Avancée dans le domaine du monitoring et de l'analyse de suspects et d'approche non ciblée par spectrométrie de masse à haute résolution et mobilité ionique.

Fan Yang

Advancement in Non-Target Analysis of Environmental Contaminants Using High-Resolution Mass Spectrometry and Ion Mobility Spectrometry

Fan YANG

Reviewers:

Pr. Gauthier Eppe	University of Liège
Dr. Sonia Cantel	Université de Montpellier

Jury:

Pr. Didier Begue	IPREM (President)
Pr. Olatz Zuloaga	Universitat del Pais Vasco

Supervisor: Dr.Hugues PREUD'HOMME

Institut des Sciences Analytiques et de Physico-Chimie pour l'Environnement et les Matériaux
IPREM
Pau, France
December, 2022

Acknowledgements

I would like to express my sincere gratitude to

Pr. Gauthier Eppe from University of Liège

Dr. Sonia Cantel from Faculté des Sciences Université de Montpellier

for their time in reading and evaluating this manuscript.

I also want to thank

Pr. Didier Begue from IPREM

Pr. Olatz Zuloaga from Universitat del Pais Vasco

Mme Laurence Sarthou from Laboratoires des Pyrénées et des Landes

for their time in attending my defense.

This research project is supervised by Dr. Hugues Preud'homme. I would like to thank Hugues for giving me this opportunity to participate in this exciting research project and for guiding me in the field of academic research. Thank you for your support and understanding when we faced difficulties during our work. Thank you for providing the opportunities for international conferences and exchange.

I also want to thank Laboratoires des Pyrénées et des Landes and SETASAR PhD Project from Région Nouvelle Aquitaine for their financial support of this project. Thank Mme Laurence Sarthou from Laboratoires des Pyrénées et des Landes for participating to my research work.

I want thank to technique support team from Bruker Daltonics, Alexandre Collgross and Yann Hebert for your continuous understanding and open discussions relative to timsTOF and software.

I want to thank the researchers from IFREMER for sharing their knowledge and providing us samples. I want to thank Dr.Randolph Singh for his patience and inspiring ideas. Thank you for sharing your experiences and showing me the novel data analysis tool. I also want to thank Dr.Yann Aminont and Prof.Catherine Munschy for patiently evaluating the abstracts and paper.

I want to thank Dr. Saer Samanipour from the University of Amsterdam for welcoming me to your team and introducing me to the field of Data Science and Machine Learning. I also want to thank Denice van Herwerden for sharing your knowledge and supporting me in coding and modeling.

I also want to thank Mariam, Soumya and Rim for all the discussions and exchanges in our work and personal lives. Nobody else can understand better than you the difficulties and challenges I met during my thesis. Thank you for listening and sharing all the interesting and nonsense stories during these three years.

I want to thank my master interns, Anaëlle and Chloé. Your curiosity and enthusiasm inspired me and helped me think more deeply about the work.

I want to thank the people I met in IPREM and at the University of Amsterdam who make my work and life more pleasant.

I want to thank my friends for discussing our common interests since childhood and afford to be stupid with you. Finally, I want to express my infinite gratitude to my family for always supporting and trusting me unconditionally. Thanks, my cat, Java, for being a cat.

Résumé

Introduction

En raison de l'industrialisation mondiale et des activités anthropiques, des millions de tonnes de produits chimiques sont consommés chaque année par l'union européenne (UE). Ces mêmes produits chimiques peuvent être à la fois rejetés dans notre environnement et à l'origine de la contamination de nos ressources. C'est pourquoi, l'exposition aux substances chimiques est un sujet critique de nos jours et qui reste intimement lié aux questions scientifiques en relation avec la biodiversité et la santé humaine. La croissance constante des activités humaines, augmente la probabilité pour l'environnement et le corps humain d'être exposé à un cocktail de composés chimiques et leurs produits de dégradation d'origine synthétique. Le concept d'*Exposome* a été introduit par Christopher Paul Wild en 2005, il définit la nature et les effets toxiques de l'exposome chimique sur la santé humaine, que cela puisse être par l'ingestion (air, nourriture et boisson) ou par le biais de processus biologiques naturels. Les méthodes d'analyses ciblées sont couramment utilisées pour rechercher de manière quantitative des contaminants connus, issus d'une liste finie et préalablement définie. La plupart d'entre eux font déjà l'objet d'une surveillance systématique et sont intégrés dans les réglementations nationales, les directives européennes ou les conventions internationales. C'est le cas par exemple de la famille des PCBs, ou de certains pesticides interdits (Atrazine, Simazine, etc,.). Bien que ces approches permettent une très grande sélectivité en parallèle d'une limite de détection basse, elles ne permettent pas de détecter des molécules hors liste et par conséquent les composés émergents (CECs).

Des approches non-ciblées sont donc actuellement en phase de développement et de normalisation de manière à pouvoir caractériser des contaminants, sans avoir connaissance au préalable de leur existence ou de leur simple présence dans l'échantillon. Avec les avancées récentes et les innovations réalisées dans le domaine de l'instrumentation moderne, la recherche sur la contamination globale et les approches non ciblées sont en pleine croissance et florissantes. La chromatographie en phase gazeuse ou la chromatographie liquide couplée à la spectrométrie de masse à haute résolution (GC/LC-HRMS) en association avec la mobilité ionique est une approche méthodologique émergeante et plutôt appliquée à des études peptidomique, lipidomique ou méta-bolomique, plutôt que pour des problématiques scientifiques en lien avec le domaine de la qualité alimentaire ou environnementale et des contaminants. De plus, seul le développement récent de méthodes analytiques et des outils informatiques associés (e.g., *Machine Learning*) ont permis d'accélérer la faisabilité de traitement de données massives et l'efficacité de l'analyse non-ciblée.

Dans un premier temps, une recherche bibliographique s'est montrée indispensable de manière à comprendre les principes et l'état de l'art des analyses non-ciblées. Une fois les étapes critiques, les points clefs et les difficultés identifiées dans le cadre d'approches non-ciblées, de premiers travaux de recherche sont présentés ici et abordent le développement en synergie de diverses méthodes ou stratégies analytiques. Elles associent la chromatographie, la mobilité ionique et la spectrométrie de masse haute résolution qui sont des pré-requis nécessaires à l'identification et l'évaluation des contaminants émergents dans diverses matrices, tout en permettant de répondre aux problématiques actuelles, de l'exposition aux contaminants émergeants dans notre environnement.

Résumé des travaux

La première partie de ce travail de recherche (Le chapitre II) est consacrée à l'optimisation d'une méthode en GC-APCI-IMS-HRMS pour caractériser des *Polluants Organiques Persistants* (POPs) notamment halogénés. La priorité a été donnée à la détection et au suivi de POPs chlorés et bromés et dont une des caractéristiques est de présenter une empreinte isotopique particulière. Parmi ces derniers, ce sont donc, des congénères de la famille des PCBs et PBDEs qui ont été utilisés ici, afin de développer une nouvelle méthode de criblage de ces familles de composés possédant à la fois plusieurs degrés de substitution et de nombreux isomères de position. Les principaux objectifs, au nombre de quatre, sont: 1) optimiser les conditions d'ionisation par APCI et réduire l'existence de déhalogénation en source; 2) introduire une dimension de séparation et de sélectivité orthogonale à l'aide de la mobilité ionique et identifier le gain relatif à la capacité de séparer des isomères de position, d'obtenir un spectre HRMS ou HRMSMS pur tout en facilitant l'interprétation; 3) établir une nouvelle base de données incluant à la fois des temps de rétention, des masse exactes, des profils isotopiques vrais et des critères de mobilité ou de section efficace (1/K0 ou CCS); 4) mettre en œuvre cette approche sur des échantillons réels en partenariat avec l' IFREMER.

Avec une source d'ionisation chimique à pression atmosphérique (APCI), on observe que les ions générés sont principalement des ions moléculaires sous forme radical+ à la fois pour les PCBs et les PBDEs, ce qui permet et facilite grandement l'élucidation structurelle d'inconnus. On observe assez eu d'adduits ou de fragments pour ces structures. Un bénéfice supplémentaire, l'IMS a permis aussi d'améliorer la séparation des isomères de position et la qualité du spectre de masse (HRMS, justesse spectrale et HRMSMS purs). La valeur de CCS expérimentale, propre à chaque conformère, y compris isobariques, donne un nouvel aspect et l'accès à un nouveau critère de sélectivité orthogonal. Cela permet aussi d'associer plus facilement et de manière univoque, l'ion parent à ses ions de fragmentation (y compris en BroadBandCID). Cela facilite d'autant l'interprétation du spectre de masse. L'hexachlorobiphényle tel que le PCB-149 a été fréquemment détecté dans les échantillons, et de nombreux faux positifs ont ainsi pu être éliminés après une revue critique des spectres obtenus et l'utilisation des valeurs de CCS comme critère de sélectivité supplémentaire. À l'heure actuelle, et en raison de la taille des données massives, les premiers résultats obtenus ont seulement pu être traités partiellement. Nous nous sommes focalisés sur des screening ciblés de ces deux grandes familles

de composés halogénés, mais une analyse non-ciblée reste néanmoins accessible via cette méthode et les données brutes pourraient être retraitées *in-silico* (à postériori) de manière non ciblée.

La seconde partie de ce travail de recherche (Le chapitre III) concerne la problématique de la détection de contaminants organiques dans des eaux naturelles (de surface) et des eaux usées. Les différentes étapes du protocole analytique requis pour une approche non-ciblée dans ce domaine particulier y ont été étudiées. La préparation des échantillons est réduite à une simple filtration et la dilution afin de préserver autant qu'il soit possible la représentativité de l'échantillon dans le cas d'analyses non ciblées. La méthode de la séparation en chromatographie a quand à elle était optimisée pour une large gamme de type de composés (allant des plus polaires ou hydrophiles au plus hydrophobes). La détection par spectrométrie de masse haute résolution est réalisée par la technologie QTOF avec en parallèle le mode *full scan* et le mode *data-independent acquisition* (DIA). Les bases des données commerciales de Bruker Daltonics ont été modifiées et enrichies par les données expérimentales acquises lors de ces développements et à partir d'échantillons réels et de nombreux standards chimiques. De nouveaux composés ont ainsi pu être ajoutés à cette base des données. La méthode analytique a quant à elle était optimisée et rendue générique; les gammes de masse ont été fixées de 100 à 1250 Da en full scan et de 80 à 1000 Da en DIA. Le traitement de données massives et son analyse critique reste la partie la plus délicate et laborieuse de l'ensemble de la procédure. Le logiciel TASQ développé par Bruker Daltonic a été utilisé pour la partie du screening ciblé. L'outil open-source MS-DIAL a quand à lui, été choisi et utilisé en seconde passe afin d'identifier les substances non-ciblées lors du premier criblage réalisé avec TASQ. Nous avons pour se faire utiliser des bases des données complémentaires disponible en ligne, (e.g., MassBank Europe) afin de faciliter le traitement de données et l'identification de composés non ciblés. Les conditions de traitement ont elles aussi étaient optimisées avec une précision en masse exacte pour l'identification des candidats fixée à $\leq$5ppm. MS-DIAL permets aussi de gérer et de filtrer les signaux, de manière à minimiser le traitement ou éviter les redondances, avec par exemple, la soustraction du blanc, la comparaison binaire, etc,. Cela facilite le processus d'identification des inconnus (*unknown-unknown*). Pour conclure, 65 substances supplémentaires ont pu être identifiés par ce biais, avec à minima un indice de confiance de niveau 3 selon la règle de Schymanski. Par contre, cela reste essentiellement qualitatif ou semi-quantitatif, et il est encore difficile de déterminer leurs structures complètes de manière univoque sans confirmation préalable à l'aide de l'injection du standard chimique correspondant. Dans un futur très proche, la spectrométrie de masse à haute résolution associée à la mobilité ionique avec l'introduction d'un nouveau critère de sélectivité orthogonal, devrait devenir approche de choix pour le screening environnemental tout en facilitant l'identification et la confirmation des suspects. Une approche similaire, en utilisant le mode d"ionisation négative pourrait également compléter cette approche méthodologique globale en permettant d'élargir le champs des familles chimiques couvertes.

La troisième partie (Le chapitre IV) se focalise sur le développement d'une méthode quantitative par dilution isotopique pour le suivi des hormones stéroïdiennes chez les alevins de poissons. Ce développement est en lien avec une problématique scientifique propre au déterminisme sexuel chez les alevins et notamment du *Bar*, en

lien avec le changement climatique et le stress induit. Le cortisol et la cortisone sont connues comme marqueurs du stress. Le développement de cette nouvelle approche analytique est basée sur une analyse directe par LC-ESI-IMS-HRMS. En raison de la petite taille d'échantillon (1 individu), l'injection directe après extraction de chacun des individus et dopage à l'aide de standards chimiques marqués (C13) est requise. Mais la complexité de la matrice, doublée d'un besoin de sensibilité font parties des challenges à résoudre. La mobilité ionique est introduite ici pour résoudre ce problème, en minimisant la complexité de l'échantillon et l'effet de matrice. Les premiers résultats le confirment dès l'usage de la solution de standard natifs et marqués (C13) à faible concentration (0.2 ng/mL). Une amélioration significative du rapport signal sur bruit a été observée dès usage des CCS comme critère de sélectivité orthogonale supplémentaire. On remarque aussi une compatibilité complète avec une quantification absolue par dilution isotopique spécifique. Cette méthode est en cours de validation, les derniers tests sur une matrice de substitution et une première cohorte d'échantillons réels doivent être menés avant l'application sur la totalité de l'étude.

Pour finir, le dernier chapitre (Le chapitre V), est quand à lui consacré à la genèse d'un outil de prédiction de CCS qui utilise le Machine Learning (ML). C'est l'algorithme *Random Forest* qui a été choisi pour apprendre la corrélation non-linéaire entre les CCS et ce que l'on dénomme les *molecular fingerprints*. Ces derniers sont en fait une manière élégante et simplifiée, pour encoder des structures moléculaire 2D en chaînes de caractères binaires (1024 bits). L'ensemble des données a été collecté à partir de différentes études basées sur différentes techniques IMS, visant à couvrir différentes classes de produits chimiques et d'instruments tout en limitant les biais. Deux approches de prédiction ont été réalisées : un premier modèle de prédiction basé sur les classes et un second modèle de prédiction directe. Les deux approches ont donné une bonne précision de prédiction. L'écart de prédiction a pu être estimé en MRE (Moyen relative error) entre 1,89% et 2,33%. Des écarts plus importants ont été observés quand l'outil développé était confronté à des macromolécules où à l'opposé à de très petites molécules (hors champ de cette étude et de nos composés d'intérêts). On peut l'expliquer par le manque de données statistiquement suffisante pour couvrir ces 2 catégories et permettre un apprentissage efficace. Comme c'est souvent le cas pour l'approche d'apprentissage automatique (ML), davantage de cas d'apprentissage peuvent être ajoutés pour améliorer les performances de la prédiction et le domaine d'usage de cette approche prédictive qui a su démontrer ses preuves et son intérêt dans nos applications ici.

Conclusion et perspectives

Divers projets ont été développés et discutés dans ce manuscrit, présentant à la fois les progrès et les défis des approches d'analyses non-ciblées sous différents points de vue et avec ou sans apport de la mobilité ionique. Au cours de ce travail de recherche, le principal objectif scientifique était d'évaluer l'avancée et les domaines d'usage de l'instrument analytique, en particulier la faisabilité, les avantages de l'utilisation de l'IMS sur le traitement des données et l'interprétation de spectre de masse afin de répondre aux problématiques de l'analyse non-ciblée. Des méthodes analytiques génériques ont été réalisées en GC-APCI(+)-IMS-HRMS et LC-ESI(+)-

IMS-HRMS. Les processus d'identification automatique des suspects et une base de données interne améliorée ont été créés dans TASQ (Bruker Daltonics). Cette dernière, pour la partie réalisée avec un mode d'introduction par GC contenait le temps de rétention (RT), la masse exacte, les profils isotopiques et les valeurs expérimentales de CCS pour 118 POPs Chlorés, Bromés. La base de données analyse par U(H)PLC contenait RT, masse exacte, spectre de masse HNRMS et HR MS / MS (mode DIA), et les profils isotopiques expérimentaux avec plus 559 produits chimiques, sans oublier la base de données constructeur (Bruker Daltonics), renfermant déjà plus 3000 entrées (sans CCS). Les valeur CCS ont pu être implémentées dans les bases de données modifiées soit via l'ajout d'une valeur expérimentale ou d'une valeur prédite de manière à améliorer l'indice de confiance pour l'identification du composé et d'offrir un critère de sélectivité supplémentaire. En terme de perspectives, toutes les méthodes et stratégies présentées tout au long de ce manuscrit et qui sont toutes décrites ici pour la première fois, peuvent et doivent être encore améliorées ou enrichies. Il s'agissait dans un premier d'établir la faisabilité et la démonstration des avantages inhérents à cette approche originale, notamment dans le cas d'études non ciblées dans le domaine de l'environnement, de l'eau et de matrices complexes ou pour toutes les travaux de recherche propres à l'émergence, au devenir et à la transformation des contaminants. Les méthodologies développées et présentées ici étaient essentiellement basées sur l'usage d'un mode d'ionisation positif. Par contre pour des contaminants émergents très hydrophiles, tels que les PFAS, les acides haloacétiques et produits secondaires, une approche similaire en mode d'ionisation négatif est nécessaire. Des essais préliminaires encourageants ont pu être menés. Pour des produits actifs ou pharmaceutiques tels que les agents de contraste (IRM) sur base de Gadolinium, une pré-étude prometteuse, en synergie avec UHPLC-ICPMSMS a pu être mise au point. Enfin, les bases de données IMS et CCS qui n'ont pas encore été appliquées pour l'analyse des eaux (échantillons réels) pourrait l'être facilement y compris de manière retrospective. La prédiction de CCS pourrait également être intégrée au cours du processus d'identification d'inconnus. Les valeurs de CCS expérimentales et prédites peuvent en effet, réduire de manière drastique, le nombre de candidats potentiels, et augmenter l'indice de confiance dans l'identification de ces derniers. D'autres exemples d'applications peuvent être envisagées dans le cadre de stratégies optimisées et présentées dans ce travail de recherche (PhD) en cas de screening GC/LC-IMS-HRMS sur des matrices de divers origines ou lors de la nécessité de lier des approches métabolomiques et l'étude de contaminants. Une approche théorie – expérience en association avec des calculs théoriques semi-empiriques pourrait démontrer son intérêt à mener de front des calculs prédictifs par ML des CCS en corollaire des mesures expérimentales.

Outline

Due to global industrialization and anthropic activities, millions of tonnes of chemicals were consumed per year in the European Union member states (EU). Chemicals can release into the environment by various resources. Thus, this research project aims to develop non-target analysis workflows for a large scale of CECs by using gas/liquid chromatography coupling with HRMS and IMS. An in-house database was built with accurate mass, isotopic pattern, fragment ions, and CCS values. This manuscript presents and discusses the major advancements in diverse aspects of NTA.

- Chapter 1: Bibliography research: The 1st chapter describes the workflow of non-target analysis. Fundamental concepts and the state-of-the-art of the HRMS and IMS are discussed.
- Chapter 2: Presents the development and application of the non-target analysis method using GC-APCI-IMS-HRMS. The method was optimized by PCB congeners and PBDE congeners. The benefits of combining APCI with IMS for Cl/Br chemicals are highlighted.
- Chapter 3: Details the development of a non-target analysis method based on UHPLC-HRMS in water analysis. Data processing strategies and workflow are described.
- Chapter 4: Steroids hormones analysis and cortisol quantification. It highlights the benefit of IMS to enhance the S/N ratio in low concentration standards.
- Chapter 5: A CCS prediction tool was developed by machine learning. The molecular fingerprint was first time used to describe chemical structure. Two modeling strategies are compared to evaluate the impact of chemical classes to prediction accuracy.
- Chapter 6: Finally, the milestones and limitations of presence research are emphasized.

Table of Contents

Acknowledgements i

Outline viii

List of Figures xiii

List of Tables xvi

1 Introduction 1

1.1 Contaminants of emerging concern (CECs) . . . 3

1.1.1 Resources and regulations of CECs . . . 4

1.1.2 Challenging CECs analysis . . . 5

1.2 Concept approaches for CEC analysis . . . 7

1.2.1 Target screening . . . 7

1.2.2 Suspect screening . . . 8

1.2.3 Non-target screening . . . 9

1.2.4 Five levels of identification confidence . . . 9

1.3 Analysis approaches of CECs . . . 10

1.3.1 Separation technique . . . 11

1.3.2 Ionization technique . . . 13

1.3.3 Mass detector . . . 14

1.3.4 Acquisition mode . . . 20

1.4 Non-target analysis strategies . . . 21

1.4.1 Data preprocessing . . . 21

1.4.2 Prioritization . . . 21

1.4.3 Structure elucidation and identification . . . 22

1.4.4 Nontarget analysis data processing software . . . 22

1.5 Ion mobility spectrometry (IMS) . . . 23
1.5.1 Principle of IMS . . . 25
1.5.2 Most commonly used IMS separation technology . . . 26
1.5.3 Use of IMS in the NTS . . . 28
1.6 Conclusion . . . 30
References . . . 31

2 Improving Halogenated POPs Analysis in Real Samples Using GC-APCI-IMS-HRMS 42
2.1 State of the art . . . 44
2.2 Objectives . . . 44
2.3 Improving Halogenated POPs Analysis in Real Samples Using GC-APCI-IMS-HRMS . . . 46
References . . . 69

3 Non-Target Analysis With UPLC-timsTOF For Water Samples 71
3.1 State-of-the-art method . . . 73
3.2 Chemicals and standard preparation . . . 74
3.2.1 Chemicals and solvents . . . 74
3.2.2 Solution preparation . . . 74
3.2.3 Standard mixtures and calibration preparation . . . 75
3.3 LC-HRMS setting and optimization . . . 75
3.3.1 QA/QC mix of standards . . . 75
3.3.2 LC separation settings . . . 76
3.3.3 HRMS settings . . . 77
3.4 Data processing . . . 80
3.4.1 Suspect screening . . . 80
3.4.2 Non-target screening . . . 82
3.5 Applications in wastewater . . . 82
3.5.1 Sample preparation . . . 82
3.5.2 Results and discussion . . . 83
3.5.3 Discussion . . . 88
3.6 Preliminary application in natural waters . . . 88
3.6.1 Sampling area and sample analysis . . . 88

3.6.2 Data processing . . . 89
3.6.3 Results and discussion . . . 90
3.7 Conclusion and perspectives . . . 91
References . . . 93

4 Method development of steroids hormones analysis in juvenile fish 98
4.1 Introduction . . . 100
4.1.1 Target hormones . . . 100
4.2 Method development strategies . . . 101
4.2.1 Separation Column . . . 101
4.2.2 The choice of solvents and additives . . . 102
4.2.3 MS detector . . . 102
4.3 Method validation . . . 102
4.3.1 Internal standard . . . 104
4.3.2 Precision and accuracy . . . 104
4.3.3 Linearity . . . 104
4.3.4 Limit of Detection (LOD) . . . 104
4.4 Chemicals and standards . . . 105
4.4.1 Preparation of solutions . . . 105
4.5 Method . . . 106
4.5.1 LC . . . 106
4.5.2 MS . . . 106
4.6 Data processing . . . 107
4.7 Method validation . . . 107
4.8 Results . . . 107
4.8.1 selectivity . . . 107
4.8.2 Enhancement of EICs and Mass spectra with IMS . . . 107
4.8.3 Repeatability and linearity . . . 113
4.9 Conclusion and perspectives . . . 114
References . . . 115

5 Collision Cross Section Prediction with Molecular Fingerprint Using Machine Learning **117**

5.1 Introduction . . . 119

5.2 Experiments . . . 119

5.2.1 Dataset . . . 119

5.2.2 Design of experiment . . . 121

5.3 Conclusion and perspectives . . . 123

5.4 Collision Cross Section Prediction with Molecular Fingerprint Using Machine Learning . . . 125

References . . . 140

6 Conclusion and perspectives **142**

Appendices **147**

List of Figures

1.1 Concept of exposome (Vermeulen et al., 2020) 4
1.2 A standard workflow of HRMS analysis (Bletsou et al., 2015) 7
1.3 Five levels of the identification confidence 10
1.4 Ionization and application range (Wei et al., 2021) 11
1.5 The mechanism of the ESI (Ho et al., 2003) 14
1.6 The mechanism of the APCI (Fang et al., 2020) 14
1.7 IMS-QTOF schematic (May and McLean, 2015) 17
1.8 Orbitrap schematic (Zubarev and Makarov, 2013) 18
1.9 Development of IMS and numbers of publications 24
1.10 Numbers of CCS values reported in the represented domain 25
1.11 Four main types of commercial IMS 26
1.12 TIMS funnels 28
1.13 Five levels of identification confidence with CCS 29
2.1 PCB and PBDE congeners 45
2.2 Isomer separation with CCS enhancement 49
2.3 MS of PBDE-Br5 generated from the same acquisition 51
2.4 Isotopic profiles fits 52
2.5 EICs of PBDEs 53
2.6 Ionization and fragmentation of PCBs 54
2.7 Ionization and fragmentation of PBDEs 55
2.8 The measurement precision for different POPs class was %RSD ¡ 0.6% 56
2.9 Protonation of PBDE-Br4-6 57
2.10 Mass-CCS trendline of the halogenated POPs. 59
2.11 Co-eluting PCBs in samples 60
2.12 Isotopic profiles deviation 61

2.13 overview of PCBs and PBDEs separation in EICs and EIMs 68
3.1 Extracted ion chromatograms obtained for pesticide standards in the optimized LC method. 78
3.2 Distribution of the standard retention time. 78
3.3 The workflow of suspect screening and non-target screening using UPLC-HRMS 80
3.4 TASQ software interface 81
3.5 Wastewater sample filtration 83
3.6 An example of TIC 83
3.7 Common substances in waste waters 85
3.8 Procedure of NTS identification: An example of Feature ID 2667 . . 86
3.9 Distribution of the detected substance class 87
3.10 Median intensity of detected substances at different sampling points . 87
3.11 Principal component analysis of contaminants and sampling site . . . 88
3.12 Natural water sampling points 89
3.13 Comparison of TICs in samples and blank 90
3.14 Percentage of samples presenting PEGs 91
4.1 Column optimization: Intensity comparison 103
4.2 Column optimization: S/N ratio comparison 103
4.3 Steroid Hormone Calibration Procedure 105
4.4 EICs of steroid hormones 108
4.5 Comparison of EIC generated by exact mass and exact mass with CCS filtering in cortisone and cortisol 110
4.6 Comparison of mass spectra generated by EIC and EIM 111
4.7 Mass spectra of cortisol 112
4.8 IMS enhancement in EIC Cortisol: Comparing chromatograms background at different mass widths and using CCS 113
4.9 IMS enhancement in EIC Cortisone: Comparing chromatograms background at different mass widths and using CCS 113
4.10 Linearity test of cortisol 114
5.1 RSD of replicated chemicals 120
5.2 Distribution of predicted chemical super class in dataset 121
5.3 An illustration of molecular fingerprints 122
5.4 An output of modeling results 122

5.5 Plotted feature importance 123
5.6 Examples of a substructure of a bit string 127
5.7 CCS Prediction Modeling workflow 128
5.8 Molecular similarity and Super class PCA 129
5.9 Confusion matrix of classification model 131
5.10 Random examples of "incorrect" predicted chemical 132
5.11 Predicted precision of class-based and direct prediction approaches . . 133
5.12 Residuals of class-based and direct models 134
5.13 Comparison of class-based and direct predicted CCS values using Norman Susdat 135

List of Tables

1.1 HRMS (Perez de Souza et al., 2021) . . . 19
2.1 MS setting of timsTOF . . . 67
3.1 Calibration plan of pesticide kit . . . 75
3.2 QA/QC mix . . . 76
3.3 LC gradient elution program . . . 77
3.4 LC-MS database of a pesticides standard mixture . . . 79
3.5 Automated Suspect screening parameter settings by TASQ® . . . 81
3.6 Natural water sample preparation . . . 89
4.1 Overview of cortisol and cortisone . . . 101
4.2 Standard list . . . 106
4.3 Steroid analysis LC gradient . . . 106
4.4 Overview of integrated EIC . . . 108
5.1 Summary of the collected dataset for CCS prediction modeling . . . 129
5.2 Results of super-class prediction modeling . . . 130
5.3 Results of CCS prediction modeling . . . 130
6.1 Detected candidate in wastewater . . . 149
6.2 Purchased Steroid Hormone Calibrators . . . 152
6.3 Purchased Steroid Hormone Internal Standard . . . 153

Chapter 1

Introduction

Summary

Environmental contamination has always been an international issue that concerns human health and environmental pollution. The growth of anthropic activities increases the release of chemicals into the environment and human body. The concept of the exposome was introduced by Christopher Paul Wild in 2005 (Wild, 2012). This concept explains the toxic effects of the chemical exposome on human health through intake (e.g., air, food and drink) and biological or natural processes, as illustrated in Figure 1.1.

With advances and innovations in modern instrumentation, research on water contamination has flourished. Moreover, developments in analytical methods and bioinformatics/cheminformatics have dramatically accelerated the efficiency of the analysis. Therefore, bibliography research is essential to understand the principles and advances in non-target analysis in aquatic ecosystems. This section is a general introduction of the important principles and milestones in non-target analysis, as well as the current challenges and perspectives in this domain.

This chapter begins with an introduction of contaminants of emerging concern (CECs) , their main resources, the related regulations and the current challenges of CECs analysis (Section 1.1).

In Section 1.2, I will summarize the concept of three common approaches in water and environmental analysis: target, suspect and non-target analysis. The advantages and limitations will be discussed.

In the following section 1.3, I will introduce the most used technique in water analysis. It consists of the separation technique, ionization technique, and high-resolution mass spectrometry (HRMS) and its acquisition mode. I will detail the techniques that are required and applied in this thesis. These techniques are gas and liquid chromatography, soft ionization, high-resolution QTOF and data-independent acquisition. Furthermore, in Section 1.4, the workflow and tools associated with data treatment will be discussed. Since water analysis is one of the main topics in this manuscript, the state-of-the-art method is summarized in suspect and non-target screening using liquid chromatography coupled with HRMS and IMS-HRMS.

In the last section 1.5, I will introduce the principle and main technique used in IMS-HRMS and, more specifically, trapped ion mobility spectrometry, its original geometry and figures of merit, which were used in this research project. The benefit of using IMS in non-target analysis will be discussed.

1.1 Contaminants of emerging concern (CECs)

Environmental pollution continues to be a popular issue as society develops. In the last 30 years, over 150,000 substances have been registered for use in Europe, the USA and Canada (Muir et al., 2019), and these substances are used in anthropic activities or are produced unintentionally during industrial processes to enhance product quality or to make our lives better. However, once a chemical is intentionally or unintentionally released into the environment, it and/or its transformation product can threaten the ecosystem and human health (Lambert and Wagner, 2018; Vandermeersch et al., 2015).

Chemical contaminants can be classified into two parts, *Legacy contaminants* and *Contaminants of Emerging Concern (CECs)*. Legacy contaminants are chemicals that were used in anthropogenic activities before their environmental effects and toxicology were discovered. Although legacy contaminants have been mostly banned or regulated for decades, some persistent chemicals are still under regular monitoring, such as lindane and polycyclic aromatic hydrocarbons (PAHs) (Hutchinson et al., 2013).

Dichlorodiphenyl trichloroethane (DDT) was once the most extensively used insecticide. Rachel Carson disclosed in her book "Silence Spring" in 1962 (Carson, 2015) that DDT and its ethylene metabolite DDE eliminated insects, resulting in a decline in the bird population. Until the 1970s, DDT was banned in the US (EPA, 2022) and Europe (legislation, 2003), and it was banned worldwide under the Stockholm Convention in 2001 (Convention, 2001). This case has since raised public concerns, and academic studies have been conducted to assess the risk of synthesized chemicals (Sauvé and Desrosiers, 2014). Hence, analytical chemistry plays a key role in the development of regulations and assessment of the evaluation of legacy contaminants.

The Stockholm Convent is focused on persistent organic pollutants (POPs), which can bioaccumulate in mammals that are part of a long food chain due to their high hydrophobicity and lipophilicity. Acts and legislative regulations have eliminated or severely restricted the use of hazardous chemicals, which effectively protect human health and the environment (Caballero-Casero et al., 2021). POPs are related to cancers, birth defects and immune dysfunctions (Xu et al., 2013). Twelve POPs, which are also known as the "dirty dozen," were initially listed in the Stockholm Convention, and new compounds have been added since 2004 (Convention, 2001). Most recently, CECs such as PFAs, OH-PCBs, and OH-PBDEs have been added to the Stockholm Convention and other regulations (Richardson and Kimura, 2019).

CECs refer to unknown xenobiotics, which can be:

- a chemical that has been released or known before but has only been recently found in the environment, food or mammals.
- a new chemical that can be a successor or replacement of withdrawn chemicals, a byproduct, or a transformation product (TP) or metabolite of a known chemical.

The detection of CECs has been increasingly reported in water and human bodies mainly due to advances in analytical methods and instrumentation (Paszkiewicz

et al., 2022; Sauvé and Desrosiers, 2014). These chemicals have diverse applications and can be potentially harmful to the ecosystem or human health. The increasing awareness of the risk to aquatic ecosystems has led to public discussion and scientific research (Hernández et al., 2019).

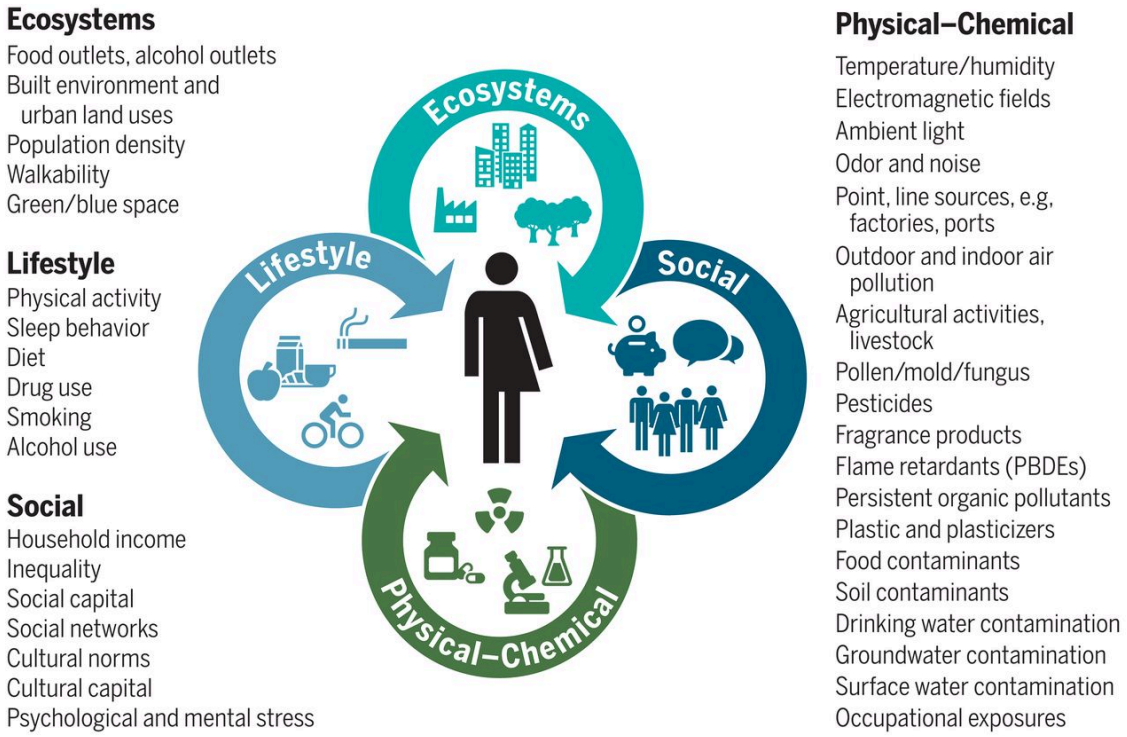


Figure 1.1: Concept of exposome (Vermeulen et al., 2020)

1.1.1 Resources and regulations of CECs

Pharmaceuticals and personal care products (PPCPs) are one of the major sources of CECs. PPCPs included diverse classifications, including antibiotics, hormones, preservatives, fragrances, UV filters from sunscreen, etc. (Liu and Wong, 2013). PPCPs can be released into the environment by domestic sewage and landfills (Yang et al., 2017). It is difficult to remove PPCPs and their metabolites through conventional urban wastewater treatment plants (UWWTPs), they have been detected at trace levels in sewage effluents, surface water or even drinking water (Yang et al., 2017). Moreover, once PPCPs and their metabolites are released into the environment, they can be degraded into so-called transformation products (TPs) through hydrolysis, photosynthesis, metabolic processes and excretion by mammals, etc.,. Moreover, the contamination of TPs becomes uncontrollable and untraceable (Wilkinson et al., 2017).

Pesticides are another major source of CECs. Pesticides are intended to kill insects or to protect crops and fruits in agriculture; therefore, pesticides are widespread in agricultural activities and personal care products. They can directly penetrate ground water. Some pesticides have been phased out or restricted due to their persistence and bioaccumulation in the environment and in mammals (Liu and Wong, 2013; Campanale et al., 2021). Although several pesticides have been banned for years, such as dieldrin and endrin, they are still widely detected in groundwater and

surface water (Affum et al., 2018). Moreover, its replacements (e.g., organophosphorus pesticides (OPPs)) share similar persistence and toxicity to the environment, and novel pesticides have become increasingly hydrophilic (Wille et al., 2012), allowing them to more easily escape from regular WWTPs and migrate to aquatic systems (Taylor et al., 2020).

CECs can be unintentionally produced through water disinfection by oxidative reactions (Richardson, 2011; Postigo et al., 2021). As some DWTPs use chlorination in a final disinfection step, chlorinated CECs have been reported in drinking water and present a potential risk to human health (Postigo et al., 2021; Tröger et al., 2021). Some disinfecting products and byproducts, such as trihalomethanes and haloacetic acids, are listed in the regular monitoring list or in the wish list (Kimura et al., 2019; Paszkiewicz et al., 2022; Tsaridou and Karabelas, 2021).

A number of studies have published the detection and toxicity of CECs, which are not in the regular monitoring list or under the detection limit in regular guidelines (Sousa et al., 2018; Schulze et al., 2020a; Wille et al., 2012). Therefore, the list of chemicals has been updated in the regulations and rules. The 3rd version of The Watch List of Water Frame Directive adds the emerging contaminants of pharmaceuticals and pesticides, for which routine monitoring in natural water is needed (Richardson and Kimura, 2019). Bisphenol A (BPA) and microcystin-LR were added to the Drinking Water Directive (EU) 2020/2184 (EU, 2021a). Additionally, in 2022, the first watch list emphasized the demand for monitoring two endocrine disrupting substances (beta-estradiol and nonylphenol) in drinking water, and a lower threshold that considers the protection of human health should be equivalently set within the whole European Union (EU, 2021b). Several new rules and regulations have also been established. For instance, the U.S. Environmental Protection Agency (EPA) requires that every 5 years, a new list of less than 30 unregulated contaminants should be monitored by public water systems. The fourth Unregulated Contaminant Monitoring Rule (UCMR-4) listed 30 contaminants to be monitored from 2018 to 2020 (EPA, 2021a). The fifth Unregulated Contaminant Monitoring Rule (UCMR 5) was published on December 27, 2021, and the standard rules about per- and polyfluoroalkyl substances (29 PFAS, so-called eternal pollutants) in drinking water need to be improved (EPA, 2021b). Similarly, the analysis method of PFAS in drinking water needs to be updated in the Directive (EU) 2020/2184 (EU, 2021a).

1.1.2 Challenging CECs analysis

The analysis of CECs in the aquatic samples mainly includes the following steps: sampling, extraction, analysis through modern instruments (typically LC-HRMS), data processing, prioritization and identification. Moreover, target, suspect and non-target screening are required as a combination analysis due to the numerous CECs present in the aquatic matrix (Menger et al., 2020). Mass spectrometry coupled with gas chromatography (GC-MS) or liquid chromatography (LC-MS) are still the classic methodologies for small molecular analyses. Although reversed-phase liquid chromatography (RPLC) is the most suitable separation technique for a wide scope screening of CECs in aquatic matrices, highly polar chemicals, such as persistent and mobile organic chemicals (PMOCs), are poorly retained in RPLC columns. In

contrast, PMOCs are highly persistent and highly mobile in aquatic systems, resulting in a significant issue with regards to the control of drinking water quality. Meanwhile, sub-ppt level concentrations in the real sample demand methods can be accurately and precisely measured at low levels (Hernández et al., 2011). On the other hand, an efficient and smart data minimizing process is required to generate a large amount of data. However, the prioritization steps result in the loss of data. False positives have been reported in several studies, and thus, a higher quality of data is needed. Eysseric et al. (2022) identified 106 transformation products and 176 congeners of industrial compounds in the Yamaska River close to wastewater treatment plants, including 28 substances that were not listed in the database. One of the frequent problems in wide-scope CEC identification is the lack of available standards for the last confirmation steps. Therefore, the predicted retention time and predicted CCS values can reduce the number of candidates (Hollender et al., 2017). HRMS can simultaneously scan thousands to tens of thousands of chemical features in a single analytical run; however, the annotation rate remains low (Vermeulen et al., 2020).

1.2 Concept approaches for CEC analysis

Modern HRMS allows the simultaneous screening of a wide mass range of molecules, dramatically improving the detection of environmental contaminants. Currently, three accepted analysis approaches (target, suspect and nontarget screening) are properly applied with regards to small molecule identification for various aquatic samples (Menger et al., 2020; Schymanski et al., 2015; Ccanccapa-Cartagena et al., 2019; Brunner et al., 2020). In Figure. 1.2, Bletsou et al. (2015) represented a flow chart of the screening procedure of target screening, suspect screening and nontarget screening. Schymanski et al. (2014) proposed five levels of identification confidence, as shown in Figure 1.3. These three approaches are often combined to obtain a global risk assessment in water (Menger et al., 2020).

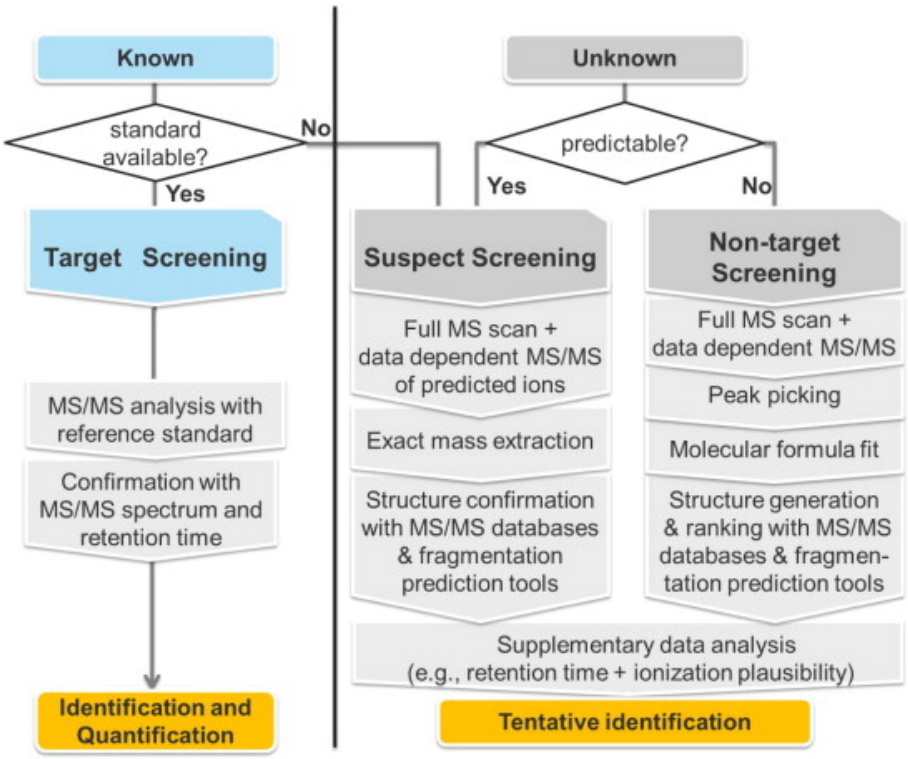


Figure 1.2: A standard workflow of HRMS analysis (Bletsou et al., 2015)
Flow chart of screening procedure of transformation products (TPs). 'Known' TPs have been confirmed or confidently identified before, other TPs are considered 'Unknown'.

1.2.1 Target screening

Target screening is a conventional way to identify contaminants in samples. It requires basic information regarding the samples prior to acquisition. The analysis methods were optimized and validated by the targeted chemical standards. The contaminants are confirmed by the retention time (RT), MS and MS/MS spectra.

The triple quadrupole (QqQ) (so-called tandem mass analyzer) has been the most routine technique for target analysis of CECs in water due to its sensitivity and robustness (Agüera et al., 2013). The selected reaction monitoring (SRM) mode

enables a quantification analysis with two specific SRM transitions. However, a relatively high mass intensity of the second transition is required to confirm and quantify a chemical. To assess the detection limit of the method, a highly selective sample preparation is required to remove matrix interference and to pre-concentrate the target compounds. Furthermore, the isotopic dilution method is recommended to ensure the certainty of the measurement. However, this approach is not applicable for a wide scope of chemical detection.

More recently, target screening has also been optimized in high-resolution mass spectrometers, enabling the simultaneous detection of thousands of chemicals in a single acquisition. A target screening and quantification method for hundreds of micropollutants, including pharmaceuticals, pesticides and PFAS, was applied in drinking water using UPLC-QTOF in MS^E-mode (Tröger et al., 2018; Ren et al., 2020). Gago-Ferrero et al. (2020) developed a quantitative method with over 2000 chemicals using QTOF in wastewater analysis. As a full-scan MS and data-independent acquisition (DIA) MS/MS, it offers the possibility to perform suspect and non-target screening at the same time(Diamanti et al., 2019) and a retrospective analysis afterward.

1.2.2 Suspect screening

Suspect screening aims to search the *known-unknown* that are expected to be present in a sample. Suspect chemicals can be referred from various sources. Chemicals are recorded in REACH, including authorized anthropic chemicals, such as pesticides, PPCPs and industrial products. TPs and metabolites that were reported in the literature are also valuable information for suspect screening (Menger et al., 2021; Wilkinson et al., 2017).

Suspect screening is commonly applied as a complementary approach to target screening using HRMS (Menger et al., 2020). It is helpful to discover unregulated chemicals by evaluating their presence in aquatic systems, leading to the delivery of regulations. Meanwhile, this approach is also implemented in wastewater analysis to estimate the removal efficiency and fate of the exposome in wastewater (Golovko et al., 2021; Wiest et al., 2021).

Suspect screening is performed by screening the detected features against a list of chemicals (Schymanski et al., 2014), and the matched structures are thereby assigned to each feature. The NORMAN Suspect List Exchange (NORMAN-SLE) is a dynamic and open access database containing over 100,000 environmental chemicals from more than 70 contributors around the world. The lists cover the chemicals under the European REACH regulation (EC: 1272/2008). CECs have been detected in real water samples (Schymanski et al., 2015). Therefore, NORMAN-SLE is an ideal resource for environmental suspect screening. Other large open-source chemical databases, such as PubChem, Massbank, and Metflag, include experimental or/and in silico spectrum information and structural descriptors (Krier et al., 2022; Menger et al., 2021).

Annotated chemicals need to be definitively confirmed by reference standards; however, analytical standards are not available in most cases. Quantitative structure-toxicity relationship (QSTR) models (Aalizadeh et al., 2016) are developed for the risk assessment of CECs and semi-quantified compounds in various water samples

(Alygizakis et al., 2019a; Sjerps et al., 2016). Recently, Aalizadeh et al. (2022) proposed a novel workflow using QSTR for the semi-quantification of CECs in GC-APCI-HRMS analysis. Moreover, it is challenging to treat a large amount of data in suspect screening. Prioritization tools facilitate the data treatment. The details of the suspect screening workflow are introduced in Section 1.4.

1.2.3 Non-target screening

Non-target screening covers all the remaining chemicals present in the samples, which are defined as "unknown-unknown". Non-target screening has a similar sample acquisition to suspect screening; however, no prior information is accessible in the samples. The features of interest are first prioritized from component lists, and then structural elucidations are carried out for masses of interest (Schymanski et al., 2014). Prioritization strategies follow a similar concept for suspect screening. Structural elucidation remains a time-consuming and massive labor, and it relies on manual and critical diagnoses by experts (Menger et al., 2020). Non-target screening sometimes comprises suspect and non-target screening, representing all the investigations of unknown structures (Menger et al., 2020). There is another term, "non-target analysis", which refers to generic HRMS acquisition methods (Menger et al., 2020).

Overall, suspect and non-target screening are pioneering studies in environmental monitoring. The origin and occurrence of organic contaminants in the environment and wastewater treatment can be assessed through suspect and non-target screening.

1.2.4 Five levels of identification confidence

Identification confidence is an important factor that is used to evaluate the detected substances. To ease communication, Schymanski et al. (2014) proposed a five-level identification confidence system, which has been extensively applied in environmental analysis. A schematic of the identification confidence system is illustrated in Figure 1.3.

Level 1 is the definitive identification level, where a substance can be confirmed by the retention time, MS and MS/MS spectra measured in a reference standard. Moreover, in IMS-HRMS analysis, the drift time or the CCS value are recommended to increase the identification point (Celma et al., 2020).

Level 2 is a high identification level in structure elucidation. The experimental spectrum is unambiguously matched to a reference spectrum in the literature, spectra library, or via structural diagnostics. The exact mass of the precursor ion, isotopic patterns and fragment ion profiles are strong evidence that can be used to propose a probable structure. Furthermore, the retention index offers a supplementary point in compound annotations. Other information, such as the experimental context and physicochemical proprieties, provide a single structure fit, but no standard is available.

Level 3 is a tentative identification level, where possible structures are assigned to an unknown; however, diagnostic information is insufficient to support a unique structure.

Level 4 contains only a simulated formula for an unknown. Generally, the MS information, including exact mass, isotopic pattern, and adduct ion, provides a possible formula. However, MS/MS matching is not clear enough or cannot predict possible structures. *Level 5* only records the exact mass, and no structure or formula exists.

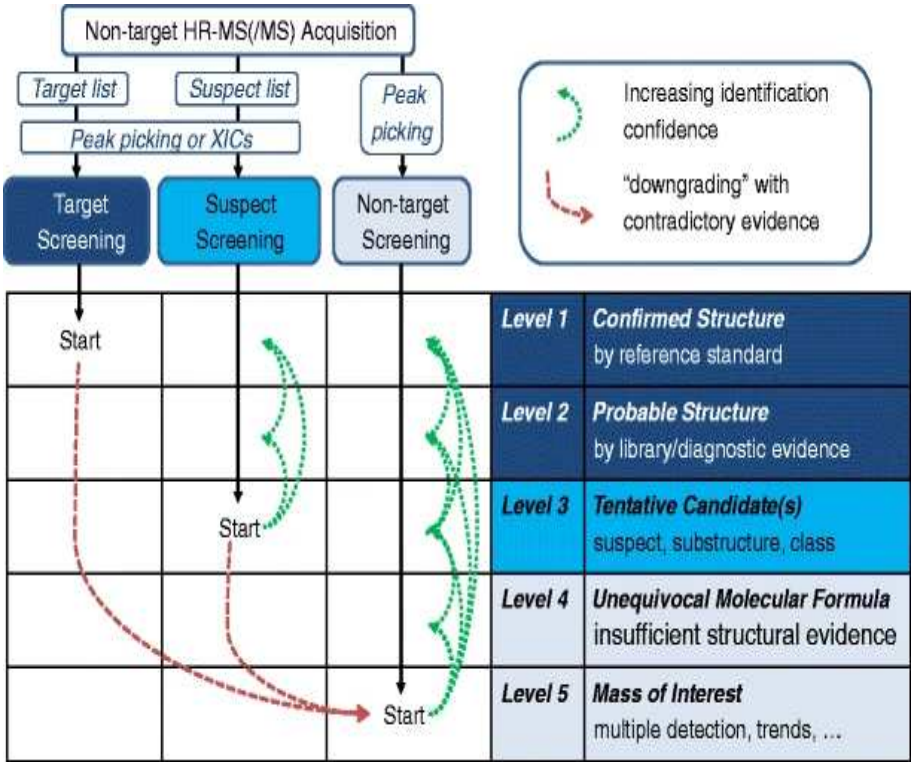


Figure 1.3: Five levels of the identification confidence proposed by(Schymanski et al., 2014)

1.3 Analysis approaches of CECs

Mass spectrometry (MS) coupled with chromatographic separation is a widespread technique in water analysis. Its high sensitivity, selectivity, accuracy and high throughput make it the first choice in many applications, including drug and food quality control, forensic science, drug discovery, and environmental analysis. Chromatographic separation consists of two major parts: gas chromatography for non-polar compound analysis and liquid chromatography for polar compound analysis. MS instruments contain three main parts: the ionization source, the mass analyzer, and the detector (El-Aneed et al., 2009).

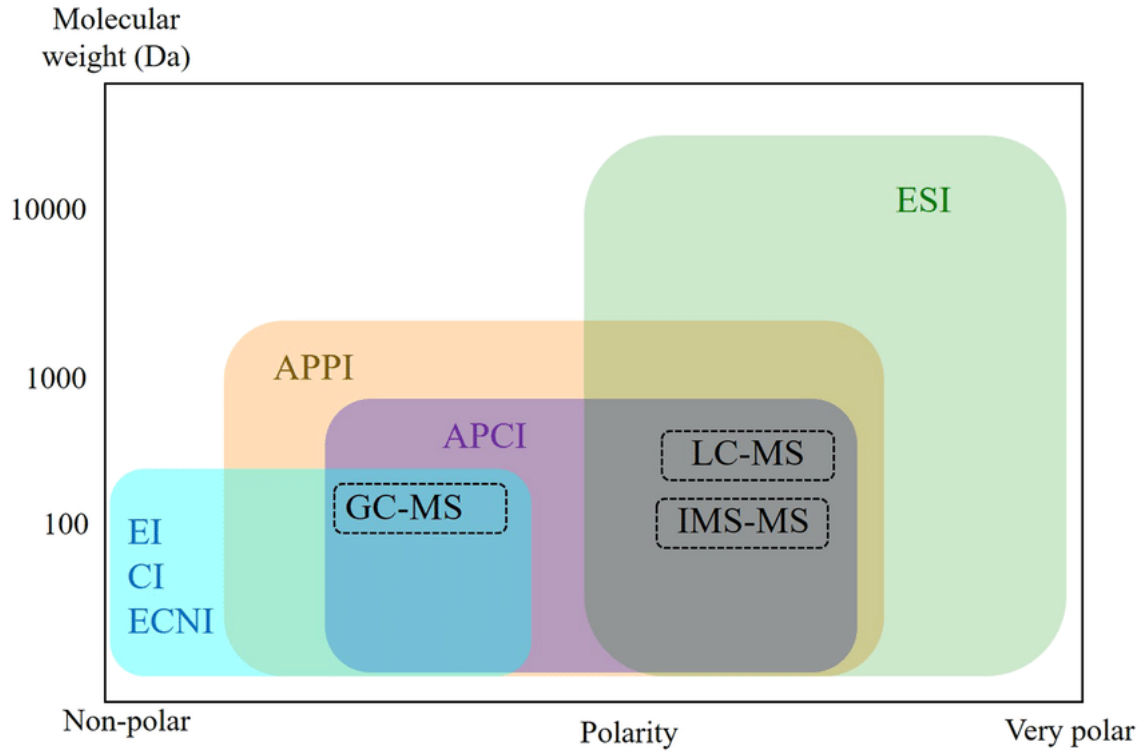


Figure 1.4: Ionization and application range (Wei et al., 2021)

1.3.1 Separation technique

Liquid chromatography (LC)

Liquid chromatography (LC) is commonly referred to as a high-performance liquid chromatography (HPLC) in modern analysis. Ultrahigh-performance liquid chromatography (UHPLC)has become an alternative to the most common HPLC. Comparing to HPLC, UHPLC operates at higher pressures that can reach up to 15,000psi, which produces a higher sensitivity . Moreover, UHPLC enable lower particle sizes (1.7 to 3 μ m) within columns and higher capacity of plate numbers. Consequently, UHPLC increases chromatographic peak resolution, and it is a high-throughput acquisition. LC is a separation technique based on the "trinity" affinity among the analyte, the stationary phase and the mobile phase. As a result, only the analyte that could be dissolved in LC solvents during use can be analyzed through UHPLC. In general, the LC system is first stabilized by the mobile phase under a constant flow, and the sample is transferred into the chromatographic column where the separation occurs at a controlled and constant temperature. The analytes are separated and then eluted out of the column and, at the end, detected by a nondestructive (e.g., UV/Vis, Refractometer) or a destructive detector (e.g., ELSD, MS).

The stationary phase and mobile phase are the key factors in LC separation. The choice of these two phases depends on the physicochemical proprieties of the analytes of interest. In classic chromatography models, the stationary phase is a polar phase (e.g., silica), and the mobile phase is a nonpolar phase (e.g., chloroform), which is called normal-phase liquid chromatography. However, NPLC was soon replaced by reversed-phase liquid chromatography (RPLC) due to its poor reproducibility and large consumption of organic solvents. RPLC has a nonpolar

stationary phase, which is commonly a bond of hydrophobic ligands, e.g., C_{18} - alkyl chains or aromatic functional groups to the surface of a rigid siliceous or polymeric support (Berruex and Freitag, 2003). The mobile phase consist of water with a water miscible solvent (such as methanol (MeOH) and acetonitrile (ACN)) and organic acids. The high proportion of water in the mobile phase ensures the dissolution of hydrophilic analytes, and the organic solvent maintains the polarity of the mobile phase and interaction between the analyte and the stationary phase. The less polar or hydrophobic analytes tend to absorb to the stationary phase due to the "Van der Waals" interaction, "dipole-dipole" interactions and "hydrogen" bonding. The polar compound has less affinity with the stationary phase and stays for a shorter time in the column. Less polar compounds can be eluted with a decreasing polarity of the mobile solvent by increasing the organic solvent. The order of the analyte eluted out of the column is recorded in terms of the retention time (RT), which is a vital parameter in compound alignment. Thus, RPLC is an ideal approach for polar CECs in water analysis, such as PCPPs and drugs.

However, highly polar chemicals are hardly retained in the RPLC column and eluted in the void, such as PMOCs. Other separation techniques include hydrophilic interaction liquid chromatography (HILIC) and supercritical fluid chromatography (SFC). HILIC has been used as a complementary analysis to RPLC for very polar analytes in water (Vaudreuil et al., 2020; Neuwald et al., 2021; Qiu et al., 2022). HILIC can be combined with RPLC (2DLC) to cover chemicals from very polar to polar compounds (Bäurer et al., 2019; Bieber et al., 2017). SFC is a separation based on a critical point of a substance, at which the substance is equal parts liquid and gas. SPC uses CO_2 as the mobile phase and MeOH as the liquid solvent. It is a cheaper and greener method than LC. In addition, SPC can separate small molecules from complex matrices and can effectively separate very polar chemicals (Schulze et al., 2020b; Bieber et al., 2017; Rice et al., 2020).

Gas chromatography (GC)

Gas chromatography (GC) is a separation technique based on the boiling point/vapor pressure and polarity of the ion species. In general, a sample is introduced into an injector through a liner, which is maintained at a high temperature to evaporate the solvent (e.g., Toluene, Acetonitrile, Hexane). The sample can be injected with different modes, including on column injection, direct injection, programmed temperature vaporizer (PTV). While the main injection mode in used for this research project were splitless and split only. In the split injection, a small fraction of the vaporized sample is pushed by a carrier gas (He or H_2) to the column. In the splitless mode, over 80% of the sample is introduced and refocused into the first part of the capillary column. After the splitless time, the carrier gas flushes the remaining samples from the liner mainly through the split. Then, the analytes are separated in the capillary with a stationary phase inside. The separation is based on the affinity between the analytes and the stationary phase. GC-MS is less dominant than LC-MS in water analysis since GC is specific for nonpolar and (semi)volatile compounds. Nonpolar CECs, including pesticides, phthalates, PAHs, PCBs, PBDEs, etc., are mainly analyzed through GC-MS or GCXGC-MS (Mazur et al., 2021; Badea et al., 2020; Peris and Eljarrat, 2021; El-Deen and Shimizu, 2021; Murrell and Dorman, 2021).

A more detailed discussion of the recent advances in GC-APCI-HRMS is presented in Section 2.1

1.3.2 Ionization technique

Mass spectrometry is an analytical technique that measures the mass-to-charge ratio (m/z) of ions. For this purpose, different kinds of ionization methods have been developed. Electron impact ionization (EI) is one of the first methods and is still popular in GC-MS. In GC-EI-MS, ion species in the gas phase are directly bombarded by high energetic electrons under a vacuum. The ionization energy is commonly set at 70 eV, and radical ions and fragment ions are produced within the ion source volume. Thus, EI is defined as or considered to be a hard ionization mode. Characteristic mass spectra can be obtained for a broad range of organic chemicals, and EI spectra are easy to search against an EI-MS spectral database. For example, the NIST20 EI library contains 350,643 EI spectra of 306,867 compounds (2020) (NIST, 2020). However, for many molecules, fragment ions are less unique or abundant than molecular ions, reducing the sensitivity (Li et al., 2015). The soft ionization mode of electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI) and other modern techniques have been developed. A summary of the current ionization mode and its application domain is plotted in Figure 1.4. ESI and APCI are the two ionization modes used for this research study.

ESI

ESI is a soft ionization mode under atmospheric pressure and transforms ions from a solution into the gaseous phase by electrical energy. Currently, it is mostly used in LC-MS for polar and thermally labile chemical analysis (Ho et al., 2003). The ESI mechanism consists of 3 steps (Figure 1.5):

1. The first step is the dispersal of a fine spray of charge droplets under a high charge capillary (± 3 to 5 kV) under atmospheric pressure. A Taylor cone is formed and stabilized by the liquid surface tension, electrostatic force and gravity

2. The second step is the evaporation of the solvent by a gas such as nitrogen. Moreover, the charge intensity on the surface of the droplet increases. Due to coulombic repulsion, droplets are split into smaller charged droplets until they have single- or multiple-charge ions.

3. The gas phase individual ions formed are attracted, trapped and moved to the mass analyzer relative to its charge and to the respective polarity of the mass spectrometer inlet. Both single charge or multiple charge ions can be created through the ESI.

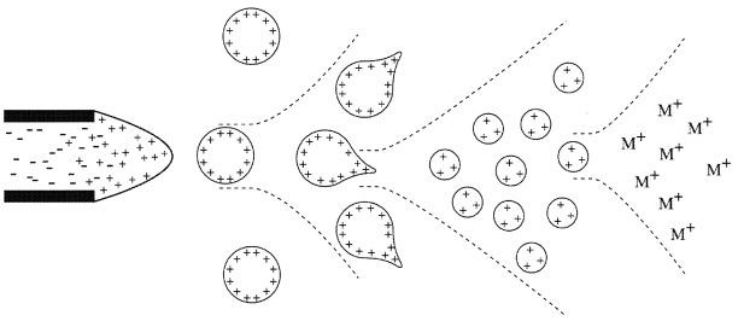


Figure 1.5: The mechanism of the ESI (Ho et al., 2003)

The ESI can be applied in both positive and negative modes. It is ideal for polar and thermolabile chemicals. Thus, LC-ESI-MS is commonly used for CECs detection in water samples.

APCI

The APCI has a similar interface as the ESI; instead of putting a voltage on the spray, the voltage is placed on a needle to generate a corona discharge under atmospheric pressures, making it more suitable for low- and medium-polarity compounds (nonpolar molecules). In the APCI, the vaporization of the solvent yields by spraying the sample solution into a heater (tunable until 400 °C) using a gas (e.g., Nitrogen). Droplets are ionized by the corona discharge needle to generate metastable ions that could, by charge transfer, softly ionize the analytes of interest. It is mainly protons that are transferred between ions to enable ionization, which leads to charge or proton transfer reactions and electrophilic addition reactions. In contrast to the ESI, the APCI involves a higher energy process and does not produce multiple charge ions.

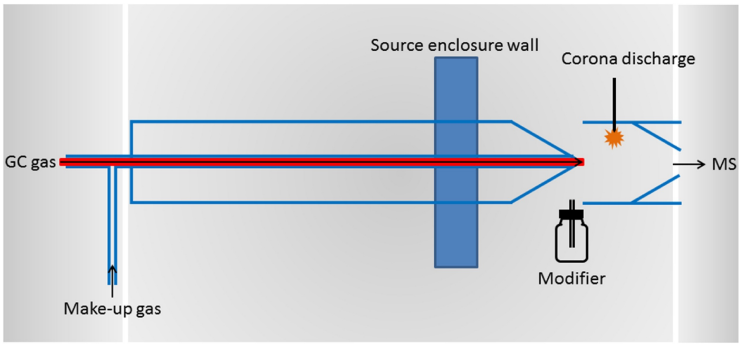


Figure 1.6: The mechanism of the APCI (Fang et al., 2020)

The APCI was initially applied in HPLC-MS. Most recently, GC-APCI-MS platforms are available, and increasing applications of GC-APCI-HRMS in non-target analysis have been published (Section 2.1 provides a reference).

1.3.3 Mass detector

Many MS systems have been used for CEC analyses in water samples. The wide scope of CECs in water analysis requires high selectivity and high precision. Hybrid

mass analyzers such as a quadrupole with a time-of-flight (Q-TOF) or an Orbitrap (Q Exactive) equipped with an ESI are currently the most promising techniques in water analysis (Menger et al., 2020). These mass spectrometers not only provide a high precision in mass measurement but also enable structure elucidation. Table 1.1 lists a modern HRMS and its key features. Several principles of the HRMS are the mass accuracy and the resolving power.

- The mass accuracy describes how the mass measured by a mass spectrometer is similar to the true mass. It is qualified by a standard deviation in parts per million *ppm* (Equation 1.1).

$$Accuracy = \frac{(m/z)_{\text{measured}} - (m/z)_{\text{theoretical}}}{(m/z)_{\text{measured}}} \times 10^6 \tag{1.1}$$

- Resolving power and mass resolution.
  The resolving power refers to the ability of a mass spectrometer to distinguish two adjacent ions of equal intensity when the overlap between two peaks is more than 10%. The resolving power is fixed for a mass range in the TOF and correlated to the flight tube length. It is expressed as Equation 1.2. The mass resolution was evaluated by the peak width and the mass for a single peak, which can be calculated by the full width at half maximum (FWHM) (Equation 1.3). However, it should be noted that the definitions of the resolving power and resolution are controversial. Thus, one should be aware of the equation used to obtain the resolving power or resolution for comparison (Hernandez et al., 2012).

$$\frac{m_1}{m_1 - m_2} \tag{1.2}$$

$$R = \frac{m}{\Delta m} \tag{1.3}$$

  where:

  m = the nominal mass of a molecule
  Δm = mass difference in FWHM

- The acquisition speed defines the number of scans per second that can be performed per second. When the MS is coupled with the LC or GC, the MS should maintain a high scan speed to have sufficient data points for a chromatographic peak and maintain the peak resolution. However, the mass resolution of the orbitrap is inversely related to the acquisition speed; thus, the orbitrap could suffer from some limitations when applied with an ultrashort transient signal, such as with the GC and UHPLC (in contrast to the TOF).

Time of Flight

The development of the TOF-MS started in the 1940s. However, with its low resolution, the TOF-MS was soon replaced by the magnetic HRMS and quadrupole MS instruments (Guilhaus, 1995; Mamyrin, 2001). Unlike MS, the principle of a TOF analyzer is: a cluster of ions are accelerated and traveled in a field-free flight

tube. Due to ions' different m/z ratios, lighter ions travel faster than heavier ones, lighter ions spend short traveling time to reach the detector. Therefore, m/z ratio is propositional to the flight time (Equation 1.4). Nowadays, modern TOF-MS is constructed with an orthogonal acceleration TOF analyzer where its drift tube is equipped with a reflectron, shown in Figure 1.7. In an oaTOF mass spectrometer, ions are first transmitted to the pusher region and accelerated orthogonally by a pusher voltage. After separating based on ions' m/z ratios, ions are focused in the reflectron region and arrive at the detector. The reflectron acts as an ion mirror which reverses the trajectory of an ion. The reflectron can minimize the spread of kinetic energy of ions with the same m/z ratio, thereby greatly enhancing resolution.

$$\begin{aligned} t &= \frac{d}{v} \\ v &= \sqrt{\frac{2KE}{m}} \\ t &= d\sqrt{\frac{m}{2KE}} \end{aligned} \tag{1.4}$$

where:

t = time of flight (s)
d = length of flight tube (m)
v = velocity of the ion (m/s)
m = mass of the ion (kg)
KE = kinetic energy of ion (J)

Hybrid quadrupole times-of-flight (QTOF) mass spectrometry has been extensively applied in environmental analysis (Menger et al., 2020). Similar to a triple quadrupole (QqQ), it consists of an MS1 and collision cell, but the last quadrupole is replaced by a TOF analyzer for MS/MS screening. Herein, the mass resolution is significantly improved. TOF analyzers provide high data acquisition rates (up to 10-20 kHz), which makes it suitable to combine UHPLC and IMS.

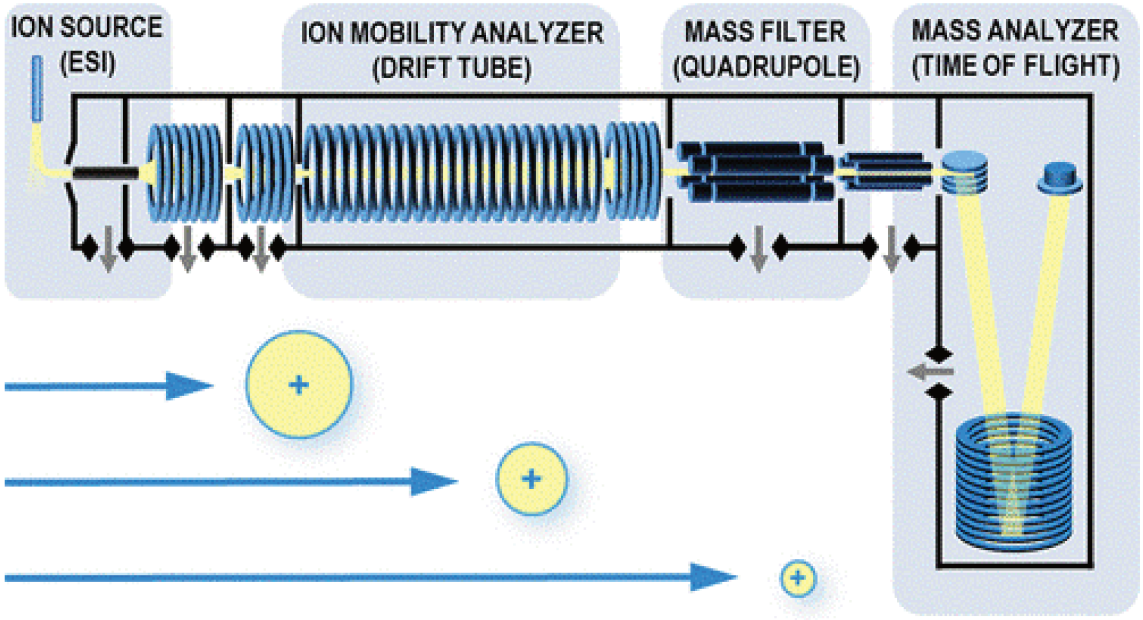


Figure 1.7: IMS-QTOF schematic (May and McLean, 2015)

Orbitrap

The first commercial orbitrap instrument was designed as a hybrid with a linear ion trap made by Thermo Fisher Scientific (LTQ Orbitrap) (Eliuk and Makarov, 2015). Since then, Orbitrap-based instruments have been produced and have become a routine technique in analytical academia and industrial laboratories. The Orbitrap mass analyzer contains three electrodes. Ions are trapped in the electrostatic fields. A key advancement in the Orbitrap is the development of an external storage device, the so-called C-trap, in which ions are accumulated before being led into the Orbitrap analyzer (Eliuk and Makarov, 2015). The Orbitrap is decoupled from the continuous ion source; thus, the Orbitrap can be implemented in any ion selected or transmit device (Zubarev and Makarov, 2013). Compared to the TOF, the resolving power R of the TOF is independent of the detection time except at a low mass, but the in-spectrum dynamic range depends on the detection time. Therefore, a higher spectrum acquisition rate results in a smaller dynamic range (Zubarev and Makarov, 2013). As illustrated in Table. 1.1, the Orbitrap Eclipse Tribid (Thermo Fisher Scientific) achieved a higher resolution (1,000,000) than QTOF (80,000 from Q-TOF maXis II, Bruker). In contrast, the resolving power of the Orbitrap depends on the acquisition rate, resulting in a limited peak resolution when coupled with the UHPLC (Menger et al., 2020).

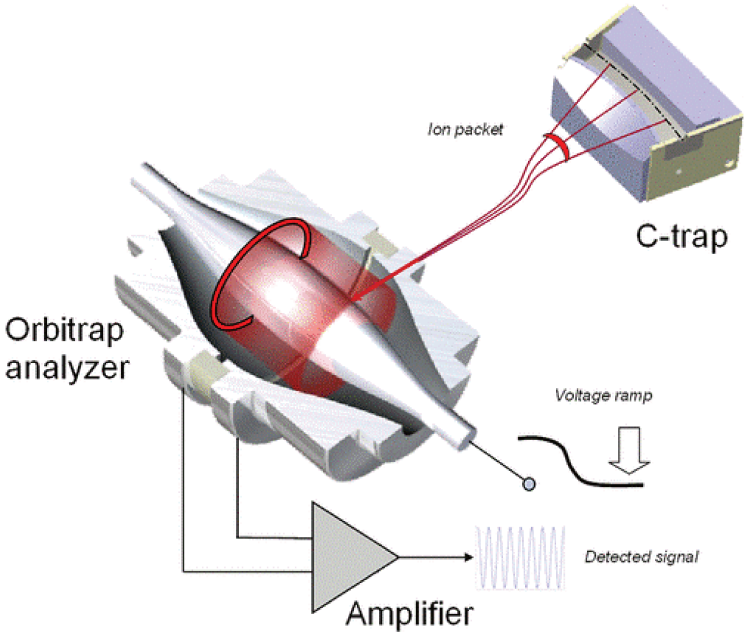


Figure 1.8: Orbitrap schematic (Zubarev and Makarov, 2013)

Table 2 | Summary of Orbitrap and Q-TOF equipment from different vendors and their main features

Instrument	Analyzer	Mass accuracy[a]	Resolution	Mass range	Dynamic range	MS/MS acquisition speed	Available ion mobility	Vendor
Q Exactive Focus	Orbitrap	<1 ppm	70,000 at *m/z* 200	50 to 3,000 *m/z*	10^3	12 Hz	-	Thermo Fisher Scientific
Q Exactive HF-X	Orbitrap	<1 ppm	240,000 at *m/z* 200	50 to 6,000 *m/z*	10^3	40 Hz	-	Thermo Fisher Scientific
Orbitrap Exploris 480	Orbitrap	<1 ppm	480,000 at *m/z* 200	40 to 8,000 *m/z*	10^3	40 Hz	FAIMS	Thermo Fisher Scientific
Orbitrap Eclipse Tribid	Orbitrap	<1 ppm	1,000,000 at *m/z* 200	50 to 8,000 *m/z*	10^3	40 Hz	FAIMS	Thermo Fisher Scientific
TripleTOF 6600	Q-TOF	<1 ppm	35,000 at *m/z* 956	5 to 40,000 *m/z*	10^5	100 Hz	FAIMS	Sciex
timsTOF	Q-TOF	<0.8 ppm	50,000 at *m/z* 1,222		10^5	>100 Hz	TIMS	Bruker
6560 Q-TOF	Q-TOF	<1 ppm	42,000 at *m/z* 2,722	100 to 10,000 *m/z*	10^5	30 Hz	DTIMS	Agilent
Xevo G2-XS	Q-TOF	<1 ppm	40,000 at *m/z* 956	20 to 16,000 *m/z*	10^4	30 Hz	-	Waters
Synapt G2-Si HDMS	Q-TOF	<1 ppm	60,000 at *m/z* 956	20 to 32,000 *m/z*	10^4	30 Hz	TWIMS	Waters

[a]Values correspond to maximum accuracy using internal calibration.

Table 1.1: HRMS (Perez de Souza et al., 2021)

1.3.4 Acquisition mode

In suspect and non-target screening, there are two main acquisition modes in the HRMS: data-dependent acquisition (DDA) and data-independent acquisition (DIA).

DDA

Data-dependent acquisition (DDA) is an ion scan mode for the selection of precursor ions in a full scan that produce a unique fragmentation spectrum in MS/MS (Samanipour et al., 2018). One of the main advantages of using DDA is that only MS/MS fragmentation generated by ions meeting a given m/z range with a predefined peak intensity threshold in the full scan will be detected. The number of precursor ions being fragmented is limited, and thereby the sensitivity of MS/MS can be increased, and the background can be eliminated (Guo et al., 2020). DDA requires a solid understanding of the experimental conditions and the precursor selection parameters. Therefore, it is a powerful tool for simultaneous target screening with MS/MS spectra pre-acquired by reference standards. However, a sufficient ion intensity and resolution of each precursor ion are enough to ensure the quality of the MS/MS spectra, resulting in a long LC run time and a good compromise between the number of precursor ions and acquisition rate (Guo et al., 2020). Although DDA has a widespread application in nontarget screening, the main shortcomings are a low amount of fragment ions and a limited number of precursor ions, resulting in false-positive hits. The presence of intense ion peaks from irrelevant compounds or sample matrices over product ions disrupts compound identification (Samanipour et al., 2018; Guo et al., 2020). Ferrer et al. (2020) developed a specific non-target screening method in the DDA mode using an LC/Q-TOF-MS instrument for water samples. Critical parameters such as the acquisition rate (MS and MS/MS), cycle time, collision energies, and ion transmission windows were optimized, and a strict selection of the common and most intense precursor ions was performed to avoid false-positive hits and reproducibility of their MS/MS spectra.

DIA

Data-independent acquisition (DIA) is another automated MS/MS acquisition mode that has recently become more popular than DDA. In DIA, all the precursor ions within a selected m/z range in the MS scan mode are fragmented in the MS/MS mode without any predefined selection criteria. DIA is defined under different names by the manufacturer: MS^E (Waters Corporation), All-Ions MS/MS fragmentation (AIF, Agilent Technologies), broadband CID (bbCID; Bruker), and multiplexed MS/MS data-independent acquisition (MSX-DIA; Thermo Fisher Scientific) (Alvarez-Rivera et al., 2019). DIA operates at a low collision energy (e.g., 6 eV) in a full scan for precursor ions and automatically switches to high collision energy (e.g., 25 eV) to obtain an MS/MS spectrum from each precursor ion. As all the ions are fragmented, the interpretation of the MS/MS spectrum is challenging, and adequate deconvolution algorithms are required to process the DIA data (Samanipour et al., 2018). The main advantage of DIA over DDA is that DIA scans all the ions and their fragments; thus, it is a suitable acquisition mode for nontarget screening to identify unknowns and perform retrospective analysis. In addition, generalized MS/MS spectral databases are easy to access for tentative identification, and several

data processing tools have already been developed for data processing automation.

1.4 Non-target analysis strategies

HRMS data generate a large amount of signal in the full-scan mode. Data processing is a crucial and time-consuming step that selects the most relevant data from mega-sized raw data. Data processing is directly related to the identification efficiency. Therefore, it is essential to understand different parameters in data treatment and explore all possible data processing tools.

1.4.1 Data preprocessing

HRMS enables a very large size and complexity of the raw data, and data preprocessing aims to eliminate irrelevant information. HRMS data are typically a collection of mass spectra over time; therefore, one of the first to minimize data size is feature detection, which includes RT alignment, pick picking and isotopes and adducts. Features represent integrated peaks for a given mass that have been aligned across samples (Hollender et al., 2017). Preprocessing parameters are essential for feature detection, which define the minimum peak intensity to be integrated. If the intensity of the threshold is too low, considerable background noise will be integrated. In contrast, if the threshold is too high, features will miss signals for further identification. Another step for data mining is to define the signal-to-noise ratio (S/N). Data preprocessing can be achieved through vendor software or open-source tools. Regarding Bruker Daltonics raw data, data preprocessing can be performed by Data Analysis and Metaboscape and Trace Finder and Compound Discoverer from Thermo Scientific. More open-source data processing tools have been introduced, such as mzMine (2 & 3) (Pluskal et al., 2010) and XCMS (Tautenhahn et al., 2012). Compared to data analysis or other vendor software, mzMine allows more flexible preprocessing parameter settings. However, Bruker TOF raw data must be first converted to mzML (Martens et al., 2011) or mzXML (Pedrioli et al., 2004) files without any losses before importing and manipulating the relative data into mzMine or XCMS.

1.4.2 Prioritization

Prioritization is aimed at selecting the compounds of high interest and is the first step in nontarget screening or suspect screening.

Binary sample comparison is a potent prioritization strategy. This strategy compares the common and different features between two samples, e.g., before and after water treatment (Schollée et al., 2018). Another strategy is trend analysis, which compares a series of samples in a time frame (Hollender et al., 2017) to investigate occurrences in the environment and identify transformation products. Detection frequency is also an efficient approach that detects highly present and intense compounds in a batch of samples, improving the identification confidence.

Halogenated compound filters can dramatically minimize the data size due to the characteristic isotopic pattern. The specific isotopic patterns of brominated and chlorinated compounds reveal obvious evidence for feature detection. Furthermore,

dehalogenation generally occurs during ionization and fragmentation, producing specific MS and MS/MS mass spectra.

Overall, it requires a holistic perception of the samples to prioritize the data.

1.4.3 Structure elucidation and identification

After prioritization, the study will focus on the formula assignment and structure elucidation within features of interest. Formula prediction relies on the mass accuracy and isotopic profiles, while the structure information is mainly acquired through the MS/MS mass spectra. The HRMS achieves a high mass accuracy (Δ mass < 5 ppm), mass resolving power $>$ 500,000 (Perez de Souza et al., 2021) and, if possible, good spectral accuracy, which is important to simulate molecular formulas. Structure elucidation is still a challenging task, and the tentative structure is typically given by searching against a mass spectra library. Therefore, the match ratio depends on the size of the library, the quality of the experimental and reference mass spectra, and user-defined thresholds and edges. Additionally, prediction tools are assessed to reduce the number of candidates, such as the predicted RT (Diamanti et al., 2019) and predicted CCS (Menger et al., 2022). In silico fragmentation tools can also facilitate structure elucidation, such as MetFrag (Ruttkies et al., 2016).

1.4.4 Nontarget analysis data processing software

Several data processing methods have been developed for nontarget analysis, including vendor and open-source software. Vendor software offers a fully automated data processing workflow, requiring minor adjustments. In contrast, open-source software allows users to optimize more parameters to fit different samples and analysis purposes. However, some software requires basic knowledge and advanced knowledge of programming to properly define the data processing parameters. Moreover, each software is initially designed for different studies, mostly lipidomic and metabolomic, and different algorithms are used for feature detection and peak picking, resulting in different output results. The choice of software is a vital step for this PhD research.

Since all the HRMS data are acquired by timsTOF (IMS-HRMS) from Bruker Daltonic, Data Analysis is the first and most accessible software to treat the raw data. It allows automated feature and peak detection with minor parameter definitions. It is easy to access other Bruker Daltonic platform software with data analysis, such as the Bruker library editor and craw finder, which allows users to easily switch from different software to perform the needed function. It is easy to build an in-house library from acquisition data.

However, similar to other vendor software, data analysis is expensive and considered to be a "black box", where the user has limited access to define the processing parameters. Moreover, data analysis is a generic raw data treatment software; therefore, it is efficient to read and process single data. MetaboScape® (from Bruker Daltonics) is ideal to run non-target analyses for a long batch of data and to perform statistical studies. However, it is more desired and dedicated for metabolite annotation than environmental contaminants, and custom-made libraries with defined formats need to be loaded for analytical purposes.

MZmine and MS-DIAL are two open-source software programs for non-target analysis data processing. They were initially developed for untargeted metabolite and lipid identification using LC-MS data (Pluskal et al., 2010). MS-DAIL was originally designed for SWATH (DIA) data (Tsugawa et al., 2015). Currently, it has become universal software for GC and LC coupled to HRMS and IMS, and both software programs have been applied in environmental analysis (Krauss et al., 2019; Beckers et al., 2020; Qian et al., 2021). The raw Bruker MS data need to be converted to mzML or mzXML before importing to MZmine, while the Bruker IMS-MS data are compatible in MZmine3. For MS-DIAL, both MS and IMS-MS data need to be converted to ABF and IBF files. MZmine allows data preprocessing, and in peak detection, MZmine can customize the scan numbers of the mass spectrum and use the characteristic mass spectrum to optimize the noise, shoulder peak and peak deconvolution. It is more flexible for different sample types. In MS-DIAL, the peak detection is automatically performed, and the user only needs to define the mass tolerance or RT tolerance. In peak detection, a reference database can be imported or searched by an online database (e.g., PubChem, HMBD, KEGG). However, MZmine searches with neutral masses, adducts and isotopic profiles. MS-DIAL first matches the exact mass, and/or RT, and isotopic profiles. Then, similar MS/MS mass spectra are calculated to estimate the probability. Peak alignment in MS-DIAL is inspired by Joint Aligner implemented in MZmine

In data treatment, MZmine and MS-DIAL have implemented statistical analysis, but MS-DIAL generates a matched compound table and summarizes the features in different sample categories. It is straightforward to visualize the detection frequency and binary sample comparison. Another main drawback of MS-DIAL over MZmine is that MS-DIAL can operate only in the Window system, and MZmine is not limited by the operating system. Overall, MZmine and MS-DIAL have well-designed user interfaces, and they do not require advanced programming knowledge. Both can efficiently process HRMS data. MS-DIAL is more automated than MZmine. CCS databases can be introduced in the peak identification workflow. However, MZmine shows advantages in processing data with higher background noise.

Moreover, NORMAN Suspect List Exchange (NORMAN-SLE) has been collaborated with over 70 contributors around the world, prioritizing non-target screening of environmental samples by mass spectrometry (Alygizakis et al., 2018). The Norman Digital Sample Freezing Platform (DSFP) was newly introduced for LC-MS data treatment. It was developed for the retrospective suspect screening of environmental pollutants (Alygizakis et al., 2019b).

1.5 Ion mobility spectrometry (IMS)

Ion mobility spectrometry (IMS) is an analytical technique that separates ions in their gaseous phases under the influence of an electric field. Its concept was first discovered by the physicist Paul Langevin in 1905 (Cumeras et al., 2015). During the 1950s and 1960s, E.W. McDaniel first coupled IMS to a magnetic sector mass spectrometer (McDaniel et al., 1962), which can be considered the beginning of IM-MS. By the 1970s, because of the commercially available IM-quadrupole mass spectrometer, IM-MS was able to analyze ions in the gas phase under ambient pres-

sure (Karasek et al., 1971). Since then, IMS has been developing into an inexpensive, powerful and portable analytical instrument to monitor gas phase samples (Cumeras et al., 2015). With the improvement in the sensitivity and miniaturization of modern IMS, its applications have been expanded to various fields beyond laboratory analysis, such as direct analysis in real time for chemical weapons monitoring, airport security, and air quality analysis (Cumeras et al., 2015). IMS-MS devices are generally constructed and used in academic research. In 2006, Waters invented the Synapt HDMS platform, and IMS-MS began to be widely applied in academic research (Morris et al., 2020; Dodds and Baker, 2019). Since then, various IMS-MS platforms have been commercialized by many instrument developers (Dodds and Baker, 2019). The history of IMS and IMS-MS development (by 2015) is plotted in Figure 1.9, and the domain of application has been flourishing and expanding as shown in Figure 1.10

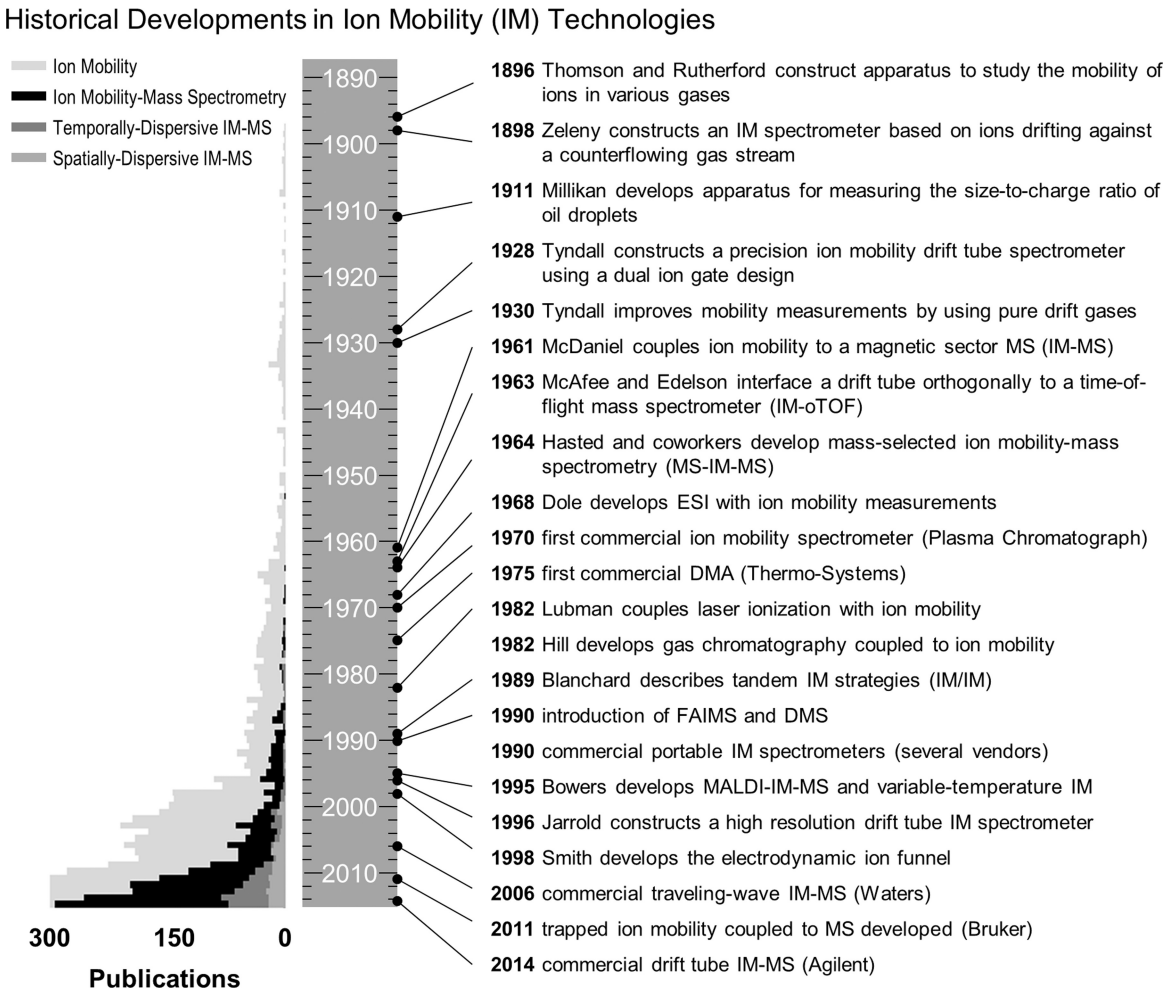


Figure 1.9: Development of IMS and numbers of publications (May and McLean, 2015)

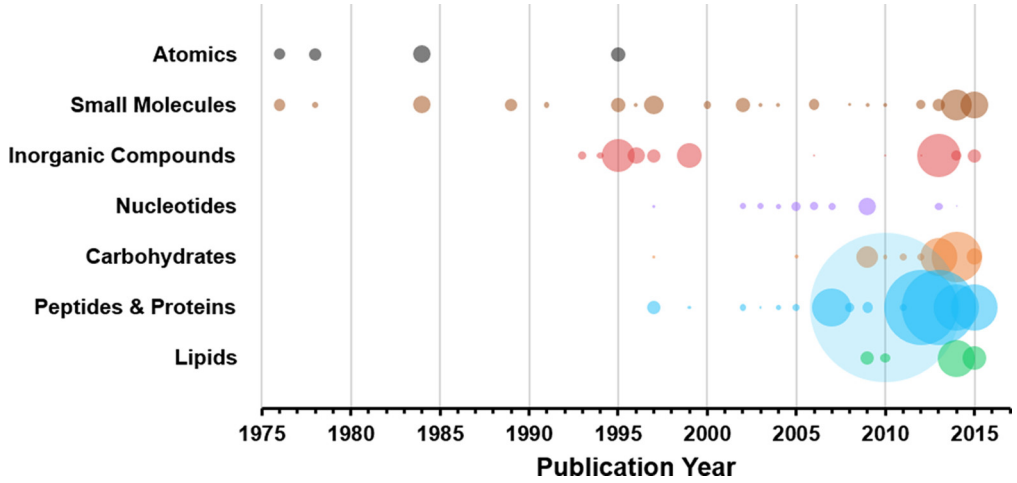


Figure 1.10: Numbers of CCS values reported in the represented domain (May et al., 2017)

IM-MS has been applied to proteomics (Zhong et al., 2012; Uetrecht et al., 2010), lipidomics (Paglia et al., 2015), metabolomics (Zhang et al., 2018), drug discovery (Lapthorn et al., 2013; Lanucara et al., 2014), etc. Most recently, IM-MS has gained increasing attention in small molecule analysis applied to biological and environmental monitoring (Lapthorn et al., 2013; Kaufmann, 2020). IM-HRMS combined with chromatographic separation has become a novel and promising approach in NTS analysis (Celma et al., 2020). In summary, IMS-MS has three main advantages: isomeric separation improvement, noise signal filtering, and feature annotation by the CCS database in NTA (Dodds and Baker, 2019).

1.5.1 Principle of IMS

IMS consists of an electric field filled with gas, called buffer gas (Gabelica, 2021). Ion species entering the IMS are accelerated by an electric force, while the collision between the ions and buffer gas causes a friction force, compensating for the acceleration of the ion speed. Thereby, the average speed, which is called the drift velocity, is constant (ν_d) (Gabelica, 2021), and it is proportional to the electric field. The mobility of the ion (K) is used to separate ion species and can be expressed as Equation 1.5:

$$K = \frac{\nu_{\mathrm{d}}}{E} \tag{1.5}$$

The collision cross section (CCS) is a physical parameter used to describe the area (cross section) in which collision between the ion and the buffer gas occurs (Delvaux et al., 2022). The CCS can be converted by the mobility K via the Mason-Schamp equation (Mason and Schamp Jr, 1958) (Equation 1.6), commonly denoted in Å^2 (square angstrom). Therefore, the CCS depends on the type of buffer gas and the gas temperature in the IMS platform. Reported CCS databases are conducted by four major IM techniques, which are summarized below.

$$CCS = \frac{3ze}{16N}\left(\frac{2\pi}{\mu k_{\mathrm{B}} T}\right)^{\frac{1}{2}} \frac{1}{K} \tag{1.6}$$

where:

z = ion charge
e = elementary charge
N = density number of the drift gas
μ = reduced mass of the ion-neutral drift gas pair
k_B = Boltzmann constant
T = gas temperature

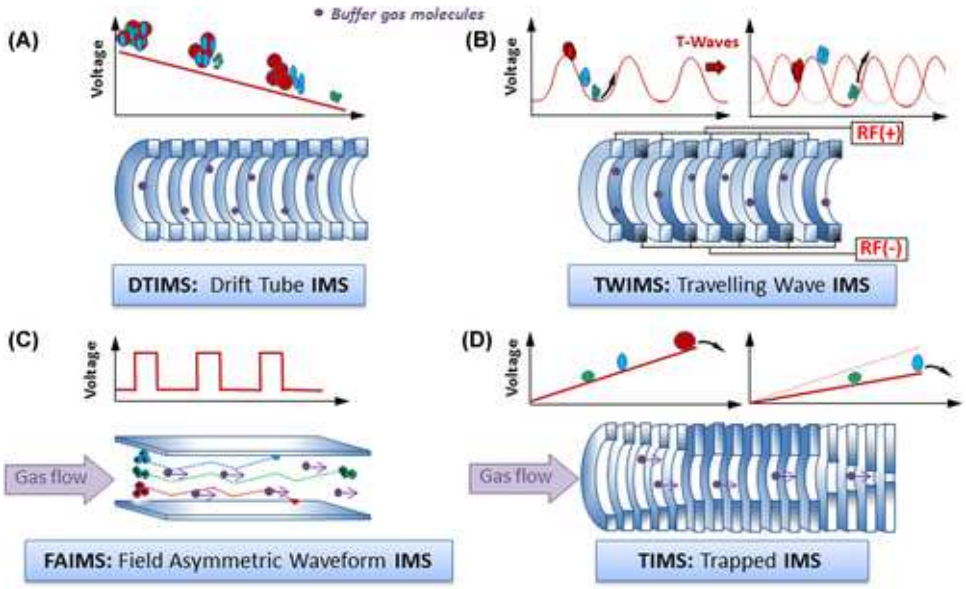


Figure 1.11: Four main types of commercial IMS

1.5.2 Most commonly used IMS separation technology

Drift Tube Ion Mobility Spectrometry (DTIMS)

DTIMS is the most traditional separation technique. The drift tube is a pressured chamber with a uniform electrical field. It measures the traveling time, which is called the drift time (t_d), that an ion needs to cross the drift tube (L_d). Since the separation of DTIMS uses a uniform electric field, the measured t_d is directly related to the mobility K; then, the mobility K can be converted to the CCS (Delvaux et al., 2022; Morris et al., 2020). Therefore, DTIMS is the only method that can directly measure the CCS (Kanu et al., 2008). In 2014, a high-performance drift tube IM-MS was released as a commercial product by Agilent Technologies (6560 ion mobility-QTOF). It was coupled with UHPLC to enable high-throughput analysis.

Traveling Wave Ion Mobility Spectrometry (TWIMS)

TWIMS was the first commercially successful IMS-MS platform with the Synapt HDMS introduced by Waters Corporation in 2006, and its next generations Synapt G2 (2011) and Synapt G2-Si (2013) are widespread in IMS-MS analysis (Celma et al., 2020; Hines et al., 2017). TWIMS has a separation principle similar to that of DTIMS (Morris et al., 2020). Both IM devices consist of a stacked ring ion guide, and ions are led by the electric field and dispersed by the different velocities between the ions and the buffer gas (Giles et al., 2004). TWIMS applies an oscillating electric

field to generate a set of traveling wave pulses that push ions out of the drift cell (Giles et al., 2004). TWIMS reaches a resolving power that is comparable to that of DTIMS (Dodds et al., 2017). However, unlike uniform field DTIMS, TWIMS does not require high voltages; therefore, with a shorter drift tube, TWIMS can reach a resolving power similar to that of DTIMS (Gabelica, 2021; Morris et al., 2020). In contrast, the traveling time of ions through the drift cell is not related to the mobility K, and a calibration protocol with ions of known mobility (e.g., polyalanine) must be performed before analysis. The precision is biased by the chemical class (May and McLean, 2015). The calibration procedures need to be improved.

Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS)

FAIMS, differential mobility spectrometry (DMS) and differential ion mobility spectrometry (DIMS) operate under the same mechanism but construct in different geometries (Dodds and Baker, 2019). Briefly, a drift region consists of two planar electrodes and an electrometer as a detector, and gas-phase ion species are separated based on their different mobilities in high (20,000 V/cm) and low (1000 V/cm) electric fields (Kolakowski and Mester, 2007). Ions entering the drift cell follow different trajectories due to the changing electric field; only analytes that match the applied compensation voltage (CV) can traverse the drift cell and be scanned (Dodds and Baker, 2019). Herein, FAIMS enables a high selectivity, but it cannot provide the CCS values.

Trapped ion mobility spectrometry (TIMS)

In contrast to DTIMS and TWIMS, ion species of TIMS are trapped by an electric field and pushed by a moving gas. The TIMS-HRMS platform was first released by Bruker Daltonics in 2014. The TIMS analyzer consists of a set of electrodes, including the entrance funnel, TIMS tunnel, and exit funnel (Ridgeway et al., 2018). The TIMS analyzer is placed after a proceeding atmospheric pressure ionization source, e.g., ESI, and before the mass detector, as shown in Table. 1.1. Ions are led by gas through a capillary and exit funnel region after separation. In the TIMS analyzer, an RF voltage is applied to the electrodes to produce a radially confining pseudopotential. Additionally, an axial electric field gradient (EFG) is generated by superimposing direct current (DC) potentials on each of the funnel and tunnel electrodes. The separation procedure includes 3 steps (Figure 1.12):

1. Ions are led to the TIMS analyzer as a gas through a capillary after ionization and accumulated/trapped within a fixed period of time. The deflection plate is set to a repulsive potential that pushes ions into the entrance funnel. After traversing the entrance funnel, ions enter the analyzer and pass through the EFG profile until the drift velocity of the ions (v_d) is equal and opposite to the velocity of the buffer gas (v_g). Thus, the ions reach an equilibrium position, and the ions with the same v_d, in other words, mobility K, accumulate in the analyzer. The ions with higher mobilities are trapped in a lower electric field E, which is near the entrance of the tunnel.

2. After ions are confined in the TIMS tunnel, the deflector switches to an attractive potential to prevent more ions from entering the funnel. Ions in the analyzer funnel are trapped under a user-defined time period.

3. The magnitude of the EFG profile decreases gradually to its initial value. Thereby, ions progressively exist in the analyzer funnel from high to low K.

4. While the separation and analysis are occurred in the second section of analyzer, new ions are collected and accumulated in the first section waiting for the next separation sequence, as shown in figure 1.12(b). The parallel accumulation/analysis promote the duty cycle up to 100%.

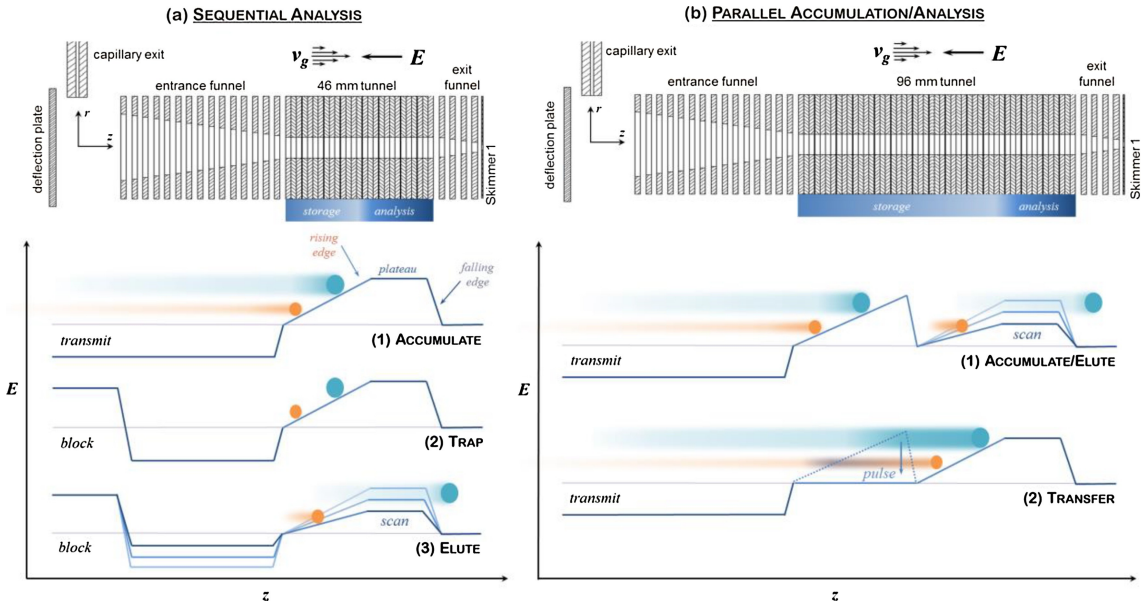


Figure 1.12: TIMS funnels
(Ridgeway et al., 2018)

1.5.3 Use of IMS in the NTS

The IMS and CCS values are recommended for use in nontarget analysis as an additional separation dimension. IMS has been demonstrated to complement gas or liquid chromatography and mass spectrometry and can be divided into 3 main applications: isomer separation, mass spectrum filtering, and annotation of unknown chemicals.

Isomer separation

Conventionally, isomers are discriminated through RT and fragmentation. Isomers have similar structures, resulting in similar polarity and mass spectra, and it is challenging to separate them by chromatographic techniques and mass analysis. Hence, IMS provides an alternative separation approach based on ion structural differences in mobility.

Per- and polyfluoroalkyl substances (PFAS) consist of a large group of synthetic organic compounds that have been extensively used in industrial products (Wang et al., 2017). de Vega et al. (2021) demonstrated that the drift time (CCS) enhances isomeric analysis for poor chromatographic separation species. The 5-trifluoromethyl isomer and 6-trifluoromethyl had ΔRT at 0.1 min, while ΔCCS was larger than 2.3 Å^2. In other halogenated POPs and their metabolites, such as PCBs and OH-PCBs,

BDE and OH-BDEs yield similar mass spectra. Furthermore, some can be better separated by IMS (such as 3-OH-BDE-47 and 2'-OH-BDE-68), but similar CCS ranges were generally observed within isomers (Adams et al., 2016; Zheng et al., 2018).

Mass spectral filtering

All ions within a selected m/z range are fragmented in the DIA mode. Therefore, not only are the compounds of interest fragmented, but also other coeluting compounds are fragmented, which complicates the mass spectrum interpretation. DIA analyses combined with IMS is a wise approach to cover nontarget analysis and structure elucidation. IMS is generally placed before the MS/MS collision cell, precursor ion and its fragments have the same CCS (drift time), enabling the association of fragmentation and precursor ions. Hence, it facilitates DIA mass spectrum interpretation and structural elucidation (Yukioka et al., 2021). Moreover, the drift time alignment produces a highly selective mass spectrum acquired by DIA, improving the identification confidence point (Celma et al., 2020).

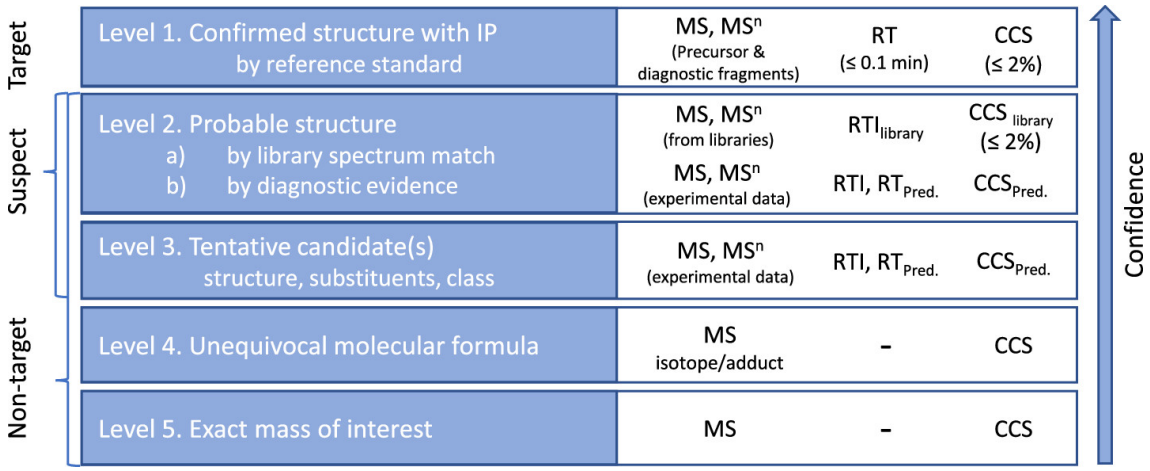


Figure 1.13: Five levels of identification confidence with CCS proposed by (Celma et al., 2020).

Annotation of the unknown

The RT can shift during sample analyses; thiabendazole was detected in the Mediterranean basin, and the deviation of the RT was + 0.24 min, while the CCS value only deviated by -0.12 % (Celma et al., 2020). This finding supports that combining the RT and CCS alignment can enhance the identification certainty and ratio (Figure 1.13. The m/z-CCS trend lines have also been discussed for various chemical families, including PFAs, PAHs, PCBs, metabolites and other xenobiotics (Foster et al., 2022; Belova et al., 2021; Zheng et al., 2017). The association of the m/z with the CCS offers a novel aspect of structural information and compound families, especially for the "unknown-unknown" structure elucidation. Experimental and predicted CCS databases are integrated into a nontarget analysis workflow to reduce candidate numbers (Menger et al., 2022). Meanwhile, the predicted CCS results in larger deviations with insufficient training of the chemical structure; therefore, manual verification is required in the annotation results (Menger et al., 2022).

1.6 Conclusion

The present chapter illustrates the expanded use of synthetic chemicals and their release into the environment. Chemicals such as POPs can be highly persistent and have accumulated in the environment and biota for decades. Polar and emerging contaminants, such as pharmaceutical products, easily enter aquatic systems via wastewater. Pesticides are widely present in surface and ground water due to agricultural runoff. Studying the legacy and emerging contaminants in samples helps to assess the contaminants occurrences and toxicity. In this context, the combination of target and nontarget analyses enables a broader screening of contaminants. Modern HRMS enables high scan speed and wide-scope analysis, which is satisfactory for environmental nontarget analysis. However, the lack of a priori knowledge of emerging contaminants leads to challenges in data treatment. The development of nontarget analysis requires an understanding of analytical methodology and instrumentation; in the meantime, it also requires knowledge of data treatment strategies and statistics. On the other hand, handling massive HRMS data is a time-consuming and labor-intensive task, and choosing suitable data treatment tools and discovering the functionality can improve the producibility of the work. Furthermore, it is worth understanding the algorithm and the user-defined parameters. Meanwhile, IMS introduces a structural separation dimension, complementing chromatographic separation and mass spectrometry. The advantages of combining IMS with HRMS are still being studied. The CCS value brings a novel identification property, which can improve isomer separation and prevent false positives. Therefore, the CCS alignment is suggested in nontarget analyses. The deviation of experimental and predicted CCS values, thereby, should be estimated to inspect matched candidates. The minor deviations of the CCS values indicate the potential of expanding the CCS database with regards to nontarget analyses.

In this context, this current PhD research addresses the following points.

- Halogenated POPs are developed using GC-APCI-HRMS and ion mobility spectrometry. Based on the specialty of PCB and PBDE congeners in structure and isotopic profiles, the benefits of combining APCI with IMS are evaluated.
- An environmental contaminant analysis workflow is developed for water samples. Target screening was optimized with an in-house database. Different data processing approaches were tested to perform nontarget analysis.
- The target method was optimized for steroid hormone analysis. The benefit of using IMS to filter noise peaks in a complex matrix was tested.
- The CCS prediction models were developed based on machine learning. This is the first time that a molecular fingerprint was used to predict the CCS value. Meanwhile, the merged dataset aimed to verify the compatibility of the cross-platform CCS database and develop an instrument-independent prediction model.

References

R. Aalizadeh, N. S. Thomaidis, A. A. Bletsou, and P. Gago-Ferrero. Quantitative structure–retention relationship models to support nontarget high-resolution mass spectrometric screening of emerging contaminants in environmental samples. *Journal of Chemical Information and Modeling*, 56(7):1384–1398, 2016.

R. Aalizadeh, V. Nikolopoulou, N. A. Alygizakis, and N. S. Thomaidis. First novel workflow for semiquantification of emerging contaminants in environmental samples analyzed by gas chromatography–atmospheric pressure chemical ionization–quadrupole time of flight–mass spectrometry. *Analytical chemistry*, 94(27):9766–9774, 2022.

K. J. Adams, D. Montero, D. Aga, and F. Fernandez-Lima. Isomer separation of polybrominated diphenyl ether metabolites using nanoesi-tims-ms. *International journal for ion mobility spectrometry*, 19(2):69–76, 2016.

A. O. Affum, S. O. Acquaah, S. D. Osae, and E. E. Kwaansa-Ansah. Distribution and risk assessment of banned and other current-use pesticides in surface and groundwaters consumed in an agricultural catchment dominated by cocoa crops in the ankobra basin, ghana. *Science of The Total Environment*, 633:630–640, 2018.

A. Agüera, M. J. Martínez Bueno, and A. R. Fernández-Alba. New trends in the analytical determination of emerging contaminants and their transformation products in environmental waters. *Environmental Science and Pollution Research*, 20 (6):3496–3515, 2013.

G. Alvarez-Rivera, D. Ballesteros-Vivas, F. Parada-Alfonso, E. Ibañez, and A. Cifuentes. Recent applications of high resolution mass spectrometry for the characterization of plant natural products. *TrAC Trends in Analytical Chemistry*, 112: 87–101, 2019.

N. A. Alygizakis, S. Samanipour, J. Hollender, M. Ibáñez, S. Kaserzon, V. Kokkali, J. A. Van Leerdam, J. F. Mueller, M. Pijnappels, M. J. Reid, et al. Exploring the potential of a global emerging contaminant early warning network through the use of retrospective suspect screening with high-resolution mass spectrometry. *Environmental science & technology*, 52(9):5135–5144, 2018.

N. A. Alygizakis, H. Besselink, G. K. Paulus, P. Oswald, L. M. Hornstra, M. Oswaldova, G. Medema, N. S. Thomaidis, P. A. Behnisch, and J. Slobodnik. Characterization of wastewater effluents in the danube river basin with chemical screening, in vitro bioassays and antibiotic resistant genes analysis. *Environment international*, 127:420–429, 2019a.

N. A. Alygizakis, P. Oswald, N. S. Thomaidis, E. L. Schymanski, R. Aalizadeh, T. Schulze, M. Oswaldova, and J. Slobodnik. Norman digital sample freezing platform: A european virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in "digitally frozen" environmental samples. *TrAC Trends in Analytical Chemistry*, 115:129–137, 2019b.

S.-L. Badea, E.-I. Geana, V.-C. Niculescu, and R.-E. Ionete. Recent progresses in analytical gc and lc mass spectrometric based-methods for the detection of emerging chlorinated and brominated contaminants and their transformation products in aquatic environment. *Science of The Total Environment*, 722:137914, 2020.

S. Bäurer, W. Guo, S. Polnick, and M. Lämmerhofer. Simultaneous separation of water-and fat-soluble vitamins by selective comprehensive hilic× rplc (high-resolution sampling) and active solvent modulation. *Chromatographia*, 82(1):167–180, 2019.

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

L. Belova, N. Caballero-Casero, A. L. van Nuijs, and A. Covaci. Ion mobility-high-resolution mass spectrometry (im-hrms) for the analysis of contaminants of emerging concern (cecs): Database compilation and application to urine samples. *Analytical Chemistry*, 93(16):6428–6436, 2021.

L. G. Berruex and R. Freitag. Separation and purification of biochemicals. 2003.

S. Bieber, G. Greco, S. Grosse, and T. Letzel. Rplc-hilic and sfc with mass spectrometry: polarity-extended organic molecule screening in environmental (water) samples. *Analytical chemistry*, 89(15):7907–7914, 2017.

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

A. M. Brunner, C. Bertelkamp, M. M. Dingemans, A. Kolkman, B. Wols, D. Harmsen, W. Siegers, B. J. Martijn, W. A. Oorthuizen, and T. L. Ter Laak. 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, 2020.

N. Caballero-Casero, L. Belova, P. Vervliet, J.-P. Antignac, A. Castano, L. Debrauwer, M. E. Lopez, C. Huber, J. Klánová, M. Krauss, et al. Towards harmonised criteria in quality assurance and quality control of suspect and non-target lc-hrms analytical workflows for screening of emerging contaminants in human biomonitoring. *TrAC Trends in Analytical Chemistry*, 136:116201, 2021.

C. Campanale, C. Massarelli, D. Losacco, D. Bisaccia, M. Triozzi, and V. F. Uricchio. The monitoring of pesticides in water matrices and the analytical criticalities: A review. *TrAC Trends in Analytical Chemistry*, 144:116423, 2021.

R. Carson. Silent spring. In *Thinking About the Environment*, pages 150–155. Routledge, 2015.

A. Ccanccapa-Cartagena, Y. Pico, X. Ortiz, and E. J. Reiner. Suspect, non-target and target screening of emerging pollutants using data independent acquisition: Assessment of a mediterranean river basin. *Science of the total environment*, 687: 355–368, 2019.

A. Celma, J. V. Sancho, E. L. Schymanski, D. Fabregat-Safont, M. Ibanez, J. Goshawk, G. Barknowitz, F. Hernandez, and L. Bijlsma. Improving target and suspect screening high-resolution mass spectrometry workflows in environmental analysis by ion mobility separation. *Environmental Science & Technology*, 54(23):15120–15131, 2020.

T. S. Convention. The stockholm convention, 2001. URL http://www.pops.int/.

R. Cumeras, E. Figueras, C. E. Davis, J. I. Baumbach, and I. Gracia. Review on ion mobility spectrometry. part 1: current instrumentation. *Analyst*, 140(5): 1376–1390, 2015.

R. G. de Vega, A. Cameron, D. Clases, T. M. Dodgen, P. A. Doble, and D. P. Bishop. Simultaneous targeted and non-targeted analysis of per-and polyfluoroalkyl substances in environmental samples by liquid chromatography-ion mobility-quadrupole time of flight-mass spectrometry and mass defect analysis. *Journal of Chromatography A*, 1653:462423, 2021.

A. Delvaux, E. Rathahao-Paris, and S. Alves. Different ion mobility-mass spectrometry coupling techniques to promote metabolomics. *Mass Spectrometry Reviews*, 41(5):695–721, 2022.

K. Diamanti, R. Aalizadeh, N. Alygizakis, A. Galani, M. Mardal, and N. S. Thomaidis. Wide-scope target and suspect screening methodologies to investigate the occurrence of new psychoactive substances in influent wastewater from athens. *Science of the Total Environment*, 685:1058–1065, 2019.

J. N. Dodds and E. S. Baker. Ion mobility spectrometry: fundamental concepts, instrumentation, applications, and the road ahead. *Journal of the American Society for Mass Spectrometry*, 30(11):2185–2195, 2019.

J. N. Dodds, J. C. May, and J. A. McLean. Correlating resolving power, resolution, and collision cross section: unifying cross-platform assessment of separation efficiency in ion mobility spectrometry. *Analytical chemistry*, 89(22):12176–12184, 2017.

A. El-Aneed, A. Cohen, and J. Banoub. Mass spectrometry, review of the basics: electrospray, maldi, and commonly used mass analyzers. *Applied Spectroscopy Reviews*, 44(3):210–230, 2009.

A. K. El-Deen and K. Shimizu. Modified μ-quechers coupled to diethyl carbonate-based liquid microextraction for pahs determination in coffee, tea, and water prior to gc–ms analysis: An insight to reducing the impact of caffeine on the gc–ms measurement. *Journal of Chromatography B*, 1171:122555, 2021.

S. Eliuk and A. Makarov. Evolution of orbitrap mass spectrometry instrumentation. *Annu. Rev. Anal. Chem*, 8(1):61–80, 2015.

EPA. *Forth Unregulated Contaminant Monitoring Rule* , 2021a. www.epa.gov/dwucmr/fourth-unregulated-contaminant-monitoring-rule [Accessed: 27. 12. 2021].

EPA. *Fifth Unregulated Contaminant Monitoring Rule* , 2021b. https://www.epa.gov/dwucmr/fifth-unregulated-contaminant-monitoring-rule [Accessed: 27. 12. 2021].

EPA. *DDT*, 2022. https://www.epa.gov/ingredients-used-pesticide-products/ddt-brief-history-and-status [Accessed: 01. 10. 2022].

EU. *Directive (EU) 2020/2184*, 2021a. https://eur-lex.europa.eu/eli/dir/2020/2184/oj [Accessed: 07. 10. 2022].

EU. *EU The Drinking Water Directive*, 2021b. https://ec.europa.eu/environment/water/water-drink/legislation_en.html [Accessed: 07. 10. 2022].

E. Eysseric, C. Gagnon, and P. A. Segura. Identifying congeners and transformation products of organic contaminants within complex chemical mixtures in impacted surface waters with a top-down non-targeted screening workflow. *Science of The Total Environment*, 822:153540, 2022.

J. Fang, H. Zhao, Y. Zhang, M. Lu, and Z. Cai. Atmospheric pressure chemical ionization in gas chromatography-mass spectrometry for the analysis of persistent organic pollutants. *Trends in Environmental Analytical Chemistry*, 25:e00076, 2020.

I. Ferrer, D. L. Sweeney, E. M. Thurman, and J. A. Zweigenbaum. Nontargeted screening of water samples using data-dependent acquisition with similar partition searching. *Journal of the American Society for Mass Spectrometry*, 31(6):1189–1204, 2020.

M. Foster, M. Rainey, C. Watson, J. N. Dodds, K. I. Kirkwood, F. M. Fernández, and E. S. Baker. Uncovering pfas and other xenobiotics in the dark metabolome using ion mobility spectrometry, mass defect analysis, and machine learning. *Environmental Science & Technology*, 2022.

V. Gabelica. Ion mobility–mass spectrometry: an overview. 2021.

P. Gago-Ferrero, A. A. Bletsou, D. E. Damalas, R. Aalizadeh, N. A. Alygizakis, H. P. Singer, J. Hollender, and N. S. Thomaidis. Wide-scope target screening of> 2000 emerging contaminants in wastewater samples with uplc-q-tof-hrms/ms and smart evaluation of its performance through the validation of 195 selected representative analytes. *Journal of hazardous materials*, 387:121712, 2020.

K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose, and R. H. Bateman. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry*, 18(20):2401–2414, 2004.

O. Golovko, S. Örn, M. Sörengård, K. Frieberg, W. Nassazzi, F. Y. Lai, and L. Ahrens. Occurrence and removal of chemicals of emerging concern in wastewater treatment plants and their impact on receiving water systems. *Science of The Total Environment*, 754:142122, 2021.

M. Guilhaus. Special feature: Tutorial. principles and instrumentation in time-of-flight mass spectrometry. physical and instrumental concepts. *Journal of mass spectrometry*, 30(11):1519–1532, 1995.

Z. Guo, S. Huang, J. Wang, and Y.-L. Feng. Recent advances in non-targeted screening analysis using liquid chromatography-high resolution mass spectrometry to explore new biomarkers for human exposure. *Talanta*, 219:121339, 2020.

F. Hernández, T. Portolés, E. Pitarch, and F. J. López. Gas chromatography coupled to high-resolution time-of-flight mass spectrometry to analyze trace-level organic compounds in the environment, food safety and toxicology. *TrAC Trends in Analytical Chemistry*, 30(2):388–400, 2011.

F. Hernandez, J. V. Sancho, M. Ibáñez, E. Abad, T. Portolés, and L. Mattioli. Current use of high-resolution mass spectrometry in the environmental sciences. *Analytical and bioanalytical chemistry*, 403(5):1251–1264, 2012.

F. Hernández, J. Bakker, L. Bijlsma, J. De Boer, A. M. Botero-Coy, Y. B. de Bruin, S. Fischer, J. Hollender, B. Kasprzyk-Hordern, M. Lamoree, et al. The role of analytical chemistry in exposure science: Focus on the aquatic environment. *Chemosphere*, 222:564–583, 2019.

K. M. Hines, D. H. Ross, K. L. Davidson, M. F. Bush, and L. Xu. Large-scale structural characterization of drug and drug-like compounds by high-throughput ion mobility-mass spectrometry. *Analytical chemistry*, 89(17):9023–9030, 2017.

C. S. Ho, C. Lam, M. Chan, R. Cheung, L. Law, L. Lit, K. Ng, M. Suen, and H. Tai. Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical Biochemist Reviews*, 24(1):3, 2003.

J. Hollender, E. L. Schymanski, H. P. Singer, and P. L. Ferguson. Nontarget screening with high resolution mass spectrometry in the environment: ready to go?, 2017.

T. H. Hutchinson, B. P. Lyons, J. E. Thain, and R. J. Law. Evaluating legacy contaminants and emerging chemicals in marine environments using adverse outcome pathways and biological effects-directed analysis. *Marine pollution bulletin*, 74(2): 517–525, 2013.

A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, and H. H. Hill Jr. Ion mobility–mass spectrometry. *Journal of mass spectrometry*, 43(1):1–22, 2008.

F. W. Karasek, M. J. Cohen, and D. I. Carroll. Trace studies of alcohols in the plasma chromatograph—mass spectrometer. *Journal of Chromatographic science*, 9(7):390–392, 1971.

A. Kaufmann. The use of uhplc, ims, and hrms in multiresidue analytical methods: A critical review. *Journal of Chromatography B*, 1158:122369, 2020.

S. Y. Kimura, A. A. Cuthbertson, J. D. Byer, and S. D. Richardson. The dbp exposome: development of a new method to simultaneously quantify priority disinfection by-products and comprehensively identify unknowns. *Water research*, 148:324–333, 2019.

B. M. Kolakowski and Z. Mester. Review of applications of high-field asymmetric waveform ion mobility spectrometry (faims) and differential mobility spectrometry (dms). *Analyst*, 132(9):842–864, 2007.

M. Krauss, C. Hug, R. Bloch, T. Schulze, and W. Brack. Prioritising site-specific micropollutants in surface water from lc-hrms non-target screening data using a rarity score. *Environmental Sciences Europe*, 31(1):1–12, 2019.

J. Krier, R. R. Singh, T. Kondić, A. Lai, P. Diderich, J. Zhang, P. A. Thiessen, E. E. Bolton, and E. L. Schymanski. Discovering pesticides and their tps in luxembourg waters using open cheminformatics approaches. *Environment International*, 158: 106885, 2022.

S. Lambert and M. Wagner. Microplastics are contaminants of emerging concern in freshwater environments: an overview. *Freshwater microplastics*, pages 1–23, 2018.

F. Lanucara, S. W. Holman, C. J. Gray, and C. E. Eyers. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature chemistry*, 6(4):281–294, 2014.

C. Lapthorn, F. Pullen, and B. Z. Chowdhry. Ion mobility spectrometry-mass spectrometry (ims-ms) of small molecules: Separating and assigning structures to ions. *Mass spectrometry reviews*, 32(1):43–71, 2013.

E. legislation. *Persistent chemicals*, 2003. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_03_219 [Accessed: 01. 10. 2022].

D.-X. Li, L. Gan, A. Bronja, and O. J. Schmitz. Gas chromatography coupled to atmospheric pressure ionization mass spectrometry (gc-api-ms). *Analytica chimica acta*, 891:43–61, 2015.

J.-L. Liu and M.-H. Wong. Pharmaceuticals and personal care products (ppcps): a review on environmental contamination in china. *Environment international*, 59: 208–224, 2013.

B. Mamyrin. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry*, 206(3):251–266, 2001.

L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W. H. Tang, A. Römpp, S. Neumann, A. D. Pizarro, et al. mzml—a community standard for mass spectrometry data. *Molecular & Cellular Proteomics*, 10(1), 2011.

E. A. Mason and H. W. Schamp Jr. Mobility of gaseous lons in weak electric fields. *Annals of physics*, 4(3):233–270, 1958.

J. C. May and J. A. McLean. Ion mobility-mass spectrometry: time-dispersive instrumentation. *Analytical chemistry*, 87(3):1422–1436, 2015.

J. C. May, C. B. Morris, and J. A. McLean. Ion mobility collision cross section compendium. *Analytical chemistry*, 89(2):1032–1044, 2017.

D. Mazur, E. Detenchuk, A. Sosnova, V. Artaev, and A. Lebedev. Gc-hrms with complementary ionization techniques for target and non-target screening for chemical exposure: Expanding the insights of the air pollution markers in moscow snow. *Science of the Total Environment*, 761:144506, 2021.

E. McDaniel, D. Martin, and W. Barnes. Drift tube-mass spectrometer for studies of low-energy ion-molecule reactions. *Review of Scientific Instruments*, 33(1):2–7, 1962.

F. Menger, P. Gago-Ferrero, K. Wiberg, and L. Ahrens. Wide-scope screening of polar contaminants of concern in water: A critical review of liquid chromatography-high resolution mass spectrometry-based strategies. *Trends in Environmental Analytical Chemistry*, 28:e00102, 2020.

F. Menger, G. Boström, O. Jonsson, L. Ahrens, K. Wiberg, J. Kreuger, and P. Gago-Ferrero. Identification of pesticide transformation products in surface water using suspect screening combined with national monitoring data. *Environmental science & technology*, 55(15):10343–10353, 2021.

F. Menger, A. Celma, E. L. Schymanski, F. Y. Lai, L. Bijlsma, K. Wiberg, F. Hernández, J. V. Sancho, and L. Ahrens. Enhancing spectral quality in complex environmental matrices: Supporting suspect and non-target screening in zebra mussels with ion mobility. *Environment International*, page 107585, 2022.

C. B. Morris, J. C. Poland, J. C. May, and J. A. McLean. Fundamentals of ion mobility-mass spectrometry for the analysis of biomolecules. *Ion Mobility-Mass Spectrometry*, pages 1–31, 2020.

D. Muir, X. Zhang, C. A. De Wit, K. Vorkamp, and S. Wilson. Identifying further chemicals of emerging arctic concern based on 'in silico'screening of chemical inventories. *Emerging Contaminants*, 5:201–210, 2019.

K. A. Murrell and F. L. Dorman. A comparison of liquid-liquid extraction and stir bar sorptive extraction for multiclass organic contaminants in wastewater by comprehensive two-dimensional gas chromatography time of flight mass spectrometry. *Talanta*, 221:121481, 2021.

I. Neuwald, M. Muschket, D. Zahn, U. Berger, B. Seiwert, T. Meier, J. Kuckelkorn, C. Strobel, T. P. Knepper, and T. Reemtsma. 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, 2021.

NIST. *DElectron Ionization Library Component of the NIST/EPA/NIH Mass Spectral Library and NIST GC Retention Index Databaseirective (EU) 2020/2184*, 2020. https://www.nist.gov/programs-projects/electron-ionization-library-component-nistepanih-mass-spectral-library-and-nist-gc [Accessed: 07. 10. 2022].

G. Paglia, M. Kliman, E. Claude, S. Geromanos, and G. Astarita. Applications of ion-mobility mass spectrometry for lipid analysis. *Analytical and bioanalytical chemistry*, 407(17):4995–5007, 2015.

M. Paszkiewicz, K. Godlewska, H. Lis, M. Caban, A. Białk-Bielińska, and P. Stepnowski. Advances in suspect screening and non-target analysis of polar emerging contaminants in the environmental monitoring. *TrAC Trends in Analytical Chemistry*, page 116671, 2022.

P. G. Pedrioli, J. K. Eng, R. Hubley, M. Vogelzang, E. W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R. H. Angeletti, R. Apweiler, et al. A common open representation of mass spectrometry data and its application to proteomics research. *Nature biotechnology*, 22(11):1459–1466, 2004.

L. Perez de Souza, S. Alseekh, F. Scossa, and A. R. Fernie. Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. *Nature Methods*, 18(7):733–746, 2021.

A. Peris and E. Eljarrat. Multi-residue methodologies for the analysis of non-polar pesticides in water and sediment matrices by gc–ms/ms. *Chromatographia*, 84(5): 425–439, 2021.

T. Pluskal, S. Castillo, A. Villar-Briones, and M. Orešič. Mzmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC bioinformatics*, 11(1):1–11, 2010.

C. Postigo, A. Andersson, M. Harir, D. Bastviken, M. Gonsior, P. Schmitt-Kopplin, P. Gago-Ferrero, L. Ahrens, L. Ahrens, and K. Wiberg. Unraveling the chemodiversity of halogenated disinfection by-products formed during drinking water treatment using target and non-target screening tools. *Journal of Hazardous Materials*, 401:123681, 2021.

Y. Qian, X. Wang, G. Wu, L. Wang, J. Geng, N. Yu, and S. Wei. Screening priority indicator pollutants in full-scale wastewater treatment plants by non-target analysis. *Journal of Hazardous Materials*, 414:125490, 2021.

J. Qiu, C. Craven, N. Wawryk, K. Carroll, and X.-F. Li. Integration of solid phase extraction with hilic-ms/ms for analysis of free amino acids in source water. *Journal of Environmental Sciences*, 2022.

H. Ren, R. Tröger, L. Ahrens, K. Wiberg, and D. Yin. Screening of organic micropollutants in raw and drinking water in the yangtze river delta, china. *Environmental Sciences Europe*, 32(1):1–12, 2020.

J. Rice, A. Lubben, and B. Kasprzyk-Hordern. A multi-residue method by supercritical fluid chromatography coupled with tandem mass spectrometry method for the analysis of chiral and non-chiral chemicals of emerging concern in environmental samples. *Analytical and bioanalytical chemistry*, 412(23):5563–5581, 2020.

S. D. Richardson. Disinfection by-products: formation and occurrence in drinking water. 2011.

S. D. Richardson and S. Y. Kimura. Water analysis: emerging contaminants and current issues. *Analytical Chemistry*, 92(1):473–505, 2019.

M. E. Ridgeway, M. Lubeck, J. Jordens, M. Mann, and M. A. Park. Trapped ion mobility spectrometry: A short review. *International Journal of Mass Spectrometry*, 425:22–35, 2018.

C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, and S. Neumann. Metfrag relaunched: incorporating strategies beyond in silico fragmentation. *Journal of cheminformatics*, 8(1):1–16, 2016.

S. Samanipour, M. J. Reid, K. Bæk, and K. V. Thomas. Combining a deconvolution and a universal library search algorithm for the nontarget analysis of data-independent acquisition mode liquid chromatography- high-resolution mass spectrometry results. *Environmental science & technology*, 52(8):4694–4701, 2018.

S. Sauvé and M. Desrosiers. A review of what is an emerging contaminant. *Chemistry Central Journal*, 8(1):1–7, 2014.

J. E. Schollée, M. Bourgin, U. von Gunten, C. S. McArdell, and J. Hollender. Non-target screening to trace ozonation transformation products in a wastewater treatment train including different post-treatments. *Water research*, 142: 267–278, 2018.

B. Schulze, Y. Jeon, S. Kaserzon, A. L. Heffernan, P. Dewapriya, J. O'Brien, M. J. G. Ramos, S. G. Gorji, J. F. Mueller, K. V. Thomas, et al. 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, 2020a.

S. Schulze, H. Paschke, T. Meier, M. Muschket, T. Reemtsma, and U. Berger. A rapid method for quantification of persistent and mobile organic substances in water using supercritical fluid chromatography coupled to high-resolution mass spectrometry. *Analytical and bioanalytical chemistry*, 412(20):4941–4952, 2020b.

E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H. P. Singer, and J. Hollender. Identifying small molecules via high resolution mass spectrometry: communicating confidence, 2014.

E. L. Schymanski, H. P. Singer, J. Slobodnik, I. M. Ipolyi, P. Oswald, M. Krauss, T. Schulze, P. Haglund, T. Letzel, S. Grosse, et al. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Analytical and bioanalytical chemistry*, 407(21):6237–6255, 2015.

R. M. Sjerps, D. Vughs, J. A. van Leerdam, T. L. Ter Laak, and A. P. van Wezel. Data-driven prioritization of chemicals for various water types using suspect screening lc-hrms. *Water research*, 93:254–264, 2016.

J. C. Sousa, A. R. Ribeiro, M. O. Barbosa, M. F. R. Pereira, and A. M. Silva. A review on environmental monitoring of water organic pollutants identified by eu guidelines. *Journal of hazardous materials*, 344:146–162, 2018.

R. Tautenhahn, G. J. Patti, D. Rinehart, and G. Siuzdak. Xcms online: a web-based platform to process untargeted metabolomic data. *Analytical chemistry*, 84(11): 5035–5039, 2012.

A. C. Taylor, G. R. Fones, and G. A. Mills. Trends in the use of passive sampling for monitoring polar pesticides in water. *Trends in Environmental Analytical Chemistry*, 27:e00096, 2020.

R. Tröger, P. Klöckner, L. Ahrens, and K. Wiberg. Micropollutants in drinking water from source to tap-method development and application of a multiresidue screening method. *Science of the total environment*, 627:1404–1432, 2018.

R. Tröger, H. Ren, D. Yin, C. Postigo, P. D. Nguyen, C. Baduel, O. Golovko, F. Been, H. Joerss, M. R. Boleda, et al. What's in the water?–target and suspect screening of contaminants of emerging concern in raw water and drinking water from europe and asia. *Water research*, 198:117099, 2021.

C. Tsaridou and A. J. Karabelas. Drinking water standards and their implementation—a critical assessment. *Water*, 13(20):2918, 2021.

H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, and M. Arita. Ms-dial: data-independent ms/ms deconvolution for comprehensive metabolome analysis. *Nature methods*, 12(6):523–526, 2015.

C. Uetrecht, R. J. Rose, E. van Duijn, K. Lorenzen, and A. J. Heck. Ion mobility mass spectrometry of proteins and protein assemblies. *Chemical Society Reviews*, 39(5):1633–1655, 2010.

G. Vandermeersch, H. M. Lourenço, D. Alvarez-Muñoz, S. Cunha, J. Diogène, G. Cano-Sancho, J. J. Sloth, C. Kwadijk, D. Barcelo, W. Allegaert, et al. Environmental contaminants of emerging concern in seafood–european database on contaminant levels. *Environmental Research*, 143:29–45, 2015.

M.-A. Vaudreuil, S. V. Duy, G. Munoz, A. Furtos, and S. Sauvé. A framework for the analysis of polar anticancer drugs in wastewater: on-line extraction coupled to hilic or reverse phase lc-ms/ms. *Talanta*, 220:121407, 2020.

R. Vermeulen, E. L. Schymanski, A.-L. Barabási, and G. W. Miller. The exposome and health: Where chemistry meets biology. *Science*, 367(6476):392–396, 2020.

Z. Wang, J. C. DeWitt, C. P. Higgins, and I. T. Cousins. A never-ending story of per-and polyfluoroalkyl substances (pfass)?, 2017.

J. Wei, L. Xiang, and Z. Cai. Emerging environmental pollutants hydroxylated polybrominated diphenyl ethers: From analytical methods to toxicology research. *Mass Spectrometry Reviews*, 40, 05 2021. doi: 10.1002/mas.21640.

L. Wiest, A. Gosset, A. Fildier, C. Libert, M. Hervé, E. Sibeud, B. Giroud, E. Vulliet, T. Bastide, P. Polomé, et al. Occurrence and removal of emerging pollutants in urban sewage treatment plants using lc-qtof-ms suspect screening and quantification. *Science of the Total Environment*, 774:145779, 2021.

C. P. Wild. The exposome: from concept to utility. *International journal of epidemiology*, 41(1):24–32, 2012.

J. Wilkinson, P. S. Hooda, J. Barker, S. Barton, and J. Swinden. Occurrence, fate and transformation of emerging contaminants in water: An overarching review of the field. *Environmental Pollution*, 231:954–970, 2017.

K. Wille, H. F. De Brabander, L. Vanhaecke, E. De Wulf, P. Van Caeter, and C. R. Janssen. Coupled chromatographic and mass-spectrometric techniques for the analysis of emerging pollutants in the aquatic environment. *TrAC Trends in Analytical Chemistry*, 35:87–108, 2012.

W. Xu, X. Wang, and Z. Cai. Analytical chemistry of the persistent organic pollutants identified in the stockholm convention: A review. *Analytica Chimica Acta*, 790:1–13, 2013.

Y. Yang, Y. S. Ok, K.-H. Kim, E. E. Kwon, and Y. F. Tsang. Occurrences and removal of pharmaceuticals and personal care products (ppcps) in drinking water and water/sewage treatment plants: A review. *Science of the Total Environment*, 596:303–320, 2017.

S. Yukioka, S. Tanaka, Y. Suzuki, S. Echigo, and S. Fujii. Data-independent acquisition with ion mobility mass spectrometry for suspect screening of per-and polyfluoroalkyl substances in environmental water samples. *Journal of Chromatography A*, 1638:461899, 2021.

X. Zhang, K. Quinn, C. Cruickshank-Quinn, R. Reisdorph, and N. Reisdorph. The application of ion mobility mass spectrometry to metabolomics. *Current Opinion in Chemical Biology*, 42:60–66, 2018.

X. Zheng, N. A. Aly, Y. Zhou, K. T. Dupuis, A. Bilbao, V. L. Paurus, D. J. Orton, R. Wilson, S. H. Payne, R. D. Smith, et al. A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. *Chemical Science*, 8(11):7724–7736, 2017.

X. Zheng, K. T. Dupuis, N. A. Aly, Y. Zhou, F. B. Smith, K. Tang, R. D. Smith, and E. S. Baker. Utilizing ion mobility spectrometry and mass spectrometry for the analysis of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, polybrominated diphenyl ethers and their metabolites. *Analytica chimica acta*, 1037:265–273, 2018.

Y. Zhong, S.-J. Hyung, and B. T. Ruotolo. Ion mobility–mass spectrometry for structural proteomics. *Expert review of proteomics*, 9(1):47–58, 2012.

R. A. Zubarev and A. Makarov. Orbitrap mass spectrometry, 2013.

Chapter 2

Improving Halogenated POPs Analysis in Real Samples Using GC-APCI-IMS-HRMS

Summary

Gas chromatography-high-resolution mass spectrometry (GC-HRMS) is a powerful non-target analysis (NTA) technique that improves the identification of environmental pollutants. Currently, most GC-HRMS instruments are equipped with electron ionization (EI). However, atmospheric pressure ionization (API) sources have demonstrated advantages over EI in NTA , such as predominantly charge-transferred ions and/or photon-transferred ions in mass spectra. These benefits promote structure elucidation, and compatibility with ion mobility mass spectrometry, with additional identification confidence by drift time/CCS value. However, as with all novel analytical approaches, the lack of spectral libraries and the reproducibility of the data are two of the main drawbacks in NTS using GC-API applications. In addition, the benefit of tandem IMS can be evaluated. Herein, a GC-APCI-timsTOF method was developed and applied to real samples, which is discussed in this chapter.

Chapter 2 contains a short review of the recent work using GC-API-IMS-HRMS. The new GC-APCI-timsTOF method is discussed in this chapter. The benefits of introducing IMS to GC-APCI-HRMS are argued with representative example in real matrix. A draft of paper is currently under preparation, and it is included at the end of this chapter.

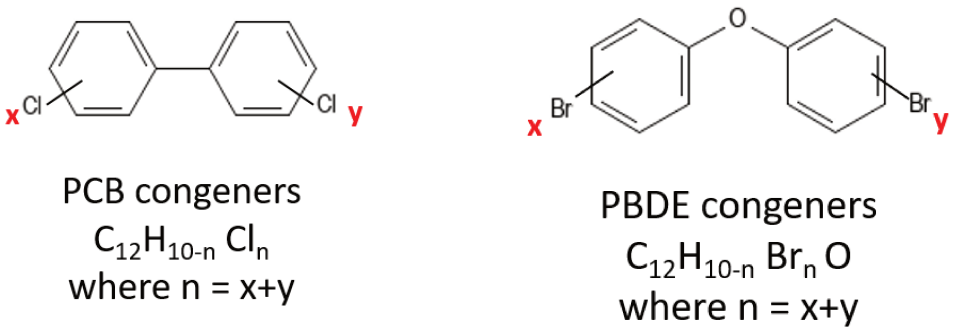
2.1 State of the art

Although atmospheric pressure chemical ionization sources were first introduced for GC-MS in the 1970s (Li et al., 2015), commercial APCI sources were first available only at the beginning of the 20th century. Since then, GC-APCI-MS has been widely applied to POPs in environmental and biological analyses (Fang et al., 2020). Moreover, APCI can be coupled with different mass detectors, such as triple quadrupole (QQQ), ion trap, QTOF/TOF and Fourier transform ion cyclotron resonance (FTICR) (Niu et al., 2020). GC-APCI-MS shows promising sensitivity and selectivity. On the other hand, as a soft ionization source, fragmentation is minimized, and it produces predominantly (quasi)molecular ions, which is a privilege in the identification of unknown compounds. Therefore, GC-APCI-MS is becoming a popular technique in environmental non-target analysis. Combined with GC-EI-MS and LC-MS, a non-target analysis covers a large scope of chemical hazards in a natural matrix, determining a complementary risk assessment. Recently, a review discussed the key parameter of the APCI source, the instrumental advanced in GC-APCI-MS, and the applications of GC-APCI-MS from target analysis to non-target analysis (Niu et al., 2020). Conventionally, GC-APCI-MS is complementary to GC-EI-MS (Cherta et al., 2015) and LC-MS (Hernández et al., 2015). Due to the lack of mass spectrum libraries, Cherta et al. (2015) used GC-APCI to detect the compound of interest and measure precursor ion mass; then, the researchers switched to an EI source for mass spectra screening. Rostkowski et al. (2019) performed non-target analysis of house dust using GC/LC coupled with different ion sources and mass analyzers. GC-APCI-HRMS enabled semi-quantification of CECs without available standards. Aalizadeh et al. (2022) applied the quantitative structure-property relationship (QSPR)-based model to estimate the APCI ionization ratio of unknowns and to estimate their quantities. Most recently, GC-APCI-HRMS that was hybridized to IMS was also reported to improve compound identification and isomer separation (Lipok et al., 2018; Izquierdo-Sandoval et al., 2022; MacNeil et al., 2022). Furthermore, GC-APCI-HRMS demonstrated advantages in the non-target analysis of halogenated compounds. Excellent isotopic matches were observed in non-target analysis of brominated and chlorinated flame retardants using Fourier transform ion cyclotron resonance mass spectrometry GC-APCI-FT-ICR-MS (Zacs et al., 2019). Hence, the researchers applied an automated isotopic profile deconvolution for high-resolution mass spectrometric data (APGC-QTOF) from biological matrices. GC-APCI-QTOF was also applied for halogenated dibenzo-p-dioxins in negative modes (Fernando et al., 2016).

2.2 Objectives

Persistent organic pollutants (POPs) refer to a group of chemicals that are highly persistent and bioaccumulative. POPs can be related to extensive anthropological activities and industrial activities (Xu et al., 2013). Halogenated POPs, such as polychlorinated biphenyls (PCBs) and polybromodiphenyl ethers (PBDEs), are an important part of POPs. PCBs (Figure 2.1a) and PBDEs (Figure 2.1b) consist of 209 congeners that can be industrial chemicals or byproducts of industrial production (e.g., electronic devices, flame retardants) (Xu et al., 2013). In France, PCBs,

PBDEs and their TPs/metabolites were detected in various matrices, such as sediments (Liber et al., 2019), water (Sarkis et al., 2021; Munschy et al., 2022), and mammals (Dron et al., 2022; Wu et al., 2021; Alfonso et al., 2019).



(a) PCB congeners (b) PBDE congeners

Figure 2.1: PCB and PBDE congeners

The aim of this study was to develop a non-target analysis method for halogenated POPs using GC-APCI-timsTOF. Due to the natural abundance of the isotopes of Br and Cl, PCBs, PBDEs, and other halogenated POPs were chosen as the compounds of interest to gain the first insight into the application of this methodology. Moreover, the benefits of combining TIMS with APCI were evaluated in standards and real samples.

-23

Improving Halogenated POPs Analysis in real samples using GC-APCI-IMS-HRMS

YANG Fan[1], AMINOT Yann[2], SINGH Randolph[2], MUNSCHY Catherine[2], PREUD'HOMME Hugues[1]

[1]Institut des Sciences Analytiques et de Physico-Chimie pour l'Environnement et les Matériaux (IPREM)-UMR5254, E2S UPPA, CNRS, Pau, France, 64000,
[2]Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)-DCN/RBE/BE/LBCO, Nantes, France, 44311.

Abstract

Atmospheric pressure chemical ionization (APCI) combined with gas chromatography (GC) produce predominately charge transfer and/or proton transfer ion, preserving the molecular information for ease identification of the unknowns. Gas chromatography (GC) coupled with the ion mobility spectrometry hybrid with high-resolution mass spectrometry (IMS-HRMS) is presented as a promising approach for the monitoring of GC amenable-POPs in complex matrices. IMS separates ions by their charge, shape and size, qualified in term of collision cross section (CCS). CCS value/IMS can clean up the spectra and discriminate the isomers. Additionally, CCS value is platform independent, the cross-platform measurement bias was below 5%. GC-APCI-IMS-HRMS indicates the potential of using online CCS database in suspect and non-target analysis due to the benefit of soft ionization. In this work, we studied the enhancement of combining IMS with APCI in peak identification and structural elucidation. The identification performance of GC-APCI-IMS-HRMS for the studied compounds was assessed in complex-matrix samples.

1. Introduction

Persistent organic pollutants (POPs) are a group of highly persistent and fat-soluble xenobiotics [1,2], resulting in stability and bio-accumulation in the food-chain transport.[1] Twelve legacy POPs were under the Stockholm Convention, such as organochlorine pesticides (OCPs), polychlorinated biphenyls (PCBs), and industrial by-products. The list has been expanded to diverse halogenated POPs, like brominated flame retardants (BFRs) and perfluorooctanoic sulfonic acid (PFOS), etc.[3] Both legacy and emerging halogenated POPs were detected in the environment and humans.[4–7] Moreover, metabolite of POPs, such as hydroxylated and methoxylated polybrominated diphenyl ethers (OH-PBDEs and MeO-PBDEs), can be persistent and accumulated in the human body and cause neurodevelopmental, thyroid dysfunctions and cancers.[4,8,9] Hence, the analysis of halogenated compounds covering from knowns to their TPs/metabolites is demanded, and retrospective analysis is also required for time-trend environmental studies.[10–12]

The characterization of contaminants and their transformation products (TPs) is vital to assess their potential risk to the environment and the biota. Suspect and non-target screening (NTS) using high-resolution mass spectrometry (HRMS) are widely accepted strategies to identify knowns.[13] Gas chromatography coupling with mass spectrometry (GC-MS) has been widespread in volatile POPs (e.g., PCBs and PBDEs) analysis in diverse samples. [6,14,15] GC-MS systems are mostly equipped with an electron ionization (EI) source (commonly at 70eV). Due to its reproducibility in a mass spectral and large amount of empirical mass spectral libraries (e.g., NIST) promotes the identification of target and suspect screening. However, it is not suitable to identify the unknown. Unlike EI, atmospheric pressure chemical ionization (APCI) is a soft ionization source that reduces

fragmentations and maintains a relatively high abundance of the (quasi)molecular ion in the spectrum.[16] Several studies highlighted the benefit of using GC-APCI-HRMS to identify unknown contaminants.[15–18] However, the reproducibility of mass spectra is one of the main drawbacks of GC-APCI- HRMS method.[15] Multiple mechanisms can coincide in APCI, resulting in signal abundance reduction.[19] And the humility of the matrix can affect the production of molecular ions or quasi molecular ions.[19] As a result, it is difficult to find a probable structure through screening experimental MS against APCI-MS library.

Ion mobility spectrometry (IMS) is a fast separation technique in the gas phase based on size, charge, and mass. Trapped ion mobility spectrometry (TIMS) is one of the IMS separation techniques, which recently can be hybridized with HRMS.[20] In TIMS analyzer, ions are trapped and accumulated by radially-confining rf voltage. Ions are subsequently propelled by a gas flow (e.g., N_2, He); an electric field against the gas flow prevents each ion from moving over time by their mobility(K).[21] IMS offers an additional separation dimension to reduce coeluting interfered mass peak and to produce a clean MS spectrum.[22,23] Moreover, collisional-cross section (CCS) values can be obtained by IMS, which is a physicochemical property of a compound. Therefore, CCS value is instrument and measurement-independent (under the same buffer gas). The cross-platform bias was calculated between Traveling wave and Drift Tube IMS (TWIMS and DTIMS), mean bias were 1.0% for $[M + H]^+$ and 1.1% for $[M + Na]^+$ ions, for few ion was at 6.2%.[24] Another study compared CCS values of 87 steroids obtained by TWIMS, DTIMS, and TIMS. The bias was within 2% for 95% of ions, with a few at 7%.[25] Additionally, predicted CCS values using machine learning demonstrated excellent accuracy (> 0.95).[26,27] Using the CCS database is a recommended way to increase identification confidence.[22] But more studies must evaluate cross-platform and interlaboratory bias for more chemical classes.

We developed a fast POPs analysis method in this study using GC-APCI-IMS-timsTOF. The ionization performance of APCI was investigated based on chemical class. An in-house database consisting of RT, accurate mass, isotope pattern, and CCS value was built. The empirical CCS values were compared to other databases and predicted CCS values to ensure compatibility between IMS techniques.

2. Materials and methods

2.1. Standards and solvents

POP standard kits purchased from Wellington Laboratories (Guelph, Ontario). A mix of the standard of 34 PCBs congeners (Cl_2 – Cl_{10}), PBDEs mix standard included Br-Br_9-PBDE, 38 congeners. Other mixtures contained 11 halogenated Carbazole, 8 Methoxylated polybrominated diphenyl ethers (MeO-PBDEs), and 17 other Organohalogen compounds (OHCs). A total of 118 compounds were used to optimize the method and build an in-house database. Toluene, nonane, and acetonitrile of LC-MS degree used for standard dilution were obtained from Biosolve (Dieuze, FRANCE)

2.2. Method

A Bruker GC SCION device (Germany) coupling with timsTOF (trapped ion mobility spectrometry coupled with time-of-flight high-resolution mass spectrometry, Bruker, Germany) was used for chromatographic separation. PCBs standards mix and PBDEs standards mix were used to optimize the GC separation program. A 30 m and a 40 m HRGC column were compared in critical pairs separation. Analyte separation was performed with a Rxi-5SilMS column (40 m x 0.18 mm × 0.18 μm). 2 μL of standard solutions and extracted samples were injected in the splitless mode. The transfer inert line temperature was set to 300 °C, and the helium carrier gas flow was set to 3.6 mL/min. GC eluant exiting the column was swept through the ion volume using a makeup flow of nitrogen (~99.999% purity) at 4 Bars. The GC inlet temperature was held at 280 °C. The analysis was performed with He carrier gas, and the GC program was: the initial temperature was set to 120 °C and held for 1.5 min; then 25 °C/min to 180 °C and 15 °C/min to 250 °C; 9 °C/min to 320 °C and held for 9 min, with a total run time of 25 min. APCI was initiated by a corona discharge (1.8 μA) in positive ion modes. timsTOF was enabled in full scan mode from 150 to 1250 m/z. The detail of the instrument setting was described in Table S1 in the SI. External-mass calibrations were performed using the Agilent ESI low-concentration tune mix by infusion before the acquisition sequences. Standard mixtures (mentioned above) were analyzed with an alkane mix to normalize the retention time. The samples were prepared and provided by IFREMER.

2.3. Data treatment

Standard identification was generated by DataAnalysis software 5.30 (Bruker Daltonics). The most abundant isotopic patterns were applied to extract compounds in chromatogram and mobilogram. Inverse Reduced Mobility($1/K_0$) was directly measured by timsTOF in N_2 at 25°C, and it was converted to CCS value by Data Analysis (Version 5.30) implemented Mobility Calculator. Thereby, we created a 4-dimensional in-house database included: (1) retention time (RT), (2) accurate mass, (3) isotopic patterns and related ratio, and (4) CCS value in the TASQ® 2021(Bruker Daltonics). The automated workflow, identification and potentially quantification is generated by TASQ®. To ensure the identification reliability, the parameter is set as: mass accuracy < 5 ppm, RT deviation <0.1 min, CCS deviation <2% from the measured standard value, and msigma < 50.

3. Results and discussion

3.1. GC separation method and isomer separation by IMS enhancement

A fast GC-APCI-IMS-HRMS method was developed using 40m Rxi-5SilMS ultra inert column, PBDEs congeners from Br-Br_9 and PCBs congeners from Cl_2- Cl_{10} in 25min (shown in Figure S1). This acquisition time is comparable to other fast GC programs.[28–30] PBDE 206 was eluted at 24.18 min and PBDE 209 at 24.82 min, two isomers were eluted at the end of the GC program (25min run time), and the separation resolution (R) was at 1.16. BDE 209 was out of the GC program. Isomeric compound discrimination is always challenging in method development and peak identification. Traditionally, isomers can be separated chromatographically or distinguished by characteristic fragments. IMS separates ions by its 3D structure, offering a novel possibility of

isomeric discrimination. Several critical pairs exist in PCBs and PBDEs. This study evaluated the chromatographic separation resolution and CCS differences specifically for these critical pairs. PBDEs 197/201 cannot be separated due to the loss of resolution at the end of acquisition time. PCB-52/PCB-49 obtained a separation resolution R =1.58, ΔCCS was 0.7 $Å^2$ (0.5% relative difference). PCB-153/PCB-132 R = 1.18, whereas IM can better discriminate the two isomers (ΔCCS = 3.1 $Å^2$). The PCB-28 and PCB-31 were coeluted; GC or IMS could not separate them. Although CCS enables an additional alignment of identification, the difference in CCS of isomers is tiny (~ 5 $Å^2$). A newer generation of IMS with better resolution can improve the separation of isomers.[31–33]

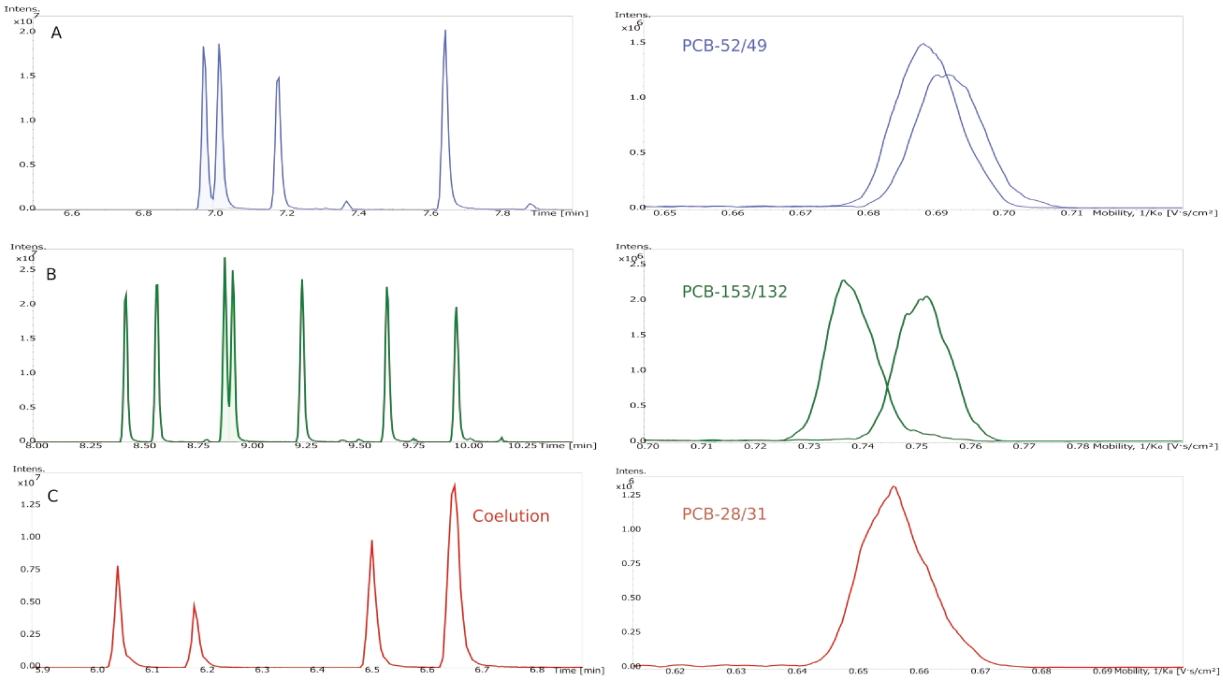


Figure 1 Isomer separation with CCS enhancement. PCB-153 and PCB-132 were coeluted in a 40 m GC column, shown on top, while adding mobility range in EIC (below), and the chromatographic peak resolution was enhanced from 0.44 to 0.59. Using a 40m GC column, the separation was increased significantly(2a) to 0.73. In contrast, IM capacity was limited by the structural similarity of PCB-153 and PCB-132, which resulted in a 22% peak mutual overlap and resolution of IMS.

3.2. Spectral Noise Filtering by CCS range

MS clean up

Structure elucidation is essential to identify the unknown, and the tentative structure is commonly assigned by screening the experimental mass spectrum against literature or spectrum libraries, like NIST[34] and MassBank.[35] Based on the defined five levels of confidence[36], candidates that unambiguously match the reference MS reach level 2. It is a concrete approach to minimize the number of hits by precursor ions and characteristic fragment ions, which request high mass accuracy and a high-quality mass spectral library. Moreover, the compounds of interest are often at trace-level, the abundance of diagnostic ions can be too low to be detected, and the interfered mass peaks can result in a false negative and false positive, especially in small molecule identification.

APCI is a soft ionization mode that produces predominantly molecular ion $[M^{\cdot}]^{+}$ or protonated ion $[M+H]^{+}$. Unlike the hard ionization sources (e.g., EI, CI), APCI yields limited in-source fragmentation, which can simplify compound identification. However, the unambiguous experimental MS database is limited.[19] In addition, multiply ionization mechanisms react simultaneously which split the (quasi) molecular ion peak abundance. Thus, a clean chromatogram and mass spectrum is essential to facility mass spectral interpretation.

IMS separation generates after ionization and before the mass detector, which adds an orthogonal characteristic dimension. As shown in Figure 2, two MS were obtained from the same acquisition and chemical. EIC was generated by mass (A), and Figure 2B was obtained from the same EIC with an IM mobility filtering. The interfered mass peak from 150 to 500 m/z was significantly eliminated by selecting the mobility range (B). Without losing peak intensity, $[M]^{+}$ became the most intense peak (m/z 563.6204: Int above= 870000 cnts vs Int below = 860000 cnts). Since two mass spectra shown in A and B were explored in BFRs standard mix, it highlights that CCS filter improves the quality of mass spectrum in the MS library.

Moreover, interferences from the sample matrix can be eliminated by mobility filtering. PCB-Cl3 congeners were identified in a sample, and the MS were shown in C/D. 257.2240 m/z was the most mass in C. Its intensity was significantly decreased by adding mobility range in D. Due to the softer ionization, APCI is prone of matrix interferences [37], this problem can be possibly potentially solve by ion mobility. Exact mass, retention time and CCS can generate a cleaner MS, making the data treatment more straightforward.

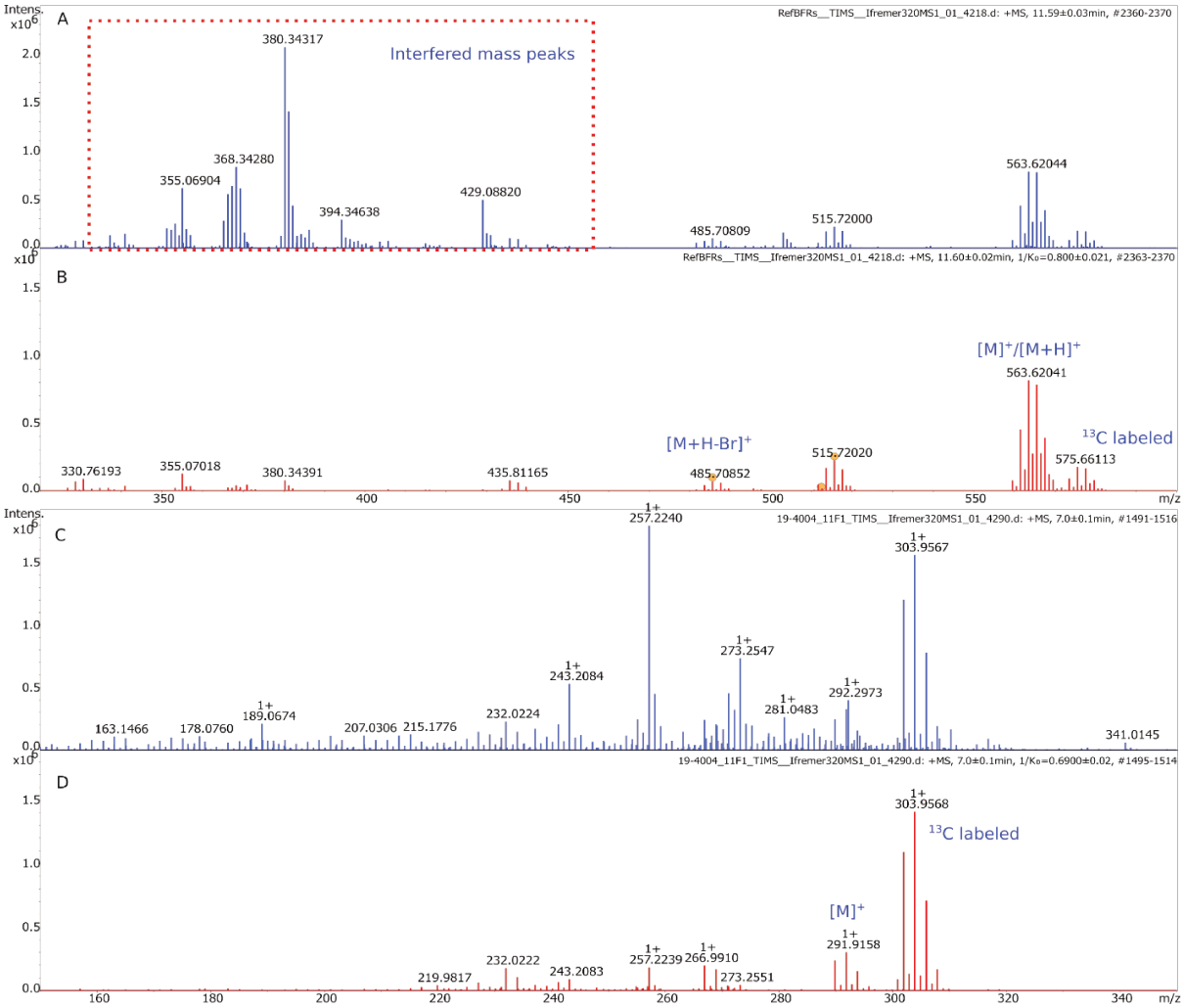


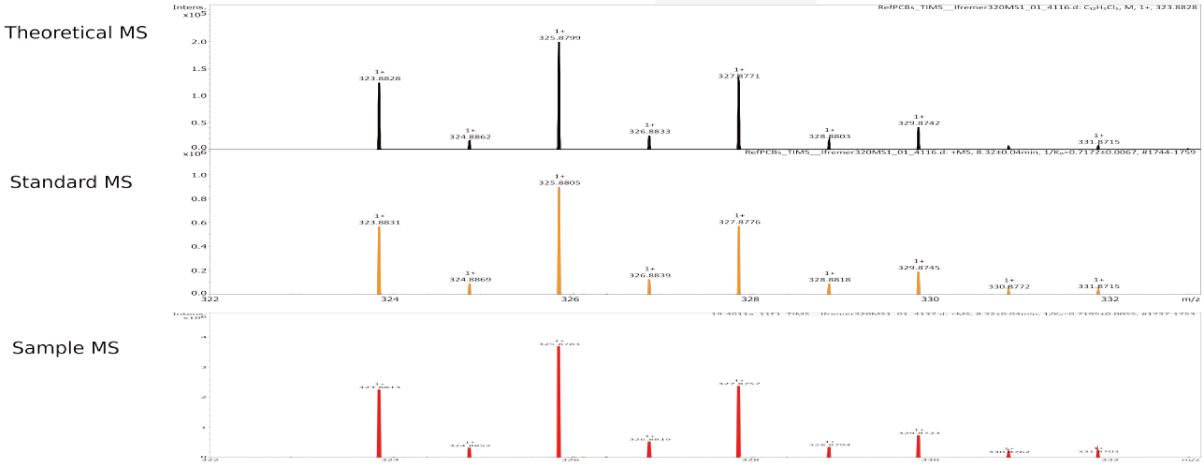
Figure 2 MS of PBDE-Br5 generated from the same acquisition. (B) added a mobility range in EIC, the noise peak such as m/z 380.34 shown above was significantly decreased compared to the one in (A), [M]+ peak intensity was in the same scale (~8.6x10^5 cnts). However, it became the most intense peak in the (B), which is essential in the APCI-MS database and chemical identification. (C) and (D) presented an identified PCB in a sample, interfered mass peaks from matrix were significantly removed when CCS range was implemented.

Isotopic profiles

Chlorinated and brominated contaminants raise environmental and human health concerns.[38] Due to the specific isotopic pattern of $^{35}Cl/^{37}Cl$ and $^{79}Br/^{81}Br$, many data processing tools are developed to filter and assign a probable molecular formula, like HaloSeeker[39], enviMass.[40]

Since the IMS filter can generate a clear MS and the high mass resolving in TOF, facilitating the formula prediction of halogenated compounds. In this study, measuring isotopic profiles of standards and samples was compared to theoretical values, and the accuracy was evaluated by mSigma embedded in Data Analysis version. The error of mass in samples was < 1.5 ppm, and the bias of the characteristic peak ratio was below 1%. The excellent fit of isotopic profile between theoretical and experimental values enhances the confidence of formula prediction in halogenated compounds. The mSigma provides a quantitative score to assess the predicted formula.

		Standard				Sample		
$C_{12}H_5Cl_5$								
	m/z	I%	Δm/z(ppm)	ΔI%	msigma	Δm/z(ppm)	ΔI%	msigma
M+2	325.8799	100	1.8		10.3	4.9		7.7
M+4	327.8771	64.5	1.5	0.8		4.3	0	
M	323.8828	62.2	0.9	0.9		4.6	1	
M+7	329.8742	20.9	0.9	0.2		5.8	0.8	



3.3. Discriminate coeluted compounds and fragments

Dehalogenation commonly occurs in halogenated POPs during ionization. Although APCI was used, we observed the loss of chlorine or bromine frequently in POPs. Different halogens and their substituents showed different fragmentation behaviors. It is mainly because meta carbocation is more stable than the ortho- or the para molecular ion. Furthermore, bromine loss is easier than chlorine, since C-Cl bond has higher bond strength than C-Br.

CCS filter removes fragment peaks in EIC

As demonstrated in Figure 3 , PBDEs ($C_{12}H_{10-n}Br_n$) were prone to loss of bromine and produced a $[M\text{-}Br+H]^+$ transition, which has the same monoisotopic mass as PBDE-Br_{n-1}. As illustrated in Figure 3, debromination was observed in mono-PBDEs to tri-PBDEs. Within isomers, the ionization behaviors were different. As an example, BDE-7 (2,6-Dichlorodiphenyl ether) was prone to produce protonated ion than molecular ion, the intensity of protonated ion was twice higher than molecular ion. BDE-10 (2,6-Dichlorodiphenyl ether produced the same level of both ions, while BDE-15 (4,4'-Dichlorodiphenyl ether) obtained twice the amount of molecular ion

than protonated ion (Figure S3). PBDEs congeners are mass and/or structural differences, resulting in ion mobility differences. CCS range can be applied to an extract ion chromatogram (EIC), where only ion traces match the exact mass, isotope patterns, and mobility range are enabled. Isomers of PCB and PBDEs have a comparable CCS value. When the number of halogenated substitutions increases, the mass and size of the molecule increase, leading to an increase in mobility (Figure S2). Thus, combining retention time, mass and mobility extracts a clear chromatogram , simplifying the interpretation of the chromatogram. Another possible application of IMS/CCS is to combine the fragment peaks since they share the same CCS.

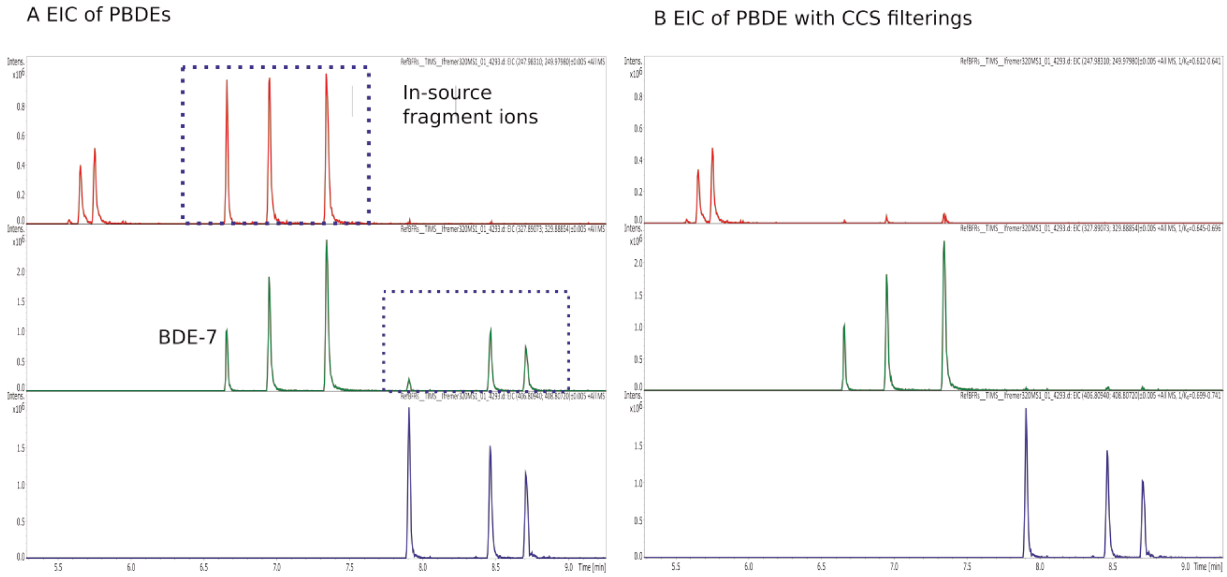


Figure 3 EICs of PBDEs (from mono to tri-PBDE). Fragmentations were commonly observed in PBDEs. However, the abundance ratio between fragment and molecular ion was varied by substitution level and position. The benefit of CCS difference, it can be applied to extract chromatograms, which 'eliminates' the fragments and produces a straightforward EIC.

CCS discriminates in-source fragments and after the IMS fragment

Furthermore, we discovered that dehalogenated fragments are prone to protonated ions. An example of PCB-Cl2 was shown in Figure 5, the chromatograms were extract by m/z :256.9686, 258.9657 ([C12H8Cl3] +); m/z: 255.9606,257.9577 ([C12H7Cl3] +); and m/z: 221.9990, 233.9961 ([C12H8Cl2] +). The mobilograms presented PCB-30 (2,4,6-Trichlorobiphenyl). Molecular and protonated ions were discriminated in the mobilogram, and molecular ion had a CCS value of 139.1 $Å^2$ and 141.0 $Å^2$. Additionally, two peaks were present in EIM of fragment ion, one at 131.4 Å2, which corresponded to [C12H8Cl2] +, and another was aligned to the CCS value of protonated ion. It indicated that fragmentation of PCB-30 occurred before passing the TIMS funnels.

In Figure 6, Ion mobility peak of BDE-7 was overlapped with the protonated ion, therefore, the fragmentation of BDE-7 was occurred after IMS separation. Since CCS value of fragment ions were consistent with CCS value of protonated ion, CCS range enable the discrimination of intact molecules and fragments. IMS analyzer is placed after ionization source and before MS detector, it can be used to discriminate in-source fragment and after IMS fragment. Since the CCS value of production ions and precursor ion were consistence, the CCS range can remove the irrelevant peaks in full scan and MS/MS to simplify the structure elucidation. Furthermore, if an in-source fragment is detected in an unknown, the CCS value of the fragment offers an extra aspect to elucidate the

structure. For example, PCB-149 (2,2',3,4',5',6-Hexachlorobiphenyl) was confirmed in a sample (in Figure 4) . The CCS and RT were matched with the reference values. in-source fragment had a similar CCS value as PCB-118 (2,3',4,4',5-Pentachlorobiphenyl), showing a potential structure of this fragment.

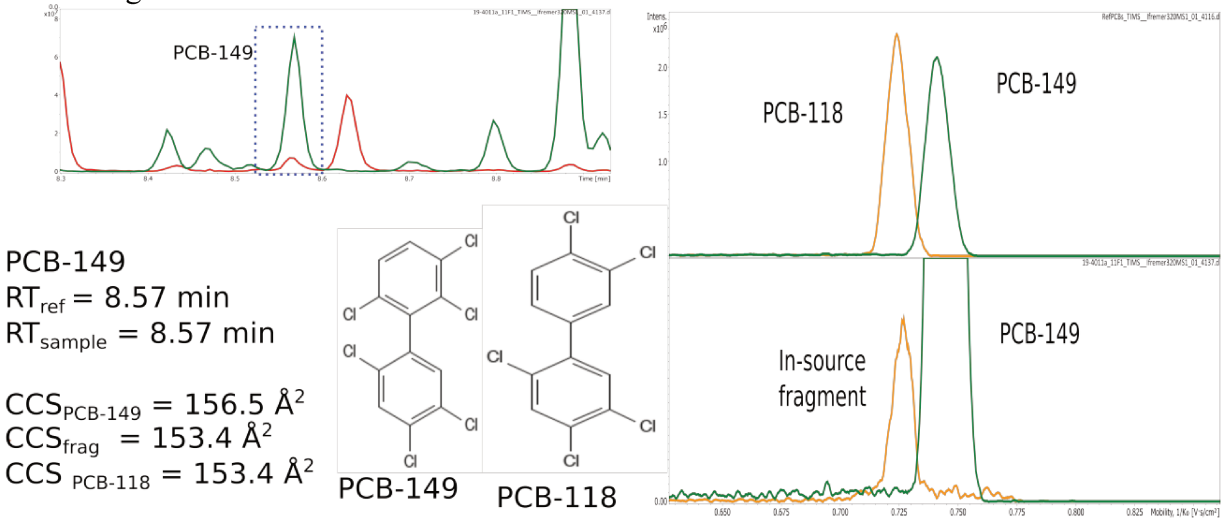


Figure 4 PCB-149 Molecular ion and its in-source fragment

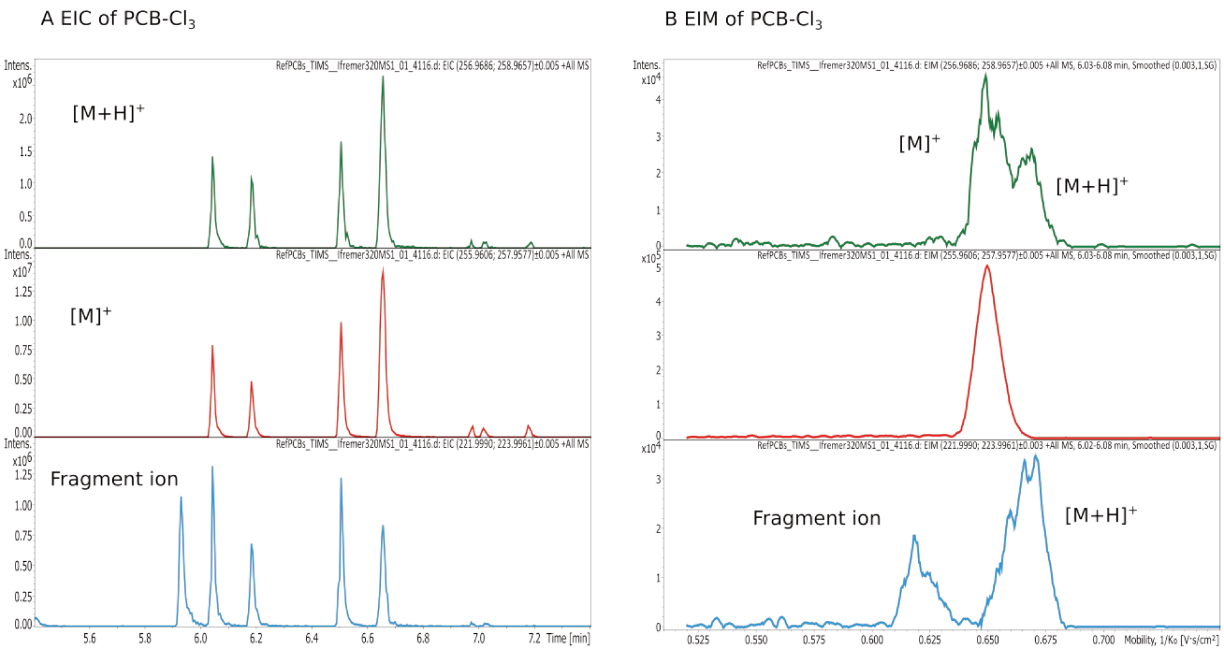


Figure 5 Ionization and fragmentation of PCBs. Dechlorination was occurred before and after the IMS analyzer,

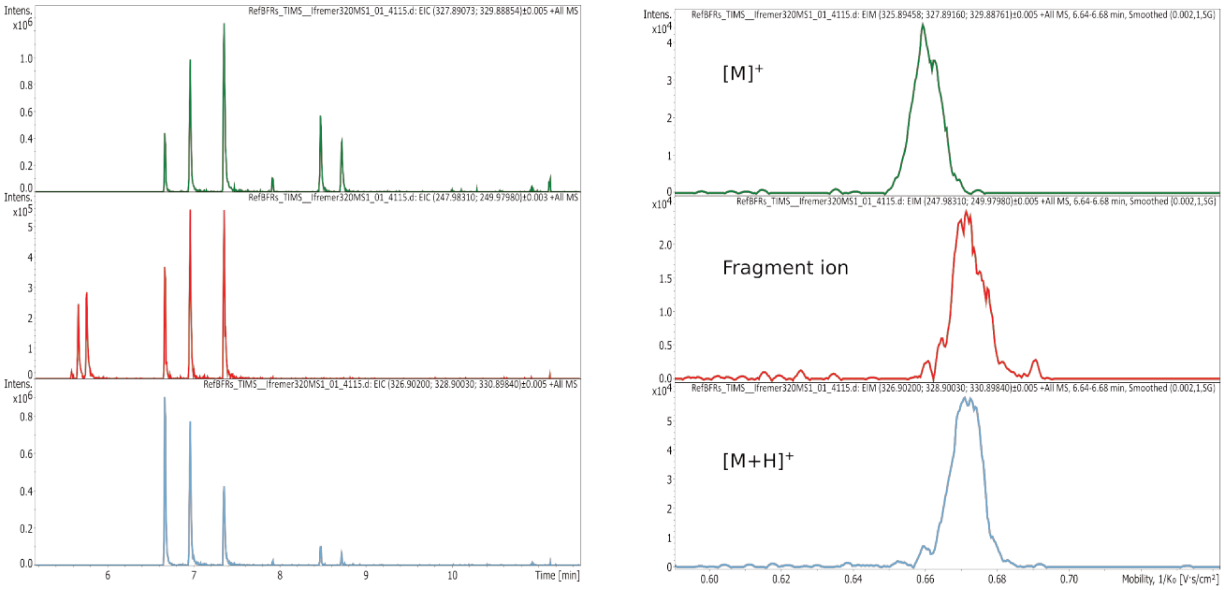


Figure 6 Ionization and fragmentation of PBDEs

3.4. GC-APCI-TIMS-HRMS Library

IMS stability and repeatability

The database consists of 195 CCS values, including $[M]^+$ and $[M+H]^+$. The repeatability precision was evaluated for each chemical class and calculated by relative standard deviation (%RSD). RSD was below 0.6% for all the chemical class (Figure 7). Furthermore, an inter-day and intra-day precision was estimated by PCBs standard solution with a 5-day repeatability test. An external calibration was performed prior to the first day of experiments, second external calibration was performed on day 3. No additional calibration was performed between acquisitions, and 57 injections of blank, standard solutions and samples were completed in 3 days. As presented in Figure 7, the repeatability test was estimated in term of absolute percentage error (APE) relative to median measured CCS value. CCS measurement precision decreased rapidly after 3 days of acquisitions. Thus, the robustness and precision of the IMS measurement is validated and the method enable a high-throughput analysis.

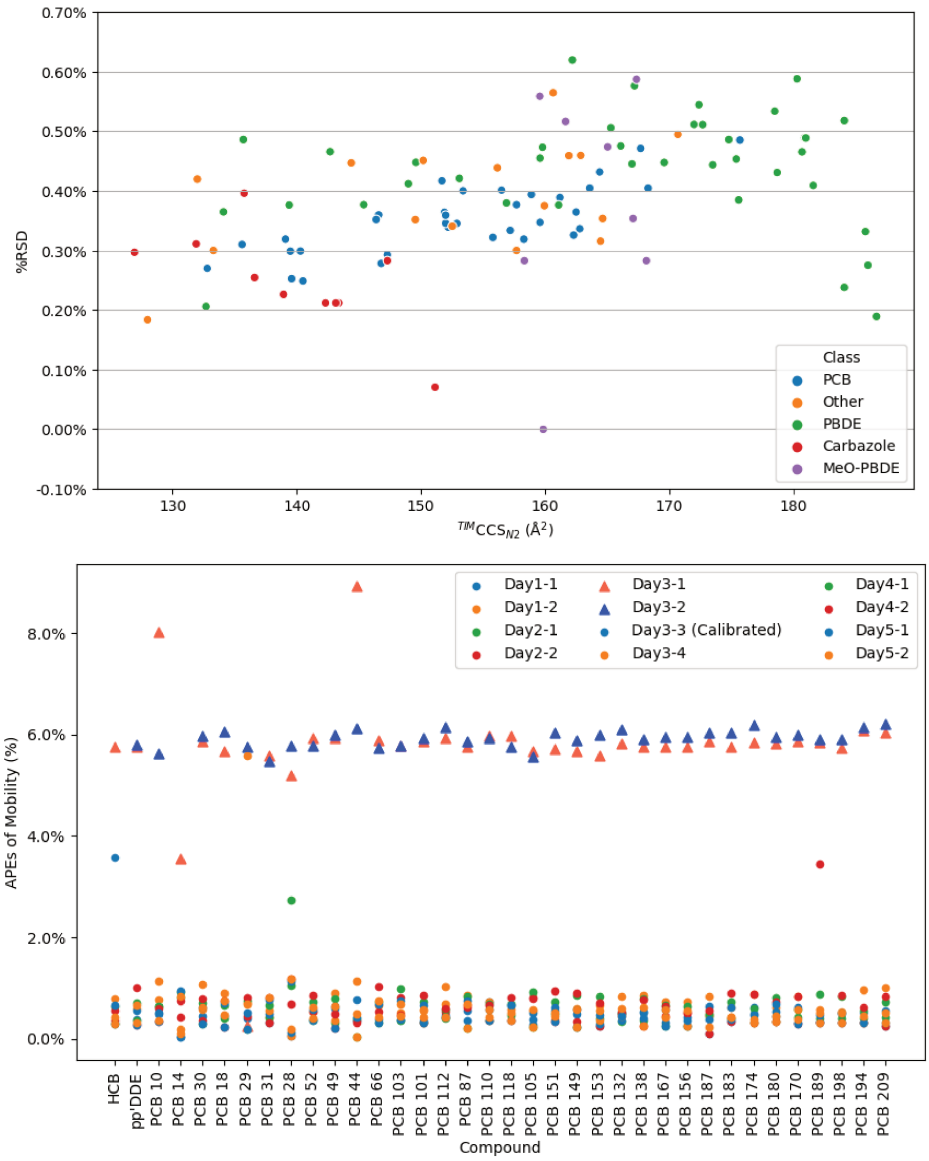


Figure 7 The measurement precision for different POPs class was %RSD < 0.6%. The inter/intra-day measurement precision of PCBs was plotted as figure below. The method enables a high throughput analysis and long sample sequences without loss measurement precision.

Charge transfer and proton transfer ions were simultaneously produced in all PCBs (Cl_2 – Cl_{10}) and PBDE (Br-Br_9). the signals of protonated ions were lower than molecular ion, and the precent of protonated ions decreases when the halogenated level increases.

Molecular ion and protonated ion and their CCS values

PCBs, PBDEs, MeO-PBDEs and carbazole produce simultaneously charge transfer and proton transfer ion. We observed that BDE-10 (2,6-dibromodiphenyl ether) produced more protonated ion than molecular ion, the ratio of chromatographic peak was 0.78 (Molecular ion intensity/Protonated ion intensity). Other PBDE congeners were prone to form molecular ion than protonated ion, the peak intensity ratio of molecular to protonated ion varied from 2 to over 30. In other words, proton transfer mechanism is easier to occur in low brominate substitution level.

Similar to PCBs, the ratio varied from 3 to over 50. MeO-PBDEs (BR_4, Br_5) and carbazole produce comparable molecular ion and protonated ions due to the higher affinity of oxygen and nitrogen. Brominated compounds have a higher tendency to produce proton transfer ion than chlorinated compounds due to higher affinity of brominated. However, the formation of $[M+H]^+$ was not stable. For example, Among 4 injection of PBDEs standards, the presence of $[M+H]^+$ was not always observed and responses of isomers were not consistent. Figure 8 compared the protonation of PBDE-$Br_{4\text{-}6}$ in the repeatability test. Niu et al.,[19] summarized the poor reproducibility of APCI in other studies. The authors argued that the simultaneous formation of charge transfer and protonation splits the intensity, thereby the detection sensitivity is reduced. Moreover, the matrix humidity affects the ionization mechanism.[19,41,42] Thereby, empirical CCS value of charge transfer and proton transfer ion were referred in the database. On the other hand, CCS values of $[M+H]^+$ were ~2% higher than the one of $[M]^+$, which is consistent with results demonstrated in Izquierdo-Sandoval et al.,[18]. Nevertheless, this difference is too tiny to reduce the interferes causing by the coexistence of [M]+ and [M+H] + and completely discriminate one from another.

Cross-platform comparison

Furthermore, we compared the in-house empirical CCS values to other studies. We searched the experimental CCS values in PubChem (version 07 September 2002)[43], CCSbase (V1.3) [44], Unified CCS Compendium (Accessed 19 October 2022)[45], and Collision Cross Section Database (Accessed 19 October 2022)[46]. Only the compounds with the same adduct ion were selected for comparison. We used the CCSbase to compare the in-house experimental values with the predicted ones. The difference between predicted values and experimental values can be biased from 37% (1,3,6,8-Tetrabromo-9H-carbazole) to 1% PCB-167 (2,3',4,4',5,5'-Hexachlorobiphenyl).

Due to the limited studies about GC-APCI-IMS, few CCS values are suitable to compare the inter-platform and interlaboratory bias, halogenated carbazole and MeO-PBDE (Br_4, Br_5) were first reported.

We compared CCS value of PAHs with 2 DTIMS[23,38], 1 TWIMS[18] and 1 TIMS[48] measured value, PCBs and PBDEs CCS values were compared to 1 TWIMS experiment.[18]

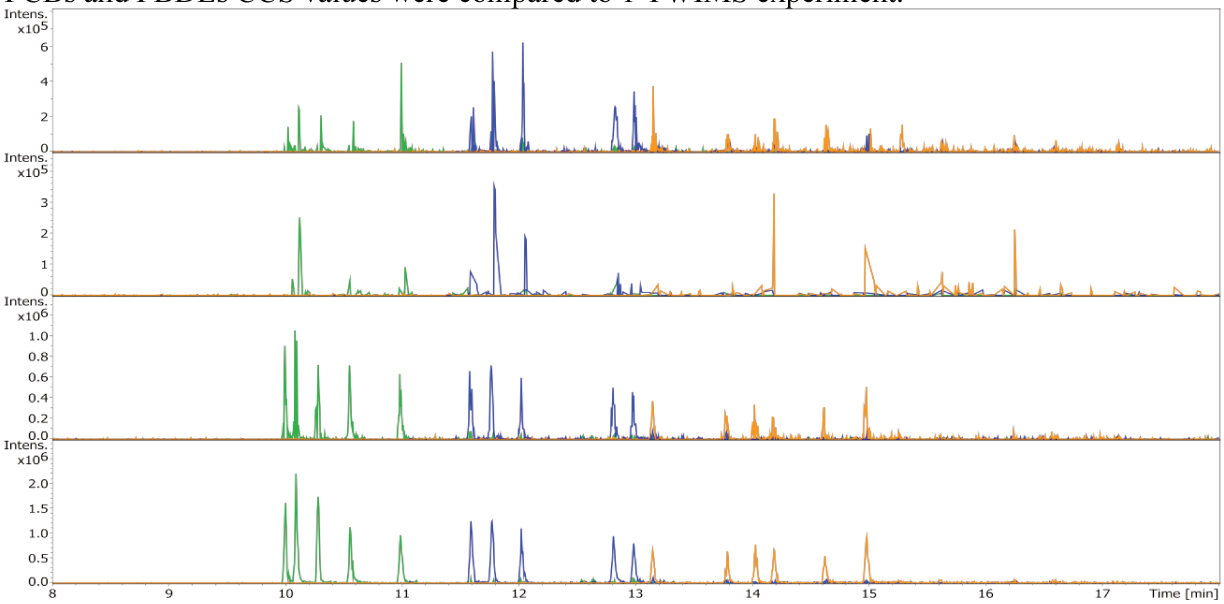


Figure 8 Protonation of PBDE- $Br_{4\text{-}6}$. It indicated an unstable and irreproducible of protonation ion under APCI source. The relative response ratios within isomers were not consistence. Not all the protonation ions can be observed by repeated injection.

The measurement bias between $^{DT}CCS_{N2}$, $^{TW}CCS_{N2}$, and $^{TIM}CCS_{N2}$; and $^{TIM}CCS_{N2}$ and $^{TIM}CCS_{N2}$ *were below* 3.5% for 88% of CCS, and the rest were smaller than 5%. Other studies compared the inter-platform bias of CCS measurement for small molecules, most of ions showed the IMS cross-platform biases within 2%, and the bias can be varied over 6.2% or 7%.[24,25,49] The low bias of inter-laboratory and inter-platform CCS measurement illustrate the potential of applying available CCS databases in unknown identification. For this reason, it is vital to establish a standardized IMS calibration procedure to ensure the quality of CCS database.

To understand the CCS values varied by mass, structure differences, the measured CCS values were plotted against the mass in Figure 9. To simplify the illustration, only the CCS values of charge transfer ions was presented in the Figure 9. The trend lines for the PCBs and PBDEs were clearly separated. PCB had higher CCS value than PBDE with similar mass. For example, PCB-Cl_3 (m/z 255.9608) have more chlorine substitution then larger structure than PBDE-Br (m/z 247.98313), therefor higher CCS value. MeO-PBDEs has higher mass than PBDEs with the same number of bromines due to a more methoxy group, while the structure does not significantly increase, thus PBDEs and its corresponding MeO-PBDEs have close CCS value. Carbazole has a tricyclic structure, including two benzene rings connected by a nitrogen containing ring, its structure is more compactable than PCBs or PBDEs, thus CCS is smaller. Overall, the mass-CCS trendline represent the structure difference of the compound, which can be used to classify the unknowns.

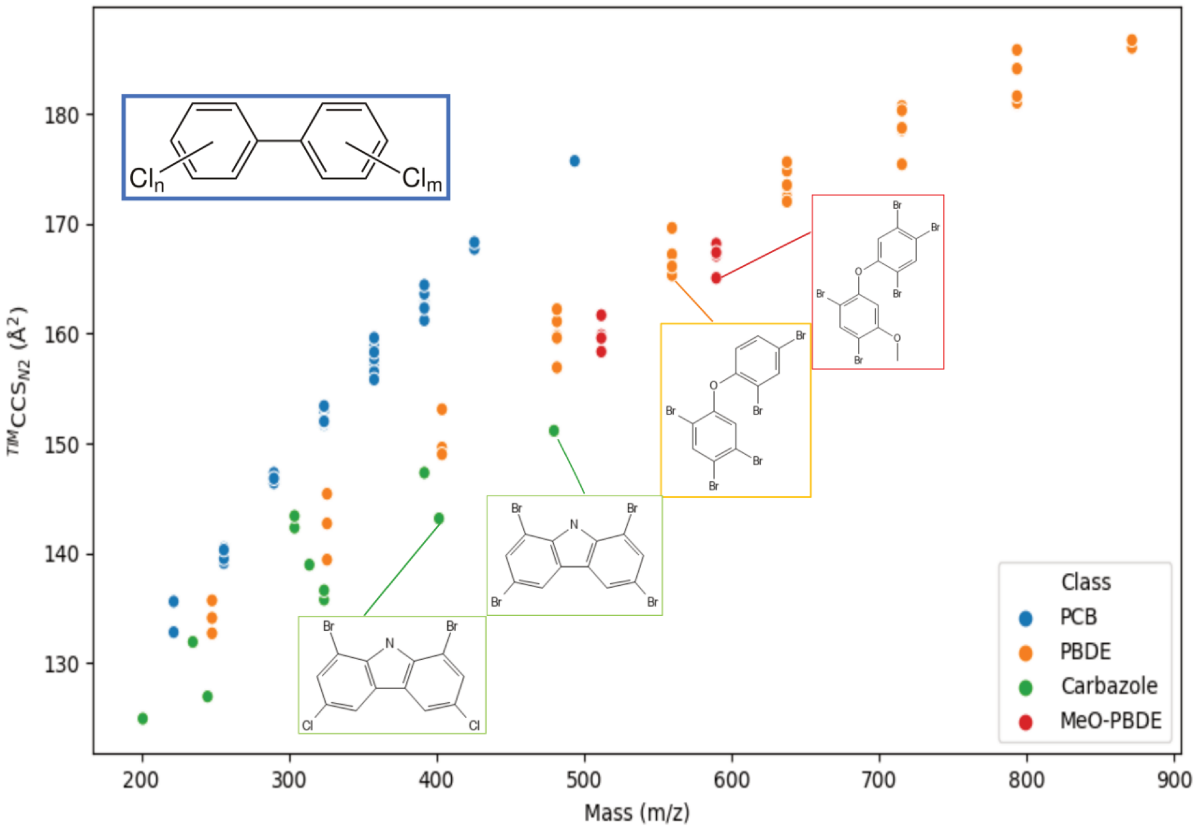


Figure 9 Mass-CCS trendline of the halogenated POPs. Two distinguished zones were separated by PCBs and PBDEs due to their structure and mass differences. For the four halogenated compound class, the CCS value was increased by the number of the chlorines or bromines. MeO-PBDE-99 presented in the red square has higher mass than PBDE-99 in the yellow square, while similar measured CCS values were obtained. It showed that methoxy group have slight impact on the structure of MeO-PBDEs

3.5. Application in real samples

An automated target screening workflow is performed by TASQ. 142 chemicals standard were referred in this database. We applied this method to sample extracted. Thirty-one halogenated POPs were identified by exact mass, retention time, isotopic profiles, and CCS values . False positive of PBDE-47/66 were observed due to the close retention time and CCS. By adding CCS filtering range in analysis, the noise peaks from matrix can be significant removed. A quantification analysis can be established later in TASQ ®.

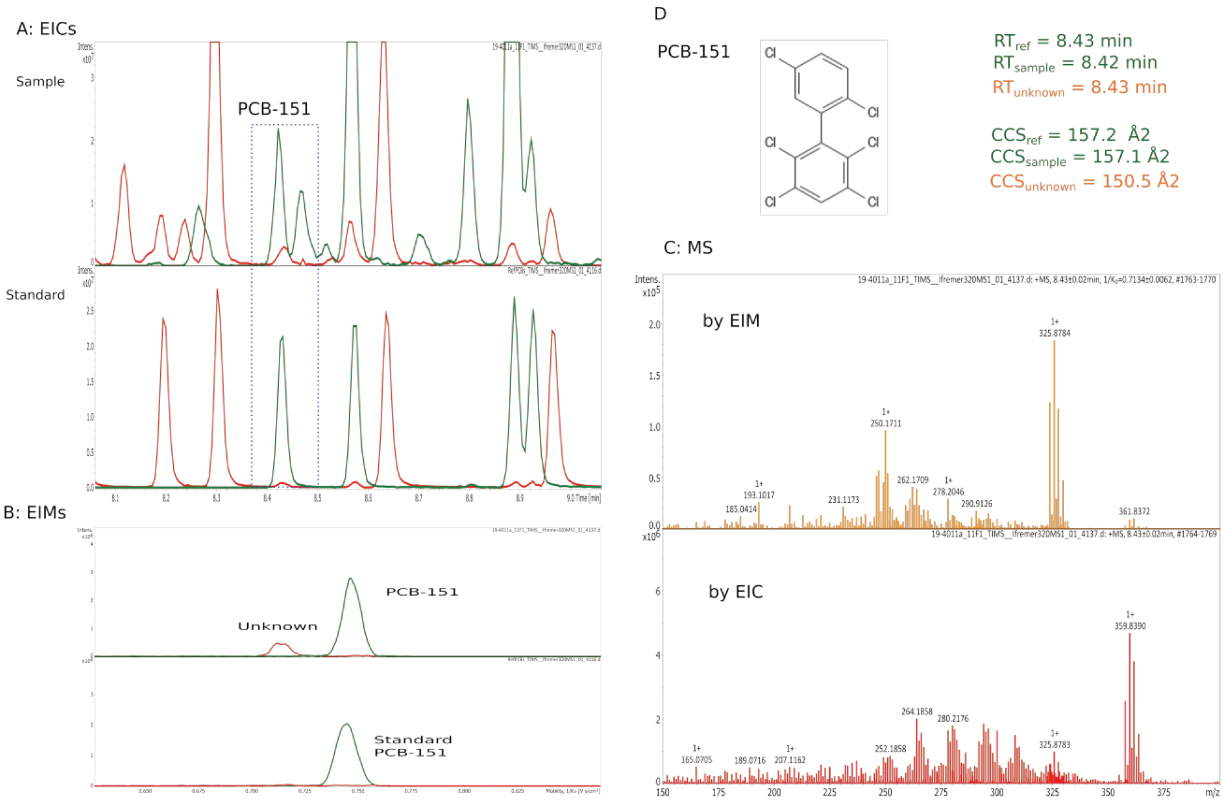
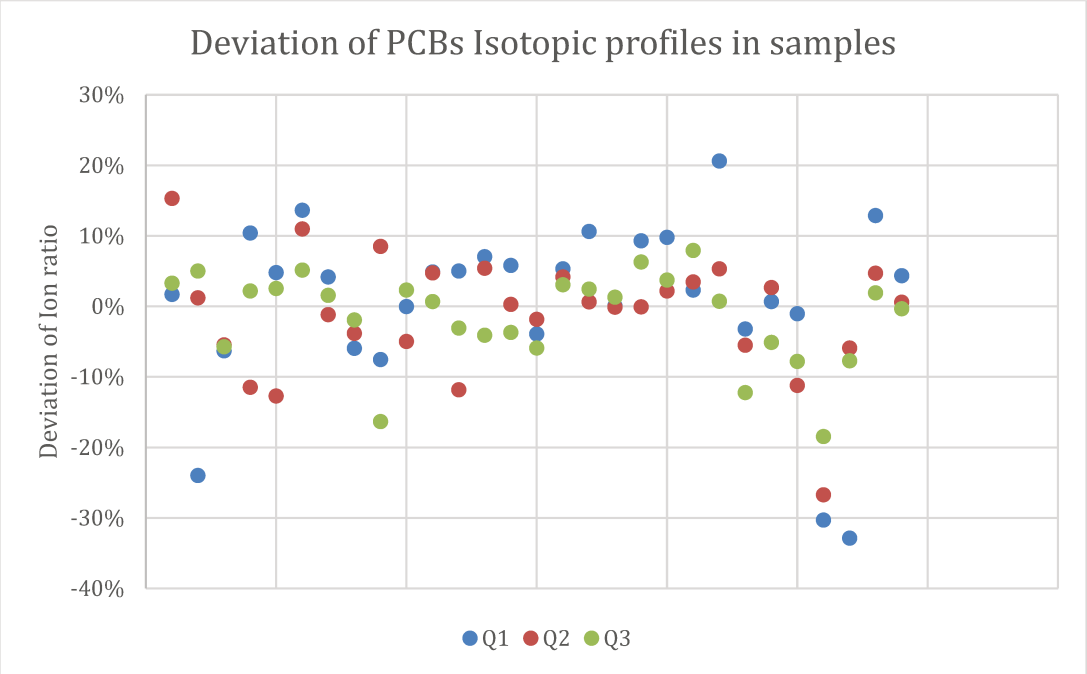


Figure 10 Co-eluting PCBs in samples

The aligned a sample to PCBs standard mix with the extracted masses of PCB-Cl5/PCB-Cl6. Except for 6 PCB-Cl5 identified in the PCB standards, 11 PCB-Cl5s were not presence in the in-house database. A co-eluting of an unknown and PCB-151 was observed as highlighted in the blue rectangle. The retention time of unknown was 8.43 min which matched the reference value of PCB-151, but in the sample, RT of PCB-151 was shifted within 0.01 min, as detailed in samples D. While the dehalogenated fragment of PCB-151 was barely present with standard solution injection, it was challenging to determine whether the trace of PCB-Cl5 was a fragment or a contaminant without reference standard. To address this problem, IMS was then introduced to collect additional evidence for structural elucidation. C illustrated two extracted ion mobilograms (EIM), similar to EIC, EIM is filtered by mass and the peak retention time range. In this case, three isotopic patterns of highest abundance were applied for EICs and EIMs, the time range was set at 8.41 to 8.45 min. Two separated peaks appeared in EIM above, ΔCCS was at 6.7 $Å^2$. Moreover, the interpretation of MS is challenge with coeluting contaminants. The high abundance of PCB-151 dominated the MS of unknown peak, which give a false negative of unknown PCBs. MS extracted by EIM removed ions that do not related to the candidate of interest and interference mass peaks, enabling formula prediction. It proposes a possible contaminant which can be confirmed once more PCB-Cl5 standards are available. The average of Δ m/z, ΔRT and Δ CCS was calculated for each identified compound.

In summary, the mass deviation was below 5 ppm. The CCS deviation was below 1.5%. Benefit of halogenated compounds, three most abundance isotopes were defined as qualifier ions, the identified compounds were confirmed by masses and ion ratios. The results were listed in SI. mSigma was between 55 to 130. Mass deviations of qualified ion were below 5 ppm for the top two isotopes, and below 14 ppm. The deviation of peak ratio was below 30%. The deviation of

isotopic profile is highly dependent on the peak intensity, the peak intensity of PCB-10 was at 2048 cnts, resulting in a deviation of isotopic ratio at 24%. The deviation was within 10% for 88% of isotopic ions.



4. Conclusion

The present work developed a robust and high throughput method using GC-APCI-TIMS-HRMS for GC-amenable POPs analysis. An automated target screening workflow was enabled in TASQ® using an in-house database and applied to real samples. The IMS showed the advantages in complex matrix analysis, including clean up the interfered mass and distinguish isomers. Moreover, CCS value offers an additional identification alignment, which can increase the identification confident level and avoid false positives. The ionization behavior of APCI was also investigated. Charge transfer and proton transfer ion were simultaneously produced in most of chemicals included in this study, however the reproducibility of protonation was unstable. The inter-platform bias of CCS measurement was below 5%, it shows the potential of compatibility of CCS database obtained by various IMS separation technique. The enrich of CCS database can compensate the shortage and irreproducibility of APCI MS spectra library and provide an efficient parameter in identification unknowns.

References

(1) Jones, K. C.; de Voogt, P. Persistent Organic Pollutants (POPs): State of the Science. *Environmental Pollution* **1999**, *100* (1), 209–221. https://doi.org/10.1016/S0269-7491(99)00098-6.

(2) Xu, W.; Wang, X.; Cai, Z. Analytical Chemistry of the Persistent Organic Pollutants Identified in the Stockholm Convention: A Review. *Analytica Chimica Acta* **2013**, *790*, 1–13. https://doi.org/10.1016/j.aca.2013.04.026.

(3) Tang, H. P. Recent Development in Analysis of Persistent Organic Pollutants under the Stockholm Convention. *TrAC Trends in Analytical Chemistry* **2013**, *45*, 48–66. https://doi.org/10.1016/j.trac.2013.01.005.

(4) Montaño, M.; Gutleb, A. C.; Murk, A. J. Persistent Toxic Burdens of Halogenated Phenolic Compounds in Humans and Wildlife. *Environ. Sci. Technol.* **2013**, *47* (12), 6071–6081. https://doi.org/10.1021/es400478k.

(5) Letcher, R. J.; Morris, A. D.; Dyck, M.; Sverko, E.; Reiner, E. J.; Blair, D. A. D.; Chu, S. G.; Shen, L. Legacy and New Halogenated Persistent Organic Pollutants in Polar Bears from a Contamination Hotspot in the Arctic, Hudson Bay Canada. *Science of The Total Environment* **2018**, *610–611*, 121–136. https://doi.org/10.1016/j.scitotenv.2017.08.035.

(6) Wu, Q.; Munschy, C.; Aminot, Y.; Bodin, N.; Vetter, W. High Levels of Halogenated Natural Products in Large Pelagic Fish from the Western Indian Ocean. *Environ Sci Pollut Res* **2021**, *28* (39), 55252–55264. https://doi.org/10.1007/s11356-021-14738-0.

(7) Aznar-Alemany, Ò.; Aminot, Y.; Vilà-Cano, J.; Köck-Schulmeyer, M.; Readman, J. W.; Marques, A.; Godinho, L.; Botteon, E.; Ferrari, F.; Boti, V.; Albanis, T.; Eljarrat, E.; Barceló, D. Halogenated and Organophosphorus Flame Retardants in European Aquaculture Samples. *Science of The Total Environment* **2018**, *612*, 492–500. https://doi.org/10.1016/j.scitotenv.2017.08.199.

(8) Dingemans, M. M. L.; van, den B. M.; Westerink, R. H. S. Neurotoxicity of Brominated Flame Retardants: (In)Direct Effects of Parent and Hydroxylated Polybrominated Diphenyl Ethers on the (Developing) Nervous System. *Environmental Health Perspectives* **2011**, *119* (7), 900–907. https://doi.org/10.1289/ehp.1003035.

(9) Qing Li, Q.; Loganath, A.; Seng Chong, Y.; Tan, J.; Philip Obbard, J. Persistent Organic Pollutants and Adverse Health Effects in Humans. *Journal of Toxicology and Environmental Health, Part A* **2006**, *69* (21), 1987–2005. https://doi.org/10.1080/15287390600751447.

(10) Wong, F.; Hung, H.; Dryfhout-Clark, H.; Aas, W.; Bohlin-Nizzetto, P.; Breivik, K.; Mastromonaco, M. N.; Lundén, E. B.; Ólafsdóttir, K.; Sigurðsson, Á.; Vorkamp, K.; Bossi, R.; Skov, H.; Hakola, H.; Barresi, E.; Sverko, E.; Fellin, P.; Li, H.; Vlasenko, A.; Zapevalov, M.; Samsonov, D.; Wilson, S. Time Trends of Persistent Organic Pollutants (POPs) and Chemicals of Emerging Arctic Concern (CEAC) in Arctic Air from 25 Years of Monitoring. *Science of The Total Environment* **2021**, *775*, 145109. https://doi.org/10.1016/j.scitotenv.2021.145109.

(11) Wit, C. A. de; Vorkamp, K.; Muir, D. Influence of Climate Change on Persistent Organic Pollutants and Chemicals of Emerging Concern in the Arctic: State of Knowledge and Recommendations for Future Research. *Environmental Science: Processes & Impacts* **2022**. https://doi.org/10.1039/D1EM00531F.

(12) 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.; Schymanski, E. L.; Slobodnik, J.;

Thomaidis, N. S.; Thomas, K. V. 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.* **2018**, *52* (9), 5135–5144. https://doi.org/10.1021/acs.est.8b00365.

(13) Hollender, J.; Schymanski, E. L.; Singer, H. P.; Ferguson, P. L. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51* (20), 11505–11512. https://doi.org/10.1021/acs.est.7b02184.

(14) Kim, L.; Lee, D.; Cho, H.-K.; Choi, S.-D. Review of the QuEChERS Method for the Analysis of Organic Pollutants: Persistent Organic Pollutants, Polycyclic Aromatic Hydrocarbons, and Pharmaceuticals. *Trends in Environmental Analytical Chemistry* **2019**, *22*, e00063. https://doi.org/10.1016/j.teac.2019.e00063.

(15) Fang, J.; Zhao, H.; Zhang, Y.; Lu, M.; Cai, Z. Atmospheric Pressure Chemical Ionization in Gas Chromatography-Mass Spectrometry for the Analysis of Persistent Organic Pollutants. *Trends in Environmental Analytical Chemistry* **2020**, *25*, e00076. https://doi.org/10.1016/j.teac.2019.e00076.

(16) Portolés, T.; Mol, J. G. J.; Sancho, J. V.; Hernández, F. Use of Electron Ionization and Atmospheric Pressure Chemical Ionization in Gas Chromatography Coupled to Time-of-Flight Mass Spectrometry for Screening and Identification of Organic Pollutants in Waters. *Journal of Chromatography A* **2014**, *1339*, 145–153. https://doi.org/10.1016/j.chroma.2014.03.001.

(17) Li, X.; Dorman, F. L.; Helm, P. A.; Kleywegt, S.; Simpson, A.; Simpson, M. J.; Jobst, K. J. Nontargeted Screening Using Gas Chromatography–Atmospheric Pressure Ionization Mass Spectrometry: Recent Trends and Emerging Potential. *Molecules* **2021**, *26* (22), 6911. https://doi.org/10.3390/molecules26226911.

(18) Izquierdo-Sandoval, D.; Fabregat-Safont, D.; Lacalle-Bergeron, L.; Sancho, J. V.; Hernández, F.; Portoles, T. Benefits of Ion Mobility Separation in GC-APCI-HRMS Screening: From the Construction of a CCS Library to the Application to Real-World Samples. *Anal. Chem.* **2022**, *94* (25), 9040–9047. https://doi.org/10.1021/acs.analchem.2c01118.

(19) Niu, Y.; Liu, J.; Yang, R.; Zhang, J.; Shao, B. Atmospheric Pressure Chemical Ionization Source as an Advantageous Technique for Gas Chromatography-Tandem Mass Spectrometry. *TrAC Trends in Analytical Chemistry* **2020**, *132*, 116053. https://doi.org/10.1016/j.trac.2020.116053.

(20) Ridgeway, M. E.; Lubeck, M.; Jordens, J.; Mann, M.; Park, M. A. Trapped Ion Mobility Spectrometry: A Short Review. *International Journal of Mass Spectrometry* **2018**, *425*, 22–35. https://doi.org/10.1016/j.ijms.2018.01.006.

(21) Michelmann, K.; Silveira, J. A.; Ridgeway, M. E.; Park, M. A. Fundamentals of Trapped Ion Mobility Spectrometry. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (1), 14–24. https://doi.org/10.1007/s13361-014-0999-4.

(22) Celma, A.; Sancho, J. V.; Schymanski, E. L.; Fabregat-Safont, D.; Ibáñez, M.; Goshawk, J.; Barknowitz, G.; Hernández, F.; Bijlsma, L. Improving Target and Suspect Screening High-Resolution Mass Spectrometry Workflows in Environmental Analysis by Ion Mobility Separation. *Environ. Sci. Technol.* **2020**, *54* (23), 15120–15131. https://doi.org/10.1021/acs.est.0c05713.

(23) Dodds, J. N.; Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (11), 2185–2195. https://doi.org/10.1007/s13361-019-02288-2.
(24) Hinnenkamp, V.; Klein, J.; Meckelmann, S. W.; Balsaa, P.; Schmidt, T. C.; Schmitz, O. J. Comparison of CCS Values Determined by Traveling Wave Ion Mobility Mass Spectrometry and Drift Tube Ion Mobility Mass Spectrometry. *Anal. Chem.* **2018**, *90* (20), 12042–12050. https://doi.org/10.1021/acs.analchem.8b02711.
(25) Feuerstein, M.; Hernández-Mesa, M.; Kiehne, A.; Bizec, B. L.; Hann, S.; Dervilly, G.; Causon, T. Comparability of Steroid Collision Cross Sections Using Three Different IM-HRMS Technologies: An Interplatform Study. **2022**. https://doi.org/10.26434/chemrxiv-2022-87k68.
(26) Plante, P.-L.; Francovic-Fontaine, É.; May, J. C.; McLean, J. A.; Baker, E. S.; Laviolette, F.; Marchand, M.; Corbeil, J. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. *Anal. Chem.* **2019**, *91* (8), 5191–5199. https://doi.org/10.1021/acs.analchem.8b05821.
(27) Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z.-J. Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2016**, *88* (22), 11084–11091. https://doi.org/10.1021/acs.analchem.6b03091.
(28) Bhaskar Reddy, A. V.; Moniruzzaman, M.; Madhavi, G.; Aminabhavi, T. M. Modern Approaches in Separation, Identification and Quantification of Polychlorinated Biphenyls. *Current Opinion in Environmental Science & Health* **2020**, *18*, 26–39. https://doi.org/10.1016/j.coesh.2020.06.003.
(29) Di Lorenzo, R. A.; Lobodin, V. V.; Cochran, J.; Kolic, T.; Besevic, S.; Sled, J. G.; Reiner, E. J.; Jobst, K. J. Fast Gas Chromatography-Atmospheric Pressure (Photo)Ionization Mass Spectrometry of Polybrominated Diphenylether Flame Retardants. *Analytica Chimica Acta* **2019**, *1056*, 70–78. https://doi.org/10.1016/j.aca.2019.01.007.
(30) Geng, D.; Kukucka, P.; Jogsten, I. E. Analysis of Brominated Flame Retardants and Their Derivatives by Atmospheric Pressure Chemical Ionization Using Gas Chromatography Coupled to Tandem Quadrupole Mass Spectrometry. *Talanta* **2017**, *162*, 618–624. https://doi.org/10.1016/j.talanta.2016.10.060.
(31) Williamson, D. L.; Nagy, G. Isomer and Conformer-Specific Mass Distribution-Based Isotopic Shifts in High-Resolution Cyclic Ion Mobility Separations. *Anal. Chem.* **2022**, *94* (37), 12890–12898. https://doi.org/10.1021/acs.analchem.2c02991.
(32) Kirk, A. T.; Bohnhorst, A.; Raddatz, C.-R.; Allers, M.; Zimmermann, S. Ultra-High-Resolution Ion Mobility Spectrometry—Current Instrumentation, Limitations, and Future Developments. *Anal Bioanal Chem* **2019**, *411* (24), 6229–6246. https://doi.org/10.1007/s00216-019-01807-0.
(33) Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* **2019**, *91* (13), 8564–8573. https://doi.org/10.1021/acs.analchem.9b01838.
(34) *Mass Spectrometry Data Center, NIST*. https://chemdata.nist.gov/ (accessed 2022-10-14).
(35) Schulze, T.; Meier, R.; Alygizakis, N.; Schymanski, E.; Bach, E.; LI, D. H.; raalizadeh; Hoffmann, N.; Tanaka, S.; Witting, M.; Treutler, H.; kohlhoff. *MassBank/MassBank-Data: Release Version 2021.02*; Zenodo, 2021. https://doi.org/10.5281/zenodo.4506480.
(36) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating

Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105.
(37) Pico, Y.; Alfarhan, A. H.; Barcelo, D. How Recent Innovations in Gas Chromatography-Mass Spectrometry Have Improved Pesticide Residue Determination: An Alternative Technique to Be in Your Radar. *TrAC Trends in Analytical Chemistry* **2020**, *122*, 115720. https://doi.org/10.1016/j.trac.2019.115720.
(38) Ayala-Cabrera, J. F.; Santos, F. J.; Moyano, E. Recent Advances in Analytical Methodologies Based on Mass Spectrometry for the Environmental Analysis of Halogenated Organic Contaminants. *Trends in Environmental Analytical Chemistry* **2021**, *30*, e00122. https://doi.org/10.1016/j.teac.2021.e00122.
(39) Léon, A.; Cariou, R.; Hutinet, S.; Hurel, J.; Guitton, Y.; Tixier, C.; Munschy, C.; Antignac, J.-P.; Dervilly-Pinel, G.; Le Bizec, B. HaloSeeker 1.0: A User-Friendly Software to Highlight Halogenated Chemicals in Nontargeted High-Resolution Mass Spectrometry Data Sets. *Anal. Chem.* **2019**, *91* (5), 3500–3507. https://doi.org/10.1021/acs.analchem.8b05103.
(40) Loos, M. *EnviMass Version 3.1*; Zenodo, 2016. https://doi.org/10.5281/zenodo.48164.
(41) Matysik, S.; Schmitz, G.; Bauer, S.; Kiermaier, J.; Matysik, F.-M. Potential of Gas Chromatography–Atmospheric Pressure Chemical Ionization–Time-of-Flight Mass Spectrometry for the Determination of Sterols in Human Plasma. *Biochemical and Biophysical Research Communications* **2014**, *446* (3), 751–755. https://doi.org/10.1016/j.bbrc.2014.01.026.
(42) Nácher-Mestre, J.; Serrano, R.; Portolés, T.; Berntssen, M. H. G.; Pérez-Sánchez, J.; Hernández, F. Screening of Pesticides and Polycyclic Aromatic Hydrocarbons in Feeds and Fish Tissues by Gas Chromatography Coupled to High-Resolution Mass Spectrometry Using Atmospheric Pressure Chemical Ionization. *J. Agric. Food Chem.* **2014**, *62* (10), 2165–2174. https://doi.org/10.1021/jf405366n.
(43) Schymanski, E.; Zhang, J.; Thiessen, P.; Bolton, E. Experimental CCS Values in PubChem, 2022. https://doi.org/10.5281/zenodo.7056297.
(44) Ross, D. H.; Cho, J. H.; Xu, L. Breaking down Structural Diversity for Comprehensive Prediction of Ion-Neutral Collision Cross Sections. *Analytical chemistry* **2020**, *92* (6), 4548–4557.
(45) Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A. Collision Cross Section Compendium to Annotate and Predict Multi-Omic Compound Identities. *Chem. Sci.* **2019**, *10* (4), 983–993. https://doi.org/10.1039/C8SC04396E.
(46) *Collision Cross Section Database*. https://brcwebportal.cos.ncsu.edu/baker/view.php (accessed 2022-10-19).
(47) Zheng, X.; Dupuis, K. T.; Aly, N. A.; Zhou, Y.; Smith, F. B.; Tang, K.; Smith, R. D.; Baker, E. S. Utilizing Ion Mobility Spectrometry and Mass Spectrometry for the Analysis of Polycyclic Aromatic Hydrocarbons, Polychlorinated Biphenyls, Polybrominated Diphenyl Ethers and Their Metabolites. *Analytica Chimica Acta* **2018**, *1037*, 265–273. https://doi.org/10.1016/j.aca.2018.02.054.
(48) Castellanos, A.; Benigni, P.; Hernandez, D. R.; Debord, J. D.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Fast Screening of Polycyclic Aromatic Hydrocarbons Using Trapped Ion Mobility Spectrometry-Mass Spectrometry. *Analytical Methods* **2014**, *6* (23), 9328–9332. https://doi.org/10.1039/c4ay01655f.
(49) Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory

Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Anal. Chem.* **2017**, *89* (17), 9048–9055. https://doi.org/10.1021/acs.analchem.7b01729.

SI

timsTOF settings in DIA mode

Table S1 Parameter settings in the timsTOF

APCI settings	
Capillary transfer	250 °C
Sampling cone	3000 V
Plate	500 V
Nebulizer gas	4 Bars
Capillary gas	2 L/min
timsTOF settings	
Mass range	150-1250 m/z
electrospray voltage	*4.5 kV*
rf funnel 1	*300 Vpp*
rf funnel 2	*400 Vpp*
multipole rf	*500 Vpp*
quadrupole ion energy	*5 eV*
collision energy	*12 eV*
collision rf	*1000 Vpp*
transfer time	*60 µs*
repulse storage	*7 µs*
ramp time	*280-350 ms*
accumulation time	*10 ms*
rolling range	*25x*
duty cycle	*3.5Hz*

Figure S1 Overview of PCBs and PBDEs

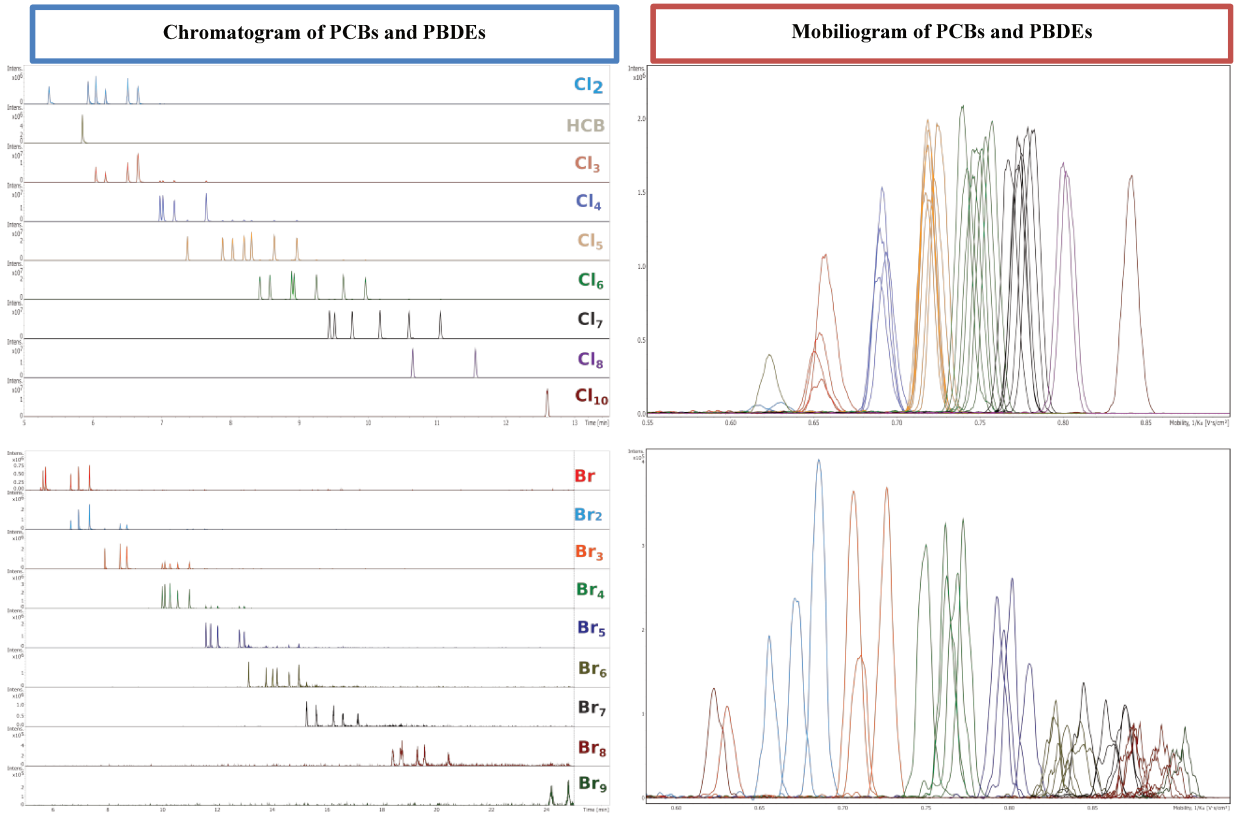


Figure S1 EIC and EIM of PCB congeners (Above) and PBDE (Below) congeners. Expect PCB-28/31, all the isomers can be separated in chromatogram or/and mobiliogram within 25 min acquisition tims.

References

R. Aalizadeh, V. Nikolopoulou, N. A. Alygizakis, and N. S. Thomaidis. First novel workflow for semiquantification of emerging contaminants in environmental samples analyzed by gas chromatography–atmospheric pressure chemical ionization–quadrupole time of flight–mass spectrometry. *Analytical chemistry*, 94(27):9766–9774, 2022.

S. Alfonso, M. Blanc, L. Joassard, S. H. Keiter, C. Munschy, V. Loizeau, M.-L. Bégout, and X. Cousin. Examining multi-and transgenerational behavioral and molecular alterations resulting from parental exposure to an environmental pcb and pbde mixture. *Aquatic Toxicology*, 208:29–38, 2019.

L. Cherta, T. Portolés, E. Pitarch, J. Beltran, F. López, C. Calatayud, B. Company, and F. Hernández. Analytical strategy based on the combination of gas chromatography coupled to time-of-flight and hybrid quadrupole time-of-flight mass analyzers for non-target analysis in food packaging. *Food chemistry*, 188: 301–308, 2015.

J. Dron, E. Wafo, P. Boissery, F. Dhermain, M. Bouchoucha, P. Chamaret, and D. Lafitte. Trends of banned pesticides and pcbs in different tissues of striped dolphins (stenella coeruleoalba) stranded in the northwestern mediterranean reflect changing contamination patterns. *Marine Pollution Bulletin*, 174:113198, 2022.

J. Fang, H. Zhao, Y. Zhang, M. Lu, and Z. Cai. Atmospheric pressure chemical ionization in gas chromatography-mass spectrometry for the analysis of persistent organic pollutants. *Trends in Environmental Analytical Chemistry*, 25:e00076, 2020.

S. Fernando, M. K. Green, K. Organtini, F. Dorman, R. Jones, E. J. Reiner, and K. J. Jobst. Differentiation of (mixed) halogenated dibenzo-p-dioxins by negative ion atmospheric pressure chemical ionization. *Analytical chemistry*, 88(10):5205–5211, 2016.

F. Hernández, M. Ibáñez, T. Portolés, M. I. Cervera, J. V. Sancho, and F. J. López. Advancing towards universal screening for organic pollutants in waters. *Journal of Hazardous Materials*, 282:86–95, 2015.

D. Izquierdo-Sandoval, D. Fabregat-Safont, L. Lacalle-Bergeron, J. V. Sancho, F. Hernández, and T. Portoles. Benefits of ion mobility separation in gc-apci-hrms screening: From the construction of a ccs library to the application to real-world samples. *Analytical Chemistry*, 94(25):9040–9047, 2022.

D.-X. Li, L. Gan, A. Bronja, and O. J. Schmitz. Gas chromatography coupled to atmospheric pressure ionization mass spectrometry (gc-api-ms). *Analytica chimica acta*, 891:43–61, 2015.

Y. Liber, B. Mourier, P. Marchand, E. Bichon, Y. Perrodin, and J.-P. Bedell. Past and recent state of sediment contamination by persistent organic pollutants (pops) in the rhône river: Overview of ecotoxicological implications. *Science of the Total Environment*, 646:1037–1046, 2019.

C. Lipok, J. Hippler, and O. J. Schmitz. A four dimensional separation method based on continuous heart-cutting gas chromatography with ion mobility and high resolution mass spectrometry. *Journal of Chromatography A*, 1536:50–57, 2018.

A. MacNeil, X. Li, R. Amiri, D. C. Muir, A. Simpson, M. J. Simpson, F. L. Dorman, and K. J. Jobst. Gas chromatography-(cyclic) ion mobility mass spectrometry: A novel platform for the discovery of unknown per-/polyfluoroalkyl substances. *Analytical Chemistry*, 94(31):11096–11103, 2022.

C. Munschy, J. Spitz, N. Bely, K. Héas-Moisan, N. Olivier, C. Pollono, and T. Chouvelon. A large diversity of organohalogen contaminants reach the meso-and bathypelagic organisms in the bay of biscay (northeast atlantic). *Marine Pollution Bulletin*, 184:114180, 2022.

Y. Niu, J. Liu, R. Yang, J. Zhang, and B. Shao. Atmospheric pressure chemical ionization source as an advantageous technique for gas chromatography-tandem mass spectrometry. *TrAC Trends in Analytical Chemistry*, 132:116053, 2020.

P. Rostkowski, P. Haglund, R. Aalizadeh, N. Alygizakis, N. Thomaidis, J. B. Arandes, P. B. Nizzetto, P. Booij, H. Budzinski, P. Brunswick, et al. The strength in numbers: comprehensive characterization of house dust using complementary mass spectrometric techniques. *Analytical and bioanalytical chemistry*, 411(10): 1957–1977, 2019.

N. Sarkis, C. Meymy, O. Geffard, Y. Souchon, A. Chandesris, M. Ferréol, L. Valette, R. Recoura-Massaquant, J. Piffady, A. Chaumot, et al. Quantification of multiscale links of anthropogenic pressures with pah and pcb bioavailable contamination in french freshwaters. *Water Research*, 203:117546, 2021.

Q. Wu, C. Munschy, Y. Aminot, N. Bodin, and W. Vetter. High levels of halogenated natural products in large pelagic fish from the western indian ocean. *Environmental Science and Pollution Research*, 28(39):55252–55264, 2021.

W. Xu, X. Wang, and Z. Cai. Analytical chemistry of the persistent organic pollutants identified in the stockholm convention: A review. *Analytica Chimica Acta*, 790:1–13, 2013.

D. Zacs, I. Perkons, and V. Bartkevics. Evaluation of analytical performance of gas chromatography coupled with atmospheric pressure chemical ionization fourier transform ion cyclotron resonance mass spectrometry (gc-apci-ft-icr-ms) in the target and non-targeted analysis of brominated and chlorinated flame retardants in food. *Chemosphere*, 225:368–377, 2019.

Chapter 3

Non-Target Analysis With UPLC-timsTOF For Water Samples

Summary

The project presented in this chapter aims to develop a generic LC-HRMS method for target and non-target analysis in divers water matrices.

An overview of the state-of-the-art method is first introduced in this chapter. The concepts and procedures of method development, including sample preparation, standard preparation, construction of in-house databases, and data treatment, are explain in the following sections.

Sample preparation is minimized to prevent loss of compounds of interest and cross-contamination through manipulation. An in-house database of small molecules and emerging contaminants was first built in this project. Target compound identification processing was critically defined in TASQ, a commercial software provided by Bruker Daltonics. Non-target analysis was processed in the open-source software, MS-DIAL, and commercial software, Data Analysis (Bruker Daltonics). To increase data treatment efficiency and identification certainty, the identification procedures are carefully described in this chapter. Afterward, the whole workflows of target and non-target analysis were evaluated by applying it to various and complex water samples. The results and perspectives are discussed at the end of this chapter.

3.1 State-of-the-art method

Suspect and non-target screening (NTS) using high-resolution mass spectrometry have diverse applications for the identification of large-scope chemicals, such as food analysis (Riedl et al., 2015), forensic analysis (Steuer et al., 2022), metabolomics analysis (Dunn et al., 2013), and environmental analysis (González-Gaya et al., 2021). In environmental analysis, water is the most common matrix for polar contaminants of emerging concern (CECs) (Gago-Ferrero et al., 2020). We could always consider that there is no life without water, and the access and the quality of drinking water and tap water are wisely recognized to have a close relationship to human health. Although particular regulations prevent or minimize the pollution of drinking and tap water as much as possible, a large number of chemicals remain unknown or nonregulated (Paszkiewicz et al., 2022). Natural waters, especially surface water and groundwater, play an important role in the risk assessment of ecological systems. Anthropic activities are the main source of aquatic environmental pollution. A number of chemicals reach the environment through agricultural activities, debris from commercial products, insufficient disposal during wastewater treatment, etc. With the development of analytical instruments and techniques, the reports of newly detected contaminants raise the global concern of ecological effects and human health. Pesticides and pharmaceutical and personal care products (PPCPs) are frequently detected in surface and drinking water (Yang et al., 2017). Persistent and mobile organic chemicals (PMOCs) are highly polar and highly persistent substances (Reemtsma et al., 2016) that can easily escape from conventional wastewater treatment and then be released into the environment. Recently, per- and polyfluoroalkyl substances (PFAS, which are known as eternal pollutants) and bisphenol A are two classes of PMOCs that have been detected through non-targeted water analysis and are added to the list for routine monitoring (Paszkiewicz et al., 2022). In recent years, thousands of CECs have been identified and tentatively identified in water samples (Menger et al., 2020; Schulze et al., 2020), highlighting the value of HRMS and non-target analysis.

The generic workflow of target and non-target analyses in water includes 1) sampling and sample treatment, 2) analysis by LC-HRMS, 3) data preprocessing, and 4) data treatment.

Since this study mainly focuses on data preprocessing and data treatment, the discussion will reference the last two steps. Moreover, water sample extractions are commonly based on solid phase extraction (SPE) with different sorbents, in which the hydrophilic-lipophilic balance (HLB) is the most used for a wide scope of screening (Menger et al., 2020). Benefiting from the increasing sensitivity of HRMS, direct injection has also been applied in water analysis (Menger et al., 2020; Hollender et al., 2019).

SPE is still dominant with regards to the sample preparation for non-target analysis, and direct injection with a large volume is an alternative(Albergamo et al., 2018; Li et al., 2018; Backe, 2021). Li et al. (2018) compared SPE with direct injection, which showed that SPE had a lower matrix effect than direct injection; however, direct injection was more suitable to evaluate the removal efficiency of wastewater treatment. Another study also used direct injection to evaluate the efficiency of

wastewater plant mitigation (Nürenberg et al., 2019). Köppe et al. (2020) injected 100 μL river water samples after a filtration step to identify unknown chemical contaminants. Large volume injection (650 μL) was also effectively used in different types of water analyses (Backe, 2021).

Direct injection reduces the sample preparation steps, preventing potential contamination and loss of substances through manipulation. In the case of (non-salty) water, minimizing the matrices without the harsh consequences of the representative of the samples and the global sample preparation and analysis time could be easily done by a simple offline or online dilution. However, it is sometimes challenging to detect ultratrace-level substances without preconcentration or desalination.

3.2 Chemicals and standard preparation

3.2.1 Chemicals and solvents

The standard solution of pesticides and pharmaceutical product kit were purchased from Agilent Technologies and Resteck. Twenty-five standard kits, with a total of 559 reference chemicals , were used to build an in-house target screening database for LC suspect analyses.

Ammonium formate, formic acid, absolute methanol, isoptropanol, and acetonitrile of the highest purity level (ULC/MS - CC/SFC) were purchased from Biosolve Chemicals (Dieuze, France). Formic acid 99% was purchased from Fisher Scientific (Geel, Belgium). Purified water was provided by a Milli-Q® purification system (Millipore, Bedford, MA).

3.2.2 Solution preparation

Concentrated buffer solution

A 500x concentrate buffer solution was prepared and stocked at 4 °C for the mobile phase. The procedure was as follows.

1. Weigh 1.58 g (25 mmol of ammonium formate) and dissolve in 7.5 mL Milli-Q water.
2. Complete the buffer solution with 2 mL MeOH and 500 μ L.
3. Obtain 2.5 mol/L ammonium formate with 5% formic acid concentrated buffer (500X concentrated). Store at 4 °C.

Mobile phases

Mobile Phase A consisted of a mix of purified water/MeOH (v/v: 99:1) with 5 mmol/L ammonium acetate and 0.1% formic acid, which was diluted from the stock concentrated buffer solution. Mobile Phase B consisted of MeOH with 5 mmol/L ammonium formate and 0.1% formic acid. MeOH was added to a fresh open bottle (1 L), and 2 mL of the concentrated buffer solution was added.

3.2.3 Standard mixtures and calibration preparation

Working solutions of standard kits were prepared at a concentration of 1 μg/mL in acetonitrile and stored at 4 °C. Mixtures of standard solutions were diluted in purified water to a final concentration of 10 ng/mL and stored at 4 °C. The standard mixture solutions were used for method optimization and database development. The original purchased standard kits were stored at -5 °C.

The calibration solutions have 7 levels, which are denoted as STD 1 - STD 7. To avoid manipulation mistakes during dilution, a succession dilution was performed using different solutions. It is described in Table 3.1:

	Calibrator (ng/mL	Vf (ul)	Vtake (ul)	Taken Solution	solvent
STD1	0.25	1000	12.5	STD5	987.5
STD2	2	500	50	STD5	450
STD3	5	500	50	STD7	450
STD4	10	500	100	STD7	400
STD5	20	500	10	work solution*	490
STD6	35	1000	35	work solution	965
STD7	50	1000	50	work solution	950

* Work solution = 1 µg / mL

Table 3.1: Calibration plan of pesticide kit

3.3 LC-HRMS setting and optimization

3.3.1 QA/QC mix of standards

Non-target analytical methods aim to cover a wide range of chemicals with broad physico-chemical properties. Therefore, a generic LC separation and MS detection method are both required to enable the sensitivity for as many compounds as possible. To assess these analytical conditions, a mixture of the reference standards at 10 ng/mL was used in method optimization (Table 3.2).

Compound	Formula	m/z (Da)	RT(min)	LogP
Methamidophos	$C_2H_8NO_2PS$	142.0086	2.94	-0.9
Acephate	$C_4H_{10}NO_3PS$	184.0192	3.22	-0.85
Omethoate	$C_5H_{12}NO_4PS$	214.0297	3.39	-0.9
Spiromesifen	$C_{23}H_{30}O_4$	371.227	3.97	5.1
Monocrotophos	$C_7H_{14}NO_5P$	224.0682	3.98	-0.2
Dicrotophos	$C_8H_{16}NO_5P$	238.0839	4.09	0
Vamidothion	$C_8H_{18}NO_4PS_2$	288.0488	4.39	0.3
Dimethoate	$C_5H_{12}NO_3PS_2$	230.0069	4.53	0.78
Mevinphos	$C_7H_{13}O_6P$	225.0523	4.85	1.2
Carbaryl	$C_{12}H_{11}NO_2$	202.086255	5.82	2.36
Isocarbophos	$C_{11}H_{16}NO_4PS$	290.0611	6.36	2.4
Dimethomorph	$C_{21}H_{22}ClNO_4$	388.131	7.23	3.9
Spirotetramate	$C_{21}H_{27}NO_5$	374.1962	7.52	3.2
Fenpropimorph	$C_{20}H_{33}NO$	304.2635	8.75	5.2
Spinosad	$C_{41}H_{65}NO_{10}$	732.4681	9.07	5.8
Spinetoram J&L	$C_{42}H_{69}NO_{10}$	748.4994	9.59	5.9
Temephos	$C_{16}H_{20}O_6P_2S_3$	466.997	9.69	6
Spirodiclofen	$C_{21}H_{24}Cl_2O_4$	411.1124	10.62	5.9

Table 3.2: QA/QC mix

3.3.2 LC separation settings

The chromatographic separation was performed with a UPLC Acquity system (Waters, USA). The system was operated with MassLynx V4.1 (Waters). An Acclaim™ 120 C18 column (100 x 2.1 mm i.d., 2.2 μm, 5-μm particle size) from Thermo Fisher Scientific (Dreieich, Germany) preceded by an Acuity UPLC BEH C18 VanGuard precolumn (1.7 μm, 2.1 mm x 5 mm, Waters, USA) was used for chromatographic separation. Mobile Phase A consisted of 5 mmol/L ammonium formate in ultrapure water, with 1% methanol and 0.1% formic acid added to Mobile Phase A to adjust the pH. Mobile Phase B was methanol with 5 mmol/L ammonium formate with 0.1% formic acid. The LC acquisition time was 20 min, and the optimized gradient is shown in Table 3.3. The column temperature was set at 40 °C, and the autosampler was set to 5 °C.

The positive ionization mode (+ESI) method previously developed for target screening of water contaminants by Bruker Daltonics was used as a basis for the present methodological work in non-target screening (Gago-Ferrero et al., 2020). The initial stationary phase was a Bruker Solo column (100 mm × 2.1 mm i.d., 5-μm particle size). An Acclaim™ 120 C18 column (100 mm × 2.1 mm i.d., 5-μm particle size,

ThermoFisher, USA) was used as an alternative. The initial elution gradient started at 4% MPB and became a 99.9% organic solvent in the wash platform. To increase the interaction of the hydrophilic analytes with the stationary phase, the percentage of the organic phase decreased to 2%. To protect the stationary phase, the presence of the organic phase was limited to 98%. The full separation gradient is presented in Table 3.3.

Positive ESI gradient					
	Retention [min]	Flow [mL/min]	%A1	%B1	Slope
1	0	0.2	96	4	
2	0.1	0.2	96	4	6
3	1	0.2	82	18	5
4	2.5	0.3	50	50	7
5	14	0.4	2	98	5
6	16	0.48	2	98	6
7	16.1	0.48	96	4	6
8	19	0.48	96	4	6
9	19.1	0.2	96	4	6
10	20	0.2	96	4	6

Table 3.3: LC gradient elution program

3.3.3 HRMS settings

Trapped ion mobility spectrometry coupled with a time-of-flight high-resolution mass spectrometer (timsTOF, Bruker Daltonics, Bremen, Germany) was used in the positive ESI mode for the acquisition in the MS mode and MSMS mode (bbCID). The instrument was controlled through OTOF (Bruker) Version 3.0.

The (+ESI) interface was set as follows: the end plate offset is 500 V; the capillary voltage is 4500 V; the nebulizer is 2.8 bar; the dry gas is 4 L/min; and the dry temperature is 230 °C. The QTOF MS system was operated in the Data Independent Acquisition (DIA) mode, which is called the broadband collision-induced dissociation (bbCID) by Bruker systems. In the DIA mode, all ions within a selected m/z range are fragments, which provides the MS and MS/MS spectra simultaneously by alternating low CID energy and high CID energy. The MS full scan was operated between 100 and 1250 m/z at 6 eV (low collision energy). The MS/MS scan was obtained at 35 eV (high collision energy) over the range m/z 80-1000. The scan speed was 2 Hz per cycle. External-mass calibrations were performed using the Agilent ESI low-concentration tune mix by infusion before the acquisition sequences.

An overview of the LC separation is presented in Figure 3.1, and the distribution of the standard retention times is summarized in the pi-chart (Figure 3.2). Ninety-six percent of the standards were eluted within 10 min, where 83% of methanol was introduced in the stationary phase and the standards covered the full LC program (RT between 2 and 14 min). The substances of interest were water phase contaminants, which have hydrophilic characteristics. Therefore, the optimized LC-HRMS

method was adapted well for a wide scope of water contamination screening. An example of a dataset for database construction is summarized in Table 3.4.

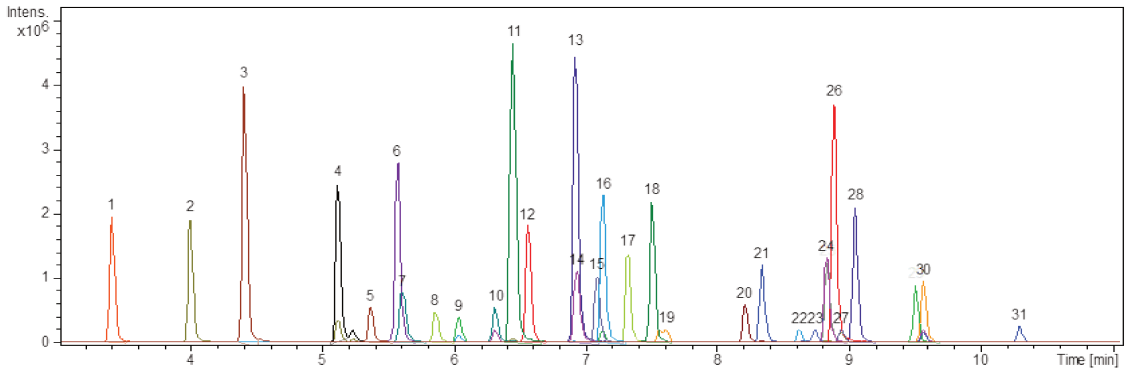


Figure 3.1: Extracted ion chromatograms obtained for pesticide standards in the optimized LC method.

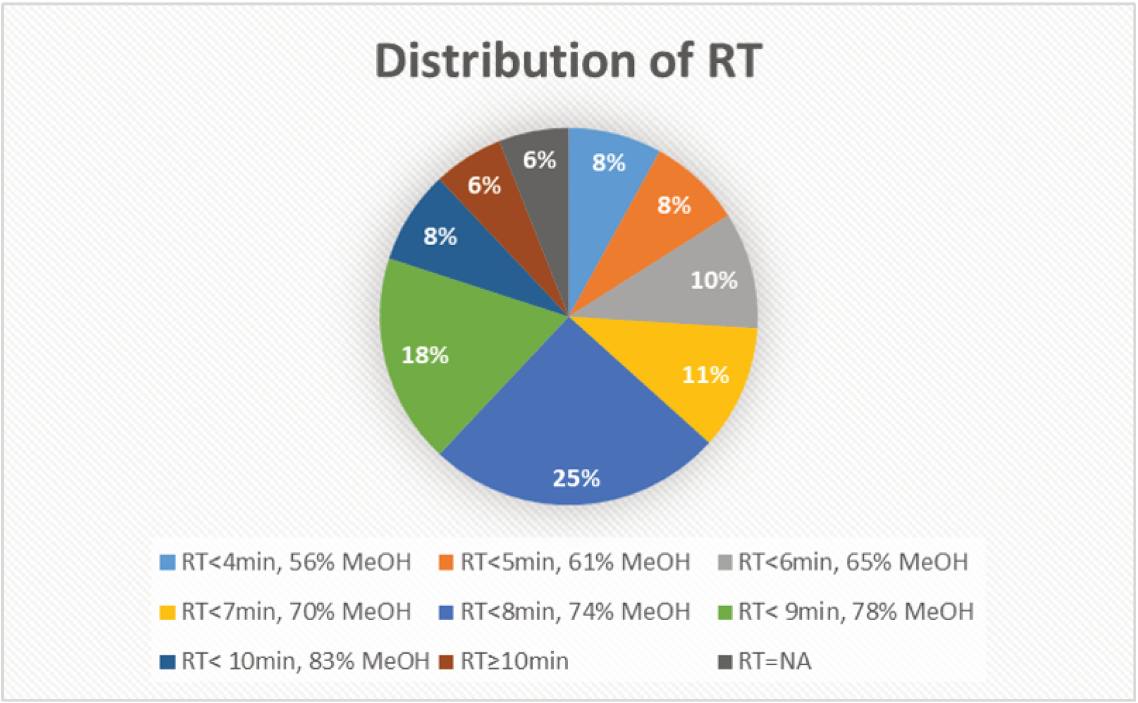


Figure 3.2: Distribution of the standard retention time.

Peak N°	Compound	formula	m/z [Da]	RT [min]	Fragment 1			Fragment 2			Fragment 3		
					ion	m/z [Da]	Ion ratio	ion	m/z [Da]	Ion ratio	ion	m/z [Da]	Ion ratio
1	Omethoate	C5H12NO4PS	214.0297	3.42	C4H8O4PS	182.9875	100	C2H6O2PS	124.9821	100			
2	Monocrotophos	C7H14NO5P	224.0682	4.02	C2H8O4P+	127.0155	100	C6H10O5P+	193.026	51.6	M+H	224.0682	11.3
3	Vamidothion	C8H18NO4PS2	288.0488	4.43	C6H12NOS+	146.0634	100	C4H8NOS+	118.0321	100	M+H	288.0488	8.7
4	Phosphamidon	C10H19ClNO5P	300.0762	5.14	M+H	300.0762	100	M+H+2	302.0737	31.4	C2H8O4P+	127.0155	100
5	Metolcarb	C9H11NO2	166.086255	5.38	C7H9O+	109.0648	100						
6	Ofurace	C14H16ClNO3	282.0891	5.59	M+H	282.0891	100	M+H+2	284.0867	29.3	C13H17ClNO2+	254.0927	91.4
7	Thidiazuron	C9H8N4OS	221.0492	5.62	M+H	221.0492	61.3						
8	Thiodicarb	C10H18N4O4S3	355.0563	5.87	C3H6NS+	88.0215	100	M+H	355.0563	43.7			
9	Ethiofencarb	C11H15NO2S	226.0896	6.05	C7H7O+	107.0491	100	C9H10NO2+	164.0706	56.3			
10	Isoprocarb	C11H15NO2	194.1176	6.32	C6H7O+	95.0491	72	C9H13O+	137.0961	59.9	M+H	194.1176	31.3
11	Methabenzthiazuron	C10H11N3OS	222.0696	6.46	C8H9N2S+	165.0481	100	M+H	222.0696	32.9			
12	Diuron	C9H10Cl2N2O	233.0243	6.57	M+H	233.0243	100	M+H+2	235.0214	62.6	C3H6NO	72.0444	86
13	Diethofencarb	C14H21NO4	268.154335	6.93	C11H16NO4+	226.1074	100	C8H10NO2+	152.0706	21.9	C6H6NO2+	124.0393	100
14	Fenobucarb	C12H17NO2	208.1332	6.95	C8H10NO2+	152.0706	24.2	C6H7O+	95.0491	11.2			
15	Linuron	C9H10Cl2N2O2	249.0192	7.1	M+H	249.0192	100	M+H+2	251.0164	64.8	C11H4NO2+	182.0241	59.3
16	Methiocarb	C11H15NO2S	226.0896	7.14	C9H13OS+	169.0682	100	C8H9O+	121.0648	47.9			
17	Malathion	C10H19O6PS2	331.043344	7.33	C6H7O3+	127.039	88.1	C8H14O5PS2+	285.0015	100	M+H	331.043344	19.5
18	Triazophos	C12H16N3O3PS	314.072276	7.51	M+H	314.072276	100	C8H8N3O+	162.0662	56.7			
19	Iprovalicarb	C18H28N2O3	321.2173	7.62	C9H19N2O3+	203.139	37	M+H	321.2173	11.2			
20	Phenthoate	C12H17O4PS2	321.037865	8.21	C9H12O2PS2	247.0011	100						
21	Phoxim	C12H15N2O3PS	299.061377	8.34	M+H	299.061377	100	C10H12N2O3PS+	271.0301	42.3			
22	Quinalphos	C12H15N2O3PS	299.061377	8.34	M+H	299.061377	100						
23	Triflumuron	C15H10ClF3N2O3	359.040481	8.63	M+H	359.040481	65.2						
24	Phosalone	C12H15ClNO4PS2	367.994141	8.75	C8H5ClNO2+	182.0003	100	M+H	367.994141	43.6			
25	Pyrazophos	C14H20N3O5PS	374.093405	8.83	M+H	374.093405	100						
26	Pencycuron	C19H21ClN2O	329.141517	8.84	M+H	329.141517	92.4	M+H+2	331.1375	29.5			
27	Thiobencarb	C12H16ClNOS	258.071389	8.94	C7H6Cl+	125.0153	100	M+H	258.071389	39			
28	Piperophos	C14H28NO3PS2	354.1321	9.05	M+H	354.1321	100						
29	Profenophos	C11H15BrClO3PS	372.942419	9.51	M+H+2	374.940163	100	M+H	372.942419	32	302.8642	C6H6BrClO3PS+	22.5
30	Quizalofop-ethyl	C19H17ClN2O4	373.094961	9.57	M+H	373.094961	100						
32	Flufenoxuron	C21H11ClF6N2O3	489.043515	10.29	M+H	489.043515	100	M+H+2	491.039	34.1			

Table 3.4: LC-MS database of a pesticides standard mixture

3.4 Data processing

The preprocessing of liquid chromatography raw data was performed in Compass Data Analysis Version 5.0. Suspect screening and non-target screening were performed, as shown in Figure 3.3. The workflow and identification confidence criteria were based on the guidelines proposed by (Schymanski et al., 2014).

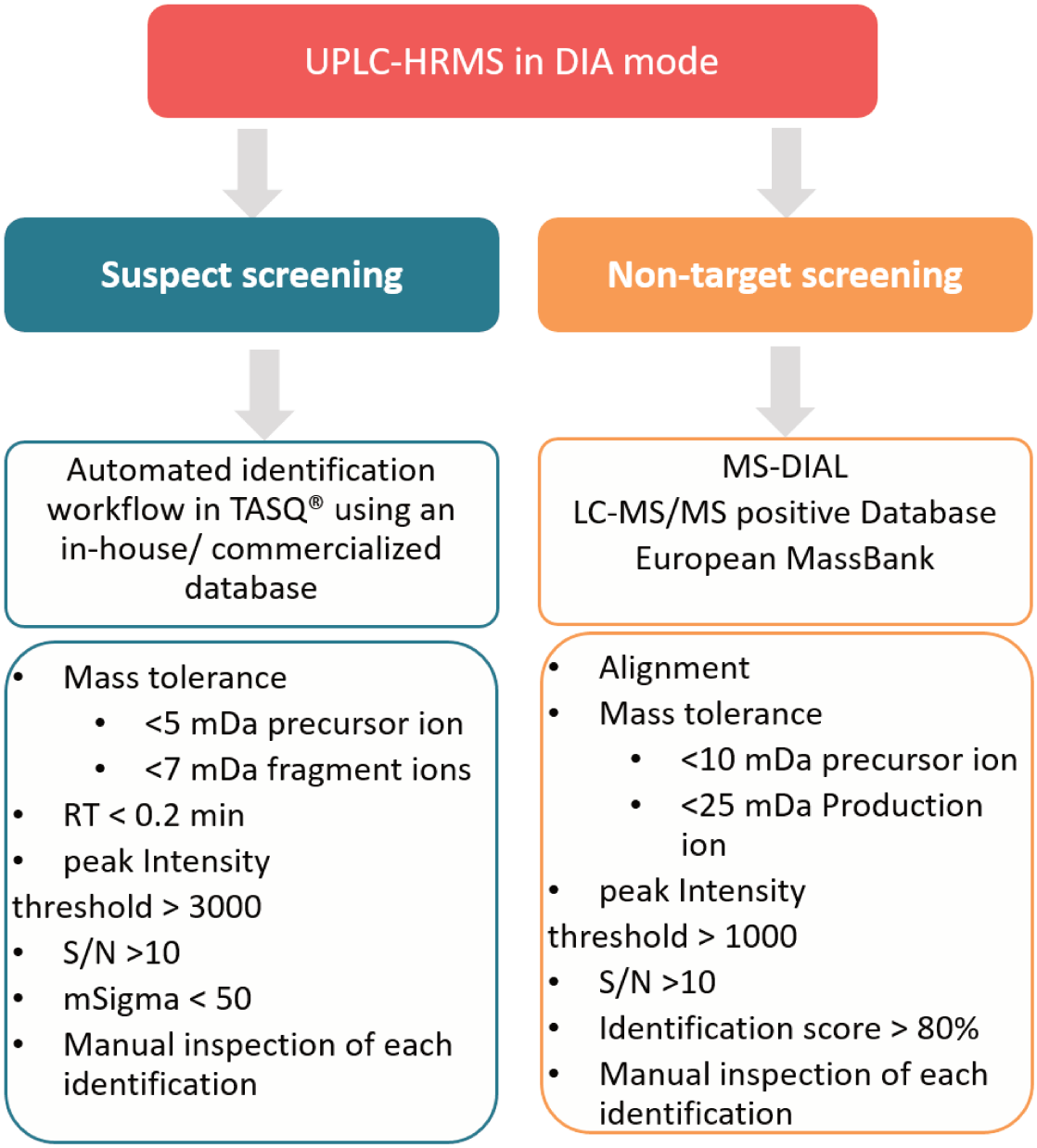


Figure 3.3: The workflow of suspect screening and non-target screening using UPLC-HRMS

3.4.1 Suspect screening

An automated suspect screening workflow was enabled in TASQ® Client 2021 (Bruker Daltonics, Bremen, Germany). The suspect list consisted of over 3000 components, including pesticides, pharmaceuticals, etc. The in-house measured ref-

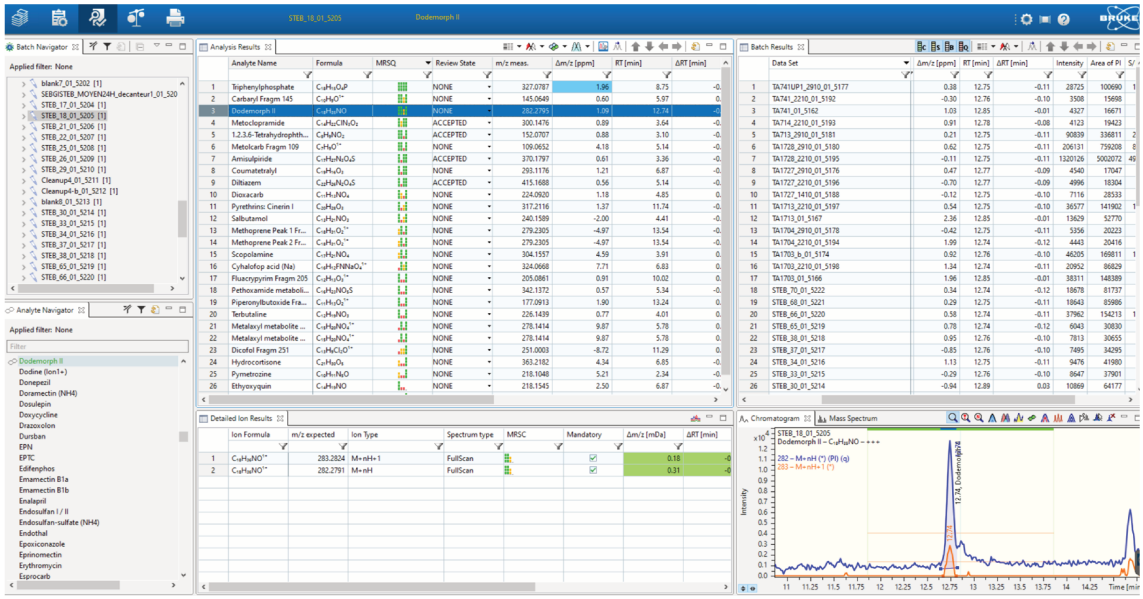


Figure 3.4: TASQ software interface

erence standards of pesticides and pharmaceuticals adapted the retention time and exact mass. Each component contained a chromatogram, mass spectra, retention time, molecular formula, exact ion mass, and fragments of DIA. The TASQ software screens against the molecular formula, the RT, fragment ions, and their ratio of each analyte, giving a score to each mapping compound. The score is named MRSQ and refers to the following 4 screening parameters: M is the exact mass of the precursor ion; R is the retention time; S is the isotopic profile (mSigma); and Q is the qualifier ions/fragment ions. For each parameter, a narrow range and a wide range were defined for scoring. If a detection compound matches the defined narrow range of a parameter, a green bar is labeled, representing a good score; an orange bar represents a value that is fit in a wide range of tolerance; and a red bar refers to the unmatched parameters. In this method, to ensure the identification confidence, the screening parameters are described in Table 3.5. Considering the background noise from the matrix and the trace level of the substance, isotopic fitting (mSigma) < 300 was acceptable. The cutoff values of the screening data were set to $S/N > 10$, and the peak intensity was set to > 3000. All the detection substances with high matching scores were confirmed through visual inspection, and then the suspected candidates were assigned to the identification confidence level as Level 2 (which is described in Section 1.2.3). A TASQ interface is shown in Figure 3.4.

Screening Parameter	narrow range	wide range
accurate mass of precursor ion	± 5 mDa	± 8 mDa
RT	± 0.2 min	± 0.5 min
mSigma	≤ 50	≤ 300
qualifier ions (accurate mass)	± 7 mDa	± 10 mDa
qualifier ion ratio	± 5%	± 10%

Table 3.5: Automated Suspect screening parameter settings by TASQ®

3.4.2 Non-target screening

Non-target screening was performed in MS-DIAL (Version 4.70) (Tsugawa et al., 2015), which is an open-source software used for HRMS data processing (Section 1.4.4). Raw Bruker MS files (.d file) were first converted in a lossless manner to the .abf format using the MS-DIAL-implemented MS convertor before data treatment. The Bruker DIA-MS raw data are profile MS data, and bbCID parameters were input into the MS method type. The minimum peak height was set to 100 cps with a 0.1 mass slice width. The retention time information was excluded from mass spectral library screening. The precursor mass tolerance was within ± 10 mDa and ± 25 mDa for fragment ions. The identification score was > 80%. Peak identification was generated by screening online mass spectral libraries: (1) ESI(+)-MS/MS database (16,481 unique compounds), downloaded from MS-DIAL (http://prime.psc.riken.jp/compms/msdial/main.html#MSP) (Accessed on 20. Oct. 2022); (2) MassBank Europe (MassBank-consortium and its contributors, 2022). The adduct ion includes $[M + H]^+$, $[M + NH_4]^+$, [M + Na], [M + K], etc. After filtering the contamination characterized from the blanks, the MS/MS of the positive matched ion was manually and individually inspected based on their peak shapes and mass spectra. Candidates that could unambiguously match a reference MS/MS were kept and assigned to an identification confidence level as Level 2a. The candidates that were frequently present in the samples with less than 3 overlapping fragment ions were defined as Level 2 or Level 3. After exporting the matched ion list, the candidate ions were analyzed through data analysis. The isotopic profiles of the proposed ions were compared to the theoretical values using the Smart Formula (Bruker Daltonics). If the deviation of mSigma was larger than 50, the proposed structure was flagged and discarded from the list.

3.5 Applications in wastewater

3.5.1 Sample preparation

Water samples were provided mainly by Laboratoires des Pyrénées et des Landes (LPL), including wastewater effluent, rainwater and storage water basins. To minimize the loss of substances during the extraction step, a simple filtration step was performed before direct injection. A standard filter vial (Thomson Instrument Company) with a PTFE membrane (0.2 μm) was selected to filter the water samples. A buffer solution consisting of 2% MeOH in 10 mmol/L ammonium formate with 0.02% formic acid (2-fold concentrated than MPA) was prepared for sample dilution. The samples were diluted 2-fold with a buffer solution before filtration. Moreover, 225 μL of the aqueous phase of the samples was added to 225 μL of the buffer solution. In total, 450 μL of the diluted sample was deposited into a shell vial, and then a plunger with an embedded filter was inserted into the outer shell and slowly and completely depressed to filter. A schematic procedure is shown in Figure 3.5.

Finally, 15 μL of the filtrated samples was directly introduced into the LC system in a partial loop (50 μL Loop).

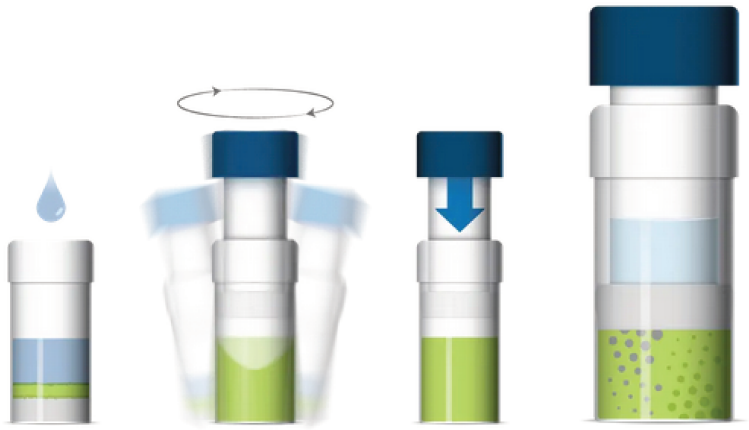


Figure 3.5: Wastewater sample filtration
https://htslabs.com/fv/standard/

3.5.2 Results and discussion

Suspect screening

The suspect screening was performed with an in-house database in TASQ. The tentative identified substances were scored by the TASQ algorithm as described in Section 3.4.1. The first step was to manually inspect the scores and classify the detected chemicals into three categories: "Accepted", "Suspect" and "Rejected".

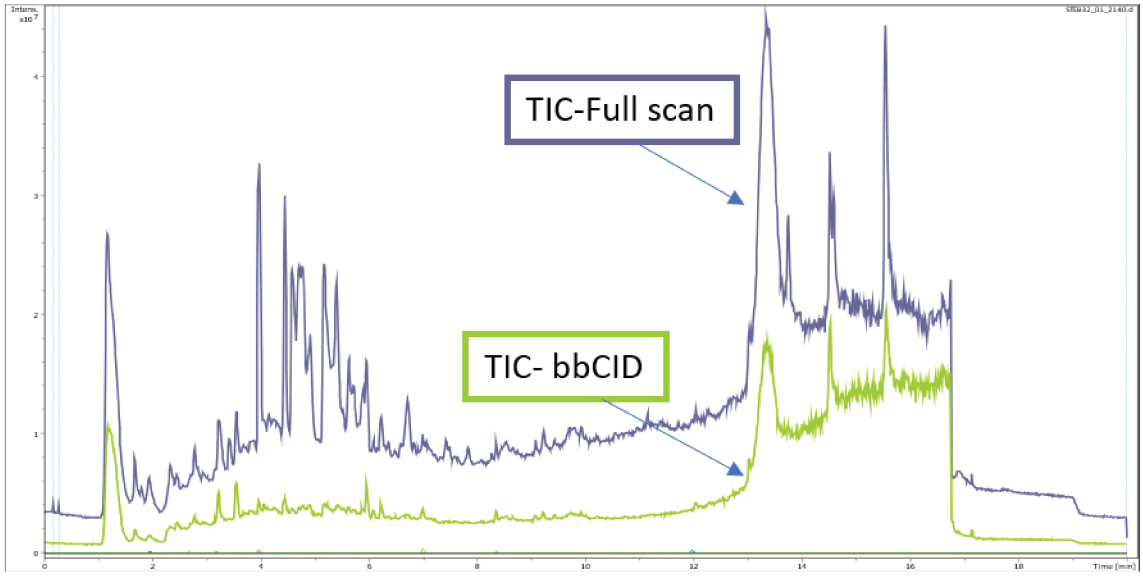


Figure 3.6: An example of TIC

If the mass error and RT of a suggested chemical are larger than the defined wide range, with ± 8 mDa and ± 0.5 min (given red bars), then this suggested structure will be rejected. In contrast, if the exact mass and isotopic profiles fit in the narrow range, qualifier ions and/or RT are within or out of the wide range, it will be labeled as a "Suspect". The complexity of the matrix can result in the deviation of the RT

and fragments. Another possibility can be an isomer or isobar of the candidate. For further confirmation, an EIC is extracted from the calculated exact mass of the candidate, and MS/MS is analyzed and mandated for structural confirmation purposes. If all the parameters were fit in a narrow range of tolerance, it needs to verify whether the candidate is detected in the blank or other samples. If the candidate is detected in the blank and its peak intensity is on a comparable scale as in the samples, the candidate will be rejected.

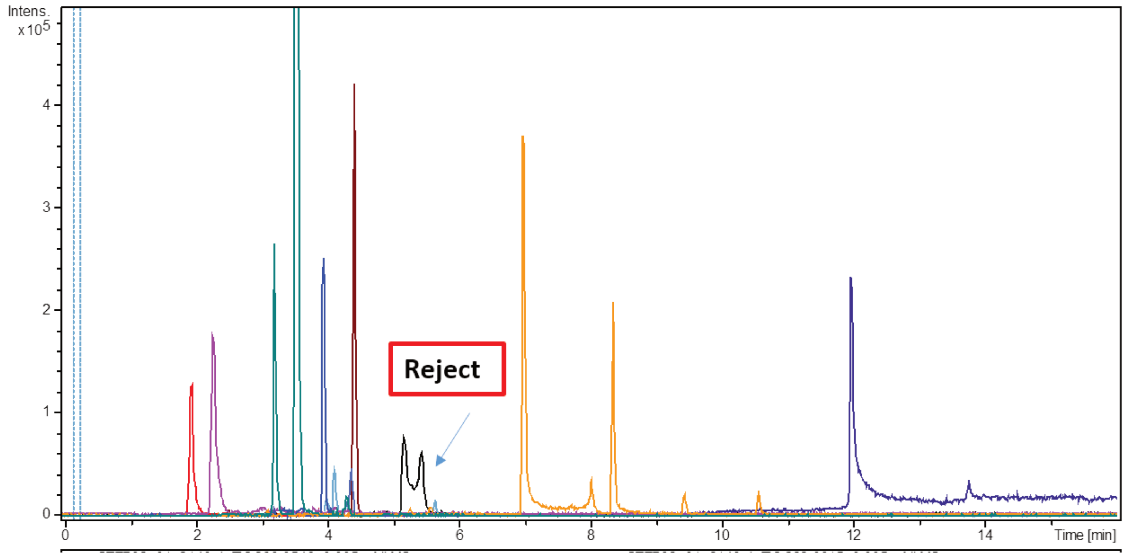
As a result, a candidate list of 22 detected chemicals was verified in the Data Analysis software. By processing EICs in the most representative samples, in which most of the candidates were detected, 18 candidates were confirmed by their peak shape and mass spectral. Figure 3.7 shows an example of the EICs of detected substances. The table below summarizes the common substances in the samples. Finally, the identified substances were searched in the in-house standard list. Aminocarb ($C_{11}H_{16}N_2O_2$) and Quinmerac ($C_{11}H_8ClNO_2$) were confirmed with reference standards, reaching a confidence level at Level 1. The other 13 substances were labeled Level 2, and 3 were in Level 3 due to the trace-level intensities.

Regarding PPCP substances, adenine and adenosine were first detected and quantified in the groundwater of Lyon (Pinasseau et al., 2019). Dumas et al. (2020) detected an increase in adenine and adenosine in male mussels exposed to effluent, resulting in purine and pyrimidine metabolism disruption. Bupivacaine is a medicine for local or regional anesthesia surgery (Burlacu and Buggy, 2008) and is commonly present in hospital wastewater (Le Corre et al., 2012; Javid et al., 2021); therefore, bupivacaine is a compound of interest to evaluate the efficiency of wastewater treatment under the Norman Suspect List Exchange (Network et al., 2020). Bupivacaine was estimated to be found in wastewater with a 67% removal efficiency using MnOx-coated coir fiber (Meza et al., 2020). Selegiline is a medication used in Parkinson treatment, and it was detected in effluent in the Netherlands (Ouyang et al., 2015).

Among 4 pesticides and herbicides, simazine has been prohibited in France since 2003 and was previously detected in two screening trials of 497 French groundwater sites (Lopez et al., 2015), in the groundwater of Lyon (Pinasseau et al., 2019) and in the effluent of wastewater treatment plants in Hérault, France (Dumas et al., 2020). The long-term observation of simazine indicates its persistence and high mobility in water systems. Dodemorph was found in influent and effluent in Spain and Italy with high frequencies and intensities (Rousis et al., 2017) and in the western Mediterranean (Novillo et al., 2017). Quinmerac was quantified in the England River (Taylor et al., 2021) and drinking water (Taylor et al., 2022) using Chemcatcher®. Quinmerac was detected in Germany after measuring a maximum concentration in lake water during heavy rainfall in the fall, which resulted from run-off (Warner et al., 2021). Aminocarb is a polar and basic herbicide that was detected in groundwater in Brazil (da Silva et al., 2021). Aminocarb was found in bivalve samples purchased from a local marketing in France (Diallo et al., 2022).

Cotinine is a well-known metabolite of nicotine and is commonly present in wastewater (Hernández et al., 2011; Andrés-Costa et al., 2017; Verovšek et al., 2022) and other aquatic systems (Branchet et al., 2021). In France, cotinine was frequently detected in groundwater (Lopez et al., 2015; Pinasseau et al., 2019). Pyridafol is classified as a pesticide and transformation product of pyridate. Pyridafol was pre-

viously reported in different foods (Goscinny et al., 2019) and quantified in surface and ground waters in Hungary (Tóth et al., 2022).



Compound	Class	Formula	RT [min]	Ion	m/z [Da]
Adenine	Drug	C5H5N5	2.72	[M+H]+	136.0618
Adenosine	Drug	C10H13N5O4	3.16	[M+H]+	268.0032
Aminocarb	Herbicide	C11H16N2O2	3.86	[M+H]+	209.1285
Chloridazon-desphenyl-methyl	Drug	C5H6ClN3O	4.17	[M+H]+	160.0273
Uracil	PPCP	C4H4N2O2	2.3	[M+H]+	113.0347
Quinmerac	Herbicide	C11H8ClNO2	4.31	[M+H]+	222.0321
Simazine	Herbicide	C7H12ClN5	5.65	[M+H]+	202.086
Dodemorph	Pesticide	C18H35NO	11.97	[M+H]+	282.2796
Cotinine	Metabolite	C10H12N2O	3.97	[M+H]+	177.0999
Pyridafol	Metabolite	C10H7ClN2O	5.11	[M+H]+	207.0311
Selegiline	Drug	C13H17N	4.46	[M+H]+	188.1429

Figure 3.7: Common substances in waste waters

Non-target analysis

Non-target screening was performed in MS-DIAL (Tsugawa et al., 2015). The data processing parameters are explained in Section 3.4.2.

A total of 124,974 features were detected. Following blank filtering and MS/MS matching, only 460 ions were annotated in the alignment results. Then, the matched ions were visually checked by their peak shapes and MS/MS spectra.

Furthermore, MS-FINDER (Tsugawa et al., 2016) was launched for the detected feature in the RT range of 1 to 16 min, with a peak intensity higher than 5000. A total of 71784 features were screened against online experimental and in-silico MS/MS databases. After excluding false positives based on the workflow described in Section 3.4.2 and excluding 4 cross-matched substances in TASQ, 47 candidates with a minimum identification confidence of Level 3 could be listed and exported. The candidates were further confirmed through a Data Analysis (Bruker Daltonics) critical review. One of the detected substances, e.g., *Feature ID 2667*, was taken as an example. As illustrated in Figure 3.8, a representative spectrum from the aligned samples was compared to a reference MS/MS spectrum in the bottom-right

panel, and the EICs of aligned results overlapped in the top-center panel. The shape of the chromatographic peak was satisfactory, and the mass error of the precursor ion was 1.2 mDa (Experimental m/z: 252.1108 Da, reference m/z: 252.1096). The proposed formula was accepted. Three hits were matched in the MS/MS library, and the highest matching hit was *2'-deoxyadenosine* ($C_{10}H_{13}N_5O_3$), which is known as a metabolite and a derivative of adenosine. Furthermore, adenosine was identified in TASQ and MS-DIAL, and a potential correlation existed between the two substances. Meanwhile, MS-FINDER provides detailed information for structure elucidation. Therefore, 2'-deoxyadenosine was identified with an identification confidence level of 2. Following the same process, the alignment results of 47 features were identified, with 26 substances included in Confidence Level 2, and 21 substances in Level 3. The classification of substances was searched in PubChem https://pubchem.ncbi.nlm.nih.gov/ and NORMAN Suspect List Exchange (Network, 2020). An overview of identified candidates in wastewater is found in Appendices. 6.

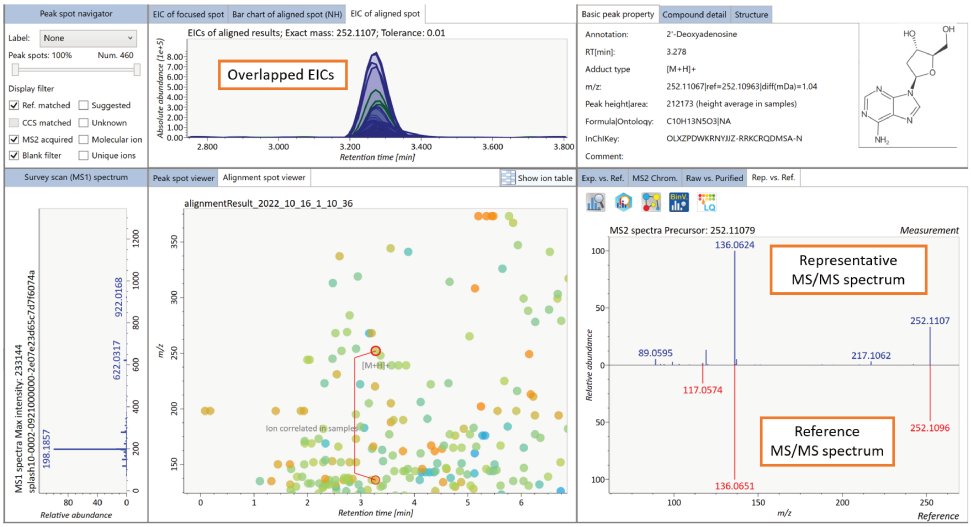


Figure 3.8: Procedure of NTS identification: An example of Feature ID 2667

The class distribution of all the detected substances is summarized in 3.9, and PPCPs, drugs and cosmetics comprised 43.9% of the contaminants. Comparing the sampling sites (Figure 3.10), 90% of the substances were detected at Sampling Site B at a relatively high level. As illustrated in Figure 3.11, the three sampling sites were discriminated from each other based on the contamination class. All 12 pesticides and 5 transformation products (TPs) of pesticides were detected only at Sampling Point B. Deethylatrazine, paired as a TP of atrazine and propazine (Kiefer et al., 2021), was found at Sampling Sites A and B. Simazine and metolachlor have been banned in France since 2003, and deisopropylatrazine is a transformation product of atrazine and simazine (Kiefer et al., 2021). Trace levels were detected in these studies. Other studies also reported withdrawn pesticides and their TPs in French ground water (Lopez et al., 2015; Pinasseau et al., 2019). Currently used pesticides, such as chlortoluron, and TPs of Fenpropimorph showed significant intensities. Chlortoluron has been widely detected in France (Lopez et al., 2015; Pinasseau et al., 2019; Mathon et al., 2022). Most PPCPs and related metabolites or TPs were frequently found at Sampling Site A, while contamination at Sampling Site B was likely associated with agricultural or industrial activities. Diethyl-phathalate and

bis(2-ethylhexyl)phthalate were detected at 3 sites, and bis(2-ethylhexyl)phthalate was found at relatively high levels at Sampling Site C.

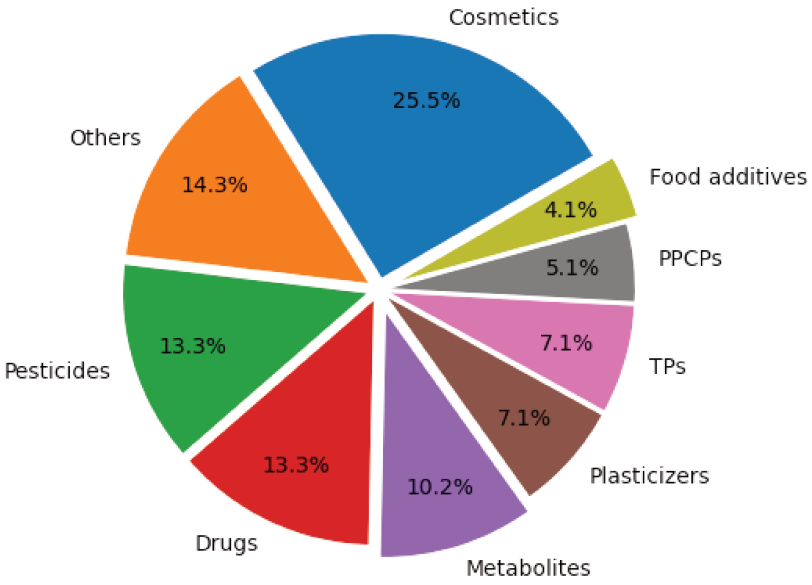


Figure 3.9: Distribution of the detected substance class

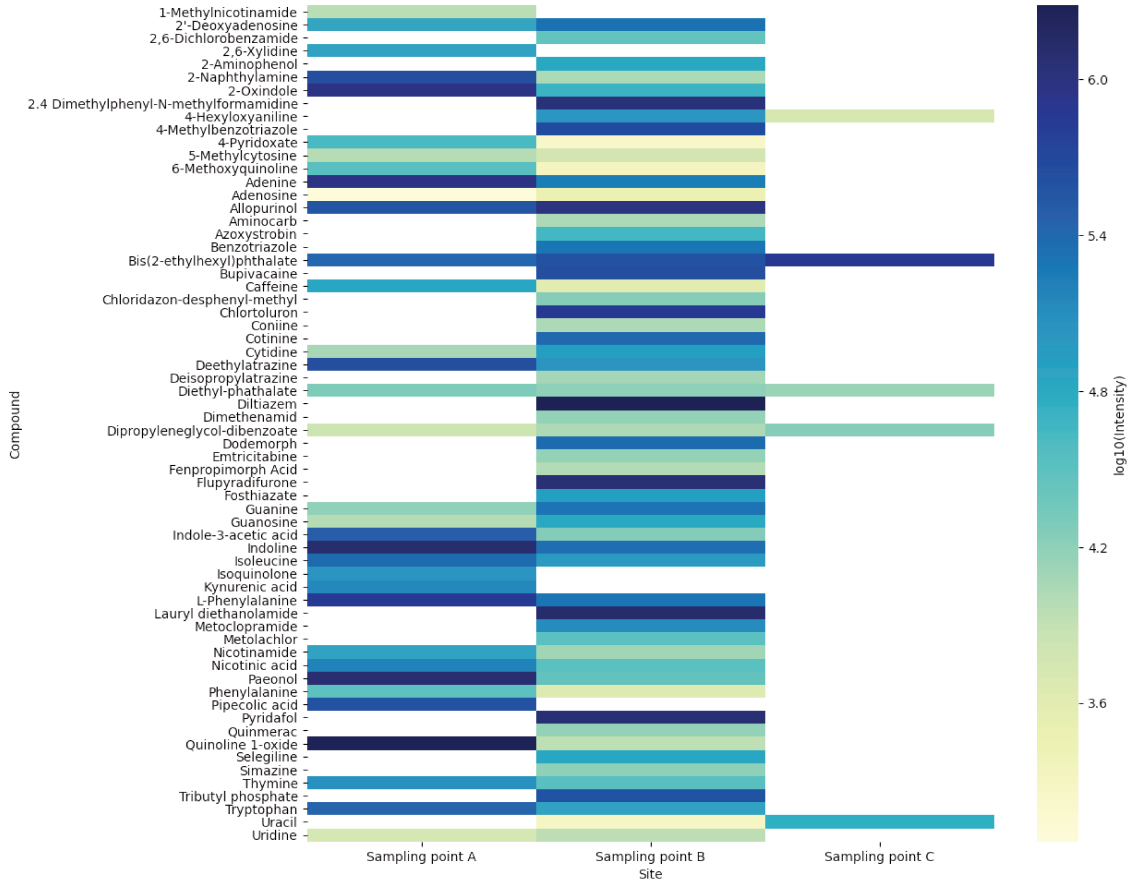


Figure 3.10: Median intensity of detected substances at different sampling points

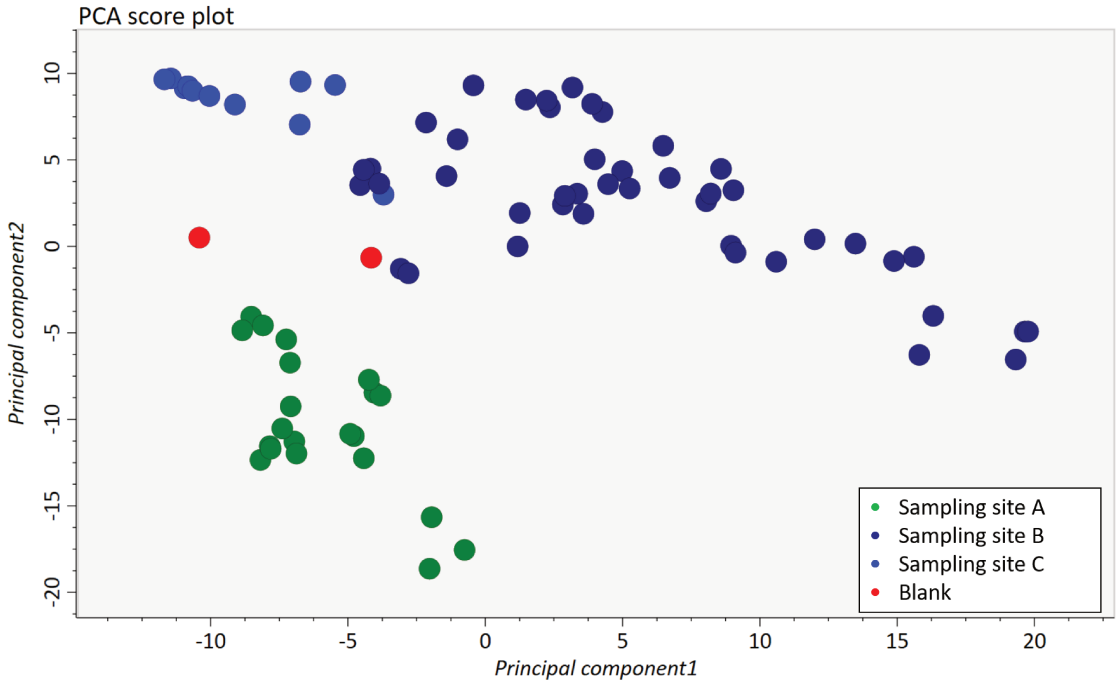


Figure 3.11: Principal component analysis of contaminants and sampling site

3.5.3 Discussion

Overall, the tentative identified substances provide a possible structure that evaluates the potential contaminants in the samples, which can only be confirmed by injecting reference standards. It is always challenging to identify small molecules in a complex matrix. DIA shows its advantages in wide scope analysis in NTS over DDA. However, in DIA, coeluting ions are fragmented simultaneously, and it is difficult to associate fragment ions to the precursor ion. Meanwhile, in this study, samples were only prepared through filtration and dilution, and mass spectra interpretation is highly interfered by the sample matrix, resulting in lower MS/MS spectrum matching scores and rates. In this context, IMS can be introduced to facilitate structural elucidation. MS-DIAL and MS-FINDER are suitable software to perform non-target analysis. Non-target analysis was achieved by screening against MassBank Europe and online databases (e.g., PubChem, HMBD). The proposed candidates were evaluated by the MS/MS matched score. On the other hand, it will be interesting to compare the peak detection results from IMS data to the results obtained here by MS-DIAL. For example, the number of candidates with regards to the peak detections and false positives are shown in these results.

3.6 Preliminary application in natural waters

3.6.1 Sampling area and sample analysis

Six tap water, 50 surface water, 13 high total dissolved solids (TDS) water and 13 seawater samples were collected from different sampling and dumping sites in southern France (Figure 3.12). The sampling was carried out during the period from

February to June 2022. The sampling sites were selected with the aim of covering the diverse types of natural waters, including sites close to anthropic/industrial activities, residential areas, and protected areas. Considering the purity of the sampling points, samples were directly introduced into the LC system with a large volume to avoid cross contamination that result from the manipulation steps. Fifty μL of the diluted samples was directly introduced into the LC system in the full loop.

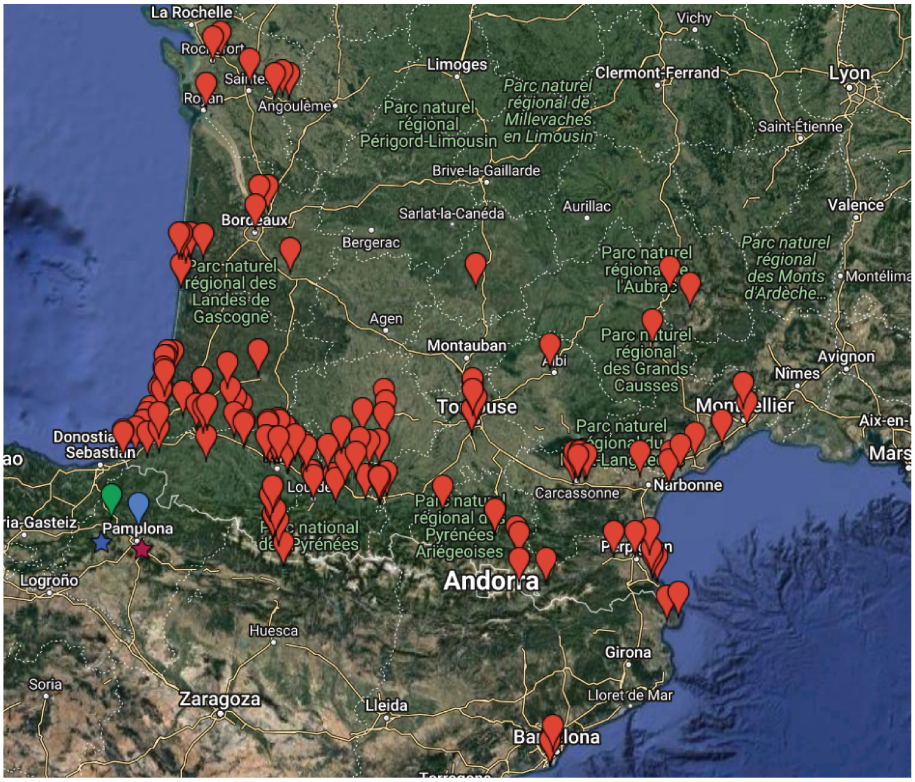


Figure 3.12: Natural water sampling points

All the natural water samples were directly injected into the LC system following different dilution factors, as described in Table 3.6.

Water type	Dilution factor	Numbers of sample
Tap Water	1.5	6
Surface Water	2	2
High TDS water	3	13
Seawater	10	13

Table 3.6: Natural water sample preparation

3.6.2 Data processing

Two data processing approaches were performed for preliminary studies. First, all the raw data were processed by the TASQ (Client 2021) automated workflow as described in Section 3.4.1. The second approach was generated by data analysis (Version 4.60). A representative sample from each water type was selected to process

automated feature detection, and then a simulated formula was calculated for each feature by the "Smart formula. The preprocessing threshold was defined as follows: $S/N > 10$; *Peak intensity* > 1000, *mSigma* < 30 for MS, and < 50 for MS/MS, heteroatoms search with $S_5P_5Cl_{10}Br_{10}F_{12}$. Deconvoluted mass spectra were screened against MassBank and PubChem.

3.6.3 Results and discussion

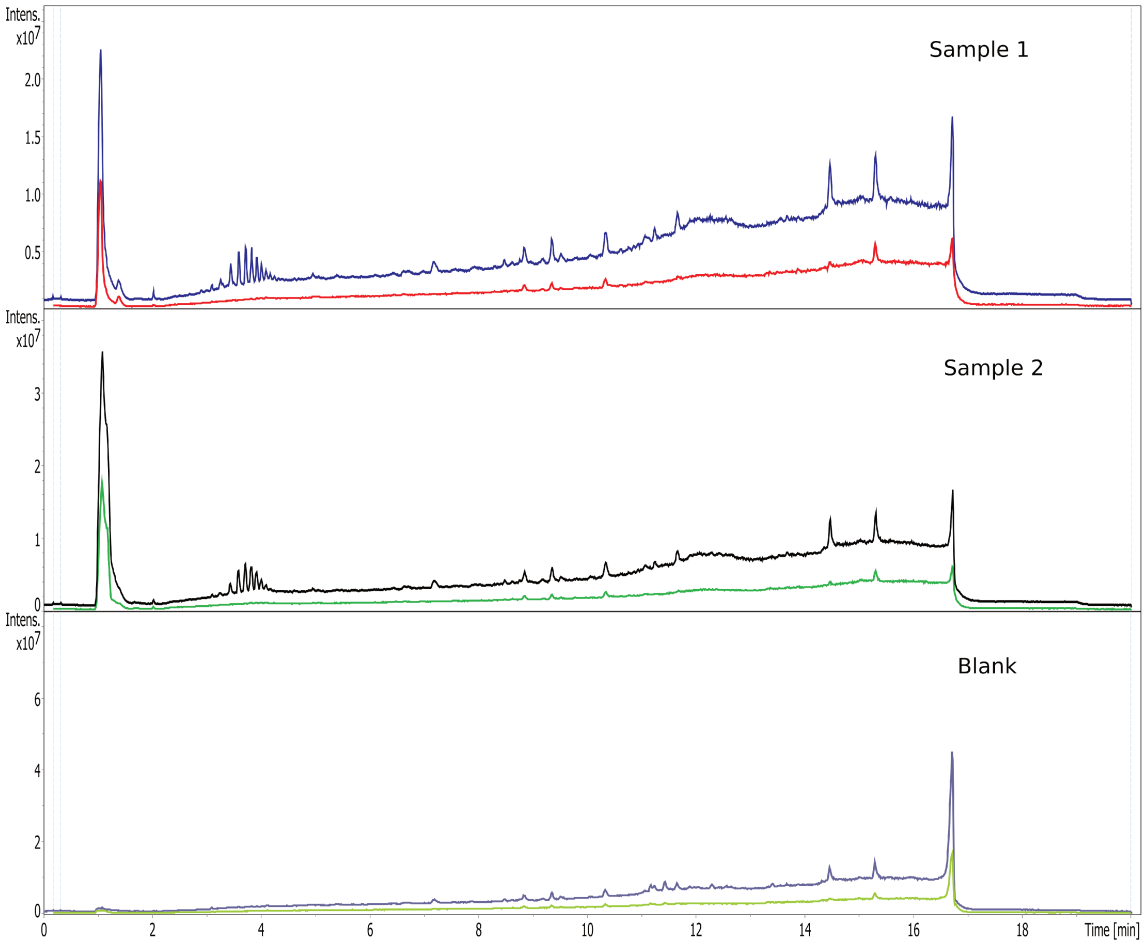


Figure 3.13: Comparison of TICs in samples and blank

As shown in Figure 3.13, a peak eluting within the void time area was observed, indicating that ultra-polar substances were barely trapped in the stationary phase. In this context, other separation methods, such as super fluid chromatography (SPC) or chromatographic columns (HILIC), should be optimized for ultra-polar compounds.

Two pharmaceutical substances, metoprolol and bisoprolol, were identified in TASQ at relatively high levels in high-TDS water samples. Metoprolol and bisoprolol are Î²1-selectivity commonly used in patients with asthma and diabetes (Benfield et al., 1986; Investigators et al., 1999). Metoprolol and bisoprolol were highly present and intractable in European water analysis (Schymanski et al., 2015). In this research, only two sampling points were close to a hospital and showed an outstanding intensity of metoprolol and bisoprolol, indicating insufficient wastewater treatment for these highly mobility compounds (Papaioannou et al., 2020).

With non-target analysis, polyethylene glycols (PEGs) were commonly identified at several sampling points. PEGs have a variety of uses in PPCPs, such as hair conditioners. PEG homologs were tentatively identified in most water samples. As shown in Figure 3.14, long-chain PEGs, such as decaethylene glycol and nonaethylene glycol, were more frequently detected in Water Type A, while short-chain PEGs were commonly found in Water Type B, such as heptaethylene glycol.

After processing with MS-DIAL, 29 peaks were annotated by MassBank Europe. However, all the candidates were poorly matched in the MS/MS spectrum due to the low intensity ($\leq$ 5,0000) and low mass ($\leq$ 150 m/z).

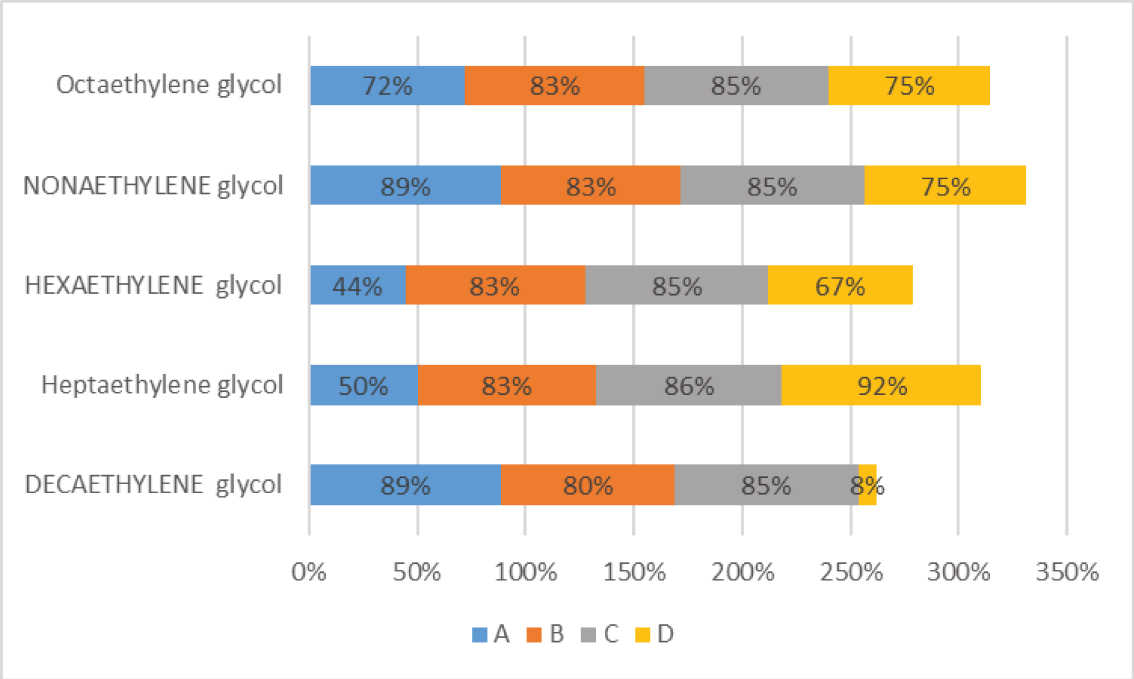


Figure 3.14: Percentage of samples presenting PEGs

3.7 Conclusion and perspectives

This chapter is dedicated to the presentation of the analysis method of various water samples and data processing strategies during this PhD research C_{18}. The column is a generic stationary phase that separates the wide polarity of chemicals; however, extrapolar compounds were barely trapped in it. Alternative analytical techniques can be further optimized to address this gap. Direct injection minimizes the sample preparation step. It is suitable for wastewater analysis and relatively high-intensity substances. However, regarding ultratrace contaminants in tap water and surface water, it is challenging. Large volume or online-SPE can be adapted to the current analysis method to enhance trace contaminant identification and to mitigate the losses of analytes of interest. Furthermore, sampling techniques, such as passive sampling and chemicatchering, can also be used to trap polar compounds in water and for long-term environmental screening. LC-HRMS is a promising analytical technique for non-target analysis. DIA provides nonselective fragmentation, which prevents the loss of information. In contrast, coeluting compounds are sources of

mass spectrum interferences and mis/bad interpretation. In this context, IMS can compensate for this drawback by additional and orthogonal selectivity and associating the precursor ion with its fragment ions by delivering a true pure MSMS spectrum. Meanwhile, it emphasizes the importance of data processing. MS-DIAL is a powerful non-target analysis tool, especially for a series of data. It illustrates straightforward results that provide binary comparison and detection frequency. Manual inspection of the identification candidates is vital to prevent false positives and to verify the identification confidence level. A careful identification process is defined for target and non-target analysis. This highlights the compromises that must be made in non-target analyses. The bottleneck remains the compound identification. The certainty of identified or tentatively identified compounds (Levels 2 and 3) needs further confirmation, such as replicated sample injection.

A preconcentration step can be processed to produce distinguishable MS/MS spectra for identification purposes. IMS will be performed in the next step, and an in-house CCS database will be created to increase the number of identification points. An automated workflow will be modified in TASQ. The Norman exchange suspect list will soon be imported into MS-DIAL. However, it was measured by DWTIM, and the deviation between TIMS and DWTIMS needs to be estimated to optimize the identification criteria. Other experimental and predicted CCS databases, such as CCSbase, are worth testing. In contrast, prediction accuracy depends on the chemical structure and class type; therefore, a careful review and awareness should be carried out systematically. However, negative ionization mode can be beneficial to the current analysis, with a time cost double.

References

V. Albergamo, R. Helmus, and P. de Voogt. Direct injection analysis of polar micropollutants in natural drinking water sources with biphenyl liquid chromatography coupled to high-resolution time-of-flight mass spectrometry. *Journal of Chromatography A*, 1569:53–61, 2018.

M. J. Andrés-Costa, V. Andreu, and Y. Pico. Liquid chromatography–mass spectrometry as a tool for wastewater-based epidemiology: Assessing new psychoactive substances and other human biomarkers. *TrAC Trends in Analytical Chemistry*, 94:21–38, 2017.

W. J. Backe. Suspect and non-target screening of reuse water by large-volume injection liquid chromatography and quadrupole time-of-flight mass spectrometry. *Chemosphere*, 266:128961, 2021.

P. Benfield, S. P. Clissold, and R. N. Brogden. Metoprolol. *Drugs*, 31(5):376–429, 1986.

P. Branchet, L. Arpin-Pont, A. Piram, P. Boissery, P. Wong-Wah-Chung, and P. Doumenq. Pharmaceuticals in the marine environment: What are the present challenges in their monitoring? *Science of the Total Environment*, 766:142644, 2021.

C. L. Burlacu and D. J. Buggy. Update on local anesthetics: focus on levobupivacaine. *Therapeutics and clinical risk management*, 4(2):381, 2008.

J. J. da Silva, B. F. da Silva, N. R. Stradiotto, M. Petrović, M. Gros, and P. Gago-Ferrero. Identification of organic contaminants in vinasse and in soil and groundwater from fertigated sugarcane crop areas using target and suspect screening strategies. *Science of The Total Environment*, 761:143237, 2021.

T. Diallo, Y. Makni, A. Lerebours, H. Thomas, T. Guérin, and J. Parinet. Development and validation according to the sante guidelines of a quechers-uhplc-qtof-ms method for the screening of 204 pesticides in bivalves. *Food Chemistry*, 386: 132871, 2022.

T. Dumas, B. Bonnefille, E. Gomez, J. Boccard, N. A. Castro, H. Fenet, and F. Courant. Metabolomics approach reveals disruption of metabolic pathways in the marine bivalve mytilus galloprovincialis exposed to a wwtp effluent extract. *Science of the Total Environment*, 712:136551, 2020.

W. B. Dunn, A. Erban, R. J. Weber, D. J. Creek, M. Brown, R. Breitling, T. Hankemeier, R. Goodacre, S. Neumann, J. Kopka, et al. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, 9(1):44–66, 2013.

P. Gago-Ferrero, A. A. Bletsou, D. E. Damalas, R. Aalizadeh, N. A. Alygizakis, H. P. Singer, J. Hollender, and N. S. Thomaidis. Wide-scope target screening of> 2000 emerging contaminants in wastewater samples with uplc-q-tof-hrms/ms and smart evaluation of its performance through the validation of 195 selected representative analytes. *Journal of hazardous materials*, 387:121712, 2020.

B. González-Gaya, N. Lopez-Herguedas, D. Bilbao, L. Mijangos, A. Iker, N. Etxebarria, M. Irazola, A. Prieto, M. Olivares, and O. Zuloaga. Suspect and non-target screening: the last frontier in environmental analysis. *Analytical Methods*, 13(16): 1876–1904, 2021.

S. Goscinny, M. McCullagh, J. Far, E. De Pauw, and G. Eppe. Towards the use of ion mobility mass spectrometry derived collision cross section as a screening approach for unambiguous identification of targeted pesticides in food. *Rapid Communications in Mass Spectrometry*, 33:34–48, 2019.

F. Hernández, L. Bijlsma, J. V. Sancho, R. Díaz, and M. Ibáñez. Rapid wide-scope screening of drugs of abuse, prescription drugs with potential for abuse and their metabolites in influent and effluent urban wastewater by ultrahigh pressure liquid chromatography–quadrupole-time-of-flight-mass spectrometry. *Analytica chimica acta*, 684(1-2):96–106, 2011.

J. Hollender, B. Van Bavel, V. Dulio, E. Farmen, K. Furtmann, J. Koschorreck, U. Kunkel, M. Krauss, J. Munthe, M. Schlabach, et al. High resolution mass spectrometry-based non-target screening can support regulatory environmental monitoring and chemicals management. *Environmental Sciences Europe*, 31(1): 1–11, 2019.

C.-I. Investigators et al. The cardiac insufficiency bisoprolol study ii (cibis-ii): a randomised trial. *The Lancet*, 353(9146):9–13, 1999.

F. Javid, T. N. Ang, S. Hanning, D. Svirskis, R. Burrell, M. Taylor, L. J. Wright, and S. Baroutian. Hydrothermal deconstruction of local anesthetics (bupivacaine and lignocaine) in pharmaceutical waste. *Journal of Environmental Chemical Engineering*, 9(5):106273, 2021.

K. Kiefer, L. Du, H. Singer, and J. Hollender. Identification of lc-hrms nontarget signals in groundwater after source related prioritization. *Water Research*, 196: 116994, 2021.

T. Köppe, K. S. Jewell, C. Dietrich, A. Wick, and T. A. Ternes. Application of a non-target workflow for the identification of specific contaminants using the example of the nidda river basin. *Water Research*, 178:115703, 2020.

K. S. Le Corre, C. Ort, D. Kateley, B. Allen, B. I. Escher, and J. Keller. Consumption-based approach for assessing the contribution of hospitals towards the load of pharmaceutical residues in municipal wastewater. *Environment international*, 45:99–111, 2012.

Z. Li, E. Undeman, E. Papa, and M. S. McLachlan. High-throughput evaluation of organic contaminant removal efficiency in a wastewater treatment plant using direct injection uhplc-orbitrap-ms/ms. *Environmental Science: Processes & Impacts*, 20(3):561–571, 2018.

B. Lopez, P. Ollivier, A. Togola, N. Baran, and J.-P. Ghestem. Screening of french groundwater for regulated and emerging contaminants. *Science of the Total Environment*, 518:562–573, 2015.

MassBank-consortium and its contributors. Massbank/massbank-data: Release version 2022.06, Oct. 2022. URL https://doi.org/10.5281/zenodo.7148841.

B. Mathon, M. Ferreol, A. Togola, S. Lardy-Fontan, A. Dabrin, I. Allan, P.-F. Staub, N. Mazzella, and C. Miège. Polar organic chemical integrative samplers as an effective tool for chemical monitoring of surface waters–results from one-year monitoring in france. *Science of The Total Environment*, 824:153549, 2022.

F. Menger, P. Gago-Ferrero, K. Wiberg, and L. Ahrens. Wide-scope screening of polar contaminants of concern in water: A critical review of liquid chromatography-high resolution mass spectrometry-based strategies. *Trends in Environmental Analytical Chemistry*, 28:e00102, 2020.

L. C. Meza, P. Piotrowski, J. Farnan, T. L. Tasker, B. Xiong, B. Weggler, K. Murrell, F. L. Dorman, J. P. V. Heuvel, and W. D. Burgos. Detection and removal of biologically active organic micropollutants from hospital wastewater. *Science of the Total Environment*, 700:134469, 2020.

N. Network. Norman suspect list exchange, 2020.

N. Network, R. Aalizadeh, N. Alygizakis, E. Schymanski, J. Slobodnik, and S. Fischer. S0 — susdat — merged norman suspect list: Susdat, Mar. 2020. URL https://doi.org/10.5281/zenodo.3695732. This is a massive collaborative effort with many contributions! This dataset merges other entries in this collection.

O. Novillo, J. Pertusa, and J. Tomás. Exploring the presence of pollutants at sea: monitoring heavy metals and pesticides in loggerhead turtles (caretta caretta) from the western mediterranean. *Science of the Total Environment*, 598:1130–1139, 2017.

G. Nürenberg, U. Kunkel, A. Wick, P. Falås, A. Joss, and T. A. Ternes. Nontarget analysis: A new tool for the evaluation of wastewater processes. *Water research*, 163:114842, 2019.

X. Ouyang, P. Leonards, J. Legler, R. van der Oost, J. de Boer, and M. Lamoree. 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, 2015.

D. Papaioannou, P. Koukoulakis, M. Papageorgiou, D. Lambropoulou, and I. Kalavrouziotis. Investigation of pharmaceutical and personal care product interactions of soil and beets (beta vulgaris l.) under the effect of wastewater reuse. *Chemosphere*, 238:124553, 2020.

M. Paszkiewicz, K. Godlewska, H. Lis, M. Caban, A. Białk-Bielińska, and P. Stepnowski. Advances in suspect screening and non-target analysis of polar emerging contaminants in the environmental monitoring. *TrAC Trends in Analytical Chemistry*, page 116671, 2022.

L. Pinasseau, L. Wiest, A. Fildier, L. Volatier, G. R. Fones, G. A. Mills, F. Mermillod-Blondin, and E. Vulliet. Use of passive sampling and high resolution mass spectrometry using a suspect screening approach to characterise emerging

pollutants in contaminated groundwater and runoff. *Science of the Total Environment*, 672:253–263, 2019.

T. Reemtsma, U. Berger, H. P. H. Arp, H. Gallard, T. P. Knepper, M. Neumann, J. B. Quintana, and P. d. Voogt. Mind the gap: Persistent and mobile organic compounds—water contaminants that slip through, 2016.

J. Riedl, S. Esslinger, and C. Fauhl-Hassek. Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Analytica Chimica Acta*, 885:17–32, 2015.

N. I. Rousis, R. Bade, L. Bijlsma, E. Zuccato, J. V. Sancho, F. Hernandez, and S. Castiglioni. Monitoring a large number of pesticides and transformation products in water samples from spain and italy. *Environmental research*, 156:31–38, 2017.

B. Schulze, Y. Jeon, S. Kaserzon, A. L. Heffernan, P. Dewapriya, J. O'Brien, M. J. G. Ramos, S. G. Gorji, J. F. Mueller, K. V. Thomas, et al. 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, 2020.

E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H. P. Singer, and J. Hollender. Identifying small molecules via high resolution mass spectrometry: communicating confidence, 2014.

E. L. Schymanski, H. P. Singer, J. Slobodnik, I. M. Ipolyi, P. Oswald, M. Krauss, T. Schulze, P. Haglund, T. Letzel, S. Grosse, et al. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Analytical and bioanalytical chemistry*, 407(21):6237–6255, 2015.

A. E. Steuer, L. Brockbals, and T. Kraemer. Untargeted metabolomics approaches to improve casework in clinical and forensic toxicology—"where are we standing and where are we heading?". *Wiley Interdisciplinary Reviews: Forensic Science*, 4(4):e1449, 2022.

A. C. Taylor, G. A. Mills, A. Gravell, M. Kerwick, and G. R. Fones. Passive sampling with suspect screening of polar pesticides and multivariate analysis in river catchments: Informing environmental risk assessments and designing future monitoring programmes. *Science of The Total Environment*, 787:147519, 2021.

A. C. Taylor, G. A. Mills, A. Gravell, M. Kerwick, and G. R. Fones. Pesticide fate during drinking water treatment determined through passive sampling combined with suspect screening and multivariate statistical analysis. *Water Research*, 222: 118865, 2022.

E. Tóth, Á. Tölgyesi, A. Simon, M. Bálint, X. Ma, and V. K. Sharma. An alternative strategy for screening and confirmation of 330 pesticides in ground-and surface water using liquid chromatography tandem mass spectrometry. *Molecules*, 27(6): 1872, 2022.

H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, and M. Arita. Ms-dial: data-independent ms/ms deconvolution for comprehensive metabolome analysis. *Nature methods*, 12(6):523–526, 2015.

H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, and M. Arita. Hydrogen rearrangement rules: computational ms/ms fragmentation and structure elucidation using ms-finder software. *Analytical chemistry*, 88(16):7946–7958, 2016.

T. Verovšek, D. Heath, and E. Heath. Occurrence, fate and determination of tobacco (nicotine) and alcohol (ethanol) residues in waste-and environmental waters. *Trends in Environmental Analytical Chemistry*, page e00164, 2022.

W. Warner, S. Zeman-Kuhnert, C. Heim, S. Nachtigall, and T. Licha. Seasonal and spatial dynamics of selected pesticides and nutrients in a small lake catchment–implications for agile monitoring strategies. *Chemosphere*, 281:130736, 2021.

Y. Yang, Y. S. Ok, K.-H. Kim, E. E. Kwon, and Y. F. Tsang. Occurrences and removal of pharmaceuticals and personal care products (ppcps) in drinking water and water/sewage treatment plants: A review. *Science of the Total Environment*, 596:303–320, 2017.

Chapter 4

Method development of steroids hormones analysis in juvenile fish

Summary

In Chapter 4, a simultaneous analysis of steroids hormones is developed using UPLC-timsTOF. The selected hormones are the key factor in sexual maturity of many species, including fish, etc.. Two approaches of analytical techniques are comparable: UPLC coupling with QTOF, and UPLC coupling with IMS-QTOF (timsTOF). The benefits of introducing IMS for complex matrix analysis are discussed.

The first part of this chapter introduces the objective of steroids studies and the advance of current steroids studies. Then the steroid hormones chosen for this study and their presence in the environment are introduced.

In the experiment, the method development strategies are discussed, including the choice of the LC column, the optimization of mobile phases, and parameter settings in the MS detector. In addition, the key notions of the method validation are introduced.

Finally, the benefits and disadvantages of using IMS compared to conventional qTOF will be discussed. The validation of the simultaneous analysis and quantification method will be evaluated.

4.1 Introduction

Steroid hormones are natural hormones that respond to a wide range of physiological processes, such as sexual maturity, stress responses, and reproduction (Ojoghoro et al., 2021). Cortisol is a glucocorticoid hormone produced in many animals, mainly by the zona fasciculata of the adrenal cortex in the adrenal gland. Cortisol is the main product of the stress response in fish (Schreck and Tort, 2016). Therefore, it is worth quantifying cortisol levels in fish after exposure to various stress resources, including light and temperature. Cortisol production can also be used to assess the effect of sexual maturity. Considering the small size of juvenile fish, preventing the loss of samples should be considered.

Gas chromatography (GC) and liquid chromatography (LC) coupled with mass spectrometry (MS) are widely used for the quantification of cortisol and other steroid hormones in various biological matrices. Most of the applications are performed by triple-quadrupole mass spectrometry, using two or three characteristic transitions, such as Cortisol ion transitions: 363 → 121, 363 → 327 in positive mode (Jia et al., 2016; Domenech-Coca et al., 2019) and 363 → 331 in negative mode (Gaudl et al., 2019). With the recent development of high-resolution mass spectrometry (HRMS), simultaneous analysis in a full scan mode that ensures both selectivity and retrospective analysis has gained interest. Thus, simultaneous quantification analysis can be achieved in a single approach. The hybrid of HRMS with ion mobility (IMS) is a novel analytical technique. By introducing IMS, the collision cross section (CCS) adds more criteria other than the retention time (RT), accurate mass, and isotope pattern. Moreover, IMS can "eliminate" noise peaks from the matrix effect, which offers a compromised approach to analyze juvenile fish after a minimum sample preparation step.

The objective of our study was to develop a sensitive and specific LC-timsTOF-MS method for absolute and accurate cortisol quantification and other steroid hormone characterization in the water or in juvenile fish.

4.1.1 Target hormones

Corticosteroids are one of the most important families of steroid hormones, and both natural and synthetic corticosteroids are commonly used in human and animal therapeutic applications, for example, inflammatory and autoimmune diseases (Wu et al., 2019). Corticosteroids can be found at trace levels (ng/L) or even sub-ng/L in aquatic samples around the world (Wu et al., 2019; Ojoghoro et al., 2021). Cortisol and cortisone are two main glucocorticoids and are mostly found in biological and environmental samples (Zhong et al., 2021). LC-MS/MS is the most used technique for steroid hormones analysis due to its high sensitivity and high selectivity (Song and Feng, 2021). Weizel et al. (2018) developed a simultaneous analytical method for wastewater and surface water analysis in Germany. The level of cortisol detected in wastewater was between 0.9 to 2.8 ng/L and 0.2 to 1.3 ng/L in surface water. Cortisone had a lower concentration level than cortisol in the same water sample, which was below 1 ng/L. Cortisol was also reported in Chinese surface water at 1.2 to 11 ng/L and much higher concentrations of cortisone (62 to 628 ng/L) due to its wide use in medicines (Shen et al., 2020).

Cortisol

Cortisol (Table 4.1) is a glucocorticoid secreted by the adrenal cortex, which is a reliable indicator of stress and is considered the major stress hormone (Sadoul and Geffroy, 2019). IIn fishes, changes in the aquaculture environment can affect the production of cortisol and the sexual maturity of the species. The level of cortisol is an important indicator that characterizes coping abilities of individual fished and evaluates how the environment can impact it (Sadoul and Geffroy, 2019). Cortisol in fish plasma can be quickly released in the environment; therefore, the analysis of cortisol in natural water samples is a promising way to assess the response of stress.

Cortisone

Cortisone (Table 4.1) is conversed by cortisol through peripheral metabolism. In contrast to cortisol, cortisone is inactive, is found particularly in the kidneys, and can converted back to cortisol through enzymatic reaction in the livers. The fast equilibrium between cortisol and cortisone can regulate glucocorticoid activity in various tissues (Nomura et al., 1996). Both natural and synthesis have a long history with regards to inflammation treatments. Cortisone is also reported in surface water in ng/L-level (Ojoghoro et al., 2021).

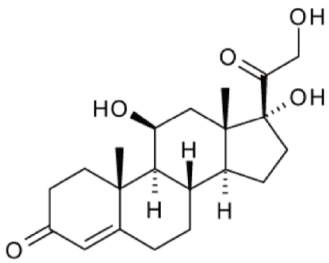
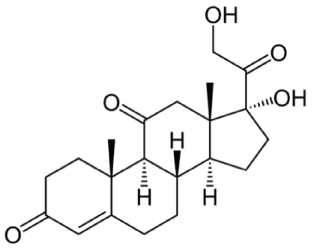
Compound	Cortisol	Cortisone
Formula	C21H30O5	C21H28O5
Exact mass	362.2093	360.1937
Structure		

Table 4.1: Overview of cortisol and cortisone

4.2 Method development strategies

4.2.1 Separation Column

The selection of the LC column includes the packing material that constructs the stationary phase, the size of the particles, and the dimensions of the column. Since separation occurs in the column, column optimization is a key factor in method development. RP C18 columns are the most frequently used for steroid hormones

(Song and Feng, 2021). In this research, Acquity BEH C18 (50 mm × 2.1 mm, 1.7 μm, Waters), Hypersil gold PFP (50 mm × 2.1 mm, 1.9 μm, Thermo Fisher Scientific), Fortis C18 (50 mm x 2.1 mm, 1.7 μm) and ACE C18-PFP (50 mm × 2.1 mm, 1.7 μm) Four columns were tested with the ACN and MeOH during the method optimization. The reported performance of the columns is based on the chromatographic resolution, peak shape, S/N and peak area.

4.2.2 The choice of solvents and additives

As chromatographic separation is performed in an RPLC column, MeOH and ACN are two major organic solvents used in RPLC. The choice of solvent is based on the 1) separation and peak shape of each analyte; 2) ionization behaviors; and 3) pH of the solvent. According to bibliographic research and experiments, MeOH significantly improved analyte ionization compared with the ACN; thus, MeOH was chosen as the mobile phase. Ammonium fluoride, ammonium acetate, and ammonium were added to the mobile phase to adjust the pH and improve ionization. Ammonium fluoride can significantly promote steroid-like molecule ionization (Gaudl et al., 2016). Therefore, 1 mmol/L ammonium fluoride was added to Mobile Phase A to enhance the ionization of cortisol

4.2.3 MS detector

MS/MS has high selectivity and sensitivity; however, it is not possible for retrospective analysis. If other compounds are added to the wish-list or even nontarget analysis, a reanalysis must be performed. Herein, we decided to develop a full scan method. In contrast, the sensitivity and selectivity will be lost by using a full scan, and IMS is introduced in the method to improve selectivity. However, the power of IMS is still being researched, and a comparison of IMS was studied. The TIMS analyzer was placed after proceeding with the ionization source and before the MS detector. The separation in IMS will not impact the separation and ionization; thus, the same separation program was applied for both approaches. The MS setting was slightly adapted for the "tims ON" method. As results, Acquity BEH C18 (50 mm × 2.1 mm, 1.7 μm, Waters) produced a higher intensity of cortisol and cortisone with MeOH, especially at lower concentrations (Figure 4.1).However, the Fortis C18 column produced a higher S/N, but at 2 ng/mL (2 ppb), the S/N ratios were comparable in BEH C18 and Fortis C18 (Figure 4.2). As a compromise, Acquity BEH C18 (50 mm × 2.1 mm, 1.7 μm, Waters) was selected for this method.

4.3 Method validation

To describe and define the limits and performances of an analytical method, it is essential to validate its standardization according to the application guidelines. The validation of the method requires the determination of several characteristics that we are going to define: the accuracy and precision and the limit of detection and linearity.

Figure 4.1: Column optimization: Intensity comparison

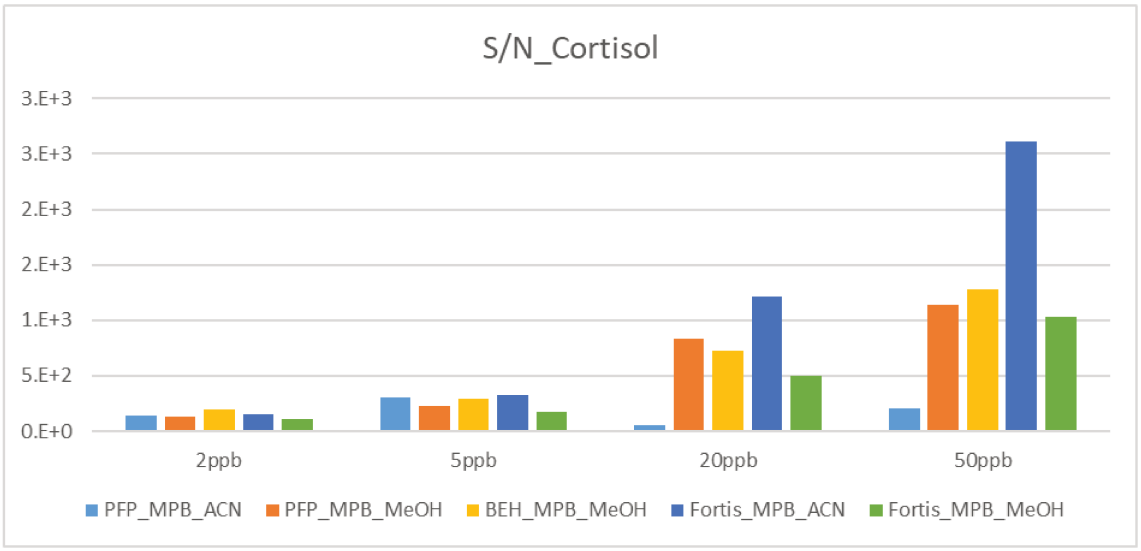


Figure 4.2: Column optimization: S/N ratio comparison

4.3.1 Internal standard

Quantification by using specific isotopic dilution (labeled surrogates) is the most accurate and precise technique for quantification analysis. Isotopically labeled standards (e.g, with ^{15}N, deuterium (D), ^{13}C) are spiked into the samples and calibrators, which are normally the same or isomer/close analog molecule as the target analyte. It introduces a known compound with a defined concentration into samples as an internal standard. The concentration and calibration curve were estimated by the response ratio between the targeted analyte and the reference internal standard, which compensated for the variation in the exact volume or quantity of samples introduced in the LC–MS system.

4.3.2 Precision and accuracy

The precision is the spread of individual measurement values from each other, while the accuracy evaluates the closeness of the measurement values to the true value. The precision can be assessed by injecting a standard solution with defined solutions multiple times. This can be calculated as the coefficient of variation (CV). The accuracy can also be defined as bias between the calculated concentration and exact concentration. The concentration of analytical standard solutions needs to be accurate to avoid quantification error. Therefore, a test solution or quality control (QC) solution must be prepared. The test solution was made by using different working solutions as standard solutions, whose concentrations should be in the calibration range. Normally, three test solutions with low, median and high concentrations are prepared and quantified through the calibration curve. The accuracy of the standard solution was then estimated by the bias between the calculated concentration using the calibration curve and the defined concentration.

4.3.3 Linearity

A calibration curve is commonly a linear regression between the concentration and response of the analyte quantity, common area or intensity of the peak. This linear regression relationship is built to estimate the unknown concentrations of the analyte in a sample. Thus, the accuracy of the predicted concentration is highly dependent on the linearity of the calibration curve. The linearity includes the weighting, regression model, and coefficient.

4.3.4 Limit of Detection (LOD)

The limit of detection (LOD) estimates the lowest analyte concentration that can be reliably distinguished from "analytical noise" in a blank sample (Armbruster and Pry, 2008). There are different approaches to calculate the LOD, and the signal-to-noise ratio (S/N) is commonly accepted to determine the LOD in pharmaceutical analysis by the European Pharmacopoeia (Desimoni and Brunetti, 2015). The S/N defines the ratio of an analytical signal to the mean background noise. The idea of using the S/N to define LOD is that the analyte responds with a sufficiently large intensity to be discriminated from the blank signal (sample without analyte). The LOD is estimated to be 3 times the S/N in the blank sample, indicating the probable presence of the analyte in the sample. The S/N is often evaluated manually;

therefore, it can sometimes be subjective. Modern analytical instruments acquire more signal values to calculate the S/N ratio.

4.4 Chemicals and standards

The steroid standards, internal standards and quality control kits were purchased from AbsoluteIDQ® BIOCRATES. The standards kit consists of 17 steroids, listed in Table 6.2.

Ammonium fluoride (NH_4F) was obtained from Fisher Scientific. All the solvents and chemicals used were in LC-MS grade. Methanol absolute (MeOH), acetonitrile (ACN), and formic acid, which are all of UHPLC–MS grade, were purchased from Biosolve Chemicals. Milli-Q® Type 1 Ultrapure Water Systems (Millipore, Bedford, MA) deionized water was used throughout.

4.4.1 Preparation of solutions

A work solution (50 mL) of 200 mmol/L ammonium fluoride (AMF, NH_4F) was prepared by weighting 370.37 mg of AMF powder. The work solution was stored at 4 °C for 1 month. The Mobile Phase A consists of 1 mmol/L AMF in water, which was fresh prepared through 200-fold dilution. The Mobile Phase B was MeOH (LC-MS grade).

The purchased Steroid Hormone kit contains 7 calibrators, named Kal1 - Kal7 (Overview in Table 6.2 and Table 6.3). To fit for our work purpose, Kal7 (Cortisol concentration at 1000 ng/mL), and Internal standard kit (d4-Cortisol concentration at 1000 ng/mL) (Overview in Table 6.3) were reconstituted with 1200 purified water (Milli-Q). Six calibrators and three QC solutions were prepared as described in Figure 4.3:

Cortisol	C[ng/mL]	Vtake[uL]	STDtake	Kal[ng/mL]	Vadd[uL]	ISTD[uL]	Vfinal[uL]
STD1	2	400	STD2	5	590	10	1000
STD2	5	250	STD3	20	740	10	1000
STD3	20	400	STD5	50	590	10	1000
STD4	33	33	Kal7	1000	957	10	1000
STD5	50	50	Kal7	1000	940	10	1000
STD6	100	1000	kal7	1000	890	10	1000
QC1	4	100	QC	40	890	10	1000
QC2	10	250	QC	40	740	10	1000
QC3	25	625	QC	40	365	10	1000

ISTD = 10 ng/mL in Calibrators

Figure 4.3: Steroid Hormone Calibration Procedure

An overview of calibration range for each steroid hormone was listed in Table 4.2.

Table 4.2: Standard list

Compound	$[M+H]^+$ (m/z)	Calibration range (ng/mL)	RT (min)
Cortisol	363.2166	2-100	2.24
Cortisone	361.201	0.2-10	2.13
Aldosterone	361.201	0.1-5	2.02
Corticosterone	347.2217	0.06-3	2.52
17α-OHP	331.2268	0.1-5	3.13
DOC	331.2268	0.03-1.5	2.89
Androstenedione	287.2006	0.064-3.2	2.77
Testosterone	289.2162	0.02-1	3.02

4.5 Method

4.5.1 LC

The chromatographic separation was performed with an UPLC Acquity system (Waters, USA). The system was operated with the MassLynx V4.1 software (Waters). An ACQUITY UPLC BEH C18 Column 1.7 μm, 2.1 mm X 50 mm with an Acuity UPLC BEH C18 VanGuard pre-column (1.7 μm, 2.1 mm X 5 mm, Waters, USA) was used for the separation in a binary gradient. The mobile phases consist of 1 mmol/L NH_4F in purified water (Milli-Q) and MeOH. The total LC acquisition time was 7 min, the gradient was shown in Table 4.3. The column temperature was set to 35 °C and the auto-sampler to 5 °C. 10 μL per sample was systematically injected in the partial loop mode.

Table 4.3: Steroid analysis LC gradient

Time(min)	Flow rate (mL/min)	% MPB	Slop
0	0.3	25	
0.25	0.3	25	6
2	0.3	65	5
4.5	0.35	95	7
5.5	0.4	95	6
5.75	0.4	25	6
7	0.3	25	6

4.5.2 MS

A trapped ion mobility (drift) cell coupled with high-resolution time-of-flight mass spectrometry (timsTOF, Bruker Daltonics, Bremen, Germany) was used in the positive ESI mode for most hormones, including cortisol, a few of them could be done only in negative mode with the same principle and approach, for acquisition in the MS mode and MSMS mode. The MS mode was set to the full scan mode between 100 and 1250 m/z with a scan speed of 2 Hz. The capillary was set to 4600 V, the nebulizer was set to 2.8 bar, a dry gas flow was set at to 6 L/min, and the dry temperature was set at to 230 °C. The collision energy of MS was applied to 13 eV,

and the collision energy in the bbCID mode was set to 35 eV. The MS analyzer was operated in the TIMS on and off mode to compare the spectral quality enhancement by IMS.

4.6 Data processing

Raw data were analyzed by Data Analysis Version 4.60 (Bruker Daltonics GmbH). The samples were acquired in the full-scan mode. The exact mass of the protonated ion [M+H]+ was used to generate an extracted ion chromatogram (EIC) with a mass width of 5 mDa. For IM-MS data, the extracted ion mobilogram (EIM) was obtained by the exact mass and retention time range of the corresponding chromatographic peak. The retention time, exact mass, and transition ions were input into TASQ for automated target screening and quantification. To ensure the confidence of compound identification, the following criteria were defined:

- the tolerance of retention time was $\leq$ 0.1 min
- mass tolerance was $\leq$ 5 ppm
- the match score of isotopic profiles, in term of mSigma, was within $\leq$ 30

4.7 Method validation

4.8 Results

4.8.1 selectivity

As the standard kit was a mix powder have different concentration levels with each hormone (Table 6.2), the responses were varied for each analyte. Hormones in the mixture, such as Dihydrotestosterone (DHT) and 11-Deoxycortisol, their concentration were 100 times lower than cortisol. In other words, we used 100 ng/mL as the most concentrated calibrator, in which 11-Deoxycortisol was at 1 ng/mL.

Eight compounds, including two pairs of isomers, were completely separated within a 7 min acquisition run under the optimized chromatography conditions. The result is shown in Table 4.2. A retention time precision of 0.01 0.025 min (CV $<$ 0.1%) was achieved throughout the whole measurements. In TOF mode and timsTOF mode, two isomers 17α-OHP (RT = 3.13 min) and DOC (RT= 2.89 min) were separated with the chromatography resolution of 6. Cortisone (RT = 2.13 min) and Aldosterone (RT = 2.02) had the chromatography resolution of 2.75. The extracted ion chromatogram mass corresponds to the protonated ion [M+H]+, with a mass width was set at 5 mDa (Figure 4.4 and Table 4.4).

4.8.2 Enhancement of EICs and Mass spectra with IMS

Background noise always interferes with peak interpretation and the sensitivity of the method. At a rather low ion concentration, noise chromatographic peaks appear more frequently. IMS aims to eliminate the noise peaks. We used a standard mix of 2 ng/mL cortisol and 0.2 ng/mL cortisol to evaluate the sensitivity of this method.

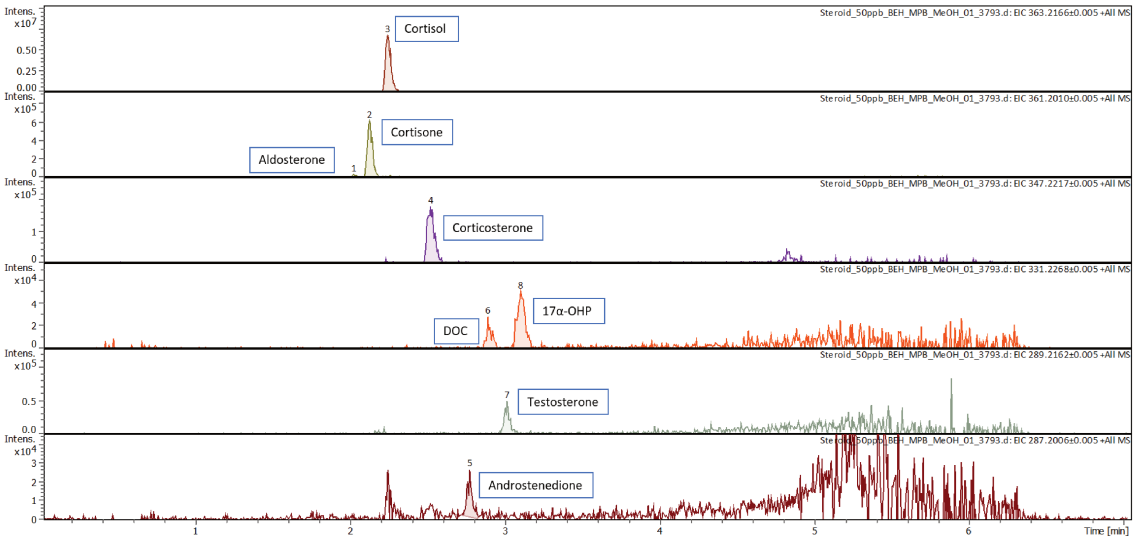
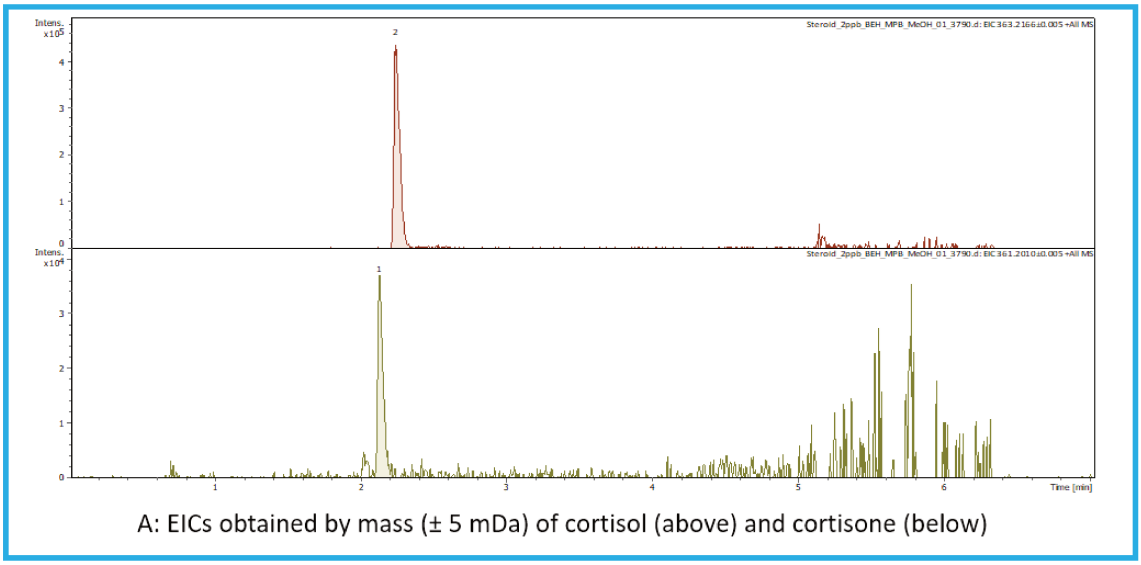


Figure 4.4: EICs of steroid hormones

Compound	RT [min]	Area	I	S/N	Concentration (ng/mL)
Aldosterone	2.02	92358	30809	10.3	1
Cortisone	2.13	1369688	607300	205.5	2
Cortisol	2.25	17153790	6711087	1274.9	20
Corticosterone	2.52	544080	177250	58.9	0.6
Androstenedione	2.77	55445	26374	5.9	0.64
DOC	2.89	63208	27979	8.5	0.6
Testosterone	3.02	112184	49507	14.3	0.2
17α-OHP	3.13	176810	49791	14.8	1

Table 4.4: Overview of integrated EIC

As plotted in Figure 4.5, the EIC generated with mass-only showed an intense noise peak, especially in cortisone, due to the low concentration (0.2 ng/mL). However, by adding the CCS range of cortisone, the background noise can be significantly reduced between 4.5 and 5.5 min. Similarly, we further compared the mass spectra obtained by EIC and EIM (extracted ion mobilogram) (Figure 4.6. Noise peaks between 300 and 360 m/z were eliminated in the mass spectrum obtained by the EIM. IMS/CCS can remove the noise peaks in the matrix to facilitate mass spectra interpretation. This is demonstrated in an example of cortisol (5 ng/mL, EIC $\pm$ 0.005 Da) shown in Figure 4.7. On the left, it represents a mass spectrum in the timsTOF mode without adding the CCS and its zoomed view with regards to the targeted m/z = 363.2175 along with simulated isotope patterns of cortisol. Moreover, a mass spectrum using a CCS filter is shown on the right. As seen in the first row, a much cleaner MS was obtained after extraction with CCS, which was more efficient in discarding the ions of large masses.



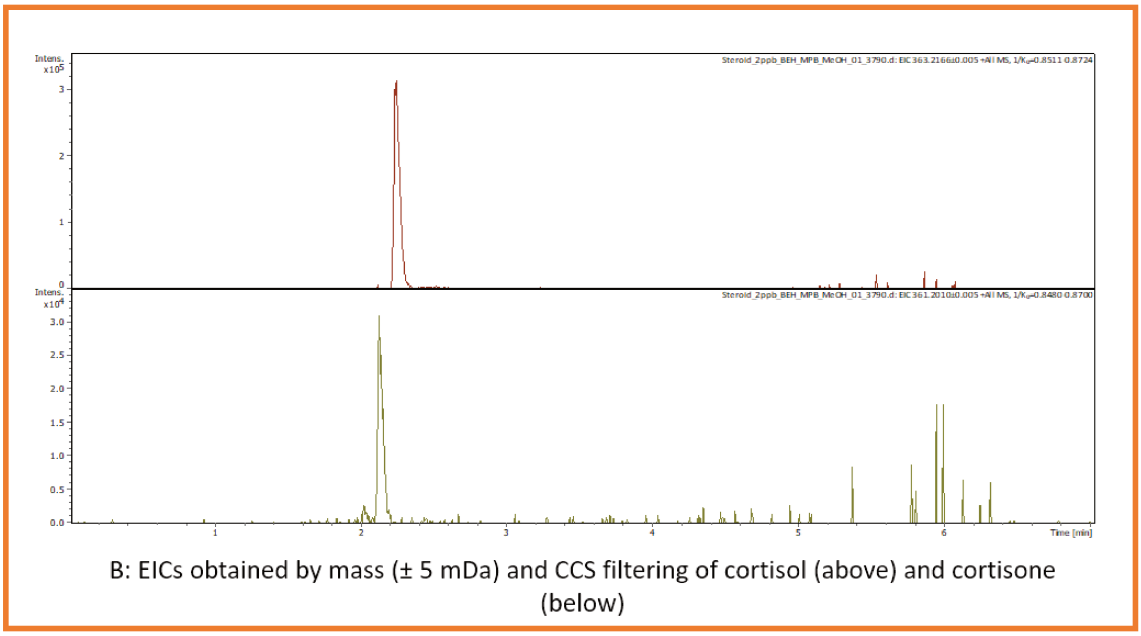
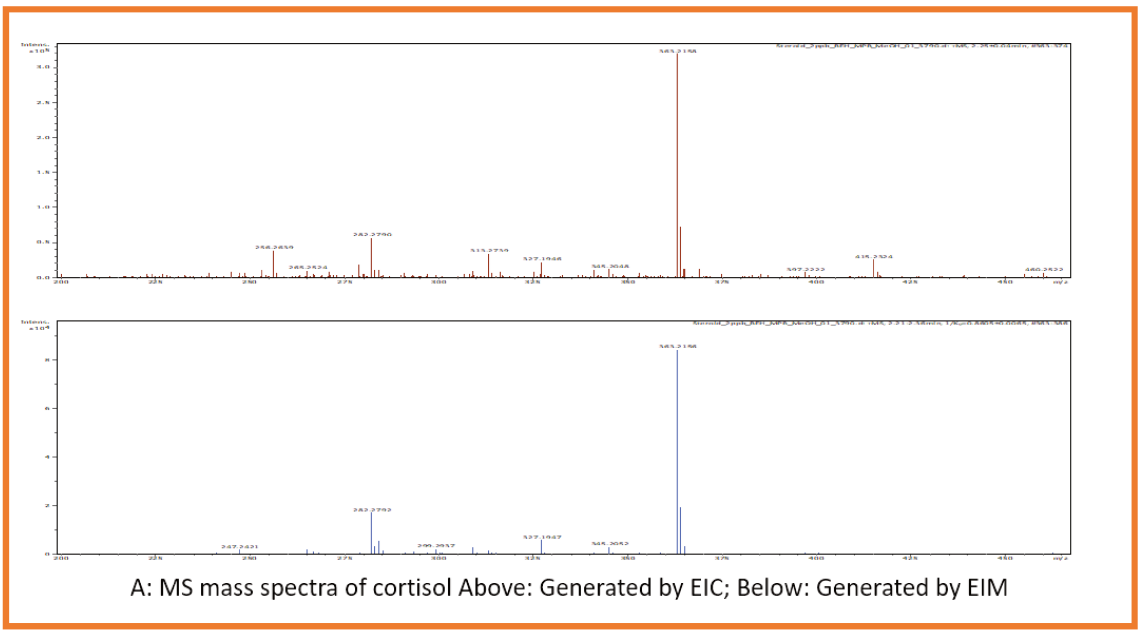


Figure 4.5: Comparison of EIC generated by exact mass and exact mass with CCS filtering in cortisone and cortisol
(A) EIC of cortisol (above) and cortisone (below) with mass; (B) EIC + CCS range of cortisol and cortisone



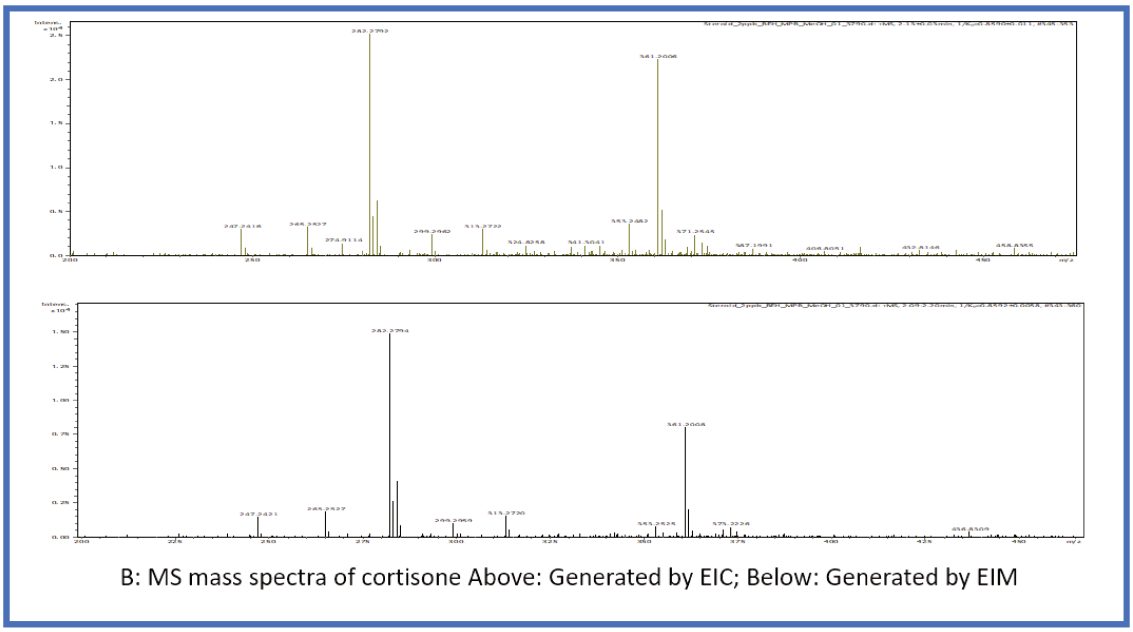


Figure 4.6: Comparison of mass spectra generated by EIC and EIM (A) Mass spectra of cortisol; (B) Mass spectra cortisone

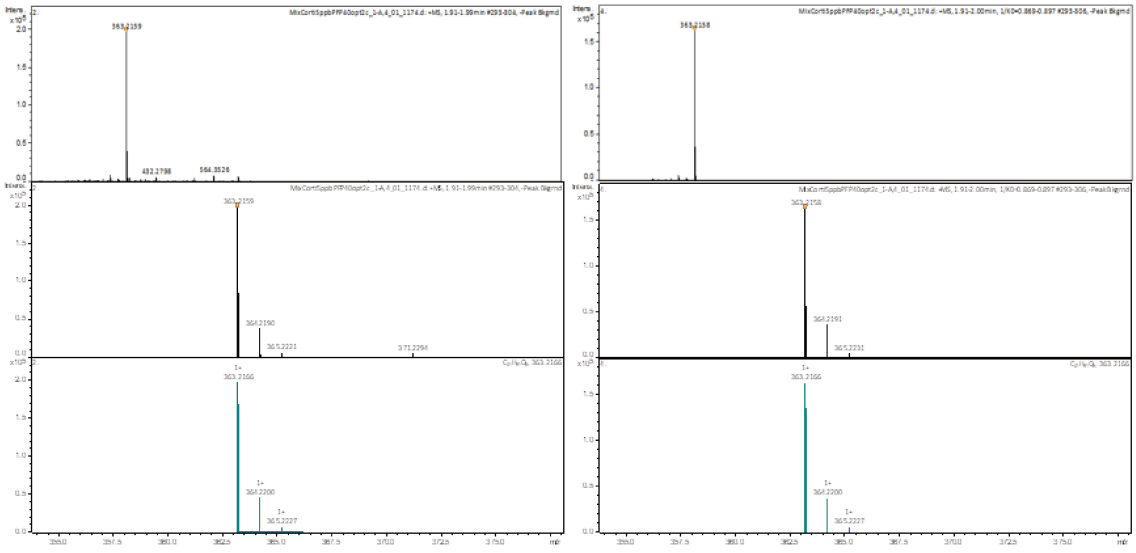


Figure 4.7: Mass spectra of cortisol
MS of Cortisol (5ng/mL), ECI = 363.2159 ± 0.005 in timsTOF mode. Left: MS extracted by accurate mass, with zoom in m/z = 363.2159 and its simulate isotope patterns. Right: MS extracted by accurate mass and CCS (1/k0 = 0.869-0.897). In the first row, m/z = 432 and 564 were eliminated by CCS, in the zoom MS, m/z = 371 was also discarded.

Furthermore, we applied the same concept with two other mass widths: ± 2 mDa and ± 10 mDa. Regarding cortisol, a sufficient concentration (2 ng/mL) was introduced in the LC-HRMS system to produce the response. Therefore, no important deviation yielded when the mass error window was modified. In contrast, with regards to cortisone, by narrowing down the mass accuracy, background noise can effectively be reduced. Compared to the CCS filter, we observed less noise with ± 10 mDa than EIC of ± 2 mDa without the CCS range, indicating the high selectivity in IMS. Moreover, the S/N ratio was estimated to be 20 if the CCS filter was applied and 14 in the EIC without the CCS filter. This result indicated that mass accuracy is an important factor in producing high-quality mass spectra and chromatographic peaks, while it can be limited by the analyte concentration. Therefore, combining IMS can enhance the mass spectrum at low concentrations and produce a cleaner mass spectrum.

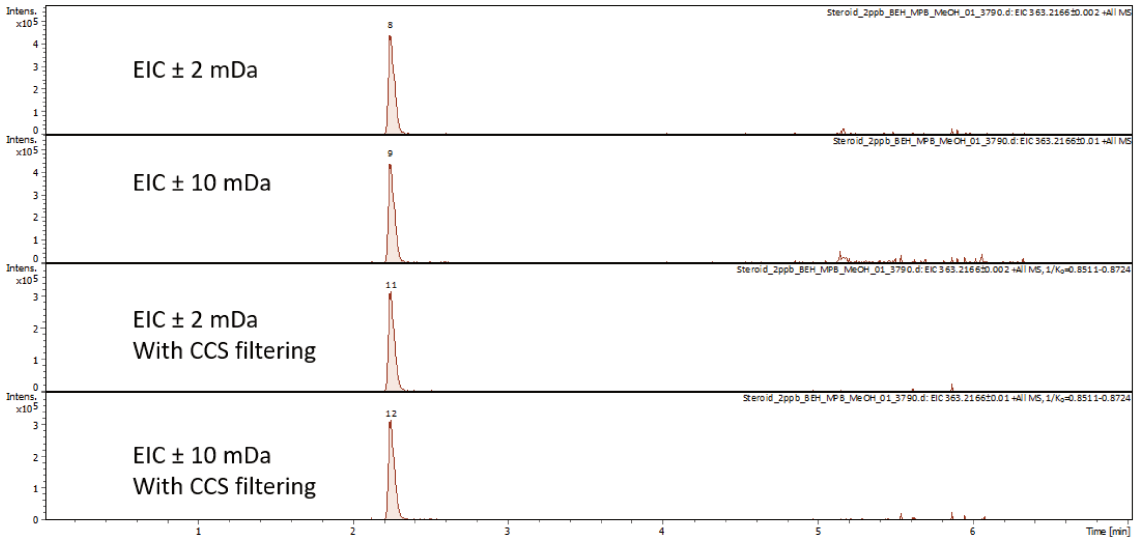


Figure 4.8: IMS enhancement in EIC Cortisol: Comparing chromatograms background at different mass widths and using CCS

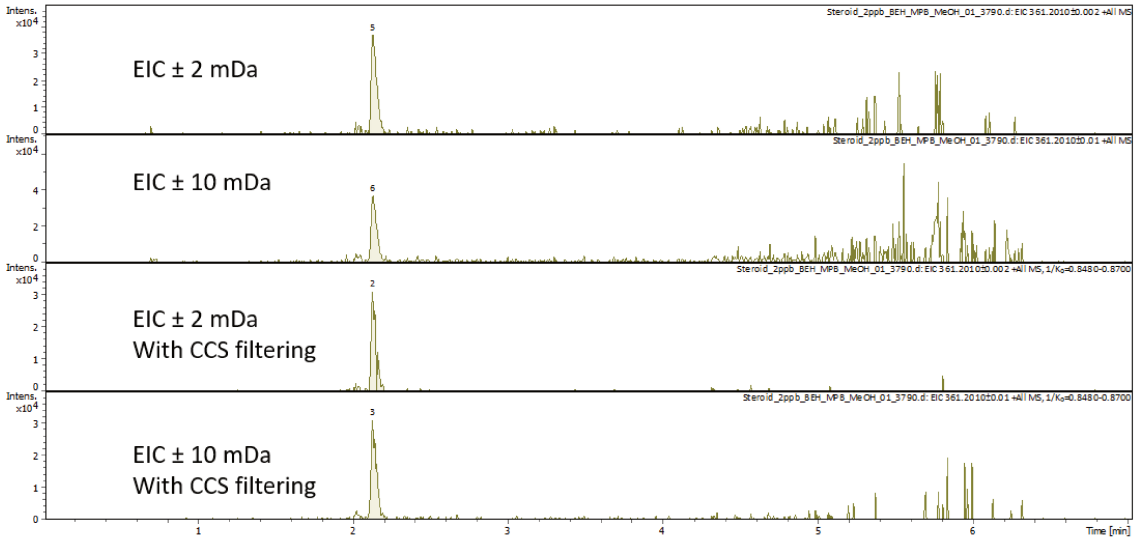


Figure 4.9: IMS enhancement in EIC Cortisone: Comparing chromatograms background at different mass widths and using CCS

4.8.3 Repeatability and linearity

The repeatability test was performed with calibration solutions in addition to a lower concentration at 0.5 ng/mL (for cortisol) with 3 replicate injections. The RT deviation was within 0.01 to 0.03 min. The mass deviation was below 3 ppm. The linearity was initially tested in cortisol concentrations ranging from 0.5 to 50 ng/mL. However, the curve started pending at the highest calibration point (Figure 4.10). Therefore, the calibration range needs further optimization for QTOF and TIMS-QTOF. Moreover, the degradation of standards was observed during method optimization, and the stability test should be undertaken in the next step.

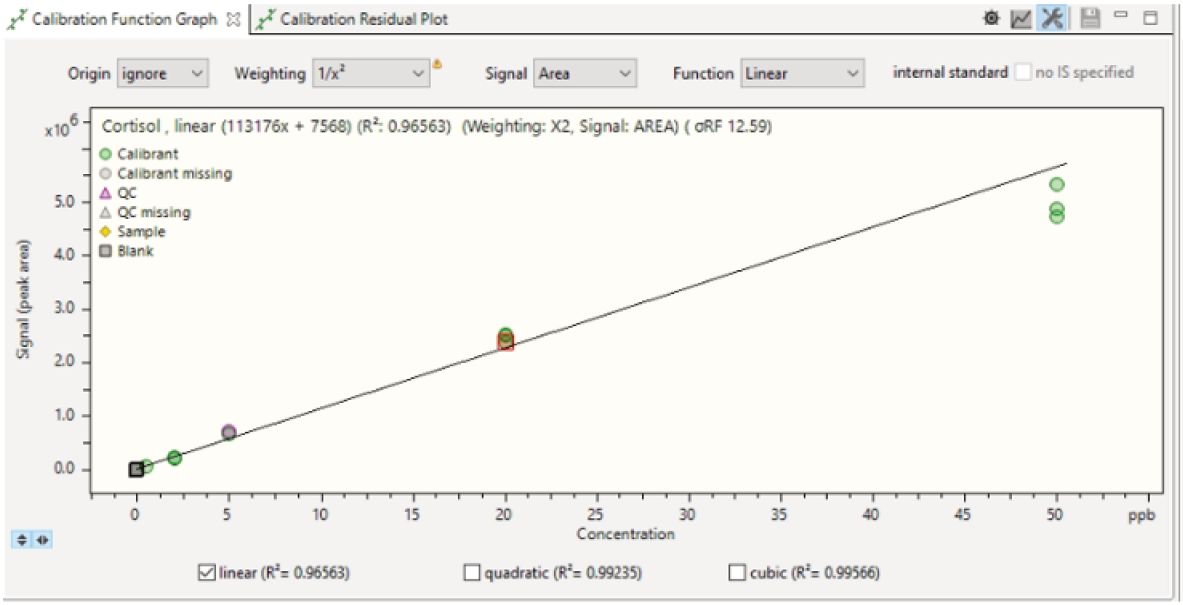


Figure 4.10: Linearity test of cortisol
Calibration ranged in 0.5 to 50 ng/mL.

4.9 Conclusion and perspectives

The study presented in this chapter is still in the early stage. We observed a loss of linearity with the calibration curve at 50 ng/mL. As this method aims to quantify cortisol in juvenile fish, a series of calibration solutions with lower concentrations can be optimized (e.g., ≤ 20 ng/mL). The LOD and LOQ are other important points that are considered in this study. A surrogate matrix can be prepared to estimate the matrix effect and LOD. The stability test of the standards, including labeled standards, needs to be undertaken. Moreover, the negative ionization mode will be a complement to analyze other steroid hormones, such as estrone and estradiol. Despite all the above, the enhancement of IMS/CCS is highlighted in this study. HRMS guarantees sensitivity with regards to the mass and a broad analysis mass range. IMS is a high-speed analytical technique, making it compatible with the TOF and UHPLC. Since IMS provides an orthogonal separation dimension, the CCS range can remove the noise peak or coeluting peak acquired in the full scan. Therefore, IMS offers a novel aspect in MS and MS/MS spectra enhancement, as presented in Chapter II. The mass spectra is cleaned, and the improvement in the S/N can be further verified in a real and complex matrix.

References

D. A. Armbruster and T. Pry. Limit of blank, limit of detection and limit of quantitation. *The clinical biochemist reviews*, 29(Suppl 1):S49, 2008.

E. Desimoni and B. Brunetti. About estimating the limit of detection by the signal to noise approach. 2015.

C. Domenech-Coca, R. Mariné-Casadó, A. Caimari, L. Arola, J. M. Del Bas, C. Bladé, and M. I. Rodriguez-Naranjo. Dual liquid-liquid extraction followed by lc-ms/ms method for the simultaneous quantification of melatonin, cortisol, triiodothyronine, thyroxine and testosterone levels in serum: Applications to a photoperiod study in rats. *Journal of Chromatography B*, 1108:11–16, 2019.

A. Gaudl, J. Kratzsch, Y. J. Bae, W. Kiess, J. Thiery, and U. Ceglarek. Liquid chromatography quadrupole linear ion trap mass spectrometry for quantitative steroid hormone analysis in plasma, urine, saliva and hair. *Journal of Chromatography A*, 1464:64–71, 2016.

A. Gaudl, J. Kratzsch, and U. Ceglarek. Advancement in steroid hormone analysis by lc–ms/ms in clinical routine diagnostics–a three year recap from serum cortisol to dried blood 17α-hydroxyprogesterone. *The Journal of Steroid Biochemistry and Molecular Biology*, 192:105389, 2019.

M. Jia, W. M. Chew, Y. Feinstein, P. Skeath, and E. M. Sternberg. Quantification of cortisol in human eccrine sweat by liquid chromatography–tandem mass spectrometry. *Analyst*, 141(6):2053–2060, 2016.

S. Nomura, M. Fujitaka, K. Jinno, N. Sakura, and K. Ueda. Clinical significance of cortisone and cortisone/cortisol ratio in evaluating children with adrenal diseases. *Clinica chimica acta*, 256(1):1–11, 1996.

J. Ojoghoro, M. Scrimshaw, and J. Sumpter. Steroid hormones in the aquatic environment. *Science of The Total Environment*, 792:148306, 2021.

B. Sadoul and B. Geffroy. Measuring cortisol, the major stress hormone in fishes. *Journal of Fish Biology*, 94(4):540–555, 2019.

C. B. Schreck and L. Tort. The concept of stress in fish. In *Fish physiology*, volume 35, pages 1–34. Elsevier, 2016.

X. Shen, H. Chang, Y. Sun, and Y. Wan. Determination and occurrence of natural and synthetic glucocorticoids in surface waters. *Environment international*, 134: 105278, 2020.

Y. Song and X.-s. Feng. Sample preparation and analytical methods for steroid hormones in environmental and food samples: An update since 2012. *Critical Reviews in Analytical Chemistry*, pages 1–19, 2021.

A. Weizel, M. P. Schlüsener, G. Dierkes, and T. A. Ternes. Occurrence of glucocorticoids, mineralocorticoids, and progestogens in various treated wastewater, rivers, and streams. *Environmental science & technology*, 52(9):5296–5307, 2018.

S. Wu, A. Jia, K. D. Daniels, M. Park, and S. A. Snyder. Trace analysis of corticosteroids (css) in environmental waters by liquid chromatography–tandem mass spectrometry. *Talanta*, 195:830–840, 2019.

R. Zhong, H. Zou, J. Gao, T. Wang, Q. Bu, Z.-L. Wang, M. Hu, and Z. Wang. A critical review on the distribution and ecological risk assessment of steroid hormones in the environment in china. *Science of The Total Environment*, 786: 147452, 2021.

Chapter 5

Collision Cross Section Prediction with Molecular Fingerprint Using Machine Learning

summary

This chapter is interested in the development and validation of CCS prediction models using machine learning. The objective of this study is to develop a CCS predictor which can be used in non-target analysis for environmental contaminants identification. It can be applied and complement to the research in Chapter II and Chapter III.

First, a short overview of the advantages of introduce IMS and CCS values in non-target screening will be discussed, following a general explanation of the experiment strategy. The limits and perspectives of this research will be emphasized in conclusion.

The experimental details and evaluations of this study are discussed in the published article "*Collision Cross Section Prediction with Molecular Fingerprint Using Machine Learning*", including at the end of this chapter Section 5.4.

5.1 Introduction

Collision-cross section (CCS) represents an observational property that averages molecular surfaces interact with a drift gas. The empirical CCS values give a unique chemical property under a defined drift gas (e.g., N_2, He), a temperature and an electric field used during the measurement (May *et al.*, 2017). Experimental CCS values are typically obtained by ion mobility spectrometry (IMS) via four separation technique, which are drift tube ion mobility spectrometry (DTIMS), traveling wave ion mobility spectrometry (TWIMS), trapped ion mobility spectrometry (TIMS), and differential mobility analyzer (DMA). The major contribution of CCS databases are made by Picache *et al.* (2019) records 3,800 experimental CCS values cross 80 classes, PubChem records over 8,965 experimental CCS (Access on April 2022) (Schymanski *et al.*, 2022). Most of IMS and CCS measurement focus on peptides and metabolites researches, Dilger *et al.* (2013) measured 1,470 peptides, Meier *et al.* (2021) reported two million CCS values for peptides. In metabolite research, CCS databases and CCS predictors are flourished, AllCCS included chemical structures with 5,000 experimental CCS (Zhou *et al.*, 2020), MetCCS is a specif CCS predictor for metabolites (Zhou *et al.*, 2016), HMDB 5.0 have added 871,680 predicted CCS values (Wishart *et al.*, 2022).

Most recently, IMS and CCS are recently introduced in environmental non-target analysis as an additional separation and identification dimension. IMS and CCS values improve the separation of isomers and mass spectrum quality by removing noise peak (Dodds and Baker, 2019). Additionally, CCS provides structure information for unknown, which can reduce numbers of candidate and elucidate molecular structure. Meanwhile, predicted CCS values have been used to validate candidates (Celma *et al.*, 2020; Hinnenkamp *et al.*, 2022; Celma *et al.*, 2021; Stephan *et al.*, 2016), proving IMS and predicted/experimental CCS values can improve identification certainty. While most of existing CCS predictors used the single-laboratory or single-instrument experimental values for modeling, the prediction accuracy can be varied by IMS separation technique (Zhou *et al.*, 2016). On the other hand, most models use molecular descriptors, such as molecular mass, number of heavy atoms, polarity, LogP, etc. Ross *et al.* (2020) introduced molecular quantum numbers (MQNs) to represent the variance between mass-CCS space. Molecular fingerprints encoding molecular structure to a binary code have been used to predict molecular properties (Heinonen *et al.*, 2012; Zang *et al.*, 2017), but molecular fingerprints have never been used in CCS prediction before.

This research is aimed to develop a novel CCS perdition model using a collected CCS dataset and molecular fingerprint to describe molecular structure.

5.2 Experiments

5.2.1 Dataset

The size, quality and the diversify of data resources are key factors in machine learning. Therefore, the first step was to collect experimental CCS values from literature and libraries. As dataset was merged from different studies, the deviation

of empirical CCS measurement can drastically affect the prediction performance. Replicated chemicals were examined before modeling for two reasons:

- to eliminate replicated data with same chemicals and same CCS values. If too many replicates are present in the entire dataset the chemicals within the test set are probably seen during the training, resulting in a higher accuracy. Moreover, too many replicates interference data diversity, prediction performance cannot cover a broad scope of chemical class. Thereby, the prediction accuracy can be biased.

- to avoid the measurement deviation in the database. Although CCS are instrument-independent values (in same drift gas and same temperature), separation techniques, experimental conditions (e.g., IMS calibration), data quality, etc, yield measurement deviation, then decrease prediction accuracy.

Because empirical CCS values cannot be verified by remeasuring with the reference standard, all the replicated chemicals with different CCS values were kept in the dataset. The replicates, matching the same chemical SMILES and CCS value were removed before modeling. As results, the deviation of CCS measurement was plotted in Figure 5.1.

Furthermore, considering of CCS values is a structure related property, molecular classes may enhance the prediction performances. An class-based model was tested. To ensure the data size of each classes, a classification prediction model was first developed. Finally, The distribution of the predicted chemicals' super class was plotted in Figure 5.2

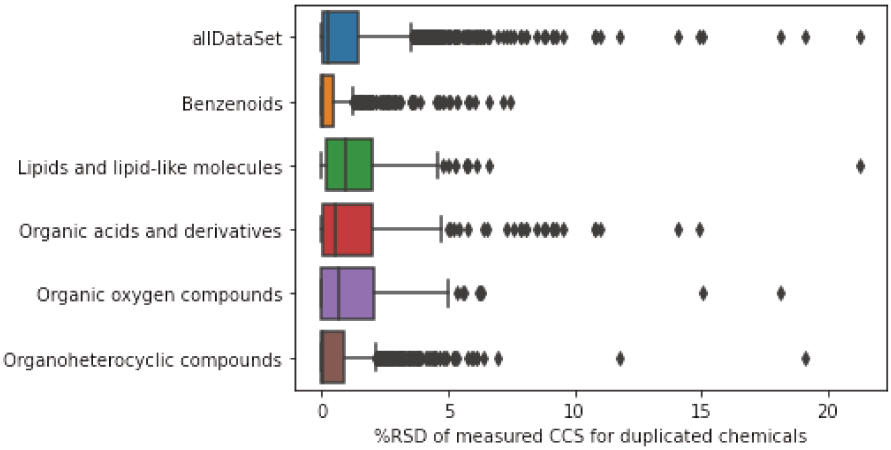


Figure 5.1: RSD of replicated chemicals
This figure showed the relative standard deviation of CCS values measured from different laboratories. The median of RSD was below 2%, which confirmed the robustness of CCS measurement cross platforms and the possibility to merge all the dataset for Machine learning.

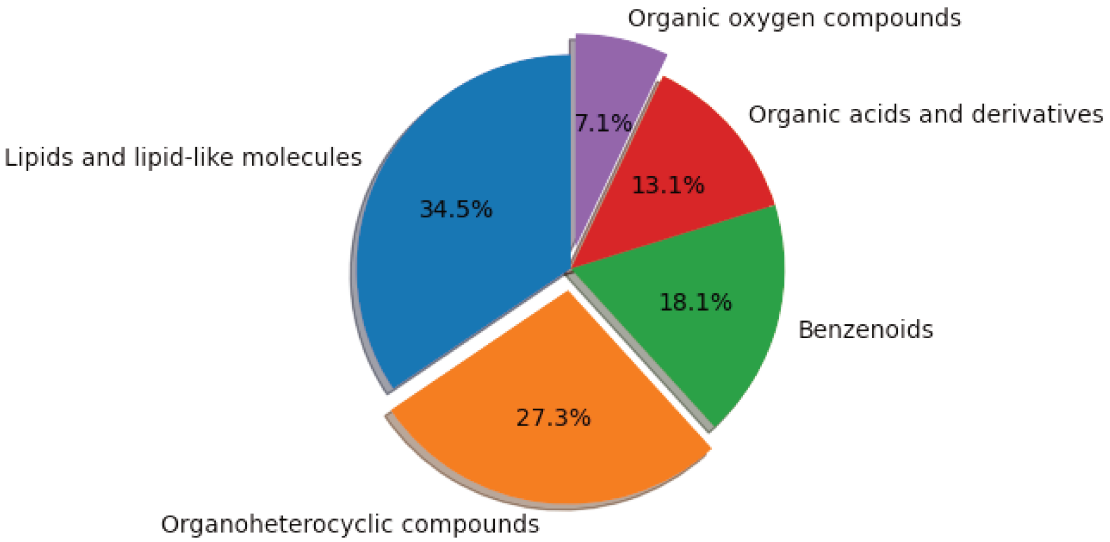


Figure 5.2: Distribution of predicted chemical super class in dataset

5.2.2 Design of experiment

Molecular fingerprint is a way to compute molecular structure to a binary string, typically 1024 bits or 2048 bits. Technically, molecular substructure is encoded to a binary string via molecular SMILES (Weininger, 1988), as 1 or 0. 1 stands for a existing substructure, 0 for the missing substructure. An example of Caffeine is shown in different types of fingerprinting algorithm have been used to represent molecular 2D structure and calculate structure similarity (Rogers and Hahn, 2010; Cereto-Massagué *et al.*, 2015), such as atom pairs (Carhart *et al.*, 1985), Morgan (Rogers and Hahn, 2010), and Topological Fingerprints (Nilakantan *et al.*, 1987). We chose topological fingerprint, which represents a short-range atom pair information of a molecule. A atom pair consist of four-atom sequences, including atom-type, topological distance and torsion angle between angle (Capecchi *et al.*, 2020). Therefore, topological fingerprint effectively represents substructure information of a small molecule.

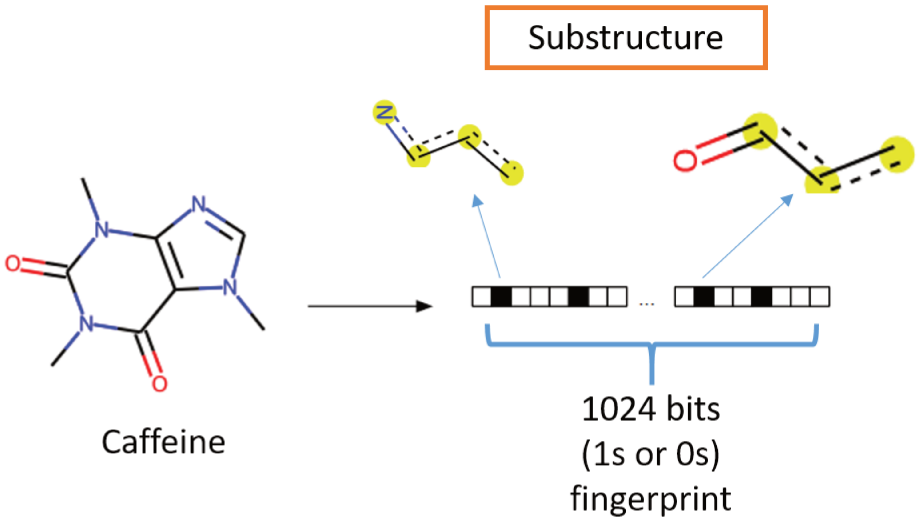


Figure 5.3: An illustration of molecular fingerprints

Random forest is one of the most poplar supervised machine learning algorithm for non-linear correlation model. The modeling of supervised random forest and this study consisted of the following steps:

1. Calculate and define the features (Molecular fingerprint) and response variable (Experimental CCS values) for the entire dataset

2. Randomly divide the whole dataset into training set (80%) and test set (20%). This procedure is repeated 5 times, meaning 5 trees grow in a forest.

3. Start with a random sample to optimize the best binary split in the features, which shows a distinguishable response. As a example, molecules contain bit-588 generally have a CCS value < 150 Å^2. A tree keep growing by descendant features.

4. run test set with the predict model to estimate accuracy. Aggregate the predictions and prediction error of each trees and output the best optimized hyperparameters (in Figure 5.4).

5. Evaluate model with Feature-Importance and its corresponding substructure to understand the models and prediction accuracy (Figure 5.5).

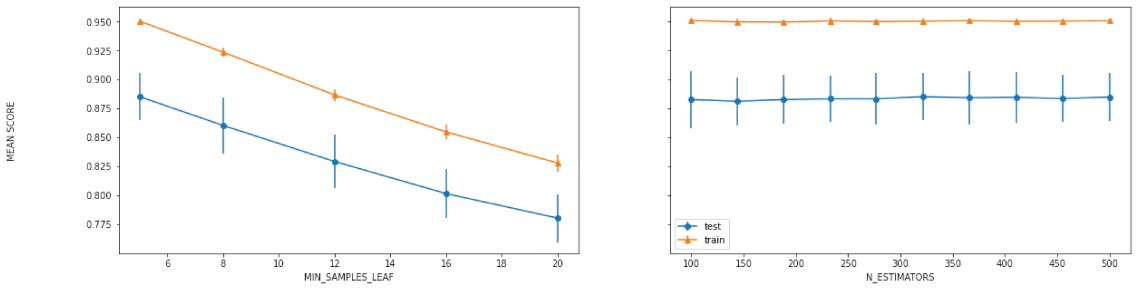


Figure 5.4: An output of modeling results

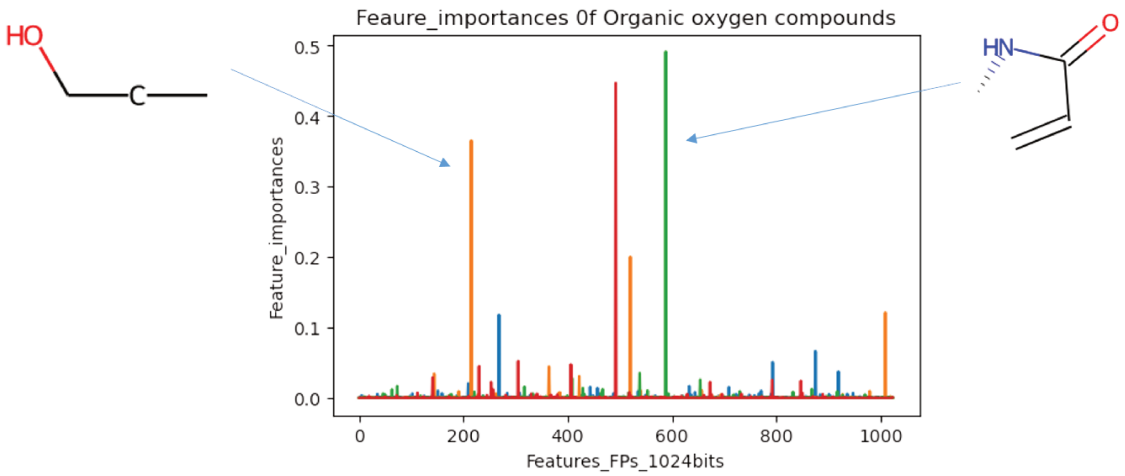


Figure 5.5: Plotted feature importance

5.3 Conclusion and perspectives

This research developed prediction modeling based on topological fingerprint to predict molecular super class and CCS values. We used a merged CCS database for training and test (DTIM and TWIM measured CCS values) models. We also compared class-based approach with a direct prediction approach. The test prediction accuracy was 0.958 by the direct prediction approach, 3 class-based prediction models more than 0.9, and over 0.86 for the remaining two classes. The MRE was between 1.89% to 2.33%. The prediction accuracy is comparable to other single-instrument and single-laboratory predictors, which indicate the compatibility of cross-platform measurement. Therefore, it can be used in non-target analysis to reduce candidate structures. As the models were learned with different sources, predicted CCS value provide a reference value to verify in-house CCS measurement. Meanwhile, predicted CCS values can compare to the in-house CCS database obtaining by TIMS to estimate the instrument and IMS technique deviation

Furthermore, topological fingerprint was first time introduced as CCS predictors. Comparing to other machine learning studies, we avoided selecting of structure-related molecular descriptors, but to use a more straightforward descriptor to learn these models. However, molecular fingerprint showed the limited to encode macromolecules and more diverse chemical classes. Because one-bit present different substructures, molecules have different structure may have the same fingerprints, other molecular descriptors can be implemented to enhance prediction performance.

Considering of the IMS technique, only DTIM and TWIM were introduced in this study, TIMS measured CCS can be potentially added to enhance the performance of predictors, and to evaluate the deviation of CCS measurement using different ion mobility separation technique. The class-based prediction approach yields comparable or even lower prediction accuracy than direct prediction model. We hypothesized that it was mainly related to the data size, and fingerprint itself might be sufficient to distinguish molecular class. However, in order to have enough data in each class, we merged 43 super class to 5, characteristic substructures of each class were lost, it might explain the comparable results of two approaches. More data with well de-

fined super class or class can help to verify this argument. One of the main drawback of this CCS predictor is that it is develop and run in Python, therefore it requires advanced knowledge of Python and raw data handling to ease perform prediction. Other CCS predictors have a user friendly interface, it can import the keys and download the results through websites, such as CCSbase (https://ccsbase.net/about), AllCCS (http://allccs.zhulab.cn/).

molecules
-15

Article

Collision Cross Section Prediction with Molecular Fingerprint Using Machine Learning

Fan Yang [1,*], Denice van Herwerden [2], Hugues Preud'homme [1,*] and Saer Samanipour [2,3,*]

[1] Institut des Sciences Analytiques et de Physico-Chimie Pour l'Environnement et les Materiaux (IPREM-UMR5254), E2S UPPA, CNRS, 64000 Pau, France
[2] Van 't Hoff Institute for Molecular Sciences (HIMS), University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[3] UvA Data Science Center, University of Amsterdam, 1098 XH Amsterdam, The Netherlands
* Correspondence: fyang@univ-pau.fr (F.Y.); hugues.preudhomme@univ-pau.fr (H.P.); s.samanipour@uva.nl (S.S.)

Abstract: High-resolution mass spectrometry is a promising technique in non-target screening (NTS) to monitor contaminants of emerging concern in complex samples. Current chemical identification strategies in NTS experiments typically depend on spectral libraries, chemical databases, and in silico fragmentation tools. However, small molecule identification remains challenging due to the lack of orthogonal sources of information (e.g., unique fragments). Collision cross section (CCS) values measured by ion mobility spectrometry (IMS) offer an additional identification dimension to increase the confidence level. Thanks to the advances in analytical instrumentation, an increasing application of IMS hybrid with high-resolution mass spectrometry (HRMS) in NTS has been reported in the recent decades. Several CCS prediction tools have been developed. However, limited CCS prediction methods were based on a large scale of chemical classes and cross-platform CCS measurements. We successfully developed two prediction models using a random forest machine learning algorithm. One of the approaches was based on chemicals' super classes; the other model was direct CCS prediction using molecular fingerprint. Over 13,324 CCS values from six different laboratories and PubChem using a variety of ion-mobility separation techniques were used for training and testing the models. The test accuracy for all the prediction models was over 0.85, and the median of relative residual was around 2.2%. The models can be applied to different IMS platforms to eliminate false positives in small molecule identification.

Keywords: collision cross section; ion mobility spectrometry; non-target screening; machine learning

Citation: Yang, F.; van Herwerden, D., Preud'homme, H.; Samanipour, S. Collision Cross Section Prediction with Molecular Fingerprint Using Machine Learning. *Molecules* **2022**, *27*, 6424. https://doi.org/10.3390/molecules27196424

Academic Editor: Thomas Letzel

Received: 30 August 2022
Accepted: 19 September 2022
Published: 29 September 2022

1. Introduction

A large number of chemicals have been released into the environment by human activities, such as agriculture, industrial productions, and their relative byproducts. Once these chemicals enter the environment, transformation products (TPs) can be produced through hydrolysis, photosynthesis, and biological metabolism [1–6]. Most of these chemicals and their TPs are missing molecular and/or structure information. Thus, these chemicals' human and environmental risk assessments remain an open question [6–12]. Although most legacy pollutants have been banned for decades in many countries, they can still be detected at trace-level in the environment [2,13–15]. The known pollution is only the tip of the iceberg compared to the number of environmental hazards [1,13,14].

Non-target screening/analysis (NTS) is considered as an appropriate methodology to identify a variety of chemicals, especially for the unknown unknowns, such as contaminants of emerging concern (CECs) [16–18]. High-resolution mass spectrometry (HRMS) coupled with gas or liquid chromatography (GC or LC) is the most commonly used analytical technique in human health and environmental assessments. Thanks to the advance of HRMS, it has been increasingly applied in NTS studies in the last decades [17,19–21]. HRMS (i.e.,

Time-of-flight (TOF) and Orbitrap) maintains a high mass accuracy within ±5 mDa m/z error, and it can be acquired in full scan MS data or plus MS/MS data [10,21–24]. The accurate mass of the parent ion and the fragments are used to identify unknowns [17,19,21]. The isotopic pattern is one of the additional criteria which can help determine the presence of hetero-elements in non-target analysis [25]. However, mass spectral information is not enough for highly confident structural elucidation [22,25,26]. Therefore, inclusion of orthogonal sources of information such as measured or predicted retention time and/or retention time indices is necessary [21,27,28]. Such measurements are complex to perform and require particular experimental conditions [29–31].

Collision cross section (CCS) is a platform-independent measure of chemical structure in the gas phase and the three-dimensional space [32–34]. Studies have demonstrated that the inter-laboratory CCS biases are within 2% for the same IMS technique [35,36]. Moreover, cross-platform biases are below 3% for over 98% of the chemicals included in their studies [37,38]. Drift tube ion mobility (DTIM) and traveling wave ion mobility (TWIM) are two of the most used IMS techniques to measure the CCS value or drift time [37,39]. CCS value and drift time have been employed in NTS as an additional source of information, to increase confidence level in structural elucidation [40–42]. In addition to experimentally defined CCS values, CCS values can be estimated/predicted via theoretical calculations or Machine Learning (ML) [43,44]. ML CCS predictions take advantage of large datasets of the experimentally defined CCS values to train, validate, and test the regression models [44]. Zhou et al. [45] reported the first CCS prediction tool using the support vector regression (SVR) ML algorithm for metabolites. Plante et al. [46] published a deep neural networks CCS prediction strategy for cross-platform CCS measurement. The currently available CCS prediction tools rely on molecular descriptors or the combination of the chemical class and the m/z value of the parent compound [44–52]. Molecular fingerprints, which are more accurate and representative of the structure of a molecule [53], have not been used for the prediction of CCS values due to the difficulties associated with variable selection.

This study proposes a novel approach for CCS prediction using molecular topology fingerprints instead of molecular descriptors. First, we built a classification model to predict the chemical super classes based on their fingerprints. This model was used to classify chemical super classes. Then, CCS prediction models were developed for each super class. Additionally, all 13,324 chemicals were combined and to build a direct CCS prediction model. We also evaluated the impact of the chemical classes on the model accuracy.

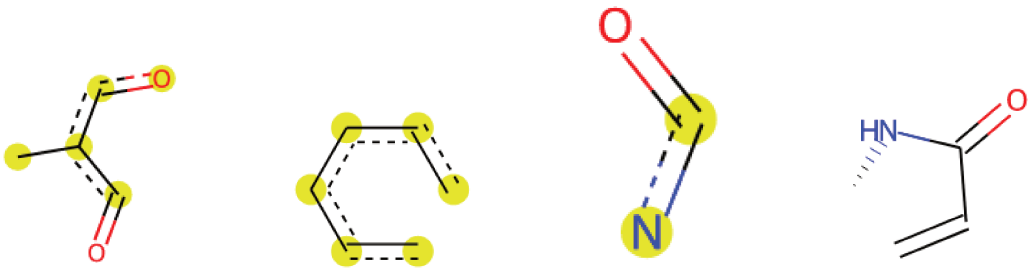
2. Materials and Methods

2.1. Datasets

Experimental CCS databases and chemical information were collected from Zenodo, PubChem, and published articles as referenced in Table 1. Firstly, we retrieved all the missing SMILES notations from PubChem by PubChem CID using the Python PubChemPy library [54]. All the datasets were concatenated, and molecular fingerprints were generated by RDKit [55] (Open-source cheminformatics https://www.rdkit.org) (accessed on 10 April 2022) modules in Python. Hence, a dataset containing PubChem CID, SMILES [56,57], and empirical CCS value was saved as a csv file ready for model development and validation. The datasets and the source codes are available at https://github.com/fyang22/CCS-Prediction-Publish (accessed on 10 April 2022). Additional details about model optimization and construction are available in the Supplementary Materials.

The merged dataset included 13,324 unique empirical CCS values from 108.4 to 450.6 Å^2, measured by TWIM and DTIM. The merged dataset of 3313 chemicals was categorized into 43 super classes, including POPs, lipids, sugars, metabolites, hormones, drugs, etc. This dataset was then used for a classification model training and testing. Topological torsion (TT) fingerprints were chosen as features to encode chemical structure. TT fingerprints were first introduced by Nilakantan et al. [58], which describe the atom type, the topological distance between two atoms within four bonds, and torsion angles [59].

Four examples of molecular substructures are shown in Figure 1. The SMILES were converted to 1024 bit-strings fingerprints (FPs) by the implemented module in RDKit. The FPs were used to calculate molecular similarity, then visualized by principal component analysis (PCA) and fit machine learning models.



(**a**) bit-string 194 (**b**) bit-string 792 (**c**) bit-string 842 (**d**) bit-string 588

Figure 1. Examples of a substructure of a bit string. The most relevant features in the prediction models.

2.2. Overall Workflow

This study consists of two major parts and three models, and the workflow is summarized in Figure 2. Firstly, we developed a classification model to categorize chemicals into five groups, so-called "super class", based on their FPs similarity. The number of the "super class" was selected to create a balanced distribution of chemicals in each class. Five class-based CCS prediction models were developed using the optimized predicted category. Meanwhile, a direct CCS prediction model was built with the complete dataset without considering chemical categories. We also compared the two strategies to assess the prediction accuracy of these two modeling approaches. Finally, we applied the models to NORMAN SusDat (i.e., 101,684 chemicals) and carried out the direct and class-based prediction of the CCS values for SusDat.

2.2.1. Dataset for Classification Model

The dataset consisted of the identified chemical super classes which were merged from three CCS libraries [60–62]. This split dataset was used for chemical classification model training, validation, and testing. Initially, 43 super classes were defined, where most super classes contained less than 20 chemicals. To avoid overfitting of the classification model, we merged different super classes based on the calculated similarity scores of the chemicals. This enabled a more balanced distribution of chemicals in each super class. First, we calculated pair-wised fingerprint similarity by the Tanimoto similarity using RDKit. Tanimoto coefficient is a way to calculate the distance metric using molecular fingerprints [53,63]. Based on the distribution of the chemicals, super classes, and the similarity scores (plotted in Figure 3a), we kept the 5 super classes with the highest population of chemicals (listed in Figure 3b) and used them as ground truth. Chemicals in other super classes were assigned to one of the referred classes based on their similarity with a minimum similarity threshold of 0.6 since around 97% of pair-wise similarities were under 0.6 (shown in Figure 3a). Chemicals (n = 118) not meeting the similarity score criteria were manually assigned to a new super class (5 super classes) based on their characterized functional groups. Meanwhile, we kept the chemicals from the same given class (43 super classes from the raw dataset) in the same new super class. The final dataset consisted of 5 super classes having around 1000 unique chemicals in each class (in Supplementary Table S1), the classification of chemicals is visualized in Figure 3b. This dataset was used for random forest classification modeling. The final dataset for classification included fingerprints with 1024 bit-strings and the assigned super classes. Our super-class reassignment strategy effectively differentiated chemical classes from each other. For example, Organic acid and derivatives (in blue) and Benzenoid (in green) are two separate clusters in the middle left and in the bottom left.

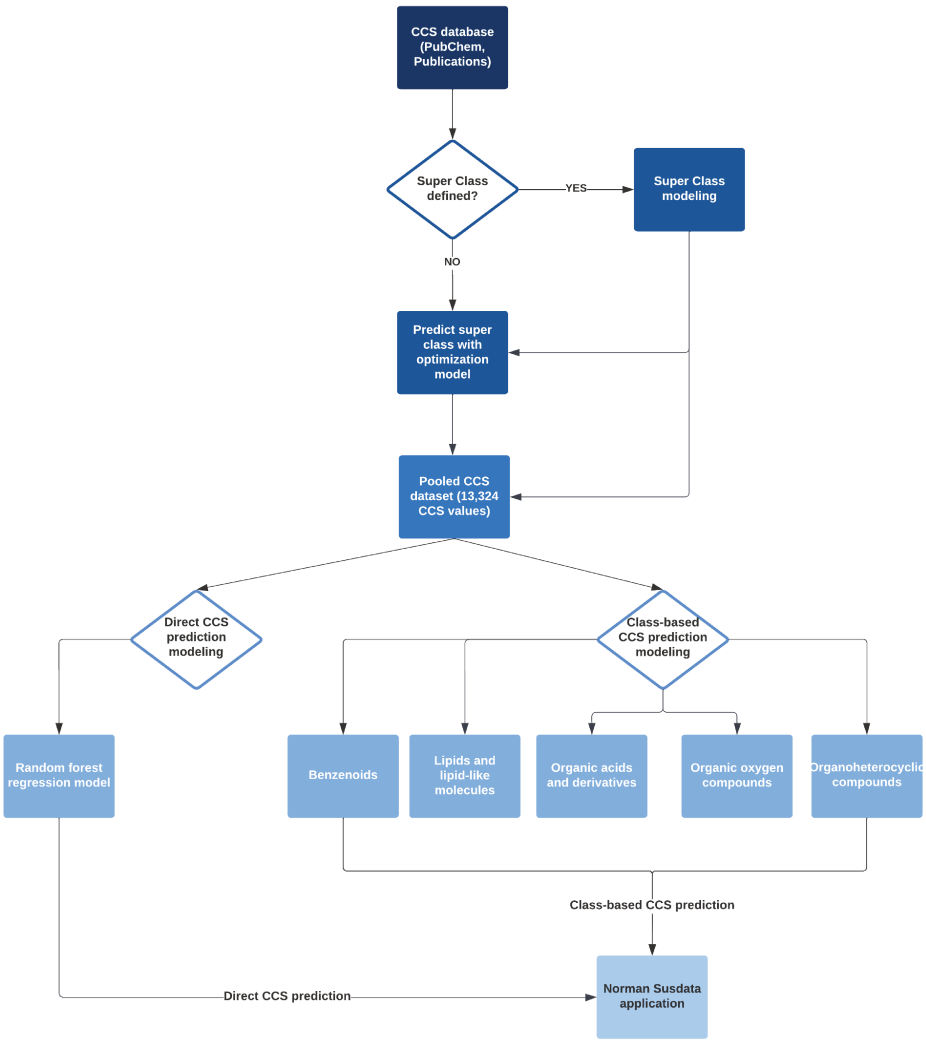
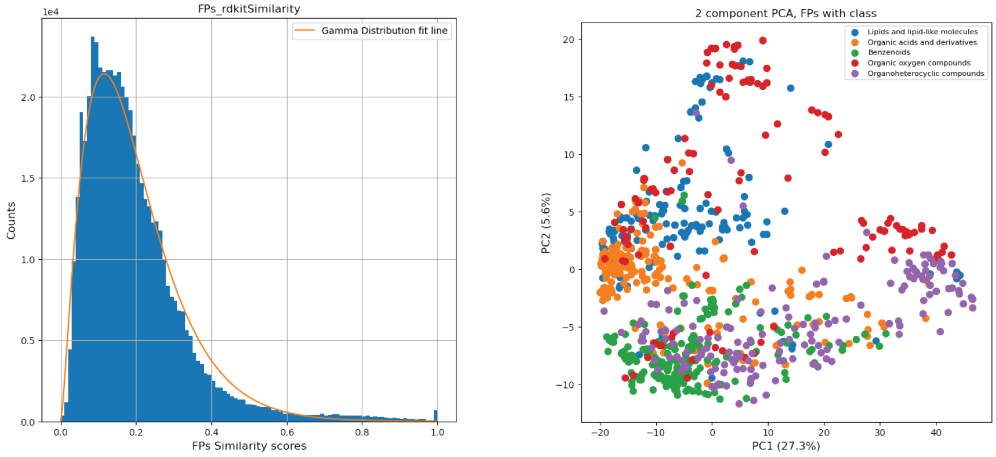


Figure 2. Modeling workflow: CCS empirical databases were collected from 6 different laboratories and PubChem. Two CCS prediction approaches were developed and validated. One model was class-based CCS prediction, and 5 super classes were defined for modeling. Another was a direct CCS prediction model. In the end, both prediction approaches were applied to the Norman Susdat list.

2.2.2. Dataset for Regression

For CCS regression modeling, we only considered protonated ions (8620 chemicals of $[M + H]^+$), deprotonated ions (4589 chemicals of $[M - H]^-$) and radical ions (115 chemicals of $[M]^{\cdot}$). Then, all the replications were removed by the SMILES, adduct ion and CCS values. Meanwhile, we calculated the standard deviation of CCS values for the same chemicals (same SMILES and adduct ion). In the training and test datasets, 642 chemicals have replications with different measured CCS values. The median of relative standard deviation (RSD) was about 1.4% (shown in Supplementary Figure S1) for both positive and negative ionization mode, and studies from multiple laboratories, which are consistent with the results reported by Hinnenkamp et al. [37] and Feuerstein et al. [38] Aspartame resulted in RSD of 12.5%, Picache et al. [60] recorded a CCS value of 127.4 $Å^2$ for Aspartame $[M + H]^+$, which is 40 $Å^2$ lower than the one measured in other references. Different Aspartame CCS values are also recorded in https://pubchem.ncbi.nlm.nih.gov/compound/134601

#section=Collision-Cross-Section (accessed on 1 June 2022). Hence, this dataset, collected from different laboratories and measured by different IM-MS platforms, was appropriate for CCS prediction. The entire dataset contained 13,324 unique empirical CCS values ranging from 108.4 to 450.6 Å^2, covering metabolites, drugs, lipids, etc., and it is available in Supplementary Table S2.



(**a**) Tanimoto similarity (**b**) Super class PCA

Figure 3. Super class distribution: A histogram of pair-wise fingerprints similarity is plotted in (**a**), and a normalized gamma distribution was fitted to the data and is shown as a red line. Based on the gamma distribution curve, similarity $\geq$ 0.6 was chosen to arrange the dataset. In (**b**), a 2D-scatter plot of PCA is generated by fingerprints.

Table 1. Summary of the dataset used in CCS prediction model optimization.

Reference	Number of Chemicals	Instrument *
Picache et al. [60]	1195	Agilent 6560 IM-QTOF MS
Hines et al. [64]	1304	Waters Synapt G2-Si HDMS
Celma et al. [40]	631	Waters VION IMS-QTOF MS
Zheng et al. [61,62]	891	Agilent 6560 IM-QTOF MS
Belova et al. [65]	145	Agilent 6560 IM-QTOF MS
Bijlsma et al. [51]	193	Waters VION IMS-QTOF MS
PubChem [66]	8965	

* Agilent: Drift tube ion mobility (DTIM), Waters: Traveling wave ion mobility (TWIM).

2.3. Modeling

In this study, we optimized three models: (1) Class prediction, (2) Class-based CCS regression model, and (3) a direct CCS regression model. A super class prediction model was first optimized using random forest classification. This model was used to assign the super class (i.e., five classes) of the whole dataset. Then, a regression model was built for each super class to predict the CCS values based on the FPs. Finally, we developed a model using only molecular FPs for CCS prediction. We compared the pros and cons of two CCS prediction approaches. All the modelings were performed using a 5-fold cross-validation by GridSearchCV build-in functionality in Scikit-learn. The details of each modeling strategy are provided below.

2.3.1. Class Prediction

The Class prediction model was first optimized using the random forest classification algorithm. The dataset was split into a training set (80%, n = 836) and a test set (20%, n = 210) with even distribution by super classes. In the random forest classifier, different hyper-parameters impact the model accuracy differently [67]. In this study, we focused on the number of trees in the random forest (n_estimators) and the minimum number of

samples required at each leaf node (min_samples_leaf). These two parameters appeared to have the highest impact on the balance between the model robustness and accuracy. We generated a grid with 25 candidates for the number of trees ranging from 100 to 200 and 2 to 15 for minimum sample leaf. For each model, we performed 5 folds of cross-validation to assess the model accuracy. The model with the highest cross-validation accuracy was chosen as the optimized classification model, and the GridSearchCV scores are plotted in Supplementary Figure S2. The accuracy and F1 scores of each class are listed in Table 2.

Table 2. Results of super-class prediction modeling.

Super Class	Training	Test	F1 Score	Accuracy
Benzenoids	181	46	0.905	0.935
Lipids and lipid-like molecules	189	47	0.909	0.889
Organic acids and derivatives	184	46	0.848	0.813
Organic oxygen compounds	142	36	0.861	0.861
Organoheterocyclic compounds	140	35	0.822	0.857

2.3.2. Class-Based CCS Regression

For class-based regression modeling, we applied the optimized classification model (mentioned above) to the entire dataset, and the results are shown in Supplementary Table S2 and Figure S3. We independently performed the CCS prediction modeling for 5 data splits based on this classification, using the random forest regression algorithm. A total of 80% of the datasets were trained and tested by the rest. Similarly, we generated a grid with 50 candidates and the number of tree fits of 100 to 500. To avoid overfitting, the minimum sample leaf was set from 5 to 20. For each model and each class, 5 folds of cross-validation were evaluated to assess the model accuracy (Supplementary Figure S4a–e).

2.3.3. Direct CCS Regression

For comparison, we developed and tested a direct CCS prediction model for the entire dataset (13,324 compounds). A total of 80% of the data was used to train the model, and 20% of the data to test with 5-fold cross-validation (Supplementary Figure S4f). Similarly to the class-based CCS prediction model, n_estimators, and min_samples_leaf were optimized. The hyper-parameter optimization followed the same steps as class-based modeling (mentioned above). The model details and accuracy are listed in Table 3.

Table 3. Results of CCS prediction modeling.

	Training		Test		
Dataset	**Data**	**R^2**	**Data**	**R^2**	**MRE (%)**
All	10,659	0.972	2665	0.958	2.20
Benzenoids	1930	0.942	483	0.869	1.89
Lipids and lipid-like molecules	3675	0.940	919	0.932	2.33
Organic acids and derivatives	1392	0.950	348	0.901	2.21
Organic oxygen compounds	754	0.925	189	0.860	2.33
Organoheterocyclic compounds	2907	0.960	724	0.933	1.96

3. Results

3.1. Random Forest Classifier and Regression Prediction Model

Random forest is a suitable supervised machine learning algorithm for categorical and nonlinear data. We used a random forest classifier model to divide chemicals into 5 super classes by their molecular fingerprints. Then, we developed two CCS prediction strategies using molecular fingerprints. One is based on molecular super classes and molecular fingerprints, and another is a direct prediction by molecular fingerprints. As a CCS value is

related to the chemical structure, we described each chemical structure by 1024 bit-strings molecular fingerprints, which were used as the prediction features. Each bit represents a substructure of a chemical, and some refer to a characteristic chemical substructure. These bits build up sets of nodes and leaves, then a decision tree.

A collection of decision trees results in a random forest model (decision trees files are available in Supplementary Materials). In order to obtain a generalized CCS prediction model, we merged 7 CCS libraries containing 13,324 unique CCS values (108.4 to 450.6 $Å^2$) measured by TWIM and DTIM platforms from multiple laboratories. Additionally, using a merged dataset for modeling allowed us to understand the variation of CCS measurement.

3.2. Evaluation of Classification Model

We obtained a classification model to separate 5 super classes with a global test accuracy (R^2) $\geq$ 0.871. In the classification model, it is crucial to have sufficient examples and similar training weights for each class. For example, if the dataset is randomly split to 80% of the training set that contains 50 organoheterocyclic compounds but over 200 chemicals of other classes, it would lead to insufficient training for organoheterocylic compounds and an overfitting problem, which can impact the overall performance of the classifier prediction. As shown in Table 2, the training and test sets were evenly distributed by super classes before modeling. The F1 score was over 0.9 for two classes and over 0.82 for the other three, indicating that the training data were balanced between classes. To further evaluate the classification model, we also generated a confusion matrix (Figure 4).

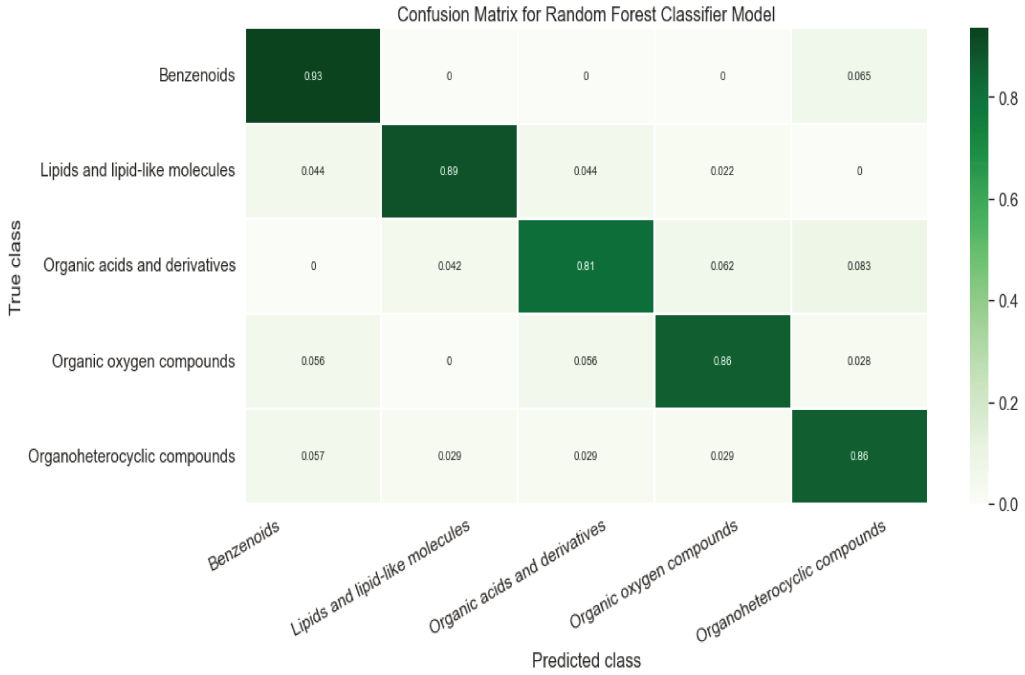
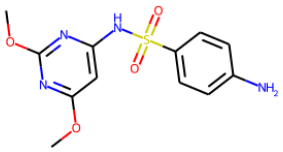


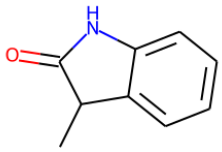
Figure 4. Confusion matrix of classification model.

Our model correctly predicted the super class of around 87% of the chemicals while around 8% of Organic acids and derivatives were classified as Organoheterocyclic compounds or Organic oxygen compounds. We noticed that errors frequently occurred in carboxylic acid compounds with phosphate esters or peptides. We randomly selected 3 incorrectly classified chemicals in each class. For instance, sulfadimethoxine (Figure 5a) was defined as Benzenoids due to an aniline. Nevertheless, it also contains pyrimidine, which was predicted as an Organoheterocyclic compound. Similarly, 3-Methyloxindole (Figure 5b) is an oxinole derivative consisting of a benzene ring and a heterocyclic with nitrogen. It was assigned to Organoheterocyclic (indole) in the collected dataset but went to Benzenoids compounds by prediction. We further investigated these incorrect

classifications by examining the feature importance, shown in Supplementary Figure S5. Figure 1 shows a possible substructure of the most relevant bit-strings. For example, bit 792 (Figure 1b) would define whether a compound is classified as a Benzenoid or Organoheterocyclic compound. On the other hand, the bit-string 842 (Figure 1c) was used to decide whether a chemical should go to Organic oxygen compounds. None of the bit-strings displayed significant importance from others, indicating that the "incorrect" classification mainly has to do with which functional groups were given the higher priority when the original training set was being compiled.



(**a**) Sulfadimethoxine



(**b**) 3-Methyloxindole

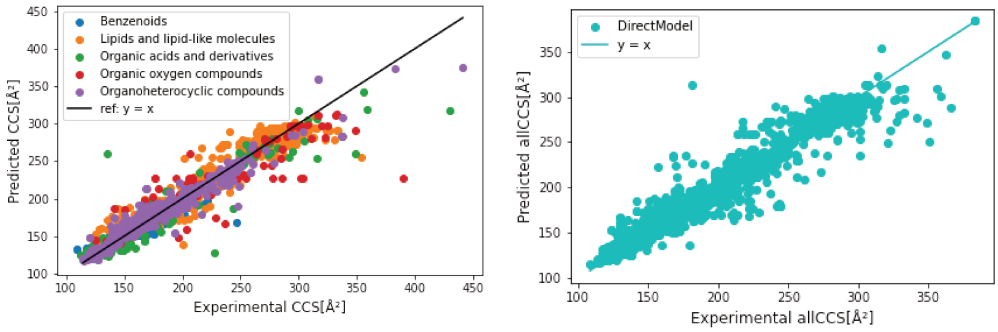
Figure 5. Random examples of "incorrect" predicted chemical.

3.3. Evaluation of Regression Models

In class-based modeling, the prediction R^2 was from 0.860 to 0.933, and the median relative error (MRE) of prediction was from 1.89% to 2.33% (Table 3). Direct CCS prediction, on the other hand, reached an R^2 of 0.95 and MRE of 2.2%, showing a good performance. Although we dropped replicated chemicals having the same CCS values before generating the modeling, considering that this dataset was merged by inter-laboratory studies, some chemicals might have been seen during training. Thus it can affect the prediction accuracy. Chemicals with less measurement deviation will increase the accuracy. On the contrary, those who have a significant deviation will bias prediction performance. We confirmed that for the direct prediction model, only 2% of the chemicals were common over 2665 test samples. The dataset was split by category in the class-based prediction, and the replications percentage was varied by chemical class. About 10% chemicals in the test set of Organic oxygen compounds were used in training before prediction, and less than 5% for other classes. Furthermore, except for a few outliers, the deviation of replications was under 6%. Therefore, we considered that the impact of replicated chemicals was negligible.

Additionally, we compared the performance of class-based models. Organic oxygen compound model obtained the lowest accuracy due to the lack of training data. Moreover, in its test split, the relative error $\geq$10% only occurred to macromolecules (e.g., maltodecaose ($C_{60}H_{102}O_{51}$)), contributing 15% to the test split, which resulted in poor prediction accuracy. Since we could not remeasure outliers' CCS values, we hypothesize that the error is associated with the compact and complex chemical structure. For instance, IMS measures the rotational-average surface of the maltodecaose ion. While a 1024 bit fingerprint is not enough to represent its complex chemical structure, resulting in a relative prediction error of 41.9% (true CCS at 390.3 while predicted 226.6 Å^2). Another possible reason can be the training weight. The dataset size of Organic oxygen compounds were almost 5 times less than Lipids and lipid-like molecules dataset, and glucose was the minority in the Organic oxygen compounds dataset. The model cannot properly generate the chemical rarely present during training. Therefore, higher accuracy was reached by Lipids and lipid-like molecules model and the direct prediction model. Outliers of other models were further investigated (shown in Supplementary Figure S6), and Figures 6 and 7a compare the predicted results of class-based models and direct prediction model. Four error cases have occurred to macromolecules (e.g., Diphenyl phosphate ($C_{39}H_{34}O_8P_2$)), which can be explained by the same hypothesis as maltodecaose (mentioned above). Metronidazole ($C_6H_9N_3O_3$) has 6 empirical CCS values measured with Waters TWIM, 5 were between 124 to 133 Å^2, while 200 Å^2 was measured by Picache et al. [60], leading a $-$61 Å^2 residual error (predicted CCS = 139.3 Å^2). L-tenuazonic acid ($C_{10}H_{15}NO_3$) was predicted to have a twice higher CCS than the measured one by the class-based model (35% higher by the direct prediction model). It might result from an inappropriate prediction by certain important

features. Predicted CCS of vinyl acetate ($C_4H_6O_2$) was 127.4 Å^2 through the class-based model, and 147.9 Å^2 by direct prediction, while the empirical one was 227.2 Å^2. We hypothesize that vinyl acetate might be polymerized leading to higher measured CCS values. Benefiting from datasets from multiple sources, class-based and direct prediction models can verify experimental CCS and evaluate the inter-laboratory and inter-platform deviation.



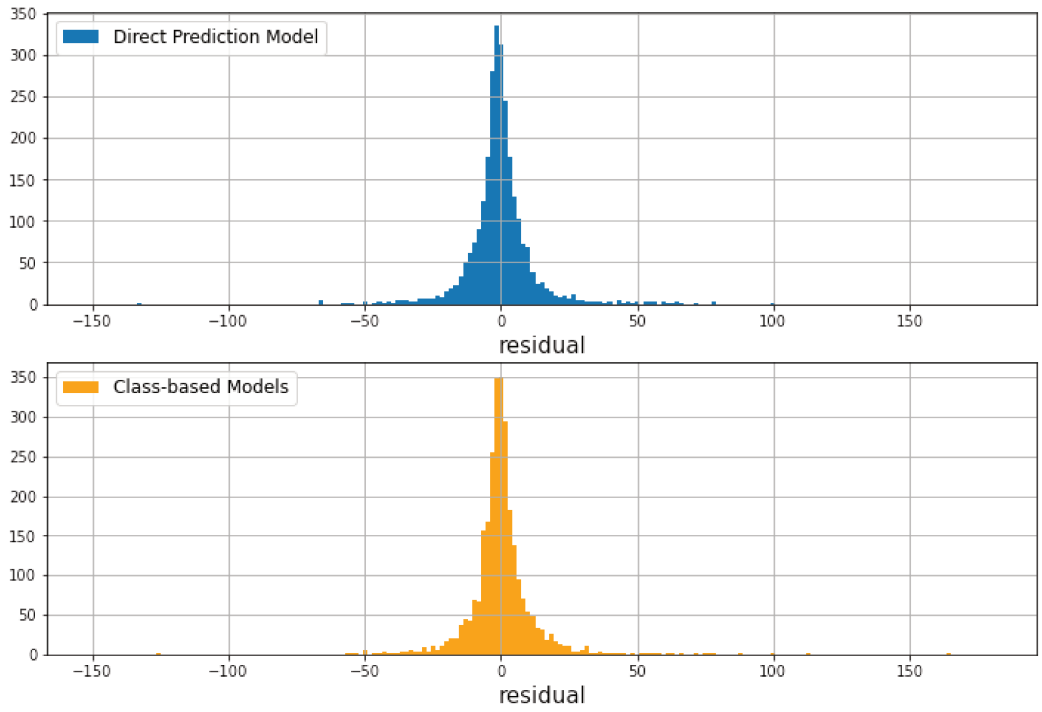
(**a**) Class-based Models (**b**) Direct prediction Model

Figure 6. Precision comparison between each predictive class and direct prediction without class. Class-based models lead to a better precision from 150 to 300 Å^2, while giving more bias by the small and macro molecules. The direct prediction model is less affected by the extreme cases.

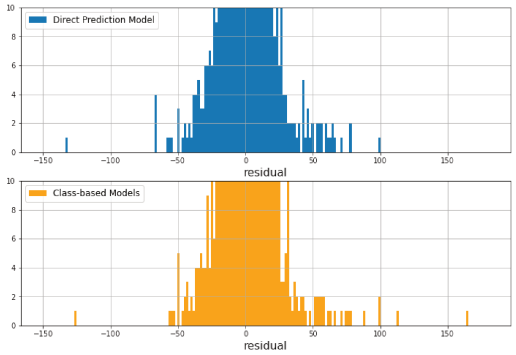
Figure 6 compares empirical and predicted CCS values generated by different models. We noticed that the direct prediction model was less biased by chemical class and structure, small and/or macro molecules, leading to higher prediction accuracy than the class-based prediction results. Although class-based models generated lower MREs (Table 3), a higher residual error was obtained in vinyl acetate and macro molecules resulting in lower R^2. As we can see in Figure 7c, over 25% of the test dataset obtained relative residual lower than 1%, and class-based models gained slightly higher, at 26.6%. All prediction models were further evaluated by feature importance (shown in Supplementary Figures S7 and S8). In both prediction approaches, the most relevant features divided chemicals into relative low CCS and high CCS. In other words, the decision tree was made of different CCS ranges based on certain substructures. For example, the most relevant feature in Organic acids and derivatives CCS prediction model was bit 588 (Figure 1d). If a chemical has its represented substructure, this chemical will be considered as CCS $>$ 150 Å^2, which might yield the prediction error for l-tenuazonic acid. Overall, the direct CCS model generated the best prediction performance, and a more extensive dataset ensured a more robust model.

MetCCS was a support vector regression (SVR) based on a prediction method only for metabolites. It achieved an excellent $R^2 > 0.96$ with the intra-laboratory and inter-laboratory measurements, relative residual was within 5% [45]. Bijlsma et al. [51] developed an artificial neural network (ANN)-based CCS prediction tool and [52] published an multivariate adaptive regression splines (MARS) CCS prediction model. Both were trained by TWIM data, and the relative error was within 6% for 95% of the chemicals. Belova et al. [68] compared experimental DTIM measured CCS values to predicted CCS values by the ANN-based and MARS-based predictors. A total of 95% of the protonated and deprotonated ions observed the relative error under 6.7%. However, only 56 compounds with 108 DTIM measured CCS values were compared in their study. We obtained comparable results by direct and class-based models, 87% of predicted results obtained the relative error within 7% (Figure 7c). DeepCCS is a more generalized CCS prediction model generated by SMILES with the deep neural network. R^2 was greater than 0.97, and MRE was below 2.6% [46]. However, only 1637 datasets were initially used to train the model, and the prediction power might be declined by chemical class [49,50]. We achieved a comparable accuracy for a wider scope of chemicals by direct prediction model (R^2 over 0.95, MRE within 2.2%). AllCCS and CCSbase generated better accuracy, with R^2 over 0.98 and MRE below 2%, since both tools used a larger and more diverse training set than DeepCCS and MetCCS.

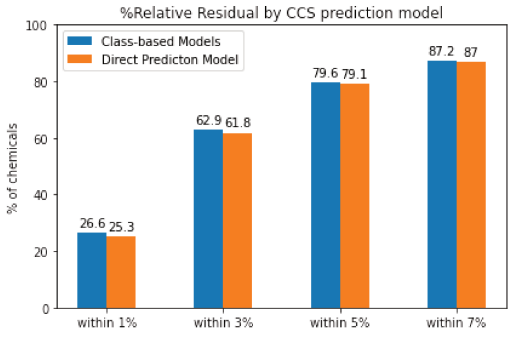
More structural-related features were emphasized in their studies. Considering our models, we reached comparable MREs with other tools and over 90% of the chemicals predicted within 8% relative residual. The results are satisfied with the CCS measurement bias via different IMS instrumentation and techniques [37].



(**a**) Residuals of class-based and direct model



(**b**) Zoom-in of (**a**)



(**c**) Comparison of relative error

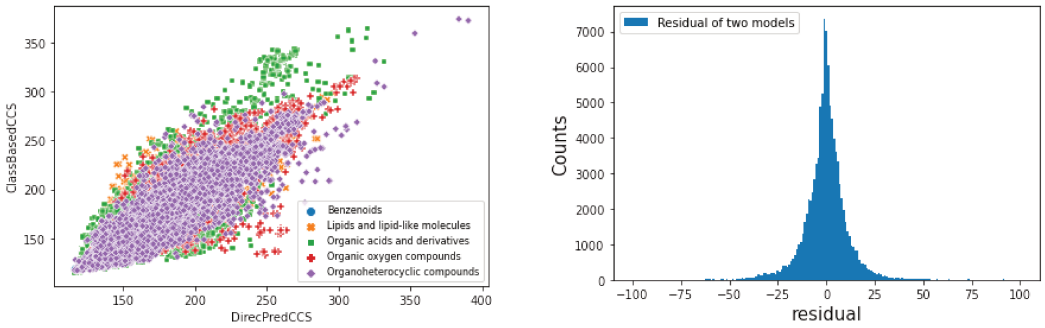
Figure 7. (**a**) compared the residuals of predicted CCS from class-based CCS prediction model and direct CCS prediction model. (**b**) is a zoomed-in of (**a**). Both approaches generate a good prediction power. A total of 98% of chemicals has a predicted difference within 25 Å^2. (**c**) Comparison of relative error in the testing set between the two approaches within 1%, 3%, 5%, and 7%.

3.4. Application on SusDat

NORMAN SusDat database contains over 111,000 environmentally relevant chemicals, with SMILES, accurate mass, and physiochemical properties [69]. We applied direct CCS prediction and class-based CCS prediction to the SusDat database, which contains chemicals that have never been seen during training and test, such as antibiotics and transformation products. A total of 96 % of the chemicals have a predicted difference within 25 Å^2 by two approaches (shown in Figure 8). The lack of true CCS values in SusDat, thus, by comparing the differences in predicted results generated by two approaches, demon-

strates the robustness of models, and the direct prediction model can discriminate different chemical classes.

Predicted CCS values are provided in Supplementary Table S3 for use in non-target screening or retrospective analysis. Moreover, these predicted CCS values can be compared to the measured CCS values by standard inter-laboratory evaluation and inter-platform deviation and improve the performances of our models.



(**a**) (**b**)

Figure 8. Comparison of direct and class-based CCS prediction model using Norman Susdat. (**a**) Scatter plot of class-based predicted CCS value against direct predicted CCS value. (**b**) Difference of predicted CCS values between class-based and direct prediction models. A total of 96 % of the chemicals have a predicted difference within 25 Å^2.

4. Discussion

In this study, we introduced topological fingerprints to categorize chemicals and generate CCS prediction models using the random forest algorithm. Our methods are generalized to TWIM and DTIM measured CCS data collected from seven sources. Prediction models were developed for five super classes of chemicals (Benzenoids, Lipids and lipid-like molecules, Organic acids and derivatives, Organic oxygen compounds, and Organoheterocyclic compounds) and the entire dataset. The test prediction accuracy was 0.958 by the direct prediction approach, 3 class-based prediction models more than 0.9, and over 0.86 for the remaining two classes. The MRE was between 1.89% to 2.33%. Additionally, models only required SMILES to encode fingerprints. A significant predicted variation was observed in macro molecules and vinyl acetate, with over 100 Å^2 residual. We noticed that the residuals were reduced through the direct prediction model due to an extensive training set and a higher presence of macro molecules in the dataset. The prediction performances are highly dependent on the collected CCS libraries. Therefore, it is emphasized that multiple and accurate empirical CCS libraries with a broad scope of chemicals are crucial to CCS machine learning studies. Moreover, this bias indicated a limited prediction performance for chemicals with unique structures. A better classification model or other structural importance features might improve the prediction accuracy. Since fingerprint was the only input feature for prediction, adduct ions (e.g., $[M + Na]^+$) were eliminated in this study. Other features can be introduced in the models to generate more ion types. Moreover, fingerprints offer a novel aspect in CCS prediction using machine learning. The generated feature importance of 1024 bits was directly related to the structures and thus easier to interpret chemically.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/molecules27196424/s1, Table S1: SuperClassModeling.csv; Table S2: CCSPredictedData.csv; Table S3: SusDatCCSprediction.csv; Decision trees files: DecisionTrees.zip. Figure S1: RSD of replicated chemicals; Figure S2: Classification GridSearchCV scores; Figure S3: Distribution of predicted super classes; Figure S4: Regression modeling scores of hyper-parameters optimization by GridSearchCV; Figure S5: Feature Importance of classification model; Figure S6: Outliers with predicted Class, predicted CCS by Class-based and direct models; Figure S7: Regression modeling Feature Importances; Figure S8: Example of most relevant features.

Author Contributions: F.Y., D.v.H. and S.S. designed the study. F.Y. performed the experiments and wrote the original draft. H.P. reviewed the manuscript. S.S. edited, reviewed the manuscript and supervised this project. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by SETASAR PhD Project from Région Nouvelle Aquitaine and LPL. The exchange project was support by E2S UPPA.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets and the source codes can be found at https://github.com/fyang22/CCS-Prediction-Publish (accessed on 10 April 2022) and https://www.mdpi.com/article/10.3390/molecules27196424/s1.

Acknowledgments: The authors thank Région Nouvelle Aquitaine and LPL for their financial support of SETASAR PhD Project, E2S UPPA for the exchange grant in the University of Amsterdam. S. Samanipour is grateful to UvA Data Science center and ChemistryNL for their financial support, projects EDIFIED and SCOPE.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Muir, D.C.; Howard, P.H. Are there other persistent organic pollutants? A challenge for environmental chemists. *Environ. Sci. Technol.* **2006**, *40*, 7157–7166. [CrossRef] [PubMed]
2. Howard, P.H.; Muir, D.C. Identifying new persistent and bioaccumulative organics among chemicals in commerce II: Pharmaceuticals. *Environ. Sci. Technol.* **2011**, *45*, 6938–6946. [CrossRef] [PubMed]
3. Escher, B.I.; Stapleton, H.M.; Schymanski, E.L. Tracking complex mixtures of chemicals in our changing environment. *Science* **2020**, *367*, 388–392. [CrossRef] [PubMed]
4. Newton, S.R.; McMahen, R.L.; Sobus, J.R.; Mansouri, K.; Williams, A.J.; McEachran, A.D.; Strynar, M.J. Suspect screening and non-targeted analysis of drinking water using point-of-use filters. *Environ. Pollut.* **2018**, *234*, 297–306. [CrossRef] [PubMed]
5. Shi, Q.; Xiong, Y.; Kaur, P.; Sy, N.D.; Gan, J. Contaminants of emerging concerns in recycled water: Fate and risks in agroecosystems. *Sci. Total Environ.* **2022**, *814*, 152527. [CrossRef] [PubMed]
6. Rizzo, L.; Gernjak, W.; Krzeminski, P.; Malato, S.; McArdell, C.S.; Perez, J.A.S.; Schaar, H.; Fatta-Kassinos, D. Best available technologies and treatment trains to address current challenges in urban wastewater reuse for irrigation of crops in EU countries. *Sci. Total Environ.* **2020**, *710*, 136312. [CrossRef]
7. Manaia, C.M. Assessing the risk of antibiotic resistance transmission from the environment to humans: Non-direct proportionality between abundance and risk. *Trends Microbiol.* **2017**, *25*, 173–181. [CrossRef]
8. López-Pacheco, I.Y.; Silva-Núñez, A.; Salinas-Salazar, C.; Arévalo-Gallegos, A.; Lizarazo-Holguin, L.A.; Barceló, D.; Iqbal, H.M.; Parra-Saldívar, R. Anthropogenic contaminants of high concern: Existence in water resources and their adverse effects. *Sci. Total Environ.* **2019**, *690*, 1068–1088. [CrossRef]
9. Ma, Y.; He, X.; Qi, K.; Wang, T.; Qi, Y.; Cui, L.; Wang, F.; Song, M. Effects of environmental contaminants on fertility and reproductive health. *J. Environ. Sci.* **2019**, *77*, 210–217. [CrossRef]
10. 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. 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.* **2018**, *52*, 5135–5144. [CrossRef]
11. Pedrazzani, R.; Bertanza, G.; Brnardić, I.; Cetecioglu, Z.; Dries, J.; Dvarionienė, J.; García-Fernández, A.J.; Langenhoff, A.; Libralato, G.; Lofrano, G.; et al. Opinion paper about organic trace pollutants in wastewater: Toxicity assessment in a European perspective. *Sci. Total Environ.* **2019**, *651*, 3202–3221. [CrossRef] [PubMed]
12. Rueda-Ruzafa, L.; Cruz, F.; Roman, P.; Cardona, D. Gut microbiota and neurological effects of glyphosate. *Neurotoxicology* **2019**, *75*, 1–8. [CrossRef] [PubMed]
13. Lohmann, R.; Breivik, K.; Dachs, J.; Muir, D. Global fate of POPs: Current and future research directions. *Environ. Pollut.* **2007**, *150*, 150–165. [CrossRef] [PubMed]
14. Samanipour, S.; Martin, J.W.; Lamoree, M.H.; Reid, M.J.; Thomas, K.V. Optimism for nontarget analysis in environmental chemistry. *Environ. Sci. Technol.* **2019**, *53*, 5529–5530. [CrossRef] [PubMed]
15. Vermeulen, R.; Schymanski, E.L.; Barabási, A.L.; Miller, G.W. The exposome and health: Where chemistry meets biology. *Science* **2020**, *367*, 392–396. [CrossRef]
16. Schymanski, E.L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.P.; Hollender, J. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098. [CrossRef]
17. Schulze, B.; Jeon, Y.; Kaserzon, S.; Heffernan, A.L.; Dewapriya, P.; O'Brien, J.; Ramos, M.J.G.; Gorji, S.G.; Mueller, J.F.; Thomas, K.V.; et al. An assessment of quality assurance/quality control efforts in high resolution mass spectrometry non-target workflows for analysis of environmental samples. *TrAC Trends Anal. Chem.* **2020**, *133*, 116063. [CrossRef]

18. Pérez-Lemus, N.; López-Serna, R.; Pérez-Elvira, S.I.; Barrado, E. Analytical methodologies for the determination of pharmaceuticals and personal care products (PPCPs) in sewage sludge: A critical review. *Anal. Chim. Acta* **2019**, *1083*, 19–40. [CrossRef]
19. Hollender, J.; Schymanski, E.L.; Singer, H.P.; Ferguson, P.L. Nontarget screening with high resolution mass spectrometry in the environment: Ready to go? *Environ. Sci. Technol.* **2017**, *51*, 11505–11512. [CrossRef]
20. Guo, Z.; Huang, S.; Wang, J.; Feng, Y.L. Recent advances in non-targeted screening analysis using liquid chromatography—High resolution mass spectrometry to explore new biomarkers for human exposure. *Talanta* **2020**, *219*, 121339. [CrossRef]
21. Hollender, J.; Van Bavel, B.; Dulio, V.; Farmen, E.; Furtmann, K.; Koschorreck, J.; Kunkel, U.; Krauss, M.; Munthe, J.; Schlabach, M.; et al. High resolution mass spectrometry-based non-target screening can support regulatory environmental monitoring and chemicals management. *Environ. Sci. Eur.* **2019**, *31*, 42. [CrossRef]
22. Knolhoff, A.M.; Callahan, J.H.; Croley, T.R. Mass accuracy and isotopic abundance measurements for HR-MS instrumentation: Capabilities for non-targeted analyses. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1285–1294. [CrossRef] [PubMed]
23. Hernandez, F.; Sancho, J.V.; Ibáñez, M.; Abad, E.; Portolés, T.; Mattioli, L. Current use of high-resolution mass spectrometry in the environmental sciences. *Anal. Bioanal. Chem.* **2012**, *403*, 1251–1264. [CrossRef] [PubMed]
24. Kaufmann, A. The current role of high-resolution mass spectrometry in food analysis. *Anal. Bioanal. Chem.* **2012**, *403*, 1233–1249. [CrossRef] [PubMed]
25. Knolhoff, A.M.; Croley, T.R. Non-targeted screening approaches for contaminants and adulterants in food using liquid chromatography hyphenated to high resolution mass spectrometry. *J. Chromatogr. A* **2016**, *1428*, 86–96. [CrossRef] [PubMed]
26. Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinform.* **2006**, *7*, 234. [CrossRef]
27. d'Atri, V.; Causon, T.; Hernandez-Alba, O.; Mutabazi, A.; Veuthey, J.L.; Cianferani, S.; Guillarme, D. Adding a new separation dimension to MS and LC–MS: What is the utility of ion mobility spectrometry? *J. Sep. Sci.* **2018**, *41*, 20–67. [CrossRef]
28. Boelrijk, J.; van Herwerden, D.; Ensing, B.; Forré, P.; and Samanipour, S. Predicting RP-LC retention indices of structurally unknown chemicals from mass spectrometry data. *ChemRxiv* **2022**. [CrossRef]
29. Celma, A.; Ahrens, L.; Gago-Ferrero, P.; Hernández, F.; López, F.; Lundqvist, J.; Pitarch, E.; Sancho, J.V.; Wiberg, K.; Bijlsma, L. The relevant role of ion mobility separation in LC-HRMS based screening strategies for contaminants of emerging concern in the aquatic environment. *Chemosphere* **2021**, *280*, 130799. [CrossRef]
30. Mairinger, T.; Causon, T.J.; Hann, S. The potential of ion mobility–mass spectrometry for non-targeted metabolomics. *Curr. Opin. Chem. Biol.* **2018**, *42*, 9–15. [CrossRef]
31. Goscinny, S.; Joly, L.; De Pauw, E.; Hanot, V.; Eppe, G. Travelling-wave ion mobility time-of-flight mass spectrometry as an alternative strategy for screening of multi-class pesticides in fruits and vegetables. *J. Chromatogr. A* **2015**, *1405*, 85–93. [CrossRef] [PubMed]
32. Hill, H.H., Jr.; Siems, W.F.; St. Louis, R.H. Ion mobility spectrometry. *Anal. Chem.* **1990**, *62*, 1201A–1209A. [CrossRef] [PubMed]
33. Borsdorf, H.; Eiceman, G.A. Ion mobility spectrometry: Principles and applications. *Appl. Spectrosc. Rev.* **2006**, *41*, 323–375. [CrossRef]
34. Eiceman, G.A.; Karpas, Z. *Ion Mobility Spectrometry*; CRC Press: Boca Raton, FL, USA, 2005.
35. Hernández-Mesa, M.; D'atri, V.; Barknowitz, G.; Fanuel, M.; Pezzatti, J.; Dreolin, N.; Ropartz, D.; Monteau, F.; Vigneau, E.; Rudaz, S.; et al. Interlaboratory and interplatform study of steroids collision cross section by traveling wave ion mobility spectrometry. *Anal. Chem.* **2020**, *92*, 5013–5022. [CrossRef]
36. Stow, S.M.; Causon, T.J.; Zheng, X.; Kurulugama, R.T.; Mairinger, T.; May, J.C.; Rennie, E.E.; Baker, E.S.; Smith, R.D.; McLean, J.A.; et al. An interlaboratory evaluation of drift tube ion mobility–mass spectrometry collision cross section measurements. *Anal. Chem.* **2017**, *89*, 9048–9055. [CrossRef] [PubMed]
37. Hinnenkamp, V.; Klein, J.; Meckelmann, S.W.; Balsaa, P.; Schmidt, T.C.; Schmitz, O.J. Comparison of CCS values determined by traveling wave ion mobility mass spectrometry and drift tube ion mobility mass spectrometry. *Anal. Chem.* **2018**, *90*, 12042–12050. [CrossRef] [PubMed]
38. Feuerstein, M.L.; Hernández-Mesa, M.; Kiehne, A.; Le Bizec, B.; Hann, S.; Dervilly, G.; Causon, T. Comparability of Steroid Collision Cross Sections Using Three Different IM-HRMS Technologies: An Interplatform Study. *J. Am. Soc. Mass Spectrom.* **2022**. [CrossRef]
39. Borsdorf, H.; Mayer, T.; Zarejousheghani, M.; Eiceman, G.A. Recent developments in ion mobility spectrometry. *Appl. Spectrosc. Rev.* **2011**, *46*, 472–521. [CrossRef]
40. Celma, A.; Sancho, J.V.; Schymanski, E.L.; Fabregat-Safont, D.; Ibanez, M.; Goshawk, J.; Barknowitz, G.; Hernandez, F.; Bijlsma, L. Improving target and suspect screening high-resolution mass spectrometry workflows in environmental analysis by ion mobility separation. *Environ. Sci. Technol.* **2020**, *54*, 15120–15131. [CrossRef]
41. Menger, F.; Celma, A.; Schymanski, E.L.; Lai, F.Y.; Bijlsma, L.; Wiberg, K.; Hernández, F.; Sancho, J.V.; Lutz, A. Enhancing Spectral Quality in Complex Environmental Matrices: Supporting Suspect and Non-Target Screening in Zebra Mussels with Ion Mobility. *SSRN Electron. J.* **2022**. [CrossRef]
42. Izquierdo-Sandoval, D.; Fabregat-Safont, D.; Lacalle-Bergeron, L.; Sancho, J.V.; Hernández, F.; Portoles, T. Benefits of Ion Mobility Separation in GC-APCI-HRMS Screening: From the Construction of a CCS Library to the Application to Real-World Samples. *Anal. Chem.* **2022**, *94*, 9040–9047. [CrossRef] [PubMed]

43. Gabelica, V.; Marklund, E. Fundamentals of ion mobility spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 51–59. [CrossRef] [PubMed]
44. Ross, D.H.; Xu, L. Determination of drugs and drug metabolites by ion mobility-mass spectrometry: A review. *Anal. Chim. Acta* **2021**, *1154*, 338270. [CrossRef] [PubMed]
45. Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z.J. Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 11084–11091. [CrossRef] [PubMed]
46. Plante, P.L.; Francovic-Fontaine, É.; May, J.C.; McLean, J.A.; Baker, E.S.; Laviolette, F.; Marchand, M.; Corbeil, J. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. *Anal. Chem.* **2019**, *91*, 5191–5199. [CrossRef]
47. Zhou, Z.; Tu, J.; Zhu, Z.J. Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. *Curr. Opin. Chem. Biol.* **2018**, *42*, 34–41. [CrossRef]
48. Mollerup, C.B.; Mardal, M.; Dalsgaard, P.W.; Linnet, K.; Barron, L.P. Prediction of collision cross section and retention time for broad scope screening in gradient reversed-phase liquid chromatography-ion mobility-high resolution accurate mass spectrometry. *J. Chromatogr. A* **2018**, *1542*, 82–88. [CrossRef]
49. Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z.J. Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nat. Commun.* **2020**, *11*, 4334. [CrossRef]
50. Ross, D.H.; Cho, J.H.; Xu, L. Breaking down structural diversity for comprehensive prediction of ion-neutral collision cross sections. *Anal. Chem.* **2020**, *92*, 4548–4557. [CrossRef]
51. Bijlsma, L.; Bade, R.; Celma, A.; Mullin, L.; Cleland, G.; Stead, S.; Hernandez, F.; Sancho, J.V. Prediction of collision cross-section values for small molecules: Application to pesticide residue analysis. *Anal. Chem.* **2017**, *89*, 6583–6589. [CrossRef]
52. Celma, A.; Bade, R.; Sancho, J.V.; Hernández, F.; Humpries, M.; Bijslma, L. Prediction of Retention Time and Collision Cross Section (CCSH+, CCSH-and CCSNa+) of Emerging Contaminants Using Multiple Adaptive Regression Splines. 2022. Available online: https://doi.org/10.21203/rs.3.rs-1249834/v1 (accessed on 13 January 2022).
53. Cereto-Massagué, A.; Ojeda, M.J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular fingerprint similarity search in virtual screening. *Methods* **2015**, *71*, 58–63. [CrossRef] [PubMed]
54. Swain, M. PubChemPy: A Way to Interact with PubChem in Python. 2014. Available online: https://pubchempy.readthedocs.io/en/latest/ (accessed on 13 January 2022).
55. Landrum, G. RDKit: Open-Source Cheminformatics. 2006. Available online: https://doi.org/10.5281/zenodo.3732262 (accessed on 13 January 2022).
56. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. [CrossRef]
57. Weininger, D.; Weininger, A.; Weininger, J.L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97–101. [CrossRef]
58. Nilakantan, R.; Bauman, N.; Dixon, J.S.; Venkataraghavan, R. Topological torsion: A new molecular descriptor for SAR applications. Comparison with other descriptors. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 82–85. [CrossRef]
59. Capecchi, A.; Probst, D.; Reymond, J.L. One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome. *J. Cheminformatics* **2020**, *12*, 1–15. [CrossRef]
60. Picache, J.A.; Rose, B.S.; Balinski, A.; Leaptrot, K.L.; Sherrod, S.D.; May, J.C.; McLean, J.A. Collision cross section compendium to annotate and predict multi-omic compound identities. *Chem. Sci.* **2019**, *10*, 983–993. [CrossRef]
61. Zheng, X.; Aly, A.N.; Zhou, Y.; Dupuis, K.T.; Bilbao, A.; Paurus, V.L.; Orton, D.J.; Wilson, R.; Payne, S.H.; Smith, R.D.; et al. A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. *Chem. Sci.* **2017**, *8*, 7724–7736. [CrossRef]
62. Zheng, X.; Dupuis, K.T.; Aly, N.A.; Zhou, Y.; Smith, F.B.; Tang, K.; Smith, R.D.; Baker, E.S. Utilizing ion mobility spectrometry and mass spectrometry for the analysis of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, polybrominated diphenyl ethers and their metabolites. *Anal. Chim. Acta* **2018**, *1037*, 265–273. [CrossRef]
63. Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **2015**, *7*, 20. [CrossRef]
64. Hines, K.M.; Ross, D.H.; Davidson, K.L.; Bush, M.F.; Xu, L. Large-Scale Structural Characterization of Drug and Drug-Like Compounds by High-Throughput Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2017**, *89*, 9023–9030. [CrossRef]
65. Belova, L.; Caballero-Casero, N.; van Nuijs, A.L.N.; Covaci, A. Ion Mobility-High-Resolution Mass Spectrometry (IM-HRMS) for the Analysis of Contaminants of Emerging Concern (CECs): Database Compilation and Application to Urine Samples. *Anal. Chem.* **2021**, *93*, 6428–6436. [CrossRef] [PubMed]
66. Schymanski, E.; Zhang, J.; Thiessen, P.; Bolton, E. Experimental CCS Values in Pubchem. *Zenodo* **2022**. [CrossRef]
67. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J.C.; Sheridan, R.P.; Feuston, B.P. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947–1958. [CrossRef] [PubMed]

68. Belova, L.; Celma, A.; Van Haesendonck, G.; Lemière, F.; Sancho, J.V.; Covaci, A.; van Nuijs, A.L.; Bijlsma, L. Revealing the differences in collision cross section values of small organic molecules acquired by different instrumental designs and prediction models. *Anal. Chim. Acta* **2022**, *1229*, 340361. [CrossRef] [PubMed]
69. Dulio, V.; Koschorreck, J.; Van Bavel, B.; Van den Brink, P.; Hollender, J.; Munthe, J.; Schlabach, M.; Aalizadeh, R.; Agerstrand, M.; Ahrens, L.; et al. The NORMAN association and the European partnership for chemicals risk assessment (PARC): Let's cooperate! *Environ. Sci. Eur.* **2020**, *32*, 100. [CrossRef]

References

May, J. C.; Morris, C. B.; McLean, J. A. Ion mobility collision cross section compendium. *Analytical chemistry* **2017**, *89*, 1032–1044.

Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A. Collision cross section compendium to annotate and predict multiomic compound identities. *Chemical science* **2019**, *10*, 983–993.

Schymanski, E.; Zhang, J.; Thiessen, P.; Bolton, E. Experimental CCS Values in PubChem. 2022; https://doi.org/10.5281/zenodo.7056297.

Dilger, J. M.; Valentine, S. J.; Glover, M. S.; Clemmer, D. E. A database of alkaline-earth-coordinated peptide cross sections: insight into general aspects of structure. *Journal of the American Society for Mass Spectrometry* **2013**, *24*, 768–779.

Meier, F.; Köhler, N. D.; Brunner, A.-D.; Wanka, J.-M. H.; Voytik, E.; Strauss, M. T.; Theis, F. J.; Mann, M. Deep learning the collisional cross sections of the peptide universe from a million experimental values. *Nature communications* **2021**, *12*, 1–12.

Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z.-J. Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nature communications* **2020**, *11*, 1–13.

Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z.-J. Large-scale prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry. *Analytical chemistry* **2016**, *88*, 11084–11091.

Wishart, D. S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L., *et al.* HMDB 5.0: the human metabolome database for 2022. *Nucleic Acids Research* **2022**, *50*, D622–D631.

Dodds, J. N.; Baker, E. S. Ion mobility spectrometry: fundamental concepts, instrumentation, applications, and the road ahead. *Journal of the American Society for Mass Spectrometry* **2019**, *30*, 2185–2195.

Celma, A.; Sancho, J. V.; Schymanski, E. L.; Fabregat-Safont, D.; Ibanez, M.; Goshawk, J.; Barknowitz, G.; Hernandez, F.; Bijlsma, L. Improving target and suspect screening high-resolution mass spectrometry workflows in environmental analysis by ion mobility separation. *Environmental Science & Technology* **2020**, *54*, 15120–15131.

Hinnenkamp, V.; Balsaa, P.; Schmidt, T. C. Target, suspect and non-target screening analysis from wastewater treatment plant effluents to drinking water using collision cross section values as additional identification criterion. *Analytical and bioanalytical chemistry* **2022**, *414*, 425–438.

Celma, A.; Ahrens, L.; Gago-Ferrero, P.; Hernández, F.; López, F.; Lundqvist, J.; Pitarch, E.; Sancho, J. V.; Wiberg, K.; Bijlsma, L. The relevant role of ion mobility separation in LC-HRMS based screening strategies for contaminants of emerging concern in the aquatic environment. *Chemosphere* **2021**, *280*, 130799.

Stephan, S.; Hippler, J.; Köhler, T.; Deeb, A. A.; Schmidt, T. C.; Schmitz, O. J. Contaminant screening of wastewater with HPLC-IM-qTOF-MS and LC+ LC-IM-qTOF-MS using a CCS database. *Analytical and bioanalytical chemistry* **2016**, *408*, 6545–6555.

Ross, D. H.; Cho, J. H.; Xu, L. Breaking down structural diversity for comprehensive prediction of ion-neutral collision cross sections. *Analytical chemistry* **2020**, *92*, 4548–4557.

Heinonen, M.; Shen, H.; Zamboni, N.; Rousu, J. Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics* **2012**, *28*, 2333–2341.

Zang, Q.; Mansouri, K.; Williams, A. J.; Judson, R. S.; Allen, D. G.; Casey, W. M.; Kleinstreuer, N. C. In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. *Journal of chemical information and modeling* **2017**, *57*, 36–49.

Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* **1988**, *28*, 31–36.

Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *Journal of chemical information and modeling* **2010**, *50*, 742–754.

Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular fingerprint similarity search in virtual screening. *Methods* **2015**, *71*, 58–63.

Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *Journal of Chemical Information and Computer Sciences* **1985**, *25*, 64–73.

Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *Journal of Chemical Information and Computer Sciences* **1987**, *27*, 82–85.

Capecchi, A.; Probst, D.; Reymond, J.-L. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *Journal of cheminformatics* **2020**, *12*, 1–15.

Chapter 6

Conclusion and perspectives

Non-target analysis is a promising strategy that can be used for a broad scope of chemicals, such as "known-unknown" and "unknown-unknown". The development of non-target screening methods is challenging at different points of the analytical workflow. Additionally, the benefit of advancements in analytical instrumentation and "cheminformatics" tools, high-throughput data acquisition and "real time" data treatment strategies accelerate the detection of contaminants in the environment. The present PhD project is aimed at addressing analytical challenges and establishing a fundamental workflow for different purposes.

Bibliographic searches were essential to understand the concepts of exposomes, exposure assessment, guidelines and regulations of environmental contaminants and, more importantly, the call for emerging contaminants. High-resolution mass spectrometry has been developed and flourished in analytical chemistry. Understanding the principal and main applications of different types of high-resolution mass spectrometers helps us to optimize the instrument used in the present PhD project. Furthermore, the data processing workflow plays an important role in this PhD project. The discovery of different data processing tools enables a more efficient workflow of data treatment. Furthermore, the enhancement of ion mobility spectrometry coupled with HRMS in the area of small molecules and emerging contaminants was a proof-of-concept regarding the bibliographic searches and application in this project.

The first method was developed with GC-APCI-IMS-HRMS. The commercial availability of GC-APCI-MS opens a door for the nontarget analysis of GC-amenable compounds. Halogenated POPs, including PCBs and PBDEs, were investigated in this method. The large numbers of PCB and PBDE congeners, along with the natural and special isotopic profiles of Cl and Br, are ideal chemicals to optimize our analytical approach. Moreover, an in-house database was built with halogenated POPs. APCI predominantly produces (quasi) molecular ions, enabling the structural elucidation of "unknown-unknown" chemicals. Meanwhile, IMS improved isomeric separation and mass spectrum quality. The CCS value gives a new aspect to associate the precursor ion to its fragmentation, facilitating mass spectrum interpretation. Hexachlorobiphenyl, such as PCB-149, was frequently detected in samples, and false positives were removed after manual inspection and the consideration of CCS values.

The second method was focused on nontarget analysis in water samples. A generic UHPLC-HRMS method was optimized, and an in-house database was created for target analysis. Direct injection with a simple filtration and dilution was tested in different types of water samples. Then, it can be significantly demonstrated that it reduces the sample preparation procedure, which is suitable in wastewater analysis. However, it showed a limitation in detecting trace-level chemicals in tap water and surface water. Furthermore, the data treatment workflow and tools were explored in this project. An automated target analysis workflow was defined in TASQ, and nontarget analysis was processed in MS-DIAL using MassBank Europe. Data analysis was used to verify the results. Visual inspections are vital to prevent false positives. In total, 65 chemicals were identified with the minimum L3. Banned pesticides, such as simazine and metolachlor, and their successors or TPs deisopropylatrazine were also detected in wastewater samples, indicating the high persistence and mobility of these emerging contaminants. Contaminants related to diverse sources, such as personal care products, pharmaceutical products, plasticizers, food additives and

industrial products, were commonly found in different water samples. In the next step, a separation method for ultra-polar chemicals can be investigated to expand the detection range. IMS and CCS can be implemented in this method to increase the identification confidence point.

The third method was developed for simultaneous steroid hormone analysis by isotopic dilution, especially cortisol quantification in juvenile fish. However, the method development and validation are still in the early stage. It was proven in standard solutions that CCS filtering can eliminate background noise in low concentrations and real complex biological matrices. It can be a privilege (key advantages) in single individual juvenile fish analysis.

In the last project, a CCS prediction tool was developed using machine learning and molecular fingerprinting. The dataset was collected from different studies using different IMS techniques that aim to cover different classes of chemicals and instruments. Two prediction approaches, class-based prediction models and direct prediction models, were developed. Both approaches provide good prediction accuracies. The prediction deviation was estimated in the MRE to be between 1.89% and 2.33%. Larger deviations were observed for macromolecules and for very small molecules. As is often the case for machine learning approaches, more learning cases can be added to enhance the prediction performance/accuracy.

Different projects were developed and discussed in this manuscript, presenting the advancements and challenges in non-target analysis workflows under different points of view. During this research work, the main topic/scientific question was to evaluate the advancements of analytical instruments, especially the benefit of IMS and its synergy in data treatment strategies for non-target analysis. Generic analytical methods have been achieved in GC-APCI(+)-IMS-HRMS and LC-ESI(+)-IMS-HRMS. The automatic target identification workflow and in-house database were built in TASQ. The GC database contained the RT, accurate mass, isotopic profiles, and experimental CCS values with 118 halogenated POPs. The LC database contained the RT, accurate mass, MS/MS (DIA mode), isotopic profile with 559 chemicals, together with the Bruker commercial water contaminants database, and over 3000 common found environmental chemicals are in the list. The CCS value will be complemented in the database with experimental and predicted values for compound identification. On the other hand, the methods and strategies presented in this manuscript, which are all first presented here, still need further improvements despite the demonstration of the feasibility and their relative benefits in global environment and water monitoring. The methods developed in the present PhD research only focuses on the positive ionization mode. Emerging contaminants, such as PFAS, haloacetic acids, and bisphenol A (BPA), have not yet been fully studied. Moreover, the IMS and CCS databases have not been applied in real clean water analysis. Samples acquired through the GC analysis method were only processed for target screening, and non-target analysis can be performed with the same raw data afterward using MS-DIAL and similar workflows as discussed in water analysis. Overall, two promising target and non-target analysis approaches and, in parallel, a machine learning CCS predictor were accomplished during this PhD research. IMS have demonstrated several advantages in non-target analysis. Structure elucidation remains a time-consuming task in non-target analyses, while IMS can improve the mass spectra interpretation. Experimental and predicted CCS values can reduce

the number of candidates and increase the candidate identification confidence. More sample applications can be assessed for the strategies optimized and presented for this research work in the case of any GC/LC-IMS-HRMS screening/monitoring.

Appendices

Appendix 1: Identified candidates by target and non-target screening in wastewater. It includes the identification approaches (TASQ and MS-DIAL), the identification levels.

Compound	RT [min]	Formula	m/z [Da]	Identification approach	Identification level
1-Methylnictinamide	4.68	C7H9N2O	138	MS-DIAL	L3
2,6-Dichlorobenzamide	7.38	C7H5Cl2NO	189.98	MS-DIAL	L2
2,6-Xylidine	3.81	C8H11N	122.096	MS-DIAL	L3
2-Aminophenol	1.85	C6H7NO	110.0587	MS-DIAL	L3
2.4 Dimethylphenyl-N-methylformamidine	3.75	C10H14N2	163.1228	TASQ/MS-DAIL	L3
2'-Deoxyadenosine	3.28	C10H13N5O3	252.1091	MS-DIAL	L2
2-Naphthylamine	4.85	C10H9N	327.0081	MS-DIAL	L2
2-Oxindole	4.54	C8H7NO	134.0713	MS-DIAL	L2
4-Hexyloxyaniline	6.7	C12H19NO	194.15	MS-DIAL	L3
4-Methylbenzotriazole	4.88	C7H7N3	134.07	MS-DIAL	L2
4-Pyridoxate	3	C8H9NO4	184.0618	MS-DIAL	L2
5-Methylcytosine	3	C5H7N3O	126.067	MS-DIAL	L2
6-Methoxyquinoline	4.1	C10H9NO	160.076	MS-DIAL	L2
Adenine	2.22	C5H5N5	136.0618	TASQ/MS-DAIL	L2
Adenosine	3.16	C10H13N5O4	268.0032	TASQ/MS-DAIL	L2
Allopurinol	2.33	C5H4N4O	137.045	MS-DIAL	L2
Aminocarb	3.86	C11H16N2O2	209.1285	TASQ	L1
Azoxystrobin	6.27	C22H17N3O5	404.1267	MS-DIAL	L2
Benzotriazole	4.2	C6H5N3	120.05	MS-DIAL	L3
Bis(2-ethylhexyl)phthalate	12.94	C24H38O4	391.2856	MS-DIAL	L2
Bupivacaine	4.43	C18H28N2O	289.0546	TASQ	L2
Caffeine	2.9	C8H10N4O2	195.088	MS-DIAL	L2
Chloridazon-desphenyl-methyl	2.34	C5H6ClN3O	160.0273	TASQ	L2
Chlorotoluron	7.06	C10H13ClN2O	213.0798	TASQ/MS-DIAL	L2
Coniine	5.11	C8H17N	128.1438	MS-DIAL	L3
Cotinine	3.97	C10H12N2O	177.0999	TASQ	L2
Cytidine	2.039	C9H13N3O5	244.0943	MS-DIAL	L3
Deethylatrazine	3.56	C6H10ClN5	188.0706	MS-DIAL	L2
Deisopropylatrazine	6.36	C5H8ClN5	174.05	MS-DIAL	L3

Compound	RT [min]	Formula	m/z [Da]	Identification approach	Identification level
Diethyl-phathalate	9.08	C12H14O4	223.0973	MS-DIAL	L3
Diltiazem	5.58	C22H26N2O4S	415.167	MS-DIAL	L3
Dimethenamid	7.12	C12H18ClNO2S	276.083	MS-DIAL	L3
Dipropyleneglycol-dibenzoate	8.95	C20H22O5	343.1555	MS-DIAL	L3
Dodemorph	11.97	C18H35NO	282.2796	TASQ	L2
Emtricitabine	3.37	C8H10FN3O3S	248.0514	MS-DIAL	L3
Fenpropimorph Carboxylic Acid	5.3	C20H31NO3	334.2384	TASQ	L3
Flupyradifurone	4.44	C12H11ClF2N2O2	289.0566	TASQ	L2
Fosthiazate	5.96	C9H108NO3PS2	284.0539	TASQ	L2
Guanine	2.9	C5H5N5O	150.057	MS-DIAL	L3
Guanosine	2.78	C10H13N5O5	284.1007	MS-DIAL	L2
Indole-3-acetic acid	4.76	C10H9NO2	176.071	MS-DIAL	L2
Indoline	3.23	C8H9N	120.0811	MS-DIAL	L2
Isoleucine	2.53	C6H13NO2	132.10178	MS-DIAL	L3
Isoquinolone	5.35	C9H7NO	146.0601	MS-DIAL	L2
Kynurenic acid	3.61	C10H7NO3	190.0501	MS-DIAL	L2
Lauryl diethanolamide	9.05	C16H33NO3	288.2548	MS-DIAL	L2
L-Phenylalanine	3.23	C9H11NO2	166.0863	MS-DIAL	L2
Metoclopramide	3.89	C14H22ClN3O2	300.1477	TASQ	L2
Metolachlor	7.96	C15H22ClNO2	248.1416	MS-DIAL	L3
Nicotinamide	2.96	C6H6N2O	123.0551	MS-DIAL	L2
Nicotinic acid	2.21	C6H5NO2	124.0393	MS-DIAL	L2
Nordiltiazem	6.5	C21H24N2O4S	401.153	TASQ	L2
Paeonol	4.32	C9H10O3	167.0703	MS-DIAL	L3
Phenylalanine	5.18	C9H11NO2	166.0867	MS-DIAL	L2
Pipecolic acid	2.3	C6H11NO2	130.0862	MS-DIAL	L3
Pyridafol	7	C10H7ClN2O	207.0311	TASQ	L2
Quinmerac	4.39	C11H8ClNO2	222.0321	TASQ	L1
Quinoline 1-oxide	4.6	C9H7NO	146.0601	MS-DIAL	L2

Compound	RT [min]	Formula	m/z [Da]	Identification approach	Identification level
Selegiline	4.46	C13H17N	188.1429	TASQ	L2
Simazine	5.65	C7H12ClN5	202.086	TASQ	L2
Tryptophan	3.56	C11H12N2O2	205.0971	MS-DIAL	L2
Thymine	3..12	C5H6N2O2	127.0505	MS-DIAL	L3
Tributyl phosphate	9.05	C12H27O4P	267.173	MS-DIAL	L2
Uracil	1.96	C4H4N2O2	113.0347	TASQ/MS-DIAL	L2
Uridine	2.47	C9H12N2O6	245.0785	MS-DIAL	L2

Appendix 2: Purchased Steroid Hormone Calibrators, it contains 17 steroid hormones and their concentrations in the mixture of powders.

		Value [ng/mL]						
	Steroid Hormone	**Kal 1**	**Kal 2**	**Kal 3**	**Kal 4**	**Kal 5**	**Kal 6**	**Kal 7**
1	Aldosterone	0.05	0.1	0.2	0.4	1.5	3	5
2	Androstenedione	0.032	0.064	0.13	0.64	2.4	4.8	8
3	Androsterone	0.06	0.12	0.24	0.48	1.8	3.6	6
4	Corticosterone	0.03	0.12	0.6	2.4	9	18	30
5	Cortisol	1	4	20	80	300	600	1000
6	Cortisone	0.1	0.4	2	8	30	60	100
7	11-Deoxycorticosterone	0.03	0.12	0.6	1.2	4.5	9	15
8	11-Deoxycortisol	0.01	0.04	0.2	0.8	3	6	10
9	DHEA	0.12	0.24	0.48	2.4	9	18	30
10	DHEAS	32	64	128	640	2400	4800	8000
11	Dihydrotestosterone (DHT)	0.012	0.024	0.048	0.24	0.9	1.8	3
12	Estradiol-17-beta (E2)	0.02	0.08	0.4	1.6	6	12	20
13	Estrone (E1)	0.03	0.12	0.6	1.2	4.5	9	15
14	Etiocholanolone	0.06	0.12	0.24	0.48	1.8	3.6	6
15	17α-Hydroxyprogesterone	0.05	0.2	1	4	15	30	50
16	Progesterone	0.06	0.12	0.24	1.2	4.5	9	15
17	Testosterone	0.01	0.04	0.2	0.8	3	6	10

Technique of Analysis
Absolute***IDQ***® Stero17 Kit (LC-MS/MS)

Table 6.2: Purchased Steroid Hormone Calibrators

Appendix 3: Purchased isotopic labeled Steroid Hormone Calibrators, it contains 17 steroid hormones and their concentrations after reconstitution.

No.	Labeled Standard	Concentration [ng/mL]	Content [ng]
1	d7-Aldosterone	**400**	**480**
2	d3-Androstenedione	**30**	**36**
3	d4-Androsterone	**100**	**120**
4	d8-Corticosterone	**100**	**120**
5	d4-Cortisol	**1000**	**1200**
6	d7-Cortisone	**50**	**60**
7	d8-11-Deoxycorticosterone	**25**	**30**
8	d5-11-Deoxycortisol	**40**	**48**
9	d6-DHEAS	**40000**	**48000**
10	d3-Dihydrotestosterone (DHT)	**120**	**144**
11	d3-Estradiol (E2)	**25**	**30**
12	d4-Estrone (E1)	**35**	**42**
13	d8-17α-Hydroxyprogesterone	**25**	**30**
14	d9-Progesterone	**30**	**36**
15	d5-Testosterone	**50**	**60**

Technique of Analysis
Absolute***IDQ***® Stero17 Kit (LC-MS/MS)

Table 6.3: Purchased Steroid Hormone Internal Standard