397	93-65-2	11,5	149,18		3,1	3,8	46,5
Mefenoxam	C15H21NO4	279,1471	70630-17-0	10,2	158,47		1,6		55,8
Melamine	C3H6N6	126,0654	108-78-1	0,6	122,17		-1,4	5,0	117,0
Metazachlor	C14H16ClN3O	277,0982	67129-08-2	10,9	158,26		2,7		38,1

Annexes

Nom	Formule moléculaire	Masse monoisotopique	CAS	TR mesuré (min)	CCS* mesuré (Å ²)	CCS* (Å ²) en librairie**	log Kow**	pKa**	Surface topologique polaire (Å ²)**
Methamidophos	C2H8NO2PS	141,0013	10265-92-6	1,6	120,64	123,20	-0,8		77,6
Methidathion	C6H11N2O4PS3	301,9619	950-37-8	11,8	160,07	160,41	2,4		143,0
Methomyl	C5H10N2O2S	162,0463	16752-77-5	4,1	141,15		0,6		76,0
MethylParaben	C8H8O3	152,0473	99-76-3	6,5	127,80	128,43	2,0	8,5	46,5
MethylParaben d4	C8H4D4O3	156,0725	362049-51-2	6,5	127,59		2,0	8,5	46,5
Metolachlor	C15H22ClNO2	283,1339	51218-45-2	13,7	159,19	159,10	3,1		29,5
Metolachlor-ESA	C15H23NO5S	329,1297	171118-09-5	7,9	168,80		1,4		92,3
Metolachlor-OA	C15H21NO4	279,1471	152019-73-3	9,5	159,12		2,5		66,8
MIT	C4H5NOS	115,0092	2682-20-4	1,5	122,80		-0,8		45,6
Monuron	C9H11ClN2O	198,0560	205-766-1	8,0	141,30		1,9		32,3
Naproxen	C14H14O3	230,0943	22204-53-1	9,5	152,83	154,83	3,2	4,2	46,5
Nicotine-d4	C10H10[2H]4N2	166,1408	350818-69-8	1,1	137,14		1,2		16,1
Nigericin	C40H68O11	724,4762	28380-24-7	23,8	267,34		5,5		142,0
Octocrylene	C24H27NO2	361,2042	6197-30-4	20,1	195,56		7,1		50,1
OIT	C11H19NOS	213,1187	26530-20-1	12,6	224,23		3,5		45,6
Omethoate	C5H12NO4PS	213,0225	1113-02-6	2,9	135,60	134,62	-0,9		89,9
Phenazine	C12H8N2	180,0687	92-82-0	9,4	132,55		2,8		25,8
Progesterone	C21H30O2	314,2246	57-83-0	14,1	179,56	181,20	3,9	-4,8 ; 18,9	34,1
Prometryn	C10H19N5S	241,1361	7287-19-6		162,22	160,60	3,5	4,1	88,0
Propazine	C9H16ClN5	229,1094	139-40-2	11,3	154,74	155,50	2,9	1,7	62,7
Propiconazole	C15H17Cl2N3O2	341,0698	60207-90-1			177,88	3,5	1,1	49,2
PropylParaben	C10H12O3	180,0786	94-13-3	10,0	143,16	142,48	3,0	8,5	46,5
propylparaben-d4	C10H8D4O3	184,1038	1219802-67-1	10,0	141,17				
Pyraclostrobin	C19H18ClN3O4	387,0986	175013-18-0	15,3	181,89	183,19	4,1		65,8
pyridinium furosemide	C12H9N2O5SCI	327,9921		0,6	168,17				
Reserpine	C33H40N2O9	608,2734	50-55-5	10,6	250,86	247,30	4,0	6,6	118,0

Annexes

Nom	Formule moléculaire	Masse monoisotopique	CAS	TR mesuré (min)	CCS* mesuré (Å ²)	CCS* (Å ²) en librairie**	log Kow**	pKa**	Surface topologique polaire (Å ²)**
Rifaximin	C43H51N3O11	785,3524	80621-81-4	13,0	270,99	267,80	2,6		198,0
Saccharin	C7H5NO3S	182,9990	81-07-2	2,9	132,83		0,9	1,3	71,6
Sebuthylazine	C9H16ClN5	229,1094	7286-69-3	11,2	154,80		3,1		62,7
Simazine	C7H12ClN5	201,0781	122-34-9	8,0	143,15		2,2	1,6	62,7
Simazine-2-OH	C7H13N5O	183,1120	2599-11-3	3,6	141,12	141,28	-0,3		77,9
Simazine-d10	C7H2ClD10N5	211,1409	220621-39-6	7,8	143,27		2,2		62,7
Simvastatin	C25H38O5	418,2719	79902-63-9	17,0	197,60		4,7		72,8
Spinosad A	C41H65NO10	731,4608	1370709-67-3	13,8	269,23		4,9		218,0
sucralose	C12H19Cl3O8	396,0146	56038-13-2	4,8	176,42	177,20	-1,0		129,0
Sudan I	C16H12N2O	248,0950	842-07-9	17,1	151,39		4,1		45,0
Sulfamethoxazole-d4	C10H7D4N3O3S	257,0772	1020719-86-1	6,0	152,98		0,9		107,0
TCMTB	C9H6N2S3	237,9693	21564-17-0	12,5	143,66		3,2		116,0
Tebuconazole	C16H22ClN3O	307,1451	107534-96-3	13,4	167,10	167,37	3,7	3,7	50,9
Terbutylazine	C9H16ClN5	229,1094	5915-41-3	11,8	154,83	153,48	3,4	2,0	62,7
Terbutryn	C10H19N5S	241,1361	886-50-0	9,7	160,68	161,23	3,7	4,3	88,0
Tetraethylammonium	C8H20N	130,1596	66-40-0	0,9	128,18		1,7		0,0
Tetrahexylammonium	C24H52N	354,4100	20256-54-6	17,2	218,29		9,6		0,0
Theophylline	C7H8N4O2	180,0647	58-55-9	3,2	133,72	137,60	0,0	8,8	69,3
Thiabendazole	C10H7N3S	201,0361	148-79-8	4,0	137,37	137,70	2,5	4,6	69,8
thiacloprid	C10H9ClN4S	252,0236	111988-49-9			158,84	2,2		77,6
Thiamethoxam	C8H10ClN5O3S	291,0193	153719-23-4				-0,1		115,0
tramadol	C16H25NO2	263,1885	27203-92-5	5,8	162,81	161,40	2,6	9,4	32,7
tramadol N-oxide	C16H25NO3	279,1834	147441-56-3	6,2	167,27		2,1		47,5
Trichlorfon	C4H8Cl3O4P	255,9226	52-68-6	5,4	135,79			0,5	136,1
Triclocarban	C13H9Cl3N2O	313,9780	101-20-2	15,7	172,24		4,3	12,7	41,1
Triclocarban d4	C13H5Cl3D4N2O	318,0032	1219799-29-7	15,7	206,49				

Annexes

Nom	Formule moléculaire	Masse monoisotopique	CAS	TR mesuré (min)	CCS* mesuré (Å ²)	CCS* (Å ²) en librairie**	log Kow**	pKa**	Surface topologique polaire (Å ²)**
Triclosan	C ₁₂ H ₇ Cl ₃ O ₂	287,9512	3380-34-5	15,9	158,57	158,94	4,8	7,9	29,5
Tylosin	C ₄₆ H ₇₇ NO ₁₇	915,5192	1401-69-0	9,6	316,03	319,79	1,6	7,7	239,0

*Sans code couleur [M+H]⁺, en vert [M-H]⁻, en orange [M+Na]⁺, en rouge [M+HCOO]⁻, en violet [M+NH₄]⁺, en bleu e-

**Données issues de Pubchem (<https://pubchem.ncbi.nlm.nih.gov/>)

Annexe 2 : Réplicats après alignement de différentes colonnes chromatographiques

	PrimaryId.x	PrimaryId.y	Id
	103.95480_3.895_2.55	103.95547_1.129_2.71	11
3	103.95480_3.895_2.55	103.95548_23.867_2.73	11
	103.95480_3.895_2.55	103.95555_0.628_2.74	11
	105.06926_4.347_2.57	105.06998_2.502_2.76	11
1	105.95332_3.992_2.56	105.95360_23.837_2.71	11
	105.95332_3.992_2.56	105.95370_0.617_2.74	11
	108.07991_4.467_2.62	108.08072_0.591_2.86	11
	108.08002_3.320_2.61	108.08072_0.591_2.86	11
	108.08007_5.989_2.62	108.08072_0.591_2.86	11
	108.08012_0.866_2.61	108.08072_0.591_2.86	11

PrimaryId.x correspond aux marqueurs de la colonne chromatographique 1

PrimaryId.y correspond aux marqueurs de la colonne chromatographique 2.

Chaque *PrimaryId* regroupe les informations suivantes relatives au marqueur : m/z _TR_DT

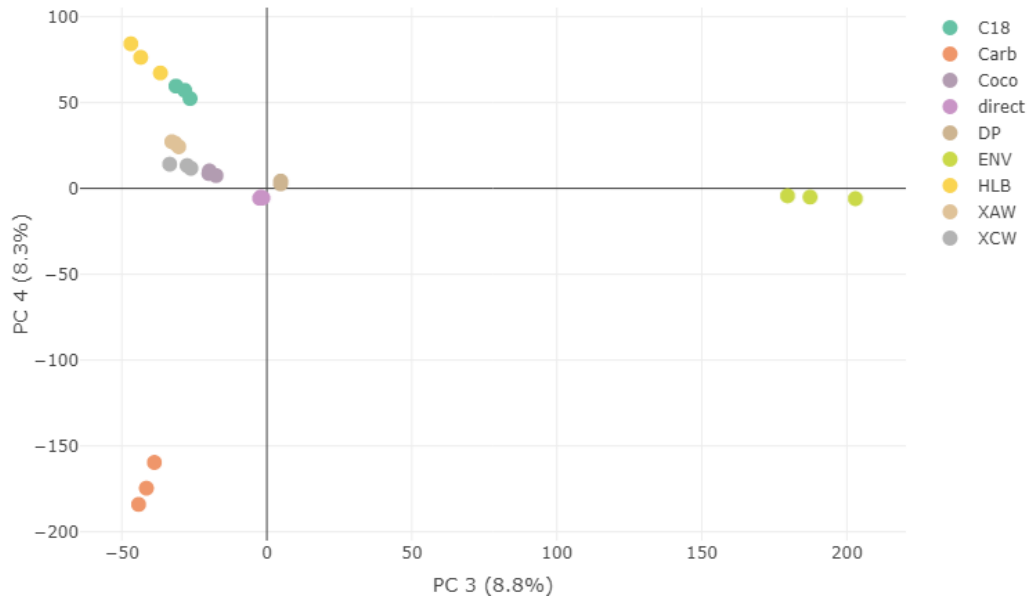
Une ligne du tableau représente l'alignement effectué en m/z puis DT.

Dans cet exemple, le marqueur de la colonne 1 désigné par 103.9548_3.895_2.55 est aligné avec 3 marqueurs différents de la colonne chromatographique 2. Ces derniers sont compris dans les fenêtres de tolérance en m/z et DT, mais leur TR respectifs sont 1,129 ; 23,867 et 0,628 min respectivement. Avec les informations à dispositions, il est impossible de savoir lequel de ces trois marqueurs correspond effectivement au marqueur de la colonne 1, qui lui, ne peut pas correspondre à ces trois marqueurs à la fois.

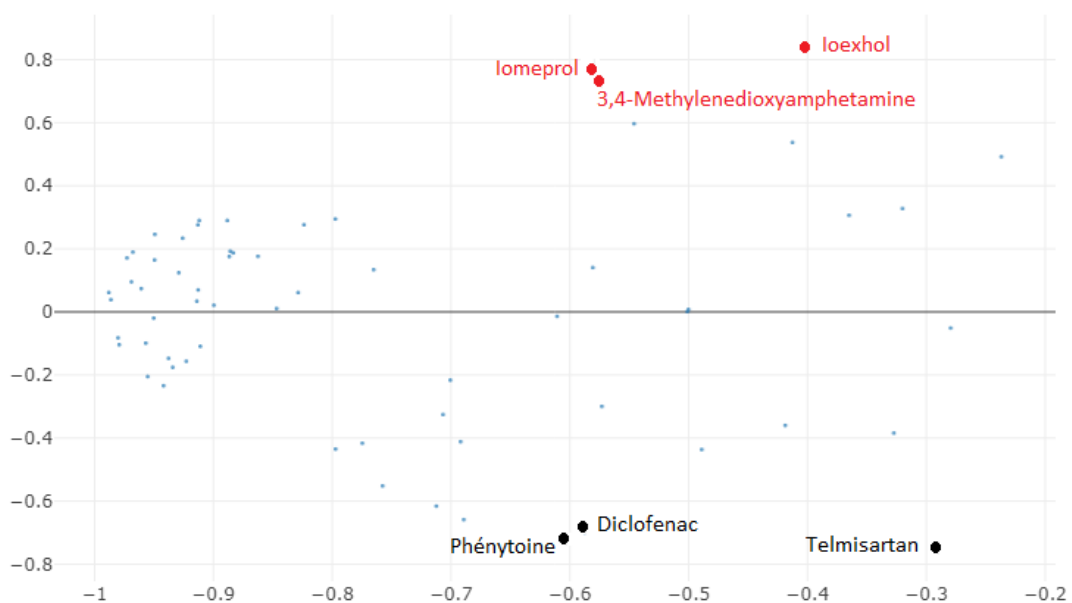
L'étape de filtration complémentaire a pour but de compter 1 seul marqueur aligné pour ce premier groupe, qui, initialement était compté comme représentant 3 marqueurs alignés.

Annexe 3 : Comparaison des méthodes analytiques

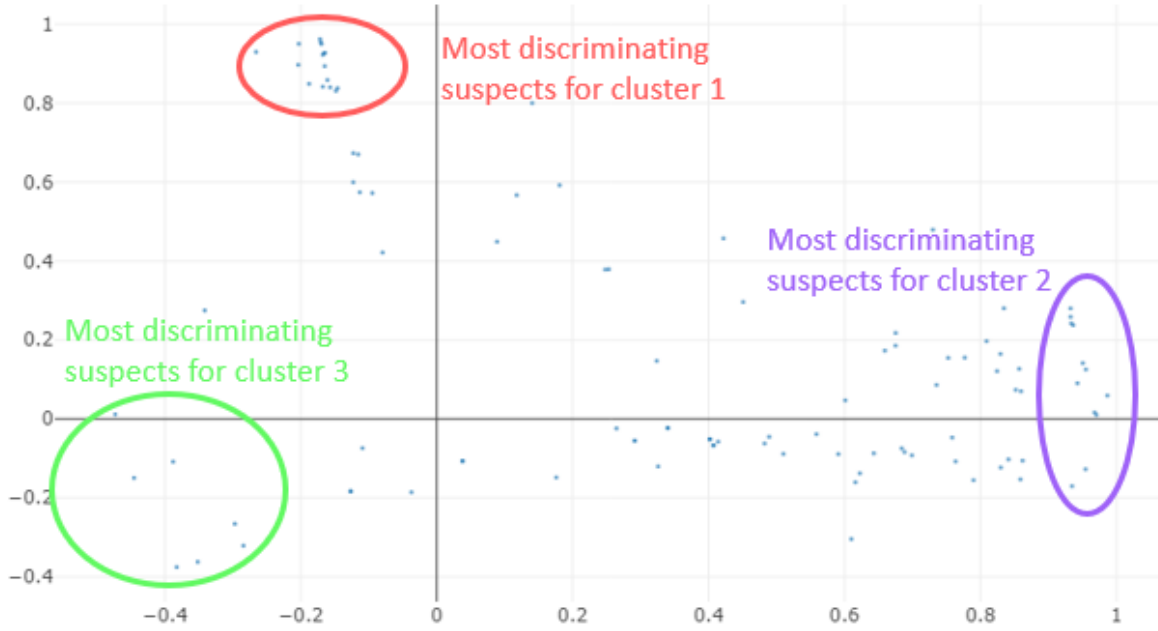
A – PC4 en fonction de PC3 de l'ACP en NTS pour la comparaison de cartouches SPE



B – *Loading plot* de l'ACP en SS pour la comparaison de cartouches SPE



C – *Loading plot* de l'ACP en SS pour la comparaison de stratégie analytique (combinaisons méthode de préparation et séparation)



Annexe 4 : Article Molecules, MDPI

Evaluation of Sample Preparation Methods for Non-target Screening of Organic Micropollutants in Urban Waters Using High-Resolution Mass Spectrometry

Nina Huynh^{1,2}, Emilie Caupos^{1,2,3}, Caroline Soares Peirera^{1,2}, Julien Le Roux^{1,2,*}, Adèle Bressy^{1,2} and Régis Moilleron^{1,2}

¹LEESU, Univ Paris Est Creteil, F-94010 Creteil, France

²LEESU, Ecole des Ponts, F-77455 Marne-la-Vallee, France

³OSU-EFLUVE, Univ Paris Est Creteil, F-94010 Creteil, France

Abstract

Non-target screening (NTS) has gained interest in recent years for environmental monitoring purposes because it enables the analysis of a large number of pollutants without predefined lists of molecules. However, sample preparation methods are diverse, and few have been systematically compared in terms of the amount and relevance of the information obtained by subsequent NTS analysis. The goal of this work was to compare a large number of sample ex-traction methods for the unknown screening of urban waters. Various phases were tested for the solid-phase extraction of micropollutants from these waters. The evaluation of the different phases was assessed by statistical analysis based on the number of detected molecules, their range, and physicochemical properties (molecular weight, standard recoveries, polarity, and optical properties). Though each cartridge provided its own advantages, a multilayer cartridge combining several phases gathered more information in one single extraction by benefiting from the specificity of each one of its layers.

Keywords

Emerging contaminants; High-resolution mass spectrometry; Micropollutant fingerprint; Non-target screening; Solid phase extraction; Statistical analysis; Urban waters

1. Introduction

The presence of organic micropollutants in the aquatic environment is a major issue in many countries due to their potential ecotoxicological impact on aquatic organisms as well as their potential risks for human health. European directives concerning the quality of surface waters advocate better control of the presence of micropollutants, but they deal with priority organic molecules that are well known by scientists for their high environmental impact (Water Framework Directive). However, a lot of emerging molecules (e.g., substitutes of forbidden molecules, metabolites, and degradation byproducts) enter the environment due to human activities through wastewater effluents. The survey of these unknown molecules using non-target screening (NTS) enables us to anticipate further regulation on water quality (Hollender et al., 2017). This analytical approach has undergone high development in the last decade thanks to the spreading of high-resolution mass spectrometry, but it requires powerful analytical systems and computational software as well as non-specific sample preparation methods (González-Gaya et al., 2021).

The increased resolving power of high-resolution mass spectrometry such as time-of-flight mass spectrometry (TOF-MS) or Orbitrap-MS allows the accurate mass measurement and isotopic pattern identification (Rodríguez-Aller et al., 2013). Therefore, these techniques are suitable for the development of three main screening strategies based on accurate-mass databases (Hernández et al., 2005; Schymanski et al., 2015) : multi-residue target screening (Cortéjade et al., 2016), suspect screening (Deeb et al., 2017; Diaz et al., 2013; Ibáñez et al., 2008), and NTS (Schollée et al., 2018; Singer et al., 2016). Contrary to target and suspect screening which aim to evaluate the presence of molecules based on a preliminary list, NTS is the characterization of substances present in a sample without any prior information, based on full scan analysis that enables the acquisition of all peaks comprised within a range of masses. It thus requires the subsequent selection (e.g., by statistical analysis) of features of interest (i.e., specific signal characterized by its retention time and m/z ratio) followed by their tentative identification with the use of databases and literature information (Krauss et al., 2010; Schulz et al., 2019).

In combination with high-resolution mass spectrometry acquisition methods, the computational technique for data processing is of primary interest. Molecular identification

depends on high-resolution data precision (Krauss et al., 2010) but also on the software possibilities to deconvolute, filter, re-align signals (Hohrenk et al., 2019), associate raw formulas to features of interest and finally compare with existing spectra libraries for identification (Hufsky and Böcker, 2017; Pluskal et al., 2012). NTS studies can be performed using proprietary software (e.g., Agilent Masshunter and MPP, Waters Unifi, Thermo Compound Discoverer, ABSciex MarkerView, SpectralWorks AnalyzerPro, associated with ChemSpider, SciFinder, METLIN, PubChem, or MassBank libraries) as well as free ones (e.g., MZmine, R, EnviMass, and MetFrag) (Hug et al., 2014) for the identification of molecules, using the information on the fragments. The qualification of a single sample requires a lot of computational treatment, and statistical studies are relevant as a complementary approach to assess organic contamination by comparing different samples. Statistical approaches in NTS have been used to differentiate samples in time (Alygizakis et al., 2019) and space (Krauss et al., 2019) or to highlight variations during processes (Schollée et al., 2018). Binary comparison, principal component analysis (PCA), and Venn diagrams are examples of methods that can be applied to a sample set in order to isolate features of interest (e.g., molecules in common between samples or specific to groups of samples) (Müller et al., 2011). Discriminated samples can be further characterized, and molecules of interest can then be identified.

Data analysis and statistical evaluation of samples being time-consuming, the quality of acquired data must be guaranteed. Sample preparation is essential to ensure the detection of as many molecules as possible and thus needs to be as unselective as possible. Different strategies and methods can be used in order to reach this objective and also to prevent samples from contamination. Direct injection of water samples can be employed to avoid potential contaminations or losses during sample preparation. However, the low concentration of molecules and interferences with other constituents of the sample (e.g., solid particles, colloids, or organic matter) can prevent the detection of trace organic compounds. Preparation techniques such as filtration (Smith, 2003), concentration (Chon et al., 2017), and solid-phase extraction (Fontanals et al., 2007) are widely used for targeted quantitative analysis and for NTS to decrease interferences and to concentrate the samples.

Most NTS studies employ a sample preparation step that is considered to be as “exhaustive” as possible, and little information is available about the efficiency of SPE for NTS purposes. For target screening purposes, knowing the properties of the targeted molecules makes it easier

to select the most appropriate cartridge and phase. Concerning NTS studies, the choice of a cartridge is trickier given that, for example, the choice of a cartridge that is recommended for polar compounds will only retain polar compounds and thus, will leave the non-polar compounds un-retained. As a result, most non-target studies employ universal phases such as Oasis HLB to retain as many compounds as possible. Validating the extraction and analysis steps is also a challenge for NTS. The choice of standards to assess cartridge efficiency or to evaluate and correct matrix effects are tricky since the compounds studied in NTS are supposed to be unknown. Therefore, a mix of internal standards composed of a large variety of molecules (in terms of molecular weight, polarity, acidity, functional groups, etc.) is often used to assess the efficiency of the method in recovering and analysing a large panel of compounds (Kern et al., 2009; Moschet et al., 2013; Nürenberg et al., 2015; Schulze et al., 2020). In addition to the injection of internal standards, various methods can be used to check the stability of the instrument or to correct the standard deviation of each feature, such as the injection of a pooled sample (consisting of a mix of all samples to be analysed), regularly during the sequence. These normalization methods were recently reviewed for NTS applications in lipidomics (Drotleff and Lämmerhofer, 2019). Finally, using replicated injections to discard signals that are not repeatable and that could lead to false-positive identifications is a commonly used step in NTS studies (Albergamo et al., 2019; Köppe et al., 2020; Nürenberg et al., 2015; Peter et al., 2019; Schulze et al., 2020). Still, only a few studies compared the efficiency of different preparation methods for NTS purposes. One study showed that non-target strategies clearly discriminated the signals in terms of number and type of features obtained with several preparation methods (liquid-liquid extraction and solid-phase extraction on two cartridges) while the extraction yields of targeted molecules were not significantly different between those techniques (Samanipour et al., 2018). In all cases, the properties of the cartridge, the pH of the sample, as well as elution solvents, and conditioning solvents will play a major role in the selection of molecules finally detected in the analysed sample. A literature review revealed that numerous phases (e.g., HLB, ENV+, C18, X-CW, X-AW, X-C, and X-A, individually or in series), different sample pH, and different conditions of elution were used in NTS studies for environmental applications (Table 1). Several studies used a multilayer cartridge (composed of HLB, ENV+, X-AW, and X-CW) developed to extract a wide range of micropollutants (Kern et al., 2009).

Table 1. Examples of SPE cartridges, pH of samples, and elution solvents used for sample preparation in environmental studies using non-target screening for the analysis of water samples. N.A. : Not Available

Matrix	pH	SPE Cartridge	Elution Solvents	Ref
Surface water and wastewater (influent and effluent)	N.A.	HLB	N.A.	(Ibáñez et al., 2008)
Surface water, groundwater and Drinking water	2	HLB, MCX	Acetonitrile (HLB) Acetonitrile, Acetonitrile + 5% ammonia (MCX)	(Hogenboom et al., 2009)
Surface water, wastewater influent and effluent,	6.7	Multilayer (HLB, ENV+, X-AW, X-CW)	Methanol/Ethyl acetate (50:50, v/v) + 2% ammonia and methanol/ Ethyl acetate (50:50, v/v) + 1.7% formic acid	(Gago-Ferrero et al., 2015; Kern et al., 2009; Moschet et al., 2013; Ruff et al., 2015; Schollée et al., 2018; Schymanski et al., 2014b)
Landfill leachate and groundwater	7 and 3	ENV+	Methanol	(Müller et al., 2011)
Wastewater effluent	2	MCX and Strata X in series	Methanol + 5% ammonia	(Nurmi et al., 2012)
Wastewater (influent and effluent)	N.A.	HLB	Methanol	(Díaz et al., 2012)
Surface water	N.A.	Multilayer (HR-X, HR-XAW, HR-XCW)	Ethyl acetate; methanol; methanol + 2% ammonia and methanol + 1% formic acid	(Schymanski et al., 2015)
Wastewater effluent	N.A.	MAX and MCX in series	Methanol/ethyl acetate/formic acid (69:29:2, v/v) (MAX) methanol/ethyl acetate/ammonia solution (67.5:27.5:5, v/v) (MCX)	(Deeb et al., 2017)
Riverbank filtration system	N.A.	HLB	Methanol	(Albergamo et al., 2019)

The overall objective of this work was to assess the influence of the extraction method, in particular the type of SPE phase/cartridge, on the non-targeted HRMS analysis of organic contaminants in urban water. To reach this objective, we pursued the following specific aims:

(i) to test different nature of phases to extract a large range of micropollutants in urban waters, based on a panel of 9 commercially-available stationary phases and de-signed for a variety of applications (*i.e.*, from non-polar/moderately polar to more polar compounds and ionic species); (ii) to develop a strategy based on relevant indicators to compare the results, such as the number of detected molecules, their range, and characteristics; and (iii) to apply the optimized SPE method and the developed data analysis strategy to urban water samples.

2. Results and Discussion

2.1. Comparison of the Cartridge Retention Ability

In this section, the retention abilities of several cartridges (ENV+, X-A, X-AW, X-CW, HLB pH 2, HLB pH 6, Multilayer, X-C, C18 ENV+, SDBL, and C18) were compared on the basis of different indicators.

2.1.1. Discrimination of Cartridges Based on Optical Properties

To get a preliminary overview of the diversity of retained organic materials, optical properties (UV absorbance at 254 nm and 3D-spectrofluorescence spectra) of one sample from a river located in an urbanized area (Marne River) were measured before and after extraction (*i.e.*, on the water phase during the SPE loading step). These optical properties are commonly used to characterize dissolved organic matter (DOM). As DOM is composed of molecules of various sizes and polarities with which micropollutants can interact, the behaviour of DOM can be related to the behaviour of micropollutants. Based on these interactions (*e.g.*, hydrogen bonding, hydrophobic, van der Waals, or dipole-dipole interactions), fluorescence can be used for a better understanding of the fate of micropollutants, for example, in advanced wastewater treatment processes such as adsorption onto activated carbon (Guillossou et al., 2021). DOM retention on SPE cartridges was evaluated to characterize their ability to retain a large variety of organic materials but also to determine if it can hinder the detection of micropollutants. DOM is indeed known to sometimes reduce the adsorption of target compounds on SPE cartridges (competition effect) and to limit the ionization of molecules for their detection in mass spectrometry (Duan et al., 2013; Wickramasekara et al., 2012). The good retention of DOM on SPE cartridges could thus be detrimental to the detection of organic molecules in non-target analyses. Figure 1 presents the retention of DOM on each SPE cartridge based on the percentage change of UV absorbance at 254 nm before and after

extraction. The highest recoveries were obtained for X-A, X-AW, X-C, and the Multilayer phases (recoveries $\geq 90\%$). Other cartridges exhibited significantly lower retention of organic materials ($<60\%$). Especially, extractions with the HLB cartridge retained fewer organic materials ($\sim 60\%$). SDBL, X-CW, and HLB (pH 6) do not seem to be appropriate phases for a large screening of organic material.

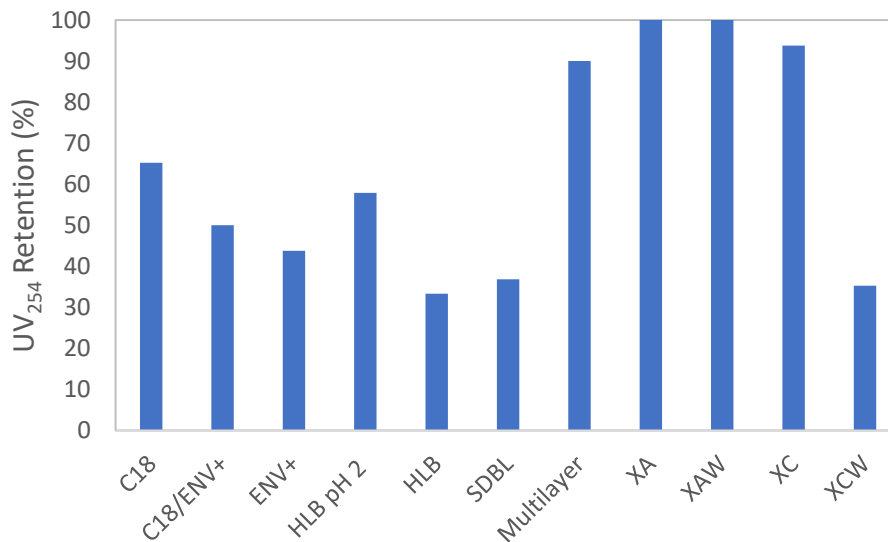


Figure 1. Retention of DOM based on UV absorbance at 254 nm (UV₂₅₄) for different SPE cartridges. UV₂₅₄ retention (%) was calculated as the percentage change between UV₂₅₄ of the Marne River sample and UV₂₅₄ measured at the outlet of each cartridge (during the loading phase).

For a better evaluation of the quality of DOM retained on the cartridges, 3D-fluorescence spectra were acquired from the Marne River sample before and after extractions. Fluorescence regional integration (FRI) was performed (Duan et al., 2013; Guillosoy et al., 2021; Sgroi et al., 2017; Wickramasekara et al., 2012), calculating the regional fluorescence intensities for each excitation-emission matrix (EEM) spectrum. The EEM spectrum of the Marne River sample before extraction (Supplementary Materials Figure S1) exhibited mainly two fluorophores: one in the region III of FRI and the other in region V. They correspond to large molecules like polysaccharides and humic-like substances (Chon et al., 2017; Coble, 1996; Parlanti, 2000). Other fluorophores (regions I and II) were also observed, representing more hydrophilic and smaller molecules, proteins, and aromatic amino acids (Chen et al.,

2003). Table 2 presents fluorophores retention depending on the phase of SPE used. Similar results were obtained with other indexes derived from (Sgroi et al., 2017) (data not shown).

Table 2. DOM retention (%) based on fluorophores for different SPE cartridges (indexes FRI derived from (Sgroi et al., 2017)).

Index	C18	C18 ENV+	ENV+	HLB pH 2	HLB	Multilayer	SDBL	XA	XAW	XC	XCW
Region I FRI	84	85	64	70	58	71	58	94	89	55	49
Region II FRI	76	69	67	68	58	86	48	96	90	59	47
Region III FRI	72	46	46	55	55	73	27	92	86	50	28
Region IV FRI	68	62	64	65	45	80	41	93	86	53	45
Region V FRI	70	47	50	56	53	55	28	93	86	48	28

As expected, the cartridges retained aromatic proteins and humic-like materials (regions I, III, and V (Guillossou et al., 2021; Zhang et al., 2008)). Higher retention was observed for XA and XAW ($\geq 85\%$). Like the results obtained with UV absorbance, the extractions on SDBL and XCW phases led to poor DOM retention (retentions generally lower than 50%). It is interesting to note that regions II and IV were well retained by XA, XAW, and Multilayer ($\geq 80\%$). They can be associated with the retention of simpler, more polar, and nitrogenous compounds (Zhang et al., 2008). Based on EEM spectra, these three cartridges seem to be the most efficient to retain hydrophobic materials as well as smaller and more polar compounds for global screening of the molecules present in the sample.

2.1.2. Discrimination of Cartridges based on HRMS Features

Samples were then analysed by HRMS, and differences between SPE cartridges were investigated through criteria classically used in omics studies. A PCA of the HRMS features found in the different extracts of the Marne River sample was performed to quickly identify similarities and differences between the feature sets obtained from each cartridge (Figure 2).

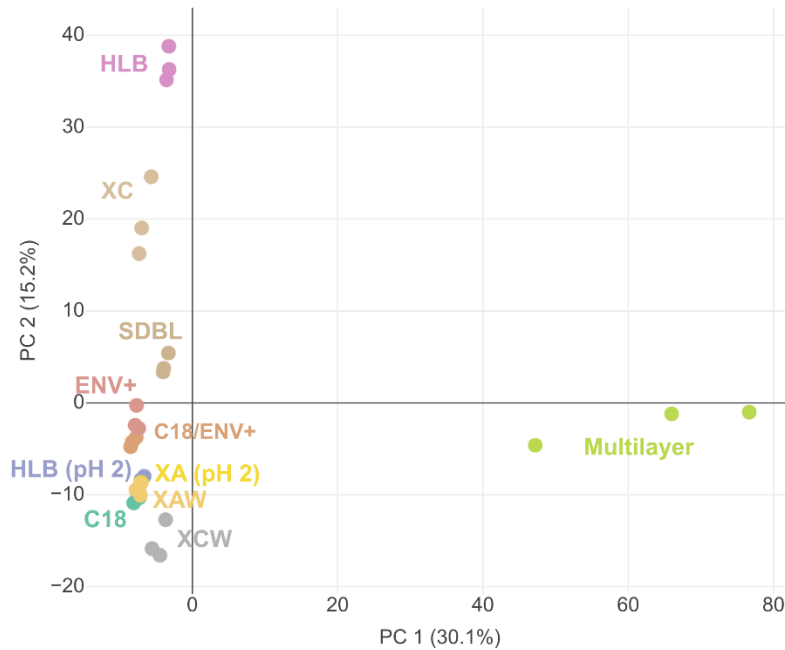


Figure 2. PCA (graph of individuals) of Marne sample extracts on different SPE cartridges.

All cartridges were clearly discriminated by the first two components of the PCA (covering 45.3% of the total variance), which indicated that each type of SPE phase extracted specific sets of HRMS features. The first component (PC 1) was mostly correlated with the features derived from the Multilayer cartridge and explained 30% of the dataset variability. The second component (PC 2) explained 15% of the variability and discriminated all individual cartridges. Polymeric cartridges (ENV+, SDBL, and C18/ENV+) were all clustered near the centre of the PCA, meaning that they all extracted similar features. Cartridges for which sample loading was performed at pH 2 (HLB pH 2 and XA) also exhibited similar behaviour. Features from anionic exchange cartridges (XA and XAW) were almost identical.

To better understand the repartition of cartridges obtained by the PCA (Figure 2), fingerprints (*i.e.*, bubble plots of all detected features) were compared between the most different cartridges (XCW, ENV+, HLB, and the Multilayer cartridge) for the Marne river sample (Figure 3).

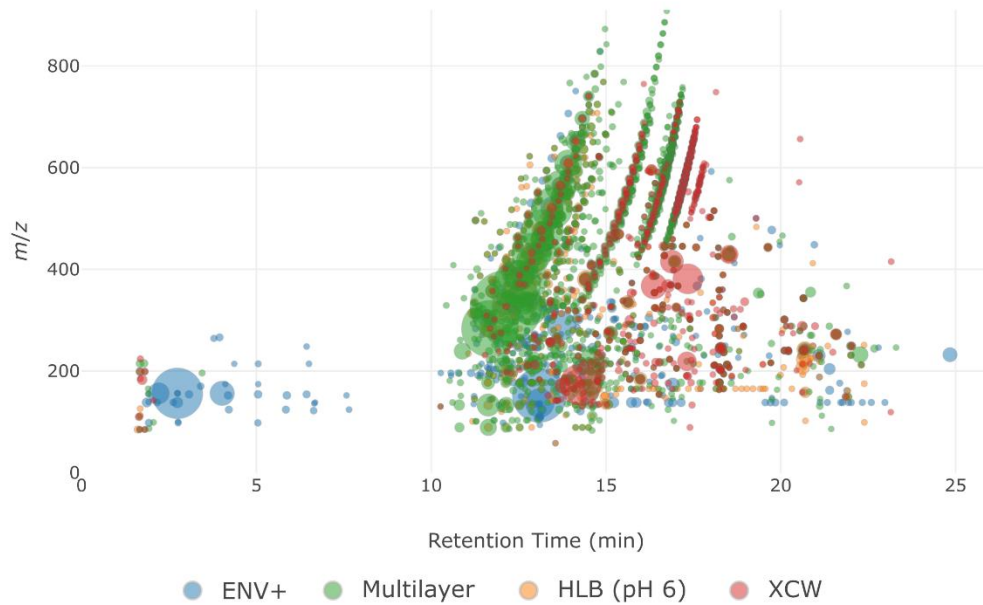


Figure 3. Fingerprints of detected features from the Marne River sample after SPE on ENV+ (blue), HLB (orange), Multilayer (green), and XCW (red) cartridges. The size of bubbles is proportional to the area of the feature.

The Multilayer cartridge displayed the most intense features on the entire m/z range, whereas ENV+ displayed more intense features for $m/z < 400$, and HLB retained more intense features with $m/z > 400$. XCW showed more intense signals at retention times greater than 15 min (*i.e.*, less polar compounds), whereas most intense signals on the two other individual cartridges were obtained at retention times before 15 min. Thus, the PC 2 component of the PCA may describe the polarity of the retained features on each cartridge. This hypothesis is supported by the fact that the C18 cartridge, which is designed for less polar compounds, was close to the XCW one on the PCA. Finally, a comparison of common features retained by the cartridges was performed (Figure 4).

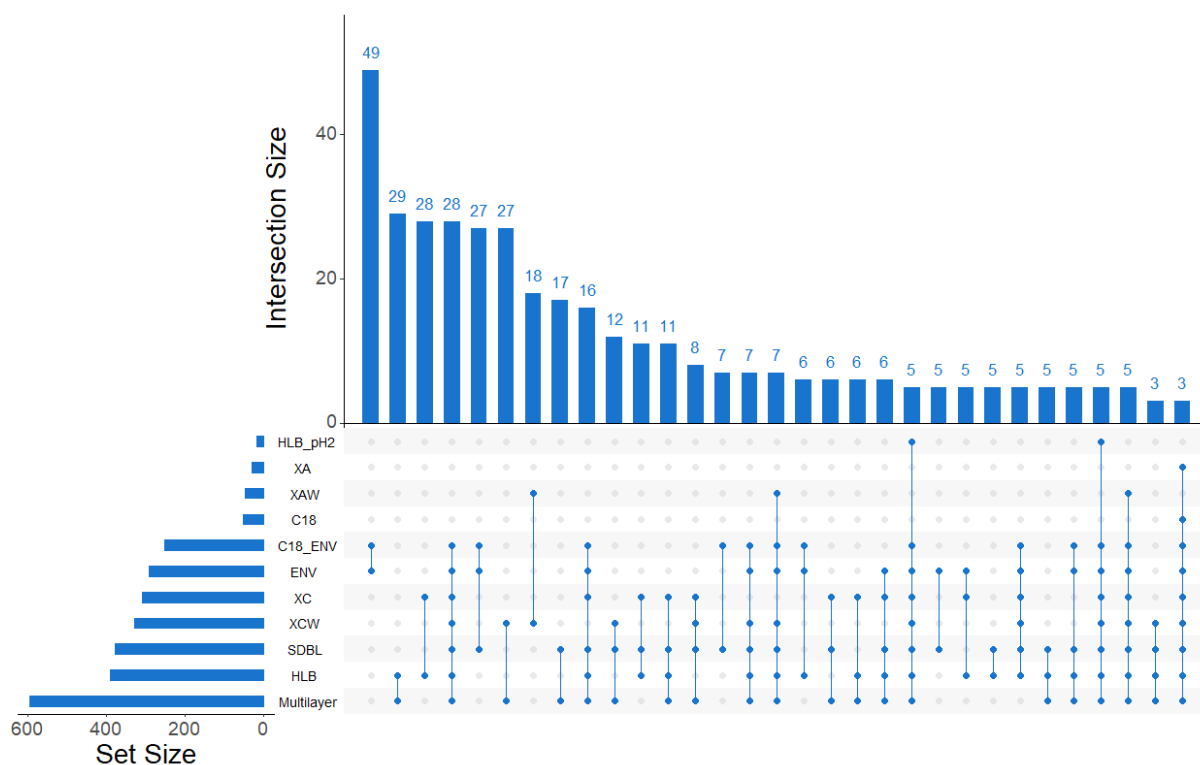


Figure 4. Upset diagram, representing the intersection of the different cartridges in terms of retained features. The “Set size” represents the total number of features from each cartridge, the “Intersection size” represents the size of the intersection (*i.e.*, number of features in the intersection designated by the dots), the dots represent the cartridges intersected (*i.e.*, number of intersections).

The first thing to notice is that the Multilayer cartridge had the biggest set size among the studied cartridges, meaning that it retained the highest number of features. The two closest cartridges in terms of retained features were the ENV+ and C18/ENV+ cartridges with 49 common features, which is not surprising given the similar nature of the phases. The cartridge involved in the highest number of intersections was the SDBL (23 intersections for 210 features intersected); however, the cartridge with the highest number of features intersected was the Multilayer one (20 intersections for 212 features intersected). Therefore, the Multilayer was the cartridge with the highest number of specific features (323 features only found in this cartridge), but also the one with the highest number of intersected features, making it the most interesting cartridge for non-target studies by retaining the most extensive set of features.

Compared to the results obtained when considering the optical properties (2.1.1), the cartridges that retained more fluorescent organic materials exhibited a lower number of HRMS features (*e.g.*, XA, XAW). On the contrary, SDBL and XCW showed low retention of fluorescing materials but a high number of HRMS features. This could be due to the competition of DOM with organic compounds for the adsorption on the SPE cartridge (Duan et al., 2013; Jeanneau et al., 2007) or to matrix effects in the HRMS analysis. Interestingly, this behaviour can also be observed for the same cartridge used at different pH. Extraction on HLB at pH 6 gave a high number of features in HRMS and low retention of fluorescing organic materials (fluorophores), contrary to the extraction on HLB at pH 2. Notably, the Multilayer cartridge exhibited both retention of fluorophores and a large number of HRMS features; this cartridge thus seemed less affected by organic content.

2.2. Characterization of the Cartridge Retention Ability with Different Matrices

Based on the previous observations (HRMS fingerprints and DOM retention), five cartridges were retained for further investigation with different matrices: the Multilayer cartridge (highest number of retained features and highest intensities), HLB at pH 6 (universal cartridge, used in many studies), ENV+ (universal cartridge, representing the polymeric cartridges clustered in the PCA), XAW (high retention of fluorophores) and XCW (retaining more apolar compounds). Three types of samples were extracted and compared, with the increasing complexity of matrix (*i.e.*, presence of organic and inorganic constituents): Ultrapure water (Milli-Q), surface water (Marne River), and wastewater effluent (WWe).

2.2.1. Recoveries of Internal Standards

To compare the different SPE cartridges and evaluate the corresponding matrix effects, the recovery of internal standards was first calculated (Figure 5).

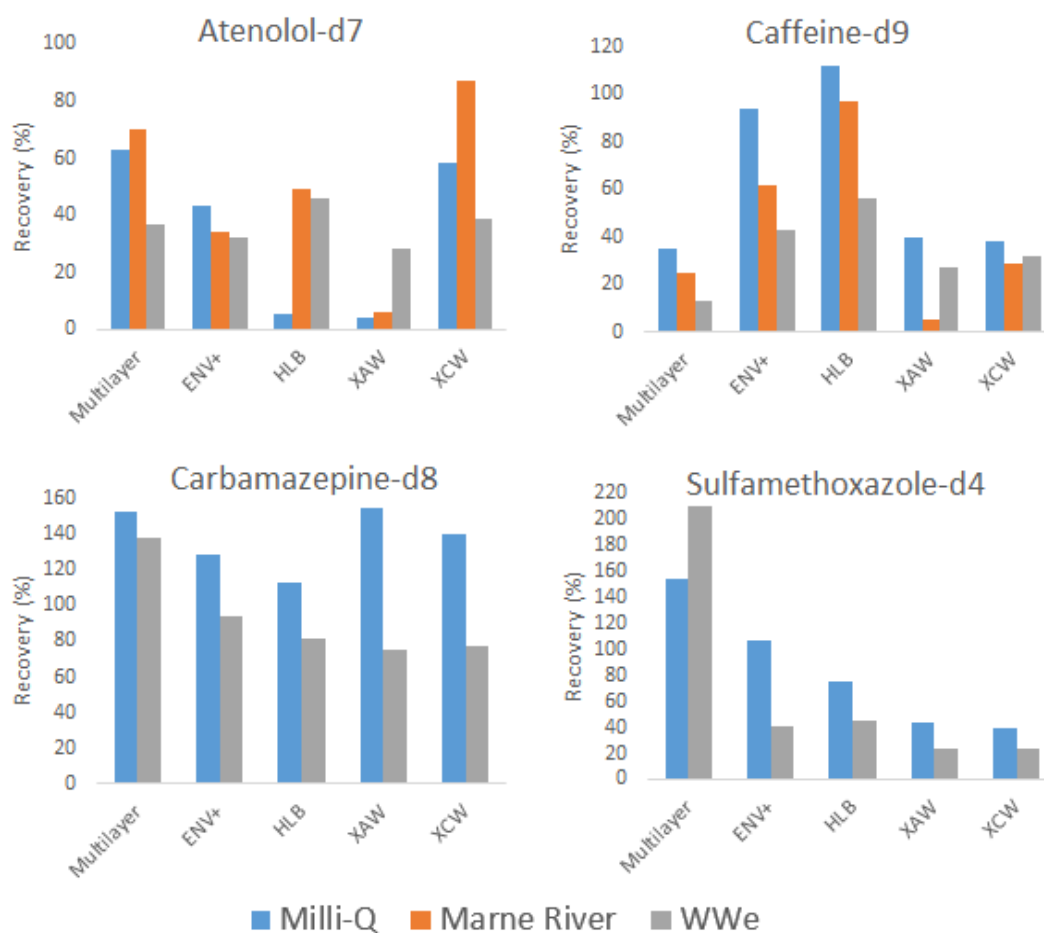


Figure 5. Recovery of internal standards after extraction by the cartridges for different matrices.

In Milli-Q, recoveries ranged from 4% to 63% for atenolol-d7, from 35% to 112% for caffeine-d9, from 113% to 154% for carbamazepine-d8 and from 40% to 155% for sulfa-methoxazole-d4. Carbamazepine-d8 exhibited an overall great recovery with all cartridges and showed the least differences between cartridges. On the contrary, atenolol-d7 gave lower recoveries, and sulfamethoxazole-d4 and caffeine-d9 exhibited large differences in recoveries between cartridges. Carbamazepine-d8 has the highest log Kow among the four molecules; its better recovery on all cartridges could be explained by its lower hydrophilicity. Extractions were all carried out at pH = 6.5–7, meaning that, for atenolol-d7, caffeine-d9 and carbamazepine-d8, the acidic form was predominant, whereas, for sulfamethoxazole-d4, its basic form is predominant. The only molecule showing a significant difference between XAW and XCW cartridges (designed for compounds with $pK_a < 5$ and $pK_a > 8$, respectively) was atenolol-d7 ($pK_a = 9.6$), with a better recovery on XCW as expected. The strong acidity (pK_a of sulfamethoxazole-d4 = 1.6) or basicity (pK_a of caffeine-d9 and carbamazepine-d8 are 14 and

13.9, respectively) of the three other compounds may explain their less contrasted retention of those two cartridges. Overall, the Multilayer cartridge gave the highest recoveries of those standards (despite a quite poor recovery of caffeine-d9).

Lower recoveries of caffeine-d9 were recorded with the Marne River and WWe samples for all five cartridges as compared to the Milli-Q water sample. The three universal cartridges (Multilayer, HLB, and ENV+) were more affected by complex matrices. The ionic XCW cartridge showed the least difference in recoveries for the three matrices. However, the XAW cartridge exhibited a very low recovery of most internal standards from the Marne River sample, while the two other samples (Milli-Q and WWe) showed similar recoveries. Better retention of caffeine-d9 was expected on the XCW cartridge because of its high pKa value of 14. Compared to other cartridges, recoveries of caffeine-d9 on XCW was low but more consistent across the different types of water. Its lower recovery can be explained by the employed elution step that was recommended by the cartridge supplier for specific compounds (Alprenolol, Acetaminophen, and Clomipramine). The elution step could be optimized with this cartridge to obtain better recoveries for internal standards such as caffeine-d9 as well as for the total number of features retained.

An unexpected high signal was obtained for carbamazepine-d8 and sulfamethoxazole-d4 in the Marne River sample after SPE, leading to excessive recovery values (> 1000%), which were therefore not included in Figure 5. The extraction of those two components on WWe, led to signal suppression as compared to the Milli-Q sample. Such matrix effects were previously reported in wastewater effluent with ion suppression of 34% for sulfamethoxazole, 5.4% for caffeine, and 23% for carbamazepine on Strata-X cartridges (Lacey et al., 2008). Atenolol-d7 exhibited various behaviours, with a signal enhancement for HLB and XAW, signal suppression for ENV+, and both signal suppression (WWe sample) and signal enhancement (Marne River sample) for the Multilayer cartridge and XCW.

Matrix effects (*i.e.*, signal suppression or enhancement) or adsorption competition on the cartridge were observed for all detected internal standards. A high suppression was especially observed for the XAW cartridge, in accordance with the impact of adsorbed organic materials already described in part 2.1.2.

2.2.2. Range and Properties of Retained Features

The number of features retained on each cartridge was compared for the WWe and the Marne River sample (Table 3). As mentioned previously, the Multilayer cartridge displayed the highest number of features compared to the other cartridges, which explains its widespread use in non-target screening studies (Kern et al., 2009 ; Schymanski et al., 2014a ; Ruff et al., 2015). HLB and ENV+, commonly considered as “universal cartridges”, were also quite efficient for the Marne River sample, retaining 32% and 47% fewer features than the Multilayer cartridge, respectively. They also retained the highest number of features from the WWe sample, ENV+ being the most efficient. XCW retained 11% more features from the Marne River sample than ENV+, which suggests the presence of cationic substances in this sample. Table 3 displays the properties of the features retained on each cartridge for both samples.

Table 3. Properties of features retained on various SPE cartridges. The average retention is given as the percentage of acetonitrile in the mobile phase (ACN) needed to elute a given feature. The weighted values correspond to the average parameter, weighted by the intensity of each individual marker.

	Number of Features	Sum of Detected Areas	Average <i>m/z</i>	Weighted Average <i>m/z</i>	Average Retention (% ACN)	Weighted Average Retention (% ACN)
Marne River						
ENV	315	4.25×10 ⁶	306.6403	300.0973	43	41
HLB	403	4.20×10 ⁶	367.9942	379.1581	44	43
XAW	49	8.69×10 ⁵	295.4186	323.9237	52	52
XCW	350	3.43×10 ⁶	404.9278	366.7695	48	46
Multilayer	594	5.84×10 ⁶	470.6467	441.9003	46	44
WWe						
ENV	301	2.72×10 ⁷	306.7059	339.2997	37	33
HLB	201	1.75×10 ⁷	337.8578	311.8444	35	27
XAW	149	9.03×10 ⁷	345.6214	384.1570	34	22
XCW	100	8.32×10 ⁶	297.6594	367.0206	28	23
Multilayer	7515	4.89×10 ⁷	461.3586	416.1775	47	35

The Multilayer cartridge retained globally bigger molecules than other cartridges, as seen from both average m/z and weighted average m/z . For the two types of samples, HLB also retained features with larger molecular sizes and ENV+ smaller ones. When those values were not weighted by the feature area, no clear trend could be observed. The polarity of retained features (as described by the average retention) was also not significantly different between the cartridges. However, it can be noted that the Multilayer cartridge average retention of features from the Marne River sample (46% ACN) was equal to the mean of the average retention of all four other cartridges (*i.e.*, the four phases used in the Multilayer cartridge). For the WWe sample, the total area of the detected features followed the same ranking order as the total number of features (*i.e.*, ENV > HLB > XAW > XCW). However, this was not the case for the Marne River sample, for which ENV+ exhibited 15% less detected features compared to HLB but a slightly more intense total signal. This observation indicates that the molecules retained on ENV+ were more ionizable, therefore different from HLB, or better retained, as it was demonstrated by the difference in their fingerprints (Figure 3). The Multilayer cartridge showed the highest total intensity among all cartridges, which demonstrates its ability to retain a large number of compounds with a high signal. This property is clearly beneficial for NTS studies to obtain clean mass spectra and thus to allow easier identifications.

The distribution of m/z values was visualized for both types of the sample matrix and for each cartridge (Figure 6). To the authors' knowledge, this representation has not yet been used for the characterization of HRMS data, although it is useful for identifying differences between samples. Comparison of the cartridges shows that ENV+ was less effective in retaining molecules with $m/z > 600$, whereas HLB covered a larger range of m/z . It is interesting to note that the Multilayer cartridge combined the ranges of m/z distribution of ENV+, HLB, XAW, and XCW.

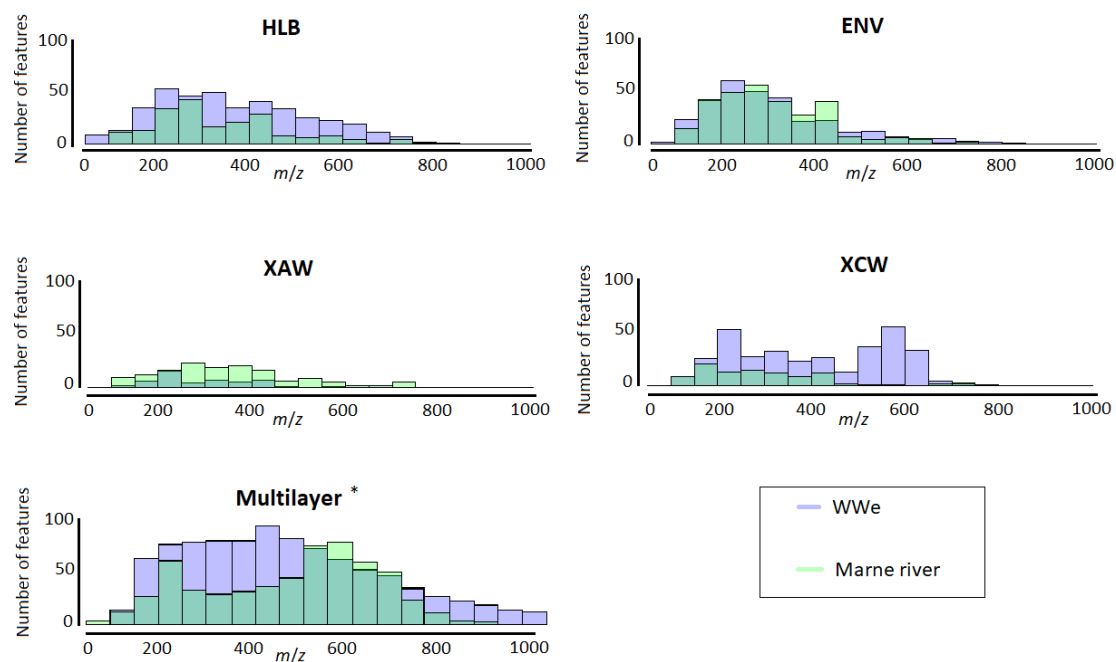


Figure 6. m/z distribution of the retained features on the different cartridges. * Number of features for Multilayer WWe were divided by eight.

2.3. Application: Evaluating a Disinfection Treatment by Performic Acid

The above-mentioned indicators showed that the Multilayer cartridge was the most efficient among the different tested cartridges and that it is the most suitable to be used as a universal cartridge. It was therefore tested in a real-case application to characterize the evolution of organic compounds during a disinfection treatment of wastewater effluents by performic acid (PFA). Raw wastewater effluent and the same effluent treated with 30 ppm of PFA (contact time of 10 min) were extracted with the Multilayer cartridge and analysed by HRMS (Figure 7).

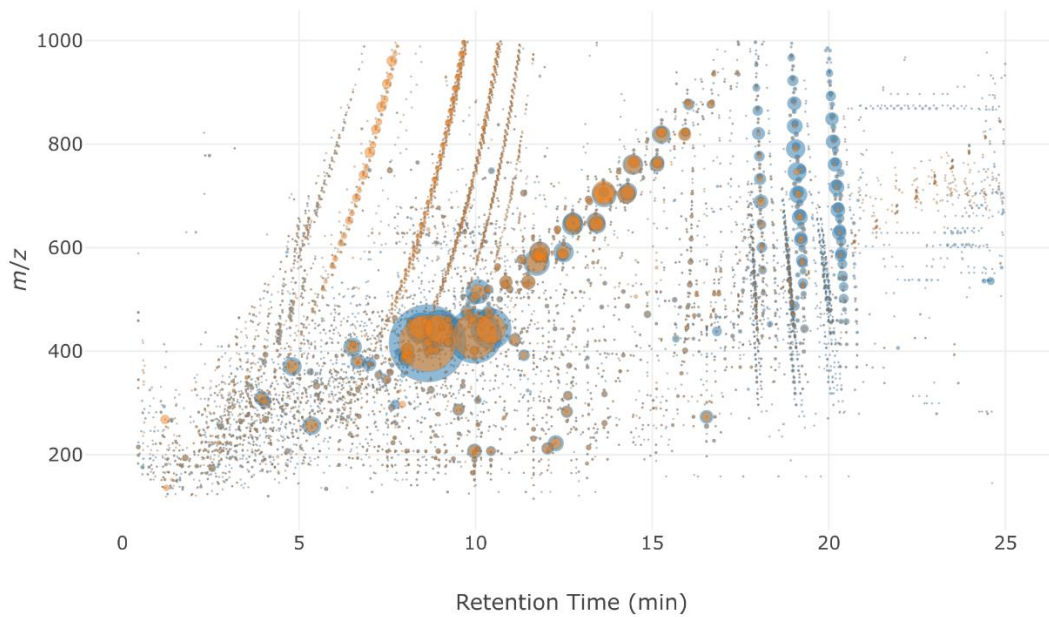


Figure 7. Bubble plot of WWe (blue) and WWe after 30 ppm PFA treatment during 10 min (orange) extracted on a Multilayer cartridge.

The Multilayer cartridge allowed the detection of a wide number of compounds covering the whole range of m/z values at low and high retention times. Sample after treatment showed a significant decrease in the area of the markers with a retention time between 17 and 21 min and an increase for the markers between 5 and 7 min compared to the non-treated wastewater.

This observation suggests that compounds with lower polarity (*i.e.*, higher retention time) were transformed into more polar ones by the treatment. Table 4 displays the distribution of the number of markers before and after treatment in different zones defined by the m/z and retention time ranges. Zone 2 (bigger and more polar molecules) exhibited a significant increase in both the number of markers and their total area. Zone 1 and zone 3 (molecules with $m/z < 500$) showed a decrease by almost a third of the number of markers after treatment, and a decrease in their total intensity as well. Finally, PFA treatment slightly increased the number of markers in zone 4 (bigger and less polar molecules), but the overall intensity of signals decreased in that zone. Globally, these results show that PFA treatment formed more polar molecules and removed a larger proportion of small molecules.

Table 4. Distribution of markers before and after PFA treatment.

	Before PFA		After 30 ppm PFA	
	Number of markers	Sum of detected areas	Number of markers	Sum of detected areas
Zone 1				
<i>m/z</i> : 0–500	4861	2.51×10^8	3781	1.85×10^8
Retention Time: 0–10 min				
Zone 2				
<i>m/z</i> : 500–1000	996	0.35×10^8	2301	1.36×10^8
Retention Time: 0–10 min				
Zone 3				
<i>m/z</i> : 0–500	2795	1.43×10^8	1728	0.75×10^8
Retention Time: 10–25 min				
Zone 4				
<i>m/z</i> : 500–1000	2770	2.93×10^7	3557	1.95×10^8
Retention Time: 10–25 min				

3. Materials and Methods

3.1. Chemicals and Standards

All standards were obtained from Sigma-Aldrich (Dr. Ehrenstorfer, Augsburg, Germany, purity > 99%). Methanol (MeOH), ammonia (35%), hydrochloric acid (37%), and formic acid (98%) were LCMS grade and purchased from Fischer Scientific (Illkirch Cedex, France). Ethyl acetate (EtAc), dichloromethane (DCM), and acetone were obtained from Merck (Darmstadt, Germany), and acetonitrile (ACN) was purchased from Biosolve (Dieuze, France), at analytical grade. HPLC water (Milli-Q) was produced from deionized water using a Millipore Milli-Q system (IQ 7000, Merck, Darmstadt, Germany) equipped with an LC-pak polisher (Merck, Darmstadt, Germany).

Deuterated standards were used to evaluate matrix effects from different samples, to evaluate the extraction efficiency on the various SPE phases, and to correct the retention time of the chromatograms. A mixed solution of 4 deuterated compounds (atenolol-d7, caffeine-d9, carbamazepine-d8, sulfamethoxazole-d4) was prepared in MeOH at a concentration of 10 mg/L for further injection in samples.

Before sampling and analysis, glassware was washed with TFD4 (Franklab, Montigny-le-Bretonneux, France), rinsed with deionized water, and calcined at 500 °C to re-move any trace of organic contamination.

3.2. Sample Collection and Preliminary Preparation

Various types of samples were used: Milli-Q water produced in the laboratory, sur-face water under anthropic pressure (Marne river, France), and wastewater effluent (WWe).

Thirty liters of water from the Marne River were collected from a bridge at Chennevières-sur-Marne (GPS coordinates: 48.79071188130754, 2.5215935793518462) down-stream the wastewater treatment plant of Marne Aval (Noisy-le-Grand, France). The sample was collected in 3 × 10 L amber glass bottles and filtered the same day, using 0.7 µm glass fibre filters (GF/F Whatman). After homogenization, aliquots of 500 mL were pre-pared and completed to 1 L with Milli-Q water in order to avoid potential matrix effects occurring during the SPE loading step. The initial pH and dissolved organic carbon (DOC) were 8.5 and 1.89 mg C/L, respectively. Some samples were acidified, either to pH 6.5 with 50 µL of formic acid or to pH 2–3 with 400 µL of formic acid according to recommendations for specific SPE cartridges.

Ten liters of treated effluent were collected in June 2019 from the wastewater treatment plant of Seine Amont (Valenton, France) in 10 L amber glass bottles and filtered on 0.7 µm glass fibre filters (GF/F Whatman). The initial pH and DOC were 7.9 and 7.1 mg C/L, respectively. One liter subsamples were acidified to pH 6.5 with 50 µL of formic acid.

3.3. SPE Cartridges

3.3.1. Selection of Cartridges

Strata-X cartridges (500 mg, 6 mL) (Phenomenex, Le Pecq, France) are ionic-based polymeric sorbents, chosen to retain compounds according to their pKa: Strata X-AW (for pKa < 5), Strata X-CW (for pKa > 8), Strata X-A (for pKa < 2), Strata X-C (for pKa > 10.5). Oasis HLB (200 mg, 6 mL) (Waters, Milford, MA, USA) is a reversed-phase cartridge used to select a wide range of substances, including neutral, basic, acidic, and polar ones. Sili-ca-based cartridges Strata C18 (200 mg, 6 mL) and Strata SDBL (500 mg, 6 mL) (Phenomenex, Le Pecq, France) are employed to retain neutral hydrophobic and non-polar molecules. ENV+ (500 mg, 6 mL) (Biotage, Glamorgan, UK) is a polymeric phase appropriate for the retention of polar analytes.

C18/ENV+ (400 mg, 6 mL) (Biotage, Glamorgan, UK) is a layered cartridge composed of the previous ENV+ phase as the bottom layer and a C18 phase as the top layer to improve the range of analytes possibly extractable. Finally, a homemade Multilayer cartridge made of Oasis HLB (200 mg), Isolute ENV+ (150 mg), Strata X-AW (100 mg), and Strata X-CW (100 mg) was prepared according to the method developed by (Kern et al., 2009).

3.3.2. Extraction Protocol

Samples were spiked with the mix of internal standards at 100 ng/L, 24 h prior to extraction and stored at 4 °C. Experimental conditions for conditioning, loading, and elution on each cartridge followed recommendations of suppliers, or the method developed by Kern et al. for the Multilayer cartridge (Kern et al., 2009), and are described in Table 5. All extractions were carried out on Visiprep (Sigma-Aldrich, Augsburg, Germany) and Autotrace (AT280, Caliper) SPE systems. All cartridges were loaded with 1 L of sample at pH 6–7, except for Strata X-A (pH 2–3 recommended) and Oasis HLB, for which both pHs were tested. Before elution, the cartridges were dried for 30 min. After elution, the extracts were stored in the dark at 4 °C prior to their analysis. SPE extracts were evaporated under a stream of nitrogen, then reconstituted in 1 mL of Milli-Q water and MeOH (80/20, v/v), and filtered through 0.2 µm PTFE filters before a 10 µL injection on the analytical system.

Table 5. Description of the methods used for water sample extraction on the different SPE cartridges.

Cartridge	Sample	Conditioning	Washing	Eluting
X-A	1 L pH 2–3	10 mL MeOH	10 mL MeOH	10 mL MeOH + 5% formic acid
	Marne	10 mL Milli-Q		
X-AW	1 L pH 6–7	10 mL MeOH	10 mL MeOH	5 mL MeOH + 5% ammonia
	Marne, MQ, WWe	10 mL Milli-Q		5 mL MeOH + 5% formic acid
X-C	1 L pH 6–7	10 mL MeOH	10 mL MeOH	5 mL MeOH + 0.1M HCl
	Marne	10 mL Milli-Q pH = 2	+ 0.1 M HCl	5 mL MeOH + 5% ammonia
X-CW	1 L pH 6–7	10 mL MeOH	10 mL MeOH	5 mL MeOH + 5% formic acid
	Marne, MQ, WWe	10 mL Milli-Q		5 mL MeOH + 5% ammonia
HLB	1 L pH 2–3	10 mL MeOH	No washing	5 mL MeOH
	Marne	10 mL AcEt		5 mL AcEt
	Marne, MQ, WWe	5 mL DCM		5 mL DCM
		10 mL Milli-Q		
ENV+	1 L pH 6–7	10 mL MeOH	5 mL Milli-Q	5 mL MeOH
	Marne, MQ, WWe	10 mL Milli-Q	/MeOH (95/5, v/v)	5 mL acetone + 5% ammonia
C18	1 L pH 6–7	10 mL MeOH	No washing	5 mL MeOH
	Marne, MQ, WWe	10 mL Milli-Q		5 mL acetone + 5% ammonia
C18/ENV+	1 L pH 6–7	10 mL MeOH	No washing	5 mL MeOH
	Marne	10 mL Milli-Q		5 mL acetone + 5% ammonia
SDBL	1 L pH 6–7	10 mL MeOH	No washing	5 mL MeOH
	Marne	10 mL Milli-Q pH = 4		5 mL acetone + 5% ammonia
Multilayer	1 L pH 6–7	10 mL MeOH	No washing	6 mL AcEt/MeOH (50/50, v/v)
	Marne, MQ, WWe	10 mL Milli-Q		+ 1.43% ammonia
				+ 1.7% formic acid

3.4. UV-Visible and 3D Fluorescence Analyses

Spectroscopic measurements were performed on samples before and after filtration on the SPE cartridges (loading phase). UV-Visible analyses were conducted on a spectrophotometer (UviLine 9400, Secomam, Aqualabo, Champigny-sur-Marne, France), and 3D-fluorescence spectra were obtained by a spectrofluorometer (FP-8300, Jasco, Pfungstadt, Germany). Those apparatuses were both equipped with a Xenon lamp and a 1 cm-quartz cell. For the fluorescence spectra, EEM were generated by scanning excitation wavelengths from 240 to 450 nm (every 5 nm), and the emission wavelengths were detected between 250 to 600 nm (every 2 nm), at a scan speed of 1000 nm/min and response at 0.1 s. To avoid any effect of the intern filter, samples with an absorbance at 254 nm higher than 0.1 were diluted.

3D-fluorescence indexes were chosen according to previous studies (Guillossou et al., 2021; Sgroi et al., 2017). Fluorescence regional integration (FRI) was performed, regions I to V were calculated and compared.

3.5. LC-HRMS Analyses

LC-HRMS analyses of the Marne River samples were conducted on SYNAPT HDMS QTOF (Waters) coupled with a Nano ACQUITY UPLC system (Waters). The column was an ACQUITY UPLC Peptide BEH C18, 130 Å (100 µm × 100 mm, 1.7 µm). The mobile phase was Milli-Q water with 0.1% formic acid (A) and ACN (B) following the gradient described in Supplementary Materials Table S1, for a total run time of 40 min. Analyses were performed in positive mode (ESI+) with screening between 50 and 1000 *m/z*. LC-HRMS analyses of Milli-Q and WWe samples were performed with a Vion–UPLC-IMS-QTOF (Waters) equipped with an ACQUITY UPLC BEH C18 (2.1 × 100 mm, 1.7 µm) column and the corresponding pre-column. The mobile phase was Milli-Q water + 0.1% formic acid (A) and ACN + 0.1% formic acid (B) following the gradient described in Supplementary Materials Table S1, for a total run time of 34 min. Analyses were performed in positive mode (ESI+) with screening between 50 and 1000 *m/z*. To ensure data quality, several steps were implemented (Supplementary Materials Text S1).

Raw HRMS data were exported directly after acquisition and converted in *.mzML format via *MSConvert* (Version 3) (Chambers et al., 2012; Kessner et al., 2008) for further treatment in R software (Version 3.6.2). Pre-treatment of raw data was performed with the *patRoan* package (Helmus et al., 2021), using *OpenMS* for peak picking and alignment. Features that were not

present in all replicates of a given sample were discarded. PCA was performed with the *FactoMineR* (Husson et al., 2020) and *Factoextra* (Kassambara and Mundt, 2020) packages, and Upset diagrams were plotted with the *UpSetR* package (Conway and Gehlenborg, 2019).

The extraction efficiency of an SPE cartridge was determined based on the area of detected internal standards in the samples compared to the area of a standard mix injected. The number of detected features and the total area of the signal (*i.e.*, sum of the areas of each feature) were compared for each cartridge. The average m/z and average retention of a sample were calculated from the m/z values and retention times of each feature detected in the sample. Retention was expressed as the percentage of ACN needed to elute a given feature at the corresponding retention time. Values of average m/z and average retention were also weighed by the area of each feature.

4. Conclusions

The goal of this work was to compare the efficiency of several commercially available SPE cartridges and of a homemade Multilayer cartridge (as developed by (Kern et al., 2009)) for non-target screening purposes. Various parameters were monitored, such as the recovery of a limited number of analytical standards, the number and properties of detected features in HRMS, or the global retention of DOM (through 3D fluorescence measurements) on each cartridge. The Multilayer cartridge was the most effective at retaining internal standards and a large range of HRMS features, and it was not as affected by the adsorption of DOM as other cartridges (*i.e.*, adsorption competition effects and/or matrix effects). It seemed to take advantage of the four phases involved in its composition (HLB, ENV+, XCW, and XAW) in terms of diversity of compounds (*i.e.*, polarity and molecular size).

Although the other cartridges had some specificity missed by this Multilayer cartridge (*e.g.*, an important number of specific features was only retained on the ENV+ cartridge), its efficiency in extracting a great number of compounds was clearly demonstrated. This observation could be further confirmed by using different analytical tools (*e.g.*, gas chromatography) or different ionization modes and sources to increase the range of molecules detected. Moreover, a further investigation involving the tentative identification of some retained features could be performed. The optimization of extraction conditions (*e.g.*, masses and combinations of sorbents, types, and volumes of eluting solvents) could also be

investigated to retain an even greater number of features. It would finally be useful to better characterize the chemical space (log Kow, pKa, chemical functions, etc.) covered by each cartridge. A larger number of internal standards covering a broader range of chemical properties could thus be studied. This could help determine the relevance of each type of cartridge towards specific purposes (*e.g.*, the detection of a larger number of polar or ionic compounds or some specific families of molecules).

5. Supplementary Materials

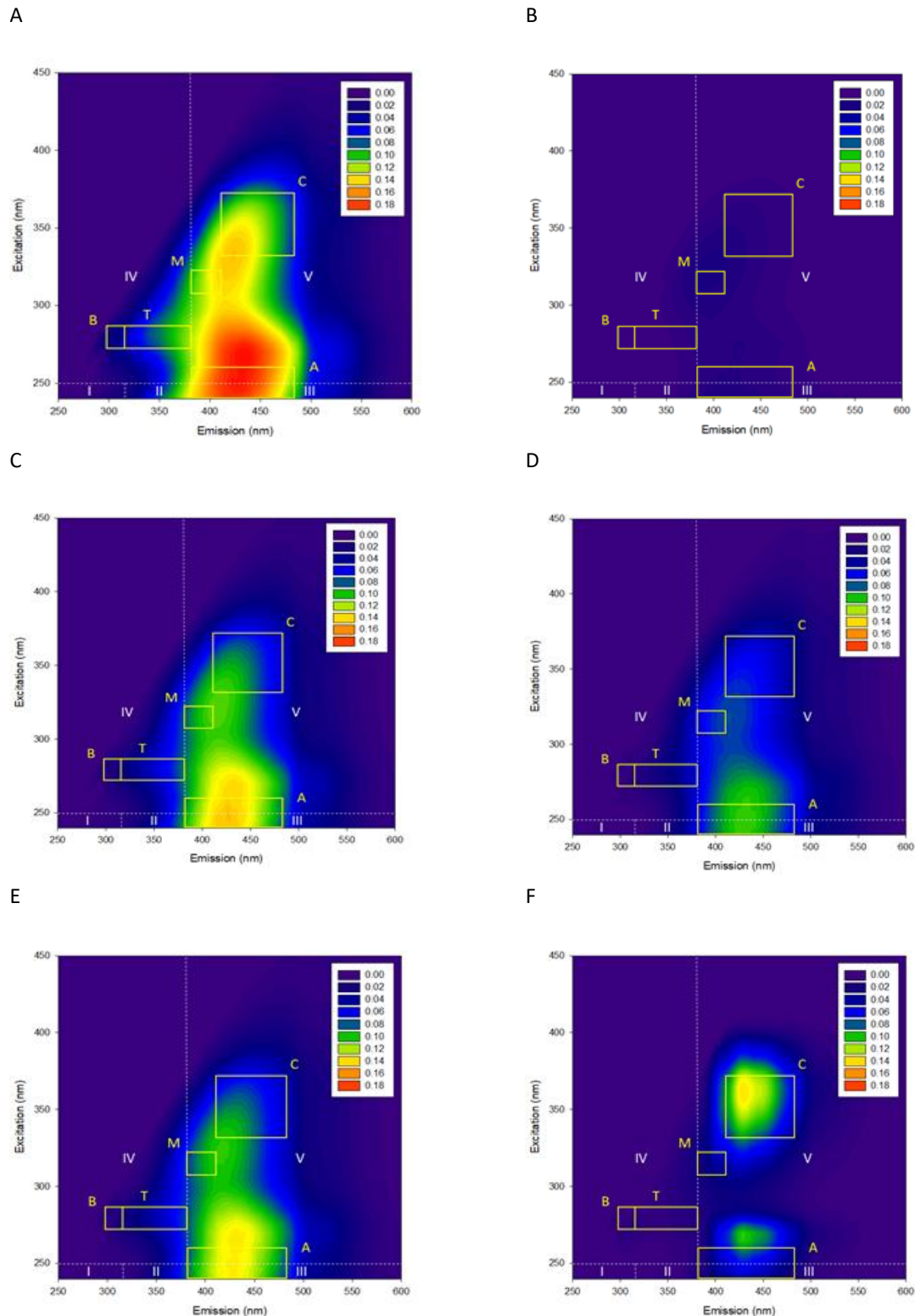


Figure S1: EEM matrices of Marne river (A) before extraction, (B) after extraction on X-A, (C) after extraction on HLB, (D) after extraction on C18/ENV+, (E) after extraction on SDBL, (F) after extraction on the multilayer cartridge

Table S1: Description of the gradient used for non-target analysis

Nano ACQUITY UPLC gradient			Vion-UPLC-IMS-QTOF gradient		
Time (min)	A (%)	B (%)	Time (min)	A (%)	B (%)
0	99	1	0	98	2
2	99	1	1	98	2
32	0	100	25	2	98
37	0	100	30	2	98
40	99	1	31.5	98	2
45	99	1	34	98	2

Text S1: Quality insurance procedure

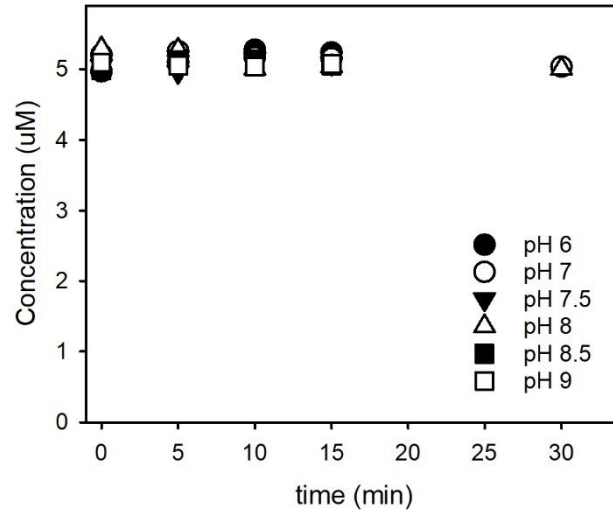
Before each analysis, a quality reference standard, consisting of 9 components (Acetaminophen, Caffeine, Leucine enkephalin, Reserpine, Sulfadimethoxine, Sulfaguanidine, Terfenadine, Val-tyr-val, Verapamil), was injected five times to check the system performance by calculating the mass error deviation, the average peak width and, when available, the CCS (collision cross section) error for those compounds. If the mass error deviation was higher than 2 ppm or the peak width was longer than 3.0 sec, a system calibration was conducted. The same procedure was repeated until the expected conditions were met. Each sampling sequence began with three blank injections to wash the column and five pool injections to stabilize the column. The pool sample consists in a mix of equal volume of each sample and is used as a quality control. Each sample was then injected in randomized triplicate to minimize the effect of instrumental deviation. For data treatment, only the features detected in every replicate of a given sample were considered. Every 10 injections, a pool was re-injected to monitor the system. The clustering of these pool samples in the PCA was checked to assess the reproducibility of the system.

Annexe 5 : Informations complémentaires sur l'étude de désinfection par le PFA

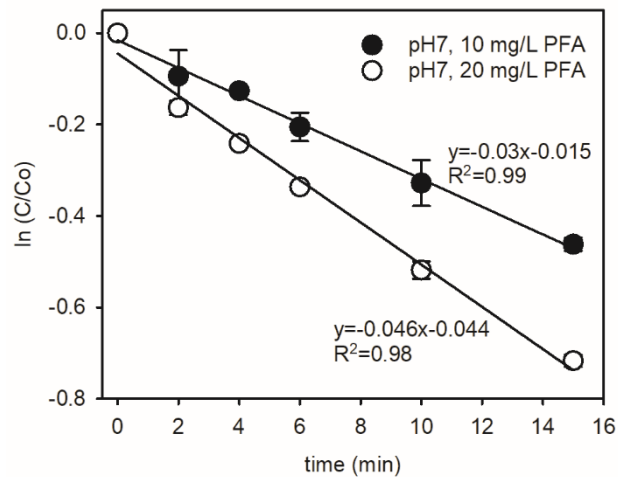
A - 5 marqueurs les plus discriminants des échantillons avant ou après désinfection (0.8 – 2.5 mg/L, 10 min)., repérés sur le S-plot obtenu après OPLS-DA

m/z	TR (min)	DT (ms)	Formule moléculaire	Identification	Niveau de confiance
Avant PFA					
230,1384	0,83	4,49	C12H20FNS	-	5
286,1479	4,30	5,31	C11H25Cl2N3O	-	4
382,1809	15,18	6,41	C23H24FNO3	-	5
246,1334	0,80	4,60	C11H19NO5	-	4
151,0211	4,67	3,17	C8H6OS	-	5
Après PFA					
246,1070	8,78	3,93	-	-	-
324,0456	7,99	5,14	C15H14ClNO3S	-	4
551,0302	7,00	6,99	-	-	5
251,1753	5,81	4,87	C14H22N2O2	lidocaine N-oxide	1
389,1531	7,73	6,22	C28H20O2	-	4

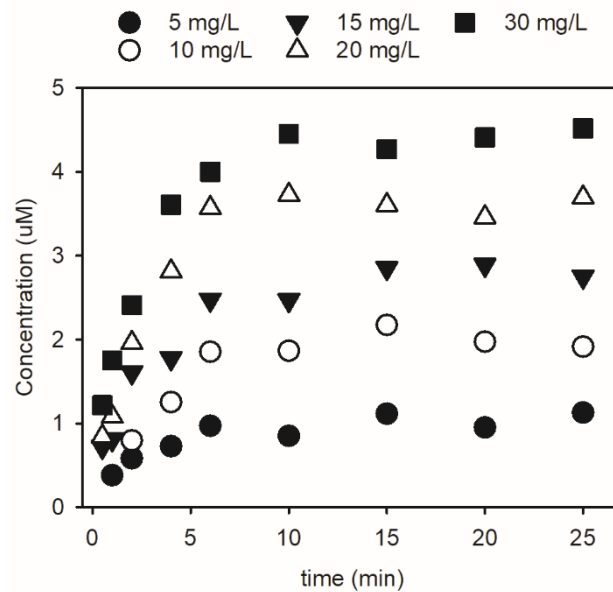
B – Effet de H₂O₂ (33 mg/L) sur la lidocaïne (5 µM) dans un tampon phosphate 10 mM (pH = 6,2-8) ou borate 10 mM (pH = 9)



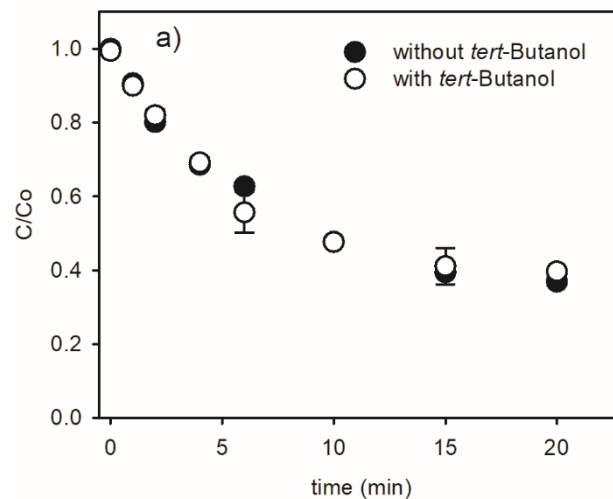
C – Cinétique de dégradation de la lidocaïne par le PFA (10 ou 20 mg/L de PFA pour 5 µM de lidocaïne, dans un tampon phosphate 50 Mm à pH = 7). Les barres d'erreur correspondent à l'écart-type entre les réplicats de manipulations



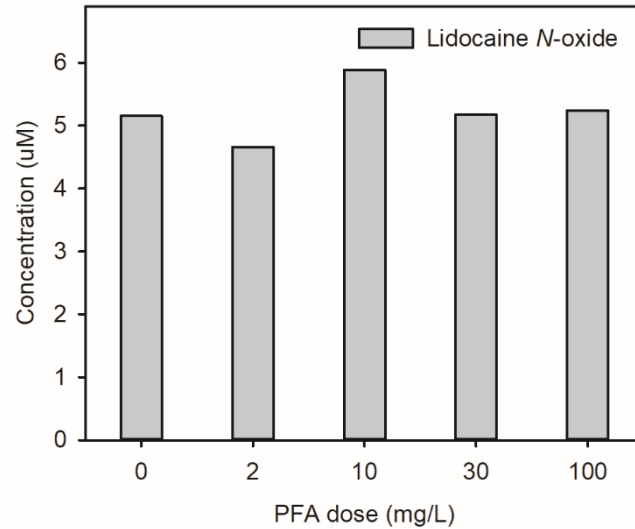
D – Formation de lidocaïne N-oxyde en fonction du temps (5-30 mg/L de PFA pour 5 μ M de lidocaïne dans un tampon phosphate à 50 Mm à pH =8).



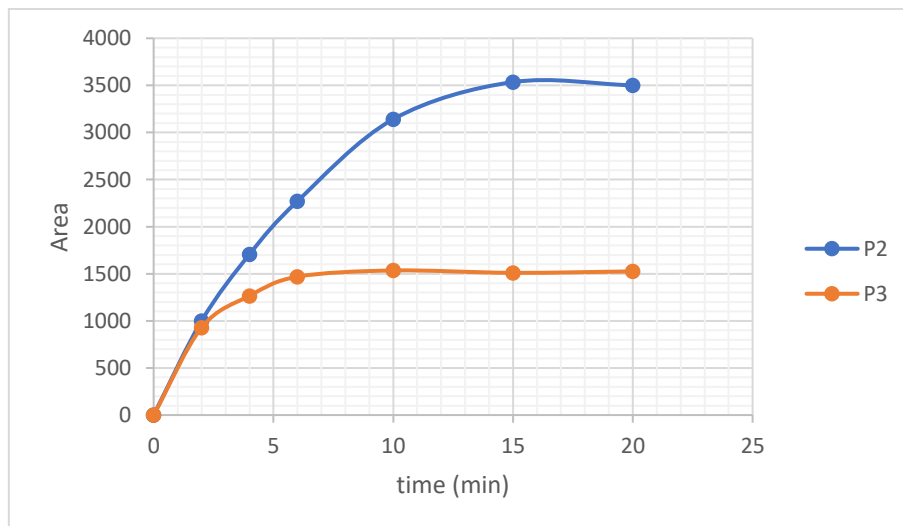
E – Influence de la présence de tert-Butanol (100 mM) sur la dégradation de la lidocaïne lors de l'oxydation par PFA (10 mg/L de PFA pour 5 μ M de lidocaïne dans un tampon phosphate à 50 Mm à pH =8). Les barres d'erreur correspondent à l'écart-type entre les réplicats de manipulations.



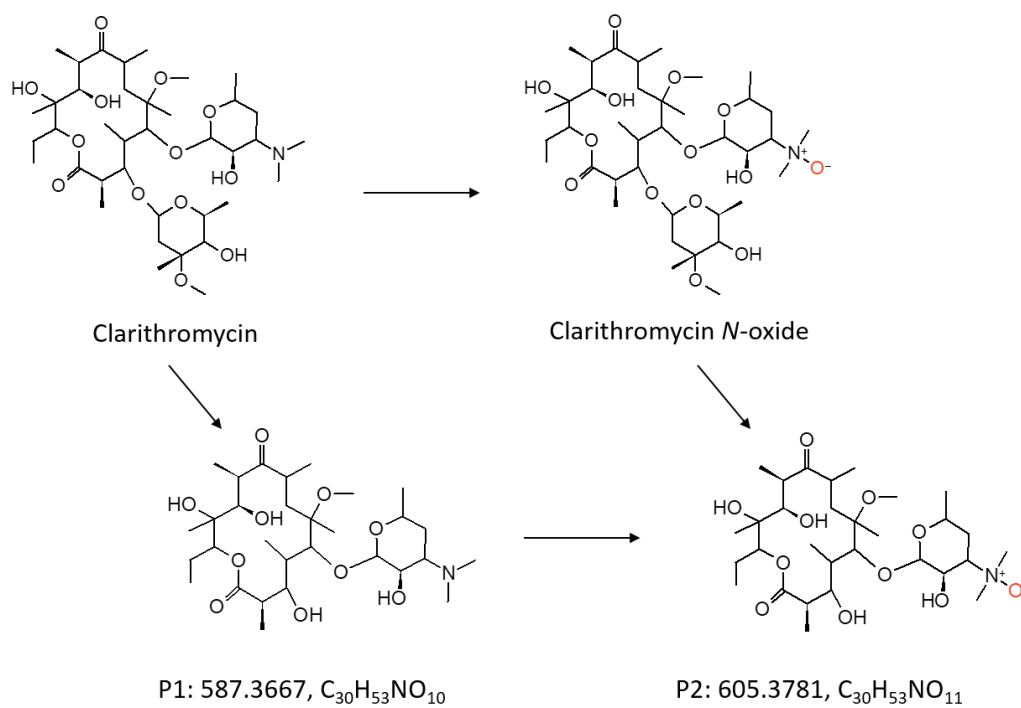
F – Stabilité de la lidocaïne N-oxyde (5 μM) en présence de PFA (0-100 mg/L pendant 10 min) dans un tampon phosphaté à 50 Mm à pH = 8.



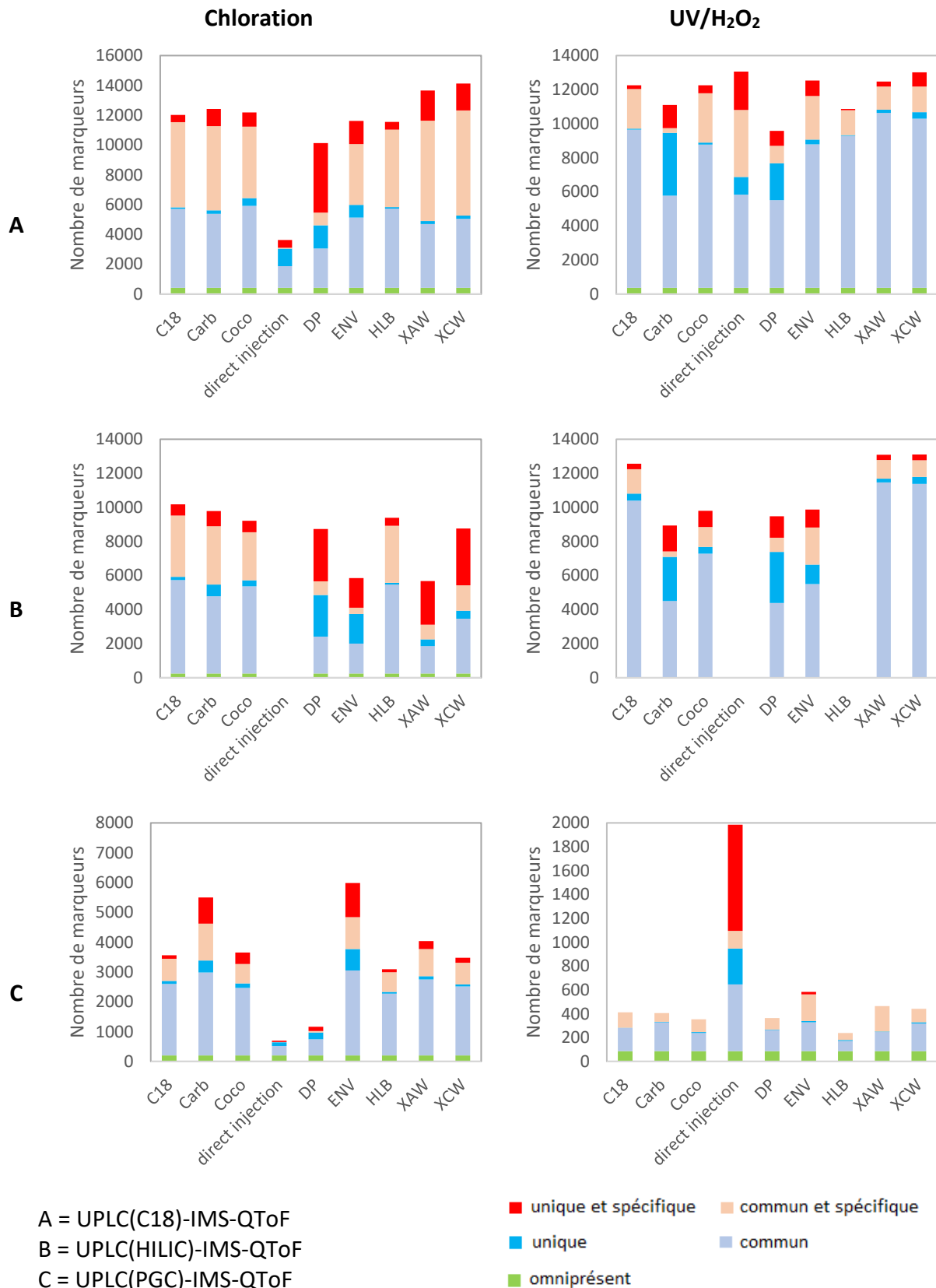
G – Evolution de l'aire des signaux de P2 et P3 en fonction du temps de réaction (10 mg/L PFA pour 5 μM de lidocaïne dans un tampon phosphate à 50 Mm à pH =8)



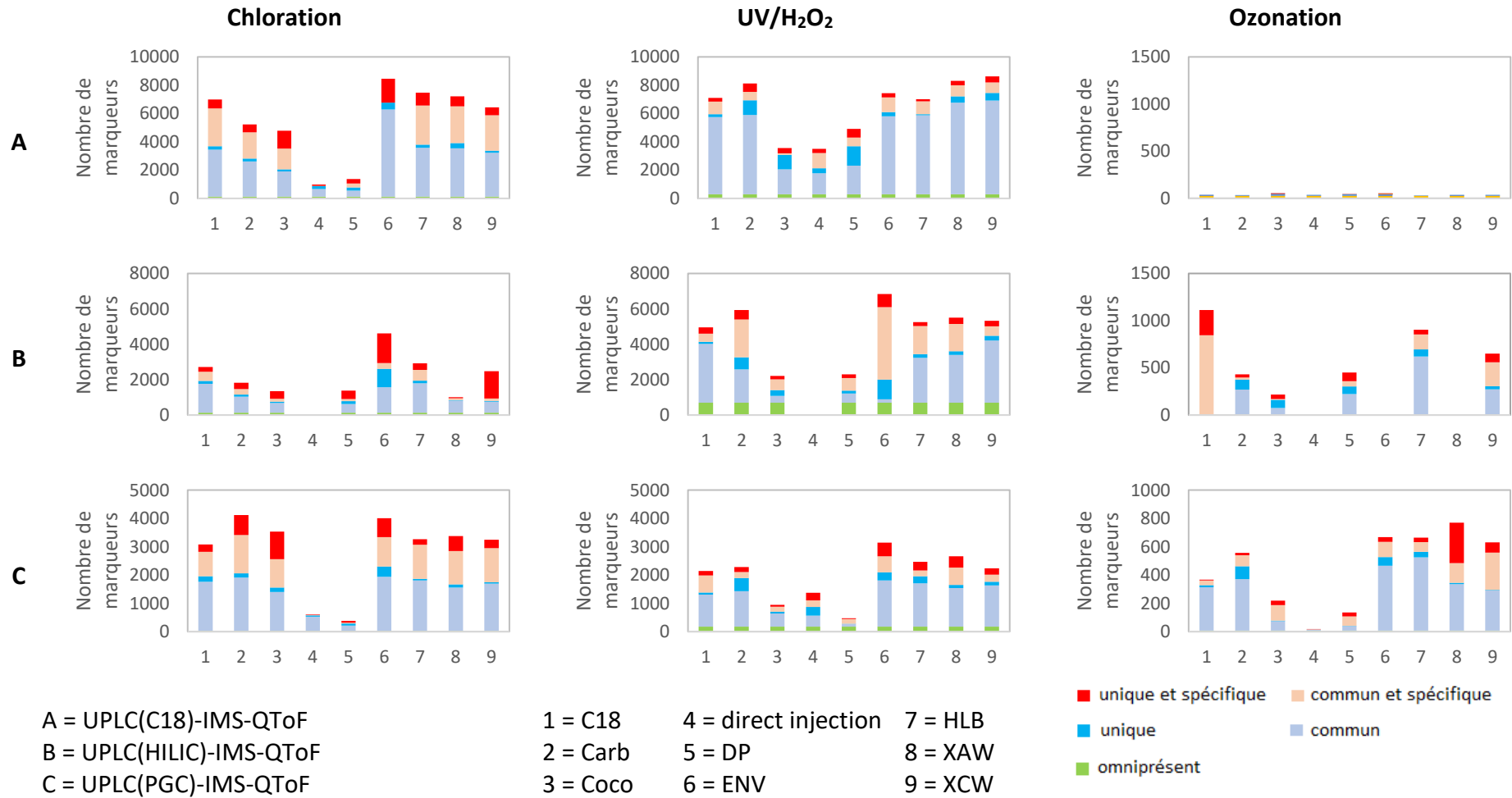
H – Mécanismes de transformation proposé pour la clarithromycine soumise à une oxydation par le PFA



Annexe 6 : Comparaison des méthodes de préparation pour l'analyse de TPs ESI+



Annexe 7 : Comparaison des méthodes de préparation pour l'analyse de TPs ESI-



Annexe 8 : Proportion de liens de biotransformation selon la méthode de préparation

A - Proportion de marqueurs *formés* (%) pouvant être expliqués par des mécanismes de transformation usuels pour l'UV/H₂O₂ en ESI- après séparation sur colonne C18, selon les différentes méthodes de préparation

	C18	Carb	Coco	DP	Direct injection	ENV	HLB	XAW	XCW
Chloration	0,00	0,00	0,26	0,00	0,00	0,00	0,00	0,00	0,00
Décarbonylation	0,00	0,47	1,55	0,00	0,00	0,00	0,00	0,51	0,48
Décarboxylation	0,00	0,47	1,04	0,10	0,00	0,00	0,00	0,00	0,00
Déchloration	0,00	0,16	0,26	0,00	0,00	0,00	0,00	0,00	0,00
Déméthylation	6,28	0,47	1,81	0,06	0,00	1,78	4,27	0,00	0,48
Désaturation	1,37	0,63	0,78	0,10	0,00	1,02	1,42	0,51	0,48
Déshydratation	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2 x Hydroxylation	0,00	0,32	0,78	0,06	0,34	0,00	0,36	0,00	0,00
Hydratation	0,00	0,00	0,26	0,03	0,00	0,00	0,00	0,00	0,00
Oxydation	1,37	0,47	0,00	0,19	0,00	0,00	0,71	0,00	0,48
Réduction	0,00	0,79	1,55	0,13	0,00	0,00	0,00	0,00	0,71
Total	9,02	3,79	8,29	0,67	0,34	2,80	6,76	1,02	2,61

Annexes

B - Proportion de marqueurs *formés* (%) pouvant être expliqués par des mécanismes de transformation usuels pour l'ozonation analysée en ESI+ après analyse par FT-ICR, selon les différentes méthodes de préparation

	C18	Carb	Coco	DP	Direct injection	ENV	HLB	XAW	XCW
Chloration	0,22	0,24	0,00	0,00	0,00	0,00	0,31	0,07	0,12
Décarbonylation	1,32	2,63	0,93	0,00	0,18	0,00	1,57	0,55	1,03
Décarboxylation	1,10	1,14	0,31	0,04	0,18	0,00	1,26	0,50	1,20
Déchloration	0,00	0,43	0,15	0,00	0,18	0,00	0,00	0,24	0,29
Déméthylation	2,21	3,04	0,77	0,00	1,11	0,00	2,51	1,14	0,69
Désaturation	1,10	2,80	0,85	0,00	0,00	0,00	1,88	0,74	0,98
Déshydratation	0,00	0,45	0,00	0,00	0,00	0,00	0,00	0,02	0,07
2 x Hydroxylation	0,44	2,03	1,54	0,00	0,00	0,00	1,10	0,50	0,59
Hydratation	0,00	0,56	0,00	0,00	0,00	0,00	0,00	0,12	0,10
Oxydation	1,10	3,36	1,31	0,00	0,18	0,07	1,88	1,33	0,81
Réduction	1,55	2,39	0,46	0,04	0,74	0,00	1,88	1,14	0,93
Total	9,05	19,07	6,33	0,08	2,58	0,07	12,40	6,36	6,80