TESIS DE LA UNIVERSIDAD DE ZARAGOZA

2022

184

Xuechao Song

Application of Ion Mobility-High Resolution Mass Spectrometry and In Silico Tools for Identifying Non-Volatile Substances in Food Contact Material

Director/es

Nerín de la Puerta, María Consolación Cristina

Canellas Aguareles, Elena Purificación

ISSN 2254-7606

Prensas de la Universidad Universidad Zaragoza

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

APPLICATION OF ION MOBILITY-HIGH RESOLUTION MASS SPECTROMETRY AND IN SILICO TOOLS FOR IDENTIFYING NON-VOLATILE SUBSTANCES IN FOOD CONTACT MATERIAL

Autor

Xuechao Song

Director/es

Nerín de la Puerta, María Consolación Cristina
Canellas Aguareles, Elena Purificación

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Ciencia Analítica en Química

2022

UNIVERSIDAD DE ZARAGOZA

Escuela de Ingeniería y Arquitectura

Departamento de Química Analítica

Application of Ion Mobility-High Resolution Mass Spectrometry and In Silico Tools for Identifying Non-Volatile Substances in Food Contact Materials

Xuechao Song

Doctoral Thesis

Directed by

Dra. Cristina Nerín de la Puerta

Dra. Elena Canellas Aguareles

2022

Departamento de Química Analítica
Universidad Zaragoza

Instituto Universitario de Investigación en Ingeniería de Aragón
Universidad Zaragoza

La **Dra. Cristina Nerín de la Puerta**, Catedrática del Departamento de Química Analítica de la Universidad de Zaragoza y la **Dra. Elena Canellas Aguareles**

CERTIFICAN:

Que la presente Memoria, titulada: **"Application of ion mobility-high resolution mass spectrometry and in silico tools for identifying non-volatile substances in food contact materials**" presentada por Don Xuechao Song para optar al grado de Doctor en Ciencia Analítica, ha sido realizada bajo nuestra codirección en el Instituto de Investigación en Ingeniería de Aragón (I3A) y la Escuela de Ingeniería y Arquitectura de la Universidad de Zaragoza (EINA) , de acuerdo a los objetivos presentados en el Proyecto de Tesis aprobado por el Departamento de Química Analítica. Por tanto, autorizamos su presentación para proseguir con los trámites oportunos y proceder a su calificación por el tribunal correspondiente.

En Zaragoza, a 10 de Mayo de 2022.

Dra. Cristina Nerín de la Puerta

Dra. Elena Canellas Aguareles

Xuechao Song (File No. 201806780031) is sponsored by the China Scholarship Council (CSC) during the PhD study in the University of Zaragoza, Spain.

Thanks are also given to the Project RTI2018-097805-B-I00 from the Spanish Ministry of Science and Innovation.

The authors also thank Waters Corporation for access to an VION IMS-QToF instrument.

Acknowledgement

It is still difficult to accept that my PhD is going to be finished. Four years is a long time but meanwhile is also short. I think this will be the most memorable days of my life, there were both happiness and sadness in these four years, but now, only gratefulness left. Many of my friends and families helped me a lot in the pursuit of my PhD, without your helps, encouragements and companies, the completion of this work could not be possible. Firstly, I want to appreciate the favorable research environment provided by my supervisor: Cristina Nerín, I am very grateful for the freedom and trust she gave to explore the secrets of ion mobility and machine learning. I would also thank her for patiently guiding my research and revising my papers.

Thanks are also given to my co-supervisor: Elena Canellas, for her selfless guidance about the equipment Vion and the identification about unknows, as well as other experimental stuffs. I also want to thank her for the helps and suggestions regarding manuscript writing. She helps me solve many problems about my studies and thesis.

I also want to express my sincere thanks to my cooperators from Waters Corporation in UK: Nicola Dreolin and Jeff Goshawk. They helped me a lot regarding the instrumental operations and analysis, building of in-house plastic additives library. Special thanks are also given to them for the exhaustive and thorough revision of my manuscripts.

I would like to thank Magdalena Wrona for teaching me how to measure the antioxidant capacity using UV-Vis spectrometry, and also for the time we spent outside the labrotory with your husband Davis Pezo. Your enthusiasm inspired and empowered me in the period of my PhD.

I also want to thank my friend Qizhi Su for his guidance on R programming, and his company in my PhD. It is still excited to recall the cycling from Spain to France in 2019, as well as the excursions in Pyrenees, these experiences enriched my doctoral life

and will be precious memories in my life.

Many thanks are also given to other members in GUIA group. I appreciated the guidance of Raquel, Silvia and Filomena about the TVB-N assay. Thanks to Esther for providing some standards of phenolics, and to Paula for her helps on my experiments. I would like to thank Jesús for the experimental supports as well as the documents for applying my NIE. Thanks to Jorge for his valuable technical support and sense of humor. The timely experimental materials bought by Pilar were also important to my project. I also give thanks to Nicolás for his helps about the exercise of my courses in 2020. I appreciated the time I spent with Sara, Jazmine, and Janira in the bars in Jueves Pinchos. I am also grateful to David for his help regarding the register and many other stuffs, hoping that you can enjoy your PhD time.

Meanwhile, thanks are given to visiting students from different countries: Cathy, Nicola, Mahdiyeh, Bianca, Elena. Although you stayed a short time in the laboratory, I knew some different cultures and have a more colorful four years.

I thank my friends as well as flatmates Monica, Daniel and Amir, they helped me a lot in my life, I enjoyed the time living with you guys. I also thank Dong for the company in the Camino de Santiago.

Last but not least, I am grateful for the scholarship provided by China Scholarship Council, this is very important for me to finish the PhD. Great thanks are given to my parents and brother, without your unconditional love and faithful support, none of these would be possible.

These four years will be unforgettable memory for me in the future work and lives. Life is short, hoping everyone can enjoy each day.

To my parents

To my brother

Table of Content

Glossary of Terms

AA: Adipic acid

AMR: Atom molar refractivity

AMR: Ghose-Crippen molar refractivity

ANN: Artificial neural network

Apol: Atomic polarizability

ATD: Arrival time distribution

BBP: Benzyl butyl phthalate

BFRs: Brominated flame retardants

BHA: Butylated hydroxyanisole

BHT: Butylated hydroxytoluene

BPI: Base peak intensity

C: Cost of constraints violation

CCS: Collision cross section

CID: Collision-induced dissociation

cIMS: Cyclic ion mobility spectrometry

CNN: Neural network model

CPPdb: Chemicals associated with Plastic Packaging Database

DIA: Data independent acquisition

DPP: Diphenyl phthalate

DT: Drift time

$^{DT}CCS_{N2}$: Drift tube collision cross section in nitrogen

DTIMS: Drift tube ion mobility spectrometry

E&L: Extractables and leachables

ECCEN: Eccentric connectivity index

EDCs: Endocrine disrupting chemicals

EG: Ethylene glycol

ESI: Electrospray ionization

FCCdb: Food Contact Chemicals Database

FCCs: Food contact chemicals

FCMs: Food contact materials

FWHM: Full width at half-maximum

GC-MS: Gas chromatography - mass spectrometry

$HDMS^E$: High definition MS^E mode

HDPE: High density polyethylene

HRMS: High resolution mass spectrometry

IAS: Intentionally added substances

IMS: Ion mobility spectrometry

K: Mobility

Kier1: First kappa shape index

Lasso: Least absolute shrinkage and selection operator

LC-MS: Liquid chromatography -mass spectrometry

LDPE: Low density polyethylene

LLDPE: Linear low density polyethylene

LOD: Limit of detections

LOQ: Limit of quantification

MDs: Molecular descriptors

MoNA: Massbank of North America

MQNs: Molecular quantum numbers

MRE: Median relative errors

MW: Molecular weight

NIAS: Non-intentionally added substances

OML: Overall migration limit

PA: Polyamide

PAA: Primary aromatic amines

PCR: Principal components regression

PDA: Photodiode array

PEG: Polyethylene glycol

PET: Polyethylene terephthalate

PFAS: Per- and polyfluoroalkyl substances

PLA: Polylactic acid

PLS: Partial least squares regression

PP: Polypropylene

PPG: Polypropylene glycol

PS: Polystyrene

PUR: Polyurethan

PVC: Polyvinylchloride

QC: Quality control

r: Correlation coefficients

***R^2*:** Determination coefficient

RBF: Radial basis function

RF: Random forest

RMSECV: Root mean squared error of cross validation

RMSEP: Root mean square error of the prediction

RSDs: Relative standard deviations

RT: Retention time

S/N: signal to noise

SLIM: Structures for lossless ion manipulations

SMILESs: Simplified molecular-input line-entry systems

SML: Specific migration limit

SMLR: Stepwise multiple linear regression

Sp: Sum of atomic polarizabilities

SPE: Solid phase extraction

SR: Sensitivity ratio

Sv: Sum of atomic Van der Waals volumes

SVM: Support vector machine

TIMS: Trapped ion mobility spectrometry

TPA: Terephthalic acid

$^{TW}CCS_{N2}$: Traveling wave collision cross section in nitrogen

TWIMS: Traveling wave ion mobility spectrometry

UHPLC-IMS-MS: Ultra-high performance liquid chromatography with an ion mobility-mass spectrometer

UPLC-QToF: Ultrahigh performance liquid chromatography coupled to a quadrupole-time-of-flight mass spectrometer

VABC: Atomic and Bond Contributions of Van der Waals volume

VIF: Variance inflation factor

Vx: McGowan volume

WPATH: Weiner path number

XGBoost: Extreme gradient boosting

γ: gamma

Presentation

The doctoral thesis present here is entitled "**Application of Ion Mobility-High Resolution Mass Spectrometry and In Silico Tools for Identifying Non-Volatile Substances in Food Contact Materials**". This work was carried out in the University Group of Analytical Research (GUIA) led by Dr. Cristina Nerín de la Puerta in the Department of Analytical Chemistry at the University of Zaragoza in Zaragoza, Spain. The group is integrated into the Aragón Institute of Engineering Research (I3A).

This thesis is focused on the application of a novel separation technique: ion mobility, and machine learning approaches in the prioritization and identification of non-volatile substances in food contact materials. In the first place, we developed a collision cross section database for extractables and leachables from food contact materials. This database is very useful and necessary for the suspect and untargeted screening analysis using ion mobility-high resolution mass spectrometry platform. Then, accurate CCS and RT prediction models were built and further applied in the identification of compounds that migrated from plastic food packaging into foodstuffs and aquatic environments. At last, the application of ion mobility was investigated in the characterization of natural phenolic compounds, as natural phenolic compounds are always used in antioxidant active food packaging.

This thesis consists of six sections.

Section I mainly introduces the basic concepts about food contact materials and migration tests, the commonly used additives and their degradation products, as well as other chemicals in food contact materials. In addition, the basic principle of ion mobility, its benefits in unknown identification, CCS databases and CCS prediction tools are also mentioned.

Section II presents the general objectives of this work as well as the specific objectives of each experimental chapter.

Section III is the experimental part, which contains 6 chapters. The brief introductions of each chapter are given below:

Development of CCS database and in-silico prediction tools for identifying non-volatile compounds in FCMs

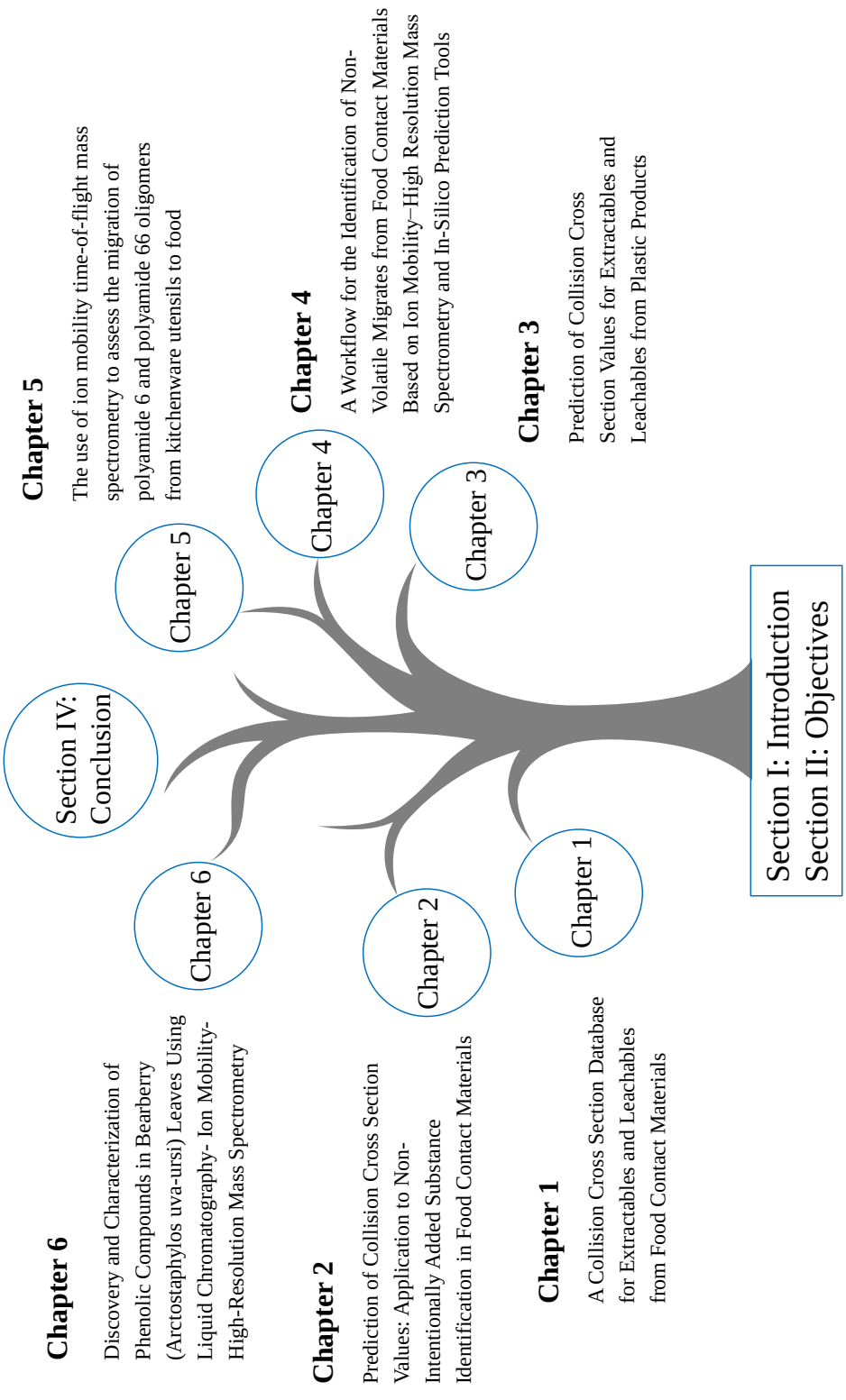
- **Chapter 1:** Developing a collision cross section database for extractables and leachables from food contact materials.
- **Chapter 2:** Preliminary prediction of collision cross section values of food contact chemicals based on the CCS database we developed.
- **Chapter 3:** Collecting experimental CCS values of plastic-related chemicals and developing an accurate model to predict the CCS values for extractables and leachables from plastic products.
- **Chapter 4:** Establishing a workflow for the prioritization and identification of food contact chemicals based on ion mobility-high resolution mass spectrometry and in-silico prediction tools.
- **Chapter 5:** Studying the migration of polyamide 6 and polyamide 66 oligomers from kitchenware utensils to foodstuffs and food simulants.

The application of ion mobility in the characterization of phenolic compounds

- **Chapter 6:** Discovery and characterization of phenolic compounds in bearberry (Arctostaphylos uva-ursi) leaves using liquid chromatography-ion mobility-high resolution mass spectrometry.

Finally, **Section IV, V** and **VI** show the conclusions, scientific publications and

list of bibliography, respectively.



Section I: Introduction

Introduction

1. Food contact materials

Food contact materials (FCMs) are all materials which are intended or reasonably to be expected to contact with food, during the production, processing, storage and food preparation and serving before its eventual consumption. Based on the service of FCMs, they can be generally classified into packaging materials, kitchenware and tableware, processing machine and storage containers. A variety of materials, such as plastic, paper and cardboard, glass, metal and ceramic, can be used to make FCMs (Muncke, 2016). Figure I-1 shows the percentages that each type of materials accounts for in FCMs. Compared to glass, metal and ceramic, the plastic, paper and carboard are considered less inert and more likely to allow chemical migration to the contacted food and surrounding environment.

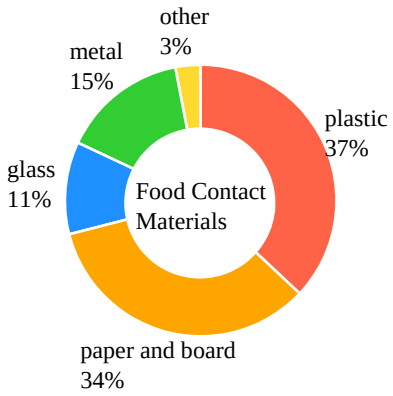


Figure I-1. Percentages of each type of materials used in FCMs, data from (Muncke, 2016).

The chemical constituents of FCMs, termed food contact chemicals (FCCs), can migrate into food and pose a potential risk to the health of consumers (Muncke et al., 2017). It is estimated that the FCCs migrated from FCMs into food can be at levels 100 times higher than pesticides and other environmental pollutants (Grob et al., 2006). There are more than 12,000 substances that are currently intentionally added in the production of FCMs (Groh et al., 2021), which are of great concern in terms of both

food safety and food quality. The FCCs that migrate into food can be harmful to human health, meanwhile, may also bring about an unacceptable change in the composition, taste and odor of the food.

2. Plastics

Plastics are one important type of FCMs, it is estimated that flexible and rigid plastics account for around 37% of FCMs market share (Muncke, 2016). Generally, plastics contain multiple polymers and various types of additives, such as plasticizers, antioxidants, flame retardants, lubricants and others. The most commonly used plastics in our daily life are low density and linear low density polyethylene (LDPE/LLDPE), high density polyethylene (HDPE), polyethylene terephthalate (PET), polypropylene (PP), polystyrene (PS), polyvinylchloride (PVC), polyurethane (PUR), as well as polyester, polyamide (PA) and acrylic fibers (Geyer et al., 2017). It was estimated that approximate 40% of all plastics are used for packaging (Geyer et al., 2017; PlasticsEurope, 2021), and among them, around 60% were used to package food and beverage products (Groh et al., 2019). The other applications of plastics include building and construction, automotive, electrical and electronics, agricultural and household products (PlasticsEurope, 2021).

The feedstocks of plastics are various petrochemicals, such as natural gas, coal and petroleum (Ilyas et al., 2018). In 2020, global plastics production almost reached 370 million tons (PlasticsEurope, 2021). The commonly used plastics show high resistance to aging and biological degradation, some of the so-called biodegradable plastics are also made from traditional petrochemical polymers blended with starch or vegetable oils (Mohd Amin et al., 2015). It was estimated that about 50% of plastic products are designed for single-use applications, such as packaging and agricultural films (Hopewell et al., 2009). These reasons lead to a rapid increase of accumulation of plastic waste in terrestrial and aquatic environment. Geyer et al. (2017) estimated that until 2015, there were generated approximately 6300 metric tons of plastic wastes,

about 9% and 12% were recycled and incinerated respectively, the other 79% was accumulated in landfills and other natural environment. Because of the high stability of plastic, the longevity of plastic waste in the environment can range from decades to centuries (Hopewell et al., 2009).

The harmful effect of plastic wastes on the environment and subsequently, on human health, mainly involves two aspects: microplastics and small molecular weight chemicals. Microplastics (<5 mm) can be derived from the fragmentation of large plastic products, as a result of photolytic, mechanical and biological degradation (Li et al., 2016). The exposure to UV radiation can lead to bond breaking, thus the plastics become brittle and break down into small pieces (Moore, 2008). Microplastics have attracted much attention from food and environmental researchers in recent years, as due to the small size of microplastics, they can be easily ingested by organisms and accumulate in the tissues. In addition to this, compared to macroplastics, microplastics have greater potential to transport contaminants to organisms due to their large ratio of surface area to volume (Browne et al., 2007).

Another type of contaminants from plastics is the small molecular weight chemicals that they release over their shelf life. Plastics contain about 93% polymeric resins and 7% of additives, on average by mass, plasticizers, flame retardant and fillers account for about 75% of total used additives (Geyer et al., 2017). In 2015 alone, about 302 metric tons of plastic waste was generated, which included 17 metric tons of plastic additives (Geyer et al., 2017). The other small molecular weight chemicals contain the degradation products of additives and polymers and side-reaction products

3. Migration

The term “migration” usually describes the transfer of matter from the packaging material to the packaged product. It is governed by partition and diffusion processes, in which the additives and other small molecules (normally below 1000 Da) are transferred from FCMs into the food or into outer environment (Arvanitoyannis and Bosnea, 2004).

The extent to which migration occurs depends on a variety of factors, such as temperature, time, the chemical characteristics of FCMs, foods and chemicals, as well as the ratio between the surface area of FCMs and the volume of food (Muncke, 2016).

The complex matrices of real food make the chemical analysis challenging, thus in practice, in order to simplify the sample pretreatment, real food is always replaced by different types of food simulants: 10% ethanol (v/v), 3% acetic acid (w/v), 20% ethanol (v/v), 50% ethanol (v/v), vegetable oil and poly(2,6-diphenyl-p-phenylene oxide (also known as Tenax) (Commission, 2011). Substitute of vegetable oil are ethanol 95% and isooctane to facilitate the specific migration studies. The first three food simulants are assigned for foods that have a hydrophilic character, 50% ethanol and vegetable oil are assigned for food with lipophilic character, while Tenax is used to simulate dry food, such as milk powder.

There are two kinds of migration test: overall migration and specific migration. The overall migration is measured by exposing FCMs to a food simulant under a specific conditions (temperature and time), after which the extracted residue is dried and weighed. The selection of food simulants and migration test conditions is based on the intended use of FCMs. Overall migration measures the total amount of all non-volatile substances that can migrate into food, which reflects on the inertness of FCMs. It should be noted that the chemical identification is not performed in overall migration. In Europe, the overall migration limit (OML) is set as 10 mg/dm^2 food contact surface area, for FCMs for infants and young children, the OML is 60 mg/kg food (Commission, 2011).

The specific migration is also performed by exposing a FCM to a food simulant for a specified and appropriate length of time, but the difference with the OM is that the individual migrates should be identified and quantified by analytical techniques. Generally, the chemical analysis is performed by gas chromatography - mass spectrometry (GC-MS) for volatile and semi-volatile compounds, and liquid chromatography -mass spectrometry (LC-MS) for nonvolatile compounds. The amount

of migration of a specific substance is usually expressed as mg/kg food simulant. This value should be compared to the specific migration limit (SML) proposed by Regulation (EU) No 10/2011 (EC, 2011).

In addition to the chemical migration into contacted food, the migration of chemicals from plastics into outer environment has also been observed. For example, photoinitiators, phenolic and organophosphite antioxidants, phthalate plasticizers and their transformation products have been detected in indoor dust (Liu and Mabury, 2019a, b, c; Liu et al., 2019), tris(2,4-di-tert-butylphenyl)phosphate and other organophosphate esters are unexpected abundant toxic pollutants in airborne particulate matters (Liu et al., 2020; Shi et al., 2020). The organophosphite antioxidants in mulch films can be oxidized into organophosphates, which are important pollutants in soil (Gong et al., 2021). The exposure to migrates in both food and environment can pose a harmful effect on human health.

4. Types of FCCs

Generally, FCCs can be divided into intentionally added substances (IAS) and non-intentionally added substances (NIAS). IAS refer to a variety of additives, which are incorporated into FCMs during the production process, in order to enhance the stability, mechanical properties and increase the service life of FCMs. In addition to IAS, FCMs also contain NIAS that may not have technical functions. NIAS can be classified into reaction by-products, degradation products of additives and polymers, and impurities from the raw materials used for their production (Nerin et al., 2013).

4.1. IAS

Approximately 12,000 intentionally added substances could be used in the manufacture of FCMs (Groh et al., 2021), based on their functional and structural characteristics, IAS can be categorized into four classes: functional additives, colorants, fillers and reinforcements (Gunaalan et al., 2020). Functional additives generally

include plasticizers, flame retardants, antioxidant, photoinitiators, slip agent, lubricants, etc., these additives can improve the physicochemical properties of polymers. Colorants include pigments and soluble azocolorants, which can be used in inks. Fillers include mica, talc, kaolin, clay, calcium carbonate, with the aim of improving coating properties of polymers. Some commonly used functional additives are introduced below.

Plasticizers are the most commonly used additives in plastic packaging, which can improve the flexibility, durability and stretchability of polymeric films. Among all plasticizers, phthalates are commonly utilized, comprising up to 85% of the total market share (Qadeer et al., 2022). However, the safety use of phthalates in plastics remains in question, as they are regarded as endocrine disrupting chemicals (EDCs) based on current scientific researches (Darbre, 2020; Grindler et al., 2018). Some alternative plasticizers have been developed, including adipates, sebacates, citrates, benzoates, phosphate esters, Currently, it is difficult to label alternative plasticizer as safe compounds as many of these don't have toxicological data (Qadeer et al., 2022).

Antioxidants can prevent the oxidation of polymers during the manufacturing and service life. Hindered phenols and phosphites are two main categories of antioxidants in plastic products, hindered phenols contain hydroxyl groups, which are used as primary antioxidants, as they can eliminate free radicals by hydrogen donation, thus terminate the radical chain reaction. Butylated hydroxytoluene (BHT) is one of the most widely used phenolic antioxidant even accepted as food additive. Other hindered phenolic antioxidants include butylated hydroxyanisole (BHA), Irganox 1010, Irganox 1076, Irganox 245, Irganox 1035, Irganox 1098, etc. Phosphites are secondary antioxidants, which can reduce hydroperoxides to the corresponding alcohols and themselves are transformed into phosphates. Irgafos 168 (tris(2,4-di-tert-butylphenyl)phosphite), a widely used secondary antioxidant, and its degradation product: tris(2,4-di-tert-butylphenyl)phosphate, have been detected in plastic films (Yang et al., 2016), soil (Gong et al., 2021), airborne particulate matters (Shi et al., 2020) and indoor dust (Liu and Mabury, 2019b). The primary and secondary antioxidants are

normally used together to achieve synergistic antioxidant effects (Ambrogi et al., 2017).

Flame retardants in plastics can prevent the start or slow the growth of fire. They can be grouped based on whether they contain bromine, chlorine, phosphorus, nitrogen, metals, or boron. Since the 1970s, brominated flame retardants (BFRs), such as polybrominated diphenyl ethers, hexabromocyclododecanes, and tetrabromobisphenol A, have been used as flame retardants in food packaging and other household products (Shaw et al., 2014), now, many of brominated flame retardants were phased out because they lead to adverse health effects in humans, such as endocrine disruption, reproductive toxicity, and cancer (Shaw et al., 2014). Organophosphate flame retardants have been proposed as replacement for BFRs, however, current data from toxicity testing and risk assessment showed that there are also health concerns at current exposure levels for organophosphate flame retardants (Blum et al., 2019). In addition, BFRs are not allowed to be used as additives for the manufacture of FCMs (Commission, 2011), however, these compounds were detected in kitchen utensils within Europe, it is likely that the plastic wastes of electrical and electronic equipment are used to produce FCMs (Puype et al., 2015; Samsonek and Puype, 2013).

Other main types of IAS include photoinitiators, lubricants, surfactants and slip agents. Photoinitiators are always used in UV curing printing inks and varnishes, they can be divided into benzophenones, amine co-initiators, phosphine oxides, thioxanthones and anthraquinones based on their structural characteristics (Liu and Mabury, 2018a, 2019a). Lubricants and slip agents can reduce the friction between surfaces of films, long-chain fatty acids (stearic acid, palmitic acid, etc.) and long-chain fatty amides (erucamide and oleamide, etc.) are widely used as lubricants and slip agents in FCMs, respectively (Getachew et al., 2017; Hahladakis et al., 2018; Hu et al., 2021).

4.2. NIAS

NIAS can be formed in FCMs during the manufacture, service life and recycling, as a result of side-reactions, degradations as well as contaminations (Nerin et al., 2013).

The side-reaction products can be formed by the reaction between starting substances of FCMs or between migrates and compositions of food (Canellas et al., 2021a). One typical type of side-reaction products observed in plastic packaging is oligomers, resulting from incomplete polymerization of starting substances. For example, many cyclic and linear oligomers were detected in PET resins, which are generated from the reaction between terephthalic acid (TPA) and ethylene glycol (EG) (Tsochatzis et al., 2020; Ubeda et al., 2018; Ubeda et al., 2020). In addition, in PUR adhesives, several cyclic oligomers (e.g., 1,6-dioxacyclododecane-7,12-dione) were produced from the reaction between adipic acid (AA) and 1,4-butanediol (Canellas et al., 2015b; Felix et al., 2012). Some common diols and dicarboxylic acids in plastics that have the potential to produce oligomers were retrieved from literature (Bauer et al., 2019; Canellas et al., 2022; Hoppe et al., 2016; Omer et al., 2018), and are shown in Figure I-2. Meanwhile, oligomers can also be formed from the reaction between carboxylic acids and amines. The most typical oligomers of this type are PA oligomers (Abe et al., 2016; Canellas et al., 2021b), where PA66 oligomers are from the reaction between 1,6-diaminohexane and AA, and PA6 oligomers originate from the polymerization of caprolactam.

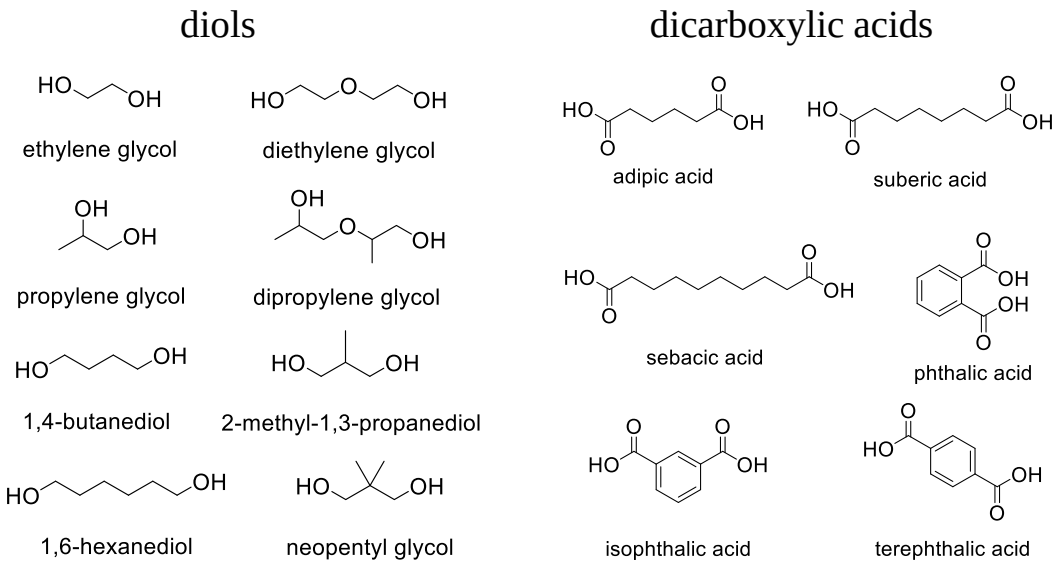


Figure I-2. Common diols and dicarboxylic acids used to produce plastic packaging.

Degradation products are another important type of NIAS in FCMs. Typically,

degradation of additives or polymers occurs in all service life of plastic packaging, however, this process can be accelerated by the UV radiation (Yang et al., 2016) and microwave heating (Alin and Hakkarainen, 2011). Some commonly observed degradation products of additives are shown in Table I-1. Sterically hindered phenolic antioxidant, such as BHT, Irganox 1010, Irganox 1330, can scavenge the free radicals in polymers by donating hydrogen atoms, meanwhile, the phenols are converted to quinoid structures (Beißmann et al., 2013). As an example of this, 2,6-di-tert-butyl-p-benzoquinone has been detected in plastics as a degradation product of BHT and Irganox 1010 (Alin and Hakkarainen, 2011; Liu and Mabury, 2018c). Some phenolic antioxidants can also undergo hydrolysis during their uses, 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propanoic acid, methyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propanoate, and benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, 1,1′-[2,2-bis(hydroxymethyl)-1,3-propanediyl] ester, have been identified as the hydrolysis products of Irganox 1010, Irganox 1076 and other 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionate antioxidants (Beißmann et al., 2013; Liu and Mabury, 2021; Vera et al., 2019).

Organophosphite antioxidants, such as Irgafos 168 and Irgafos 126, are used as secondary antioxidants in plastics, where they can reduce hydroperoxides into the corresponding alcohols and themselves are transformed into phosphates. The proposed degradation mechanism of Irgafos 168 is shown in Figure I-3. Tris(2,4-di-tert-butylphenyl)phosphate and bis(2,4-di-tert-butylphenyl) pentaerythritol diphosphate are degradation products of Irgafos 168 and Irgafos 126, respectively, which have been detected in farmland, as a result of degradation of plastic mulch film (Gong et al., 2021). Moreover, tris(2,4-di-tert-butylphenyl)phosphate can be further oxidated into bis(2,4-di-tert-butylphenyl)phosphate, by losing a 2,4-di-tert-butylphenol unit (Dorival-Garcia et al., 2018).

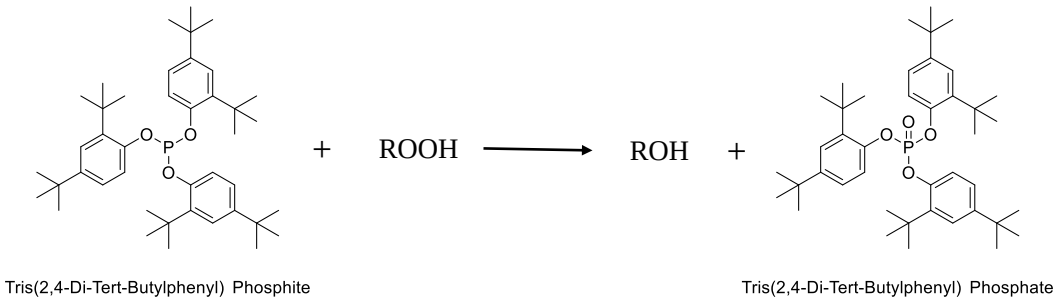


Figure I-3. Degradation of tris(2,4-di-tert-butylphenyl)phosphite (Irgafos 168) into tris(2,4-di-tert-butylphenyl)phosphate.

Table I-1. Commonly observed degradation products of additives in FCMs.

Degradation products	Additives	Reference
2,4-di-tert-butylphenol tris(2,4-di-tert-butylphenyl)phosphate	Irgafos 168	(Yang et al., 2016)
2,4-di-tert-butylphenol 2,6-di-tert-butyl-p-benzoquinone 7,9-di-tert-butyl-1-oxaspiro(4,5)deca-6,9-diene-2,8-dione methyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl) propanoate	Irgafos 168 Irganox 1010	(Alin and Hakkarainen, 2011)
2,4-di-tert-butylphenol 2,6-di-tert-butyl-p-benzoquinone 3,5-di-tert-butyl-4-hydroxybenzaldehyde 3,5-di-tert-Butyl-4-hydroxyacetophenone methyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propanoate	Irgafos 168 Irganox 1076	(Denberg et al., 2009)
Irganox 1330, with hydroxyl groups oxidized into quinone	Irganox 1330 Irganox 1010	(Beißmann et al., 2013)
2,6-di-tert-butyl-4-methylphenol 2,6-di-tert-butyl-4-ethylphenol 2,6-di-tert-butyl-4-(1-hydroxyethyl)phenol 2,6-di-tert-butyl-4-(1-hydroxyethylidene)cyclohexa-2,5-dienone 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propanoic acid methyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propanoate	Irgafos 168	
tris(2,4-di-tert-butylphenyl)phosphate bis(2,4-di-tert- butylphenyl) pentaerythritol diphosphate	Irgafos 168 Irgafos 126	(Gong et al., 2021)

Degradation products	Additives	Reference
2,6-di-tert-butyl-4-(hydroxymethyl)phenol 3,5-di-tert-butyl-4-hydroxybenzaldehyde 2,6-di-tert-butyl-p-benzoquinone 2,6-di-tert-butyl-4-hydroxy-4-methyl-2,5-cyclohexadienone	BHT	(Liu and Mabury, 2018c)
methyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl) propanoate ethyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl) propanoate benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, 1,1′-[2,2-bis(hydroxymethyl)-1,3-propanediyl] ester benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, 1,1′-[2-[[3-[3,5-bis(1,1-dimethylethyl)-4-hydroxyphenyl]-1-oxopropoxy]methyl]-2-(hydroxymethyl)-1,3-propanediyl] ester	Irganox 1010 Irganox 1076	(Vera et al., 2019)
2,4,6-trimethylbenzaldehyde 1-phenyl-2-butanone.	UV-Photoinitiators	(Scarsella et al., 2019)
tris(2,4-di-tert-butylphenyl)phosphate bis(2,4-di-tert-butylphenyl)phosphate 2,4-di-tert-butylphenol 1,3-di-tert-butylbenzene 2,6-di-tert-butyl-p-benzoquinone 2,6-di-tertbutyl-4-methoxyphenol 3,5-di-tert-butyl-4-hydroxybenzoic acid methyl 3,5-di-tert-butyl-4-hydroxybenzoate	Irgafos 168 Irganox 1010	(Dorival-Garcia et al., 2018)
alkylphenols (nonylphenol, octylphenol, etc.)	tris(nonylphenyl)phosphite alkylphenol ethoxylates	(Nerin et al., 2020)
primary aromatic amines	azo dyes in printing inks	(Aznar et al., 2009)

If the FCMs are made from recycled materials, some kinds of NIAS can be present in the final FCMs due to the low efficiency of cleaning processes. These NIAS include the chemical compounds coming from previously packaged food, or from the misuse of food packaging prior to recycling (Nerin et al., 2013). Su et al. (2021a) detected three pesticides (biopermethrin, chlorpyrifos and bifenthrin) in recycled HDPE materials for food use, indicating that the recycled plastics from agricultural field may enter the plastic flow of FCMs.

It is estimated that about 12,000 IAS and 30,000-100,000 NIAS are present in various FCMs (Muncke et al., 2020), many of these still remain unidentified. For those FCCs whose chemical structures have been characterized, if there are no commercial reference standards or pure substances available, hazard assessment is still impractical, based on current approach for chemical risk assessment (Muncke, 2016). In addition, obtaining the exposure data of FCCs is challenging, as FCMs is not the only source of FCCs. For example, tris(2,4-di-tert-butylphenyl)phosphate was detected in food-contact PP Films as a degradation product of Irgafos 168, meanwhile, this compound was also detected in indoor dust (Liu and Mabury, 2018b) and airborne particulate matters (Shi et al., 2020). Exposure to tris(2,4-di-tert-butylphenyl)phosphate through both food intake and breath seems possible, this make it difficult to obtain the real exposure data for this compound.

5. Legislation

Regulation (EC) No. 1935/2004 (EC, 2004) establishes a general principle that the substances in FCMs cannot migrate into food in levels high enough to endanger human health, or to bring about an unacceptable change in the composition of the food, or lead to a deterioration of organoleptic characteristics of the food.

Regulation (EC) No. 10/2011 (EC, 2011) provides detailed descriptions about the procedures of migration test, including food simulations, migration time and temperature. In addition, a list of substances authorized to be used in the manufacture of plastic FCMs, as well as their SMLs are also provided in Annex I. Some crucial statements of Regulation No. 10/2011 are shown below:

(1) Substances that are mutagenic, carcinogenic or toxic to reproduction should not be used in food contact materials or articles

(2) A maximum level of 0.01 mg/kg in food should be established for the migration of non-authorized substances

(3) For authorized substances, if their SMLs have been set out in Annex I, then their migration into foods should not exceed the SMLs. For authorized substances that have no SMLs or other restrictions in Annex I, a generic SMLs of 60 mg/kg should be applied. For authorized substances whose SMLs were marked by "ND" in Annex I, the substance shall not migrate in detectable quantities.

(4) The overall migration limit for FCMs is 10 mg/dm^2 food contact surface or 60 mg/Kg food.

6. Ion mobility spectrometry

LC-MS is the most popular instrument for analyzing the non-volatile FCCs in FCMs (Wrona and Nerin, 2020). High resolution mass spectrometry (HRMS), such as time-of-flight and orbitrap, can provide accurate mass for precursor and fragment ions, the elemental composition can be derived from precursor ions and isotopic pattern distributions. Subsequently, the obtained possible molecular formulas can be searched in public or private databases, such as ChemSpider and PubChem, the fragment ions of unknowns are also valuable for the further confirmation of chemical identities. However, this identification workflow is not applicable for isomeric compounds, as some molecular identifiers (precursor ion, isotopic pattern, and fragment ions) are often shared between isomeric species. In addition, the presence of co-eluted analytes or background ions bring a challenge to the interpretation of mass spectra of interest. These issues are possibly alleviated by the incorporation of another separation dimension into the conventional LC−HRMS systems (Nerin et al., 2022).

Ion mobility spectrometry (IMS) is a rapid gas-phase separation technique (normally in the timescale of millisecond), which has obtained growing interests from researchers in the last decades. The basic principle of IMS is the separation of ions in buffer gas under the influence of an electric field (D'Atri et al., 2018). Compact molecules always move faster than elongated ones in drift cell, as the formers have less interaction with buffer gas. As IMS allows the separation of ions based on their shape,

size and charge, it can provide partially orthogonal structural information in addition to those of retention time (RT) and *m/z*. Since the first commercial LC-IMS-MS platform: Synapt HDMS, was introduced by Waters in 2006 (Pringle et al., 2007), the last 16 years have seen an extensive application of IMS in the characterization of small molecules in different fields, such as metabolites (Luo et al., 2020a; Zhang et al., 2018), phenolic compounds (Masike et al., 2020; Song et al., 2021; Stander et al., 2017), food and environmental contaminants (Belova et al., 2021; Celma et al., 2020; Dodds et al., 2020; Fan et al., 2022), and extractables and leachables (E&L) from the pharmaceutical or food industries (Canellas et al., 2019; Canellas et al., 2022; Canellas et al., 2021b; Vera et al., 2019). The coupling of IMS with LC-HRMS has proved to be a powerful tool for the separation and identification of small molecules. The fundamental principle of IMS, different types of IMS instruments, and the benefits of IMS in targeted and untargeted screening analysis are described in detail in the following sections.

6.1. Fundamental principle of IMS

IMS and IMS-MS have been widely applied in the analyses of complex samples in the last 20 years, however, it is not a new separation technique. The origin of IMS can be traced back to 1896, in which the mobility of ions in various gases was investigated by Thomson and Rutherford (Thomson and Rutherford, 1896). The fundamental principle of IMS is "packets of gas-phase ions pass through a drift tube filled with buffer gas under the influence of a weak electric field (E)" (D'Atri et al., 2018). Mobility (K) of an ion is defined as the proportionality coefficient between its velocity (v_d) and E intensity:

$$K = \frac{v_d}{E}$$

K is a parameter depending on the experimental conditions (i.e., temperature and pressure), which are usually normalized into reduced mobility K_0, with p_0 equal to 760 Torr and T_0 equal to 273.15 K:

$$K_0 = K \cdot \frac{p}{p_0} \cdot \frac{T_0}{T}$$

Collision cross section (CCS or Ω) can be directly derived from K_0 in case all the parameters involved are accurately known and carefully controlled, using Mason-Schamp equation (Revercomb and Mason, 1975):

$$\Omega = \frac{3ze}{16N_0K_0}\sqrt{\frac{2\pi}{\mu k_b T}}$$

where ze is ionic charge; N_0 is buffer gas number density; $\mu = \frac{m_{ion} \times m_{gas}}{m_{ion} + m_{gas}}$, m_{ion} and m_{gas} are the mass of each ion and drift gas (normally N_2), respectively; k_b is Boltzmann's constant; T is temperature of drift region.

All types of IMS techniques share the same core principle, but they make use of slightly different operational principles. In the following section, different IMS techniques are introduced and their advantages and limitations are discussed.

6.2. Different types of IMS technology.

Drift tube IMS (DTIMS) is a historical and the simplest IMS technique. The drift tube is made of a series of piled electrodes and can provide uniform, static and weak electric field, normally in the range of 13–18 V/cm (D'Atri et al., 2018). The ionized molecules are pushed by electric field force into drift tube and collide with stationary buffer gas (normally nitrogen or helium), compact molecules would move faster as they have fewer collisions with buffer gas. The time taken to cross the drift cell is called drift time (t_d), which can be used to calculate the mobility *K*, and further CCS values of ions, using the Mason-Schamp equation. The ability of directly calculating CCS values from the first principle seems to be the most significant advantage of DTIMS, and the CCS values obtained in this method are considered more accurate and are always used for CCS calibration in other types of IMS (Hines et al., 2016).

Traveling wave IMS (TWIMS) is another widespread IMS technique, which was firstly commercialized by Waters Corporation in 2006 (Pringle et al., 2007). Since then several impactful Synapt enhancements have been introduced into commercial variants, such as Synapt G2 in 2009 (Giles et al., 2011), Synapt G2-S in 2011 and Synapt G2-Si in 2013 (May and McLean, 2015). TWIMS consists of a stacked ring ion guide where a series of voltage pulses are applied, creating a travelling wave that ions can 'surf' along (Shvartsburg and Smith, 2008). A radio frequency voltage is applied to adjacent ring electrodes to radially confine ions, in order to keep high transmission. Unlike DTIMS that CCS values can be directly calculated from drift time, in TWIMS, the IMS cell has to be calibrated with a set of reference compounds with known CCS values, the reference CCS values of calibrates are obtained from DTIMS using a step-field method.

Trapped IMS (TIMS) is a very new technique, which was commercialized by Bruker Daltonics in 2011 (Fernandez-Lima et al., 2011). Unlike conventional DTIMS and TWIMS that the ions move in stationary buffer gas forced by a low electric field, TIMS holds the ions stationary in a moving buffer gas by applying an electric field in the opposite direction. The field strength in TIMS is slowly decreased allowing the ejection of the ions with specific mobility (Ridgeway et al., 2018). It has been reported that TIMS can offer a resolving power exceeding 250, with ~3 to 8 times higher than current DTIMS and TWIMS techniques (Silveira et al., 2014). In TIMS, experimental CCS values can also be obtained by appropriate calibration using reference compounds with known CCS values.

6.3. CCS calibration of TWIMS technique

The CCS values can be directly determined by measuring drift time in DTIMS. The velocity of ions can be derived from drift time ($v_d = L/t_d$), where L is the length of drift tube, typically 78 cm. Then, mobility and CCS of an ion can be calculated from its velocity.

In TWIMS, due to the nonlinearity of applied electric field, CCS values cannot be

directly calculated using equations in section 6.1, they can be obtained after appropriate calibration using ions with known CCS values. A TWIMS calibration protocol has been proposed by Bush and coworkers (Bush et al., 2010), and the detailed procedures are described hereinafter. Firstly, the normalized CCS Ω_n was calculated from the reference CCS values Ω of the calibrates, the reference CCS values of calibrates are detected by DTIMS using a step-field method:

$$\Omega_n = \Omega \times \frac{\sqrt{\mu}}{z}$$

where $\mu = \frac{m_{ion} \times m_{gas}}{m_{ion} + m_{gas}}$, m_{ion} and m_{gas} are the mass of each calibrate and drift gas (N_2), respectively. Then the arrival time t_A of the calibrates were recorded and the corrected arrival time t_d was calculated:

$$t_d = t_A - c\sqrt{\frac{m}{z}}$$

where $c = \frac{EDC}{1000}$, *EDC* is the enhanced duty cycle (EDC) delay coefficient of the instrument. After this, the normalized CCS Ω_n versus t_d are fitted using an empirically-derived power-law function:

$$\Omega_n = A(t_d - t_0)^B$$

t_0 is the undetermined time off-set, the coefficient *A* and exponential factor *B* are calculated, which will be used for the derivation of CCS of unknown compounds. This function was then used to calculate the CCS values of calibrates and compared with their reference CCS. The acceptance criteria of the calibration is that CCS deviation of each calibrate is within ±2%.

The selection of appropriate calibrates for TWIMS is still under heavy debates in this field. Generally, the calibrates should cover the range of CCS and *m/z* values of the compounds to be analyzed, meanwhile display the similar physical properties and gas-phase conformations to analytes. Hines and coworkers compared the effect of different

types of calibrates on the accuracy of calibrated CCS values of phospholipids, and demonstrated that structurally matched calibrants provide highly accurate and reproducible traveling wave CCS values in nitrogen ($^{TW}CCS_{N2}$) (Hines et al., 2016). Masike found that the $^{TW}CCS_{N2}$ values of phenolic compounds using polyalanine as calibrates are always lower than the corresponding drift tube CCS in nitrogen ($^{DT}CCS_{N2}$) in the range below 200 $Å^2$ (Masike et al., 2022). Recently, an improved CCS calibration approach has been proposed for the TWIMS system by Richardson and coworkers, which has the potential to further improve the accuracy the $^{TW}CCS_{N2}$ measurements (Richardson et al., 2021).

6.4. Benefits of IMS in suspect and untargeted screening

6.4.1. Increasing peak capacity

The first advantage of IMS is increasing the peak capacity of conventional LC-MS by adding another separation dimension (Michelmann et al., 2015; Schroeder et al., 2019). IMS can separate ionized molecules based on their size and shape, this indicates that IMS can provide comprehensive information to LC-MS. An increase of peak capacity about 10% was observed in a TWIMS platform for analyzing small molecules (Song et al., 2022b). Belova et al. (2021) showed that the plasticizers, organophosphate flame retardants and per- and polyfluoroalkyl substances (PFAS) have different CCS versus *m/z* trendlines. PFAS contain carbon-fluorine bonds, and their CCS values are much lower in general than other compounds of similar *m/z* values. The increase of peak capacity was also observed in the analysis of large biomolecules; at a given mass, CCS values tend to increase in the order of oligonucleotides, carbohydrates, peptides, and lipids (Fenn et al., 2009). The improvement of resolving power of IMS technique could further increase the peak capacity of LC-IMS-MS systems.

6.4.2. Providing cleaner mass spectra by drift time alignment

Another advantage of IMS is that it can provide cleaner mass spectra. As IMS cell

is located before collision cell in most IMS-HRMS platforms, the precursor ion and its corresponding fragment ions share the same drift time. The alignment of precursor and fragment ions based on both retention time and drift time can eliminate many of the interfering ions, resulting in cleaner mass spectra. This advantage of IMS has been shown in many researches, which greatly simplifies the spectral interpretation (Canellas et al., 2019; Celma et al., 2021; Celma et al., 2020; Regueiro et al., 2016). Figure I-4 shows the mass spectra of tricyclazole with or without drift time alignment. As can be seen, the most abundant ion with *m/z* of 242.2839 and some other interfering ions were removed after the drift time alignment; the drift time aligned mass spectra (Figure I-4D) are more understandable. In addition to this, the elimination of background noise can lead to an improvement in signal to noise (S/N) ratio for analytes of interest, thus lower limit of detections (LOD) can be obtained (Kirk et al., 2017).

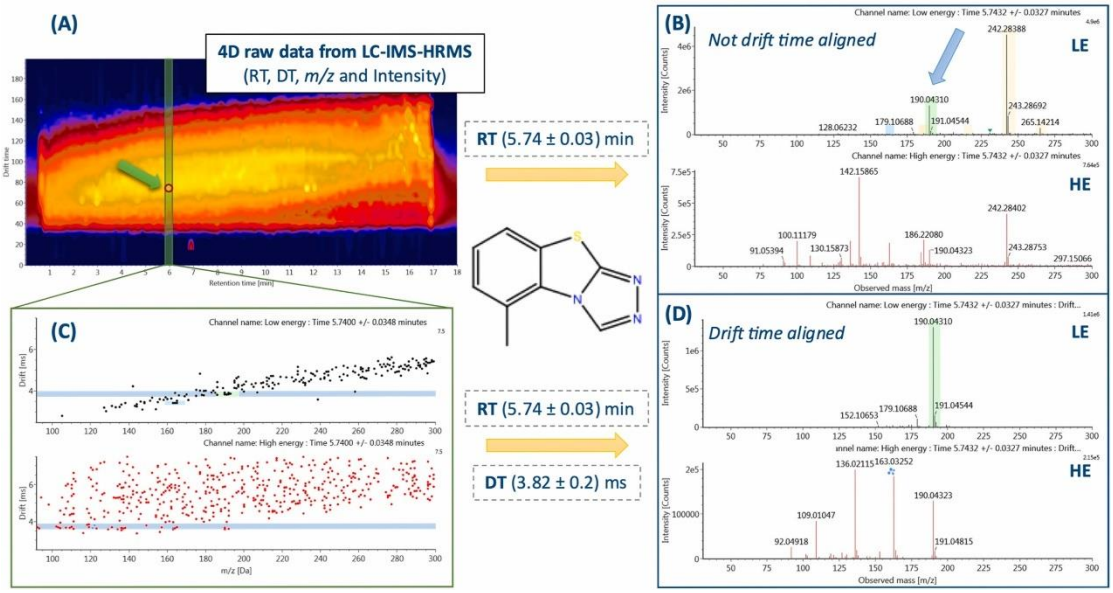


Figure I-4. Comparison of the mass spectra of tricyclazole ($[M+H]^+$, *m/z* 190.0431) with or without drift time alignment (Celma et al., 2021).

6.4.3. Isomers separation and identification

Isomers separation and identification is another attractive benefit of IMS technique. Isomeric species often share the same molecular mass and fragment ions,

which are indistinguishable by MS alone. However, their distinct spatial conformations give a chance for the separation and identification by IMS. Currently, the commercial IMS systems enable to give an interlaboratory and interplatform reproducibility around 2% (Hernandez-Mesa et al., 2020; Righetti et al., 2020). Therefore, in theory, the isomers can be identified as long as the difference between their CCS values is higher than 2%. Up to now, several pairs of isomers have been separated by IMS and identified by their different CCS values. Figure I-5 shows the structures and CCS values of three deprotonated flavones, CCS values of luteolin-6-C-glucoside (i.e., isoorientin) and luteolin-8-C-glucoside (i.e., orientin) are from (McCullagh et al., 2019b), CCS value of luteolin-7-O-glucoside is from (Song et al., 2021). Taken the CCS of luteolin-6-C-glucoside as a reference value, the CCS deviations are -5.2% and 5.0%, respectively, when comparing to the CCS values of luteolin-8-C-glucoside and luteolin-7-O-glucoside. Similar phenomena were also observed for other deprotonated flavonoid compounds, such as quercetin-3-O-glucoside (197.9 $Å^2$) and quercetin-4'-O-glucoside (204.4 $Å^2$) (Gonzales et al., 2016); quercetin 3-O-glucuronide (200.1 $Å^2$), quercetin-3'-O-glucuronide (196.1 $Å^2$) and quercetin-4'-O-glucuronide (208.8 $Å^2$) (Chalet et al., 2018); kaempferol-3-O-glucoside (199.6 $Å^2$) and kaempferol-7-O-glucoside (208.3 $Å^2$) (Schroeder et al., 2019). All these data indicate that the position of glycosyl groups on aglycones can affect the molecular size and conformations of flavonoids; it has the potential to achieve a complete identification for flavonoids by the incorporation of CCS values into HRMS-based screening strategies. More pairs of isomers and their CCS values are shown in Table I-2.

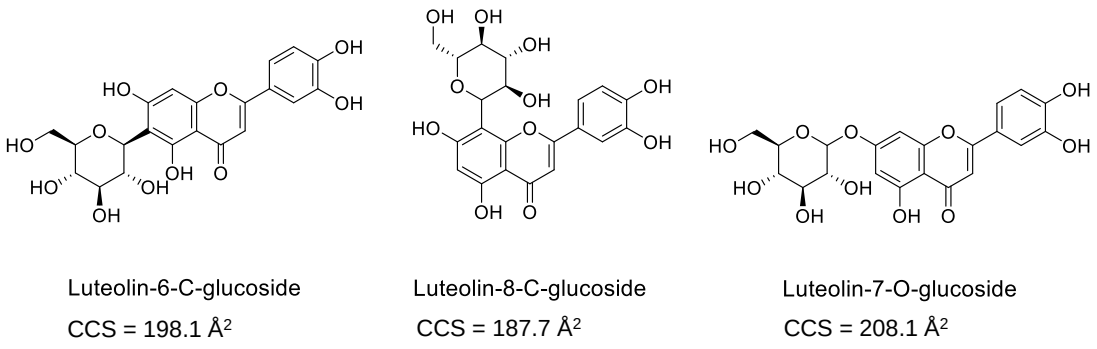


Figure I-5. Experimental collision cross section values of deprotonated luteolin glycosides.

Table I-2. Several isomeric pairs and their CCS values found in literature.

Isomer pairs	Adducts	Experimental CCS ($Å^2$)	Reference
isoorientin and orientin;	$[M-H]^-$	198.1 $Å^2$/187.7 $Å^2$	(McCullagh et al., 2019b)
isovitexin and vitexin		195.5 $Å^2$/188.8 $Å^2$	
morphine and piperine	$[M+H]^+$	164.0 $Å^2$/176.2 $Å^2$	(Lian et al., 2018)
15-acetyldeoxynivalenol and 3-acetyldeoxynivalenol	$[M+Na]^+$	176.7 $Å^2$/183.4 $Å^2$	(Righetti et al., 2020)
ethiofencarb sulfoxide	$[M+H]^+$	146.5 $Å^2$/156.9 $Å^2$	(Celma et al., 2021)
methiocarb sulfoxide;		221.5 $Å^2$/204.7 $Å^2$	
testosterone glucuronide			
epitestosterone glucuronide			
quercetin-3-O-glucoside	$[M-H]^-$	197.9 $Å^2$	(Gonzales et al., 2016)
quercetin-3'-O-glucoside		200.8 $Å^2$	
quercetin-4'-O-glucoside		204.4 $Å^2$	
kaempferol 3-glucuronide	$[M+H]^+$	199.7 $Å^2$/208.0 $Å^2$	(Connolly et al., 2021)
kaempferol 7-glucuronide;		197.7 $Å^2$/206.1 $Å^2$	
genistein-4'-Glucuronide		193.5 $Å^2$/205.9 $Å^2$	
genistein 7-Glucuronide;		248.5 $Å^2$/254.2 $Å^2$	
naringenin 4'-O-glucuronide			
naringenin-7-O-glucuronide;			
raloxifene 4'-Glucuronide			
raloxifene 6-Glucuronide			
kaempferol-3-O-glucoside	$[M-H]^-$	199.6 $Å^2$/208.3 $Å^2$	(Schroeder et al., 2019)
kaempferol-7-O-glucoside			
quercetin 3-O-glucuronide	$[M-H]^-$	200.1 $Å^2$/196.1 $Å^2$	(Chalet et al., 2018)
quercetin-3'-O-glucuronide		/208.8 $Å^2$	
quercetin-4'-O-glucuronide;		197.0 $Å^2$/ 211.0 $Å^2$	
kaempferol 3-glucuronide			
kaempferol 4'-glucuronide			
tri-m-tolyl phosphate	$[M+H]^+$	188.6 $Å^2$	(Belova et al., 2021)
tri-o-tolyl phosphate		182.4 $Å^2$	
tri-p-tolyl phosphate		190.0 $Å^2$	
PA 6 dimer and PA66 monomer;	$[M+H]^+$	150.4 $Å^2$/157.8 $Å^2$	(Schweighuber et al., 2021)
PA6 tetramer and PA66 dimer;		209.6 $Å^2$/214.7 $Å^2$	
PA6 hexamer and PA66 trimer		255.1 $Å^2$/261.1 $Å^2$	

Some structural isomers in FCCs were also identified by IMS, such as tri-m-tolyl

phosphate ($[M+H]^+$ 188.6 $Å^2$, $[M+Na]^+$ 198.6 $Å^2$), tri-o-tolyl phosphate ($[M+H]^+$ 182.4 $Å^2$, $[M+Na]^+$ 192.4 $Å^2$) and tri-p-tolyl phosphate ($[M+H]^+$ 190.0 $Å^2$, $[M+Na]^+$ 200.0 $Å^2$) (Belova et al., 2021). However, in some cases, the CCS deviations between isomers are less than 2%, which cannot effectively resolved by current IMS, such as tri-n-butyl phosphate ($[M+H]^+$ 165.4 $Å^2$) and triisobutyl phosphate ($[M+H]^+$ 166.7 $Å^2$) (Belova et al., 2021); diisoalkyl phthalates and dialkyl phthalates (Song et al., 2022b). The definitive identification of such isomers may require an IMS device with higher resolving power and better reproducibility (providing CCS deviations < 0.5%).

6.4.4. Reducing false positive candidates

CCS can also be used for reducing the false positive identifications in suspect screenings. In the study of Zhou et al. (2020), 76% of candidates were reduced with the addition of CCS match with a tolerance of 1%. On opening up the tolerance to 2%, approximately 54% of candidates were reduced. Bijlsma and coworkers proved that predicted CCS values from machine learning approaches permitted to reduce 5%-39% of false positives in suspect screening analysis, using a filter tolerance of 6% (Bijlsma et al., 2017). In principle, the number of false positives reduced by applying a CCS filter depending on the tolerance selected. Currently, a tolerance of 2% is always used for experimental CCS match in suspect and untargeted screening, based on its interlaboratory and interplatform reproducibility (Hernandez-Mesa et al., 2020). On the other hand, a tolerance of 5% or 6% is used for predicted CCS match given the current prediction accuracy (Bijlsma et al., 2017; Bijlsma et al., 2019).

6.4.5. Improving confidence level of identification

CCS can provide an additional identification point for analytes in addition to RT, *m/z* and fragment ions. This is rather important when the analytes are in low abundance where there is insufficient ion intensity for fragment ions formation, or when reference RT cannot be obtained due to the unavailability of commercial standards, or when the RT deviations between samples and standards are higher than 0.1 min due to the matrix

effect. As an illustrative example, in the study of Celma et al. (2021), relatively high RT deviations in the range of 0.14 min to 0.30 min were observed for imazalil. However, its CCS deviations were consistently less than 2%, giving higher confidence for the identification. In the study of Bijlsma et al. (2017), a feature was tentatively identified as valifenalate based on its *m/z*, characteristic chlorine isotopic pattern, and fragment ions. However, no reference standard was available for this compound, and a predicted CCS value obtained from machine learning approach was compared to the measured CCS values, obtaining more confidence in its identity.

6.5. Experimental CCS database

CCS has been demonstrated to be extremely reproducible across different instrumentations and labs. The study of Righetti et al. (2020) indicated that the $^{TW}CCS_{N2}$ measurements between two Vion platforms from different laboratories showed the deviations less than 1.5%. Additionally, 96.4% of $^{TW}CCS_{N2}$ values measured on Vion and Synapt platforms have deviation within 2%. An average of 0.29% relative standard deviations (RSDs) and an average, absolute bias of 0.54% were also observed for $^{DT}CCS_{N2}$ in an interlaboratory study (Stow et al., 2017). In addition, Hinnenkamp et al. (2018) compared the CCS values determined by TWIMS and DTIMS, finding that 93% of $[M+H]^+$ ions and 87% of $[M+Na]^+$ ions have CCS deviations lower than 2%. The high reproducibility of CCS makes it a reliable molecular identifier to be incorporated into HRMS-based screening workflows. Until now, there are several experimental CCS databases available for different types of small molecules, such as phenolics, mycotoxins, steroids, food and environmental contaminants. Table I-3 introduces several open-access experimental CCS databases.

Table I-3. Several published experimental CCS databases.

Compound type	Technology	Number of CCS records	Reference
Steroids	$^{TW}CCS_{N2}$	1080 CCS values from 300 steroids	(Hernandez-Mesa et al., 2018)
Drugs and drugs-like compounds	$^{TW}CCS_{N2}$	1440 CCS values from	(Hines et al., 2017)

Compound type	Technology	Number of CCS records	Reference
		1425 drug-like compounds	
Human and veterinary drugs	$^{TW}CCS_{N2}$	173 CCS values from 92 human and veterinary drugs	(Tejada-Casado et al., 2018)
Abused drugs and toxic compounds	$^{DT}CCS_{N2}$	124 CCS values from 124 abused drugs, of which 122 drugs are detected as $[M+H]^+$ adducts	(Lian et al., 2018)
Pesticides	$^{TW}CCS_{N2}$	205 CCS values from 205 pesticides, with $[M+H]^+$ adducts	(Bijlsma et al., 2017)
Pesticides	$^{TW}CCS_{N2}$	224 CCS values from 224 pesticides	(Regueiro et al., 2016)
Organic micropollutants in environmental samples	$^{TW}CCS_{N2}$	970 CCS values from 556 compounds, including illicit drugs, pesticides, hormones, mycotoxins, psychoactive substances, pharmaceuticals	(Celma et al., 2020)
Mycotoxins	$^{TW}CCS_{N2}$	106 CCS values from 61 mycotoxins	(Righetti et al., 2018)
Plant metabolomics	$^{TIMS}CCS_{N2}$	343 CCS values from 146 plant natural products	(Schroeder et al., 2019)
Compounds of emerging concern in human matrices	$^{DT}CCS_{N2}$	311 CCS values from 148 compounds, including plasticizers, flame retardants, bisphenols, etc.	(Belova et al., 2021)
Pollutants in indoor dust	$^{TW}CCS_{N2}$	29 CCS values from 29 compounds, including flame retardants, pesticides	(Mullin et al., 2020)
Chemicals in food contact materials	$^{TW}CCS_{N2}$	1038 CCS values from 675 standards, including plasticizers, antioxidants, photoinitiators, etc.	(Song et al., 2022b)

In these 12 CCS databases, 9 databases are obtained from TWIMS, 2 databases are from DTIMS, the last one is measured by TIMS. It should be mentioned that in the CCS database of steroids (Hernandez-Mesa et al., 2018), the CCS values of dimeric ionic species, such as $[2M + H]^+$, $[2M + Na]^+$, and $[2M - H]^-$, were also included. Some databases are only developed under positive electrospray ionization (ESI) mode (Bijlsma et al., 2017; Hines et al., 2017; Lian et al., 2018; Regueiro et al., 2016). These databases mainly contain pesticides, drugs and drug-like compounds, with carbonyl oxygen, amine, or ether oxygen in their structure. The CCS database of plant metabolomics was established in negative ESI mode; this database mainly contains plant natural products, such as phenolic compounds and other plant metabolites (Schroeder et al., 2019).

Although several CCS databases have been established for small molecules, there remain many compounds that are not included in such libraries. In this case, theoretical CCS values from machine learning approaches can be alternatives to be used for suspect and untargeted screening analysis. In the following section, some public CCS prediction tools are introduced.

6.6. CCS prediction

The fundamental principle of CCS prediction is to use machine learning approaches to correlate the experimental CCS values with corresponding molecular descriptors, then this correlation is used to predict the CCS values of unknows with the help of calculated molecular descriptors. The predicted CCS values are valuable and necessary due to the unavailability of experimental CCS values for many compounds. Figure I-6 shows the general workflow for CCS prediction using machine learning approaches. The first step, as well as the most challenging one, is collecting and curating all the experimental CCS data of interest. If the goal of model is to achieve accurate CCS predictions for diverse compounds, then the training set of model should contain compounds of diverse chemical classes, and cover a wide chemical space. A total of 3539, 7325, 7405 and 2439 CCS records belonging to different chemical classes were

used to develop AllCCS (Zhou et al., 2020), CCSondemand (Broeckling et al., 2021), CCSbase (Ross et al., 2020) and DeepCCS (Plante et al., 2019), respectively. On the other hand, if the goal is to build a CCS prediction model for specialized compounds, the structurally similar compounds should be included in the training set. For example, 458 CCS values from lipids were used to develop CCS prediction tool for lipid and lipid-like compounds (Zhou et al., 2017). After the curation of experimental CCS data, the molecular descriptors are calculated, then the dataset is split into training and testing set, training set is used for the calibration and optimization of the model, and testing set is used for the external validation and avoiding data overfitting.

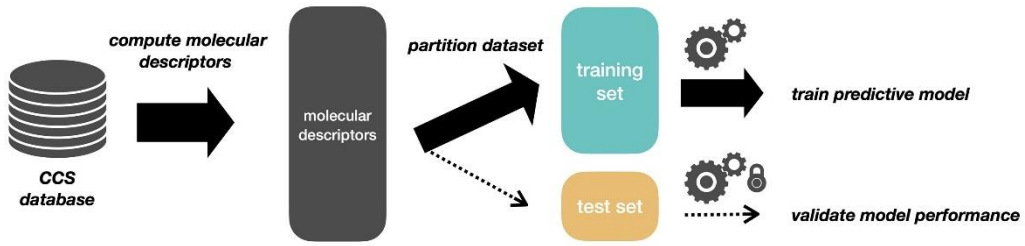


Figure I-6. General workflow for CCS prediction using machine learning approaches (Ross and Xu, 2021).

In recent few years, several comprehensive or specialized CCS prediction tools have been developed by different laboratories and groups (see Table I-4). The first CCS prediction tool: MetCCS, was developed by Zhou and coworkers in 2016 (Zhou et al., 2016), of which 796 CCS values of metabolites were correlated with 14 molecular descriptors, with the help of support vector machine (SVM). After that, the same researcher continuously introduced a specialized CCS prediction tool for lipids in 2017: LipidCCS (Zhou et al., 2017), and a more accurate CCS prediction webserver for metabolites in 2020: AllCCS (Zhou et al., 2020). However, high prediction errors were found when AllCCS model was applied to the prediction of CCS values of mycotoxins (Righetti et al., 2020), and FCCs (Song et al., 2022b). This is possibly due to the structural dissimilarity between the training set of AllCCS and the predicted molecules. Considering that the comprehensive CCS prediction tools may provide less accurate

prediction results for a specific class of molecules, some laboratories have developed their own CCS prediction tools for the compounds of interest to them, such as pesticides (Bijlsma et al., 2017), phenolics (Gonzales et al., 2016) and drugs (Mollerup et al., 2018).

Table I-4. Several public CCS prediction tools.

Tools	Training data	Algorithm	Application	Reference
CCSondemand	7325 $^{TW}CCS_{N2}$ values from 3775 compounds	XGBoost	$[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M-H]^-$, $[M+HCOO]^-$,	(Broeckling et al., 2021)
AllCCS	3539 CCS values from 2193 compounds	SVM	$[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M+H-H_2O]^+$, $[M-H]^-$, $[M+HCOO]^-$, $[M+Na-2H]^-$	(Zhou et al., 2020)
CCSbase	7405 CCS entries from 3526 small molecules, 1041 lipids, 91 peptides, 84 carbohydrates	SVM	$[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M+K]^+$, $[M-H]^-$, $[M+Na-2H]^-$, $[M]^+$, $[M]^-$	(Ross et al., 2020)
LipidCCS	458 CCS values	SVM	$[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M-H]^-$, $[M+HCOO]^-$	(Zhou et al., 2017)
MetCCS	796 CCS values	SVM	$[M+H]^+$, $[M+Na]^+$, $[M+H-H_2O]^+$, $[M-H]^-$, $[M+Na-2H]^-$	(Zhou et al., 2016)
DeepCCS	2439 CCS values	DNN	$[M+H]^+$, $[M+Na]^+$, $[M-H]^-$, $[M-2H]^{2-}$	(Plante et al., 2019)
CCS prediction for pesticides	205 CCS values of pesticides	ANN	$[M+H]^+$, $[M+Na]^+$	(Bijlsma et al., 2017)
CCS prediction for phenolics	56 CCS values of phenolics	PLS	$[M-H]^-$	(Gonzales et al., 2016)
CCS prediction for drugs	357 CCS values of pharmaceuticals, drugs	ANN	$[M+H]^+$	(Mollerup et al., 2018)
CCS prediction for food contact chemicals	635 CCS values of food contact chemicals	SVM	$[M+H]^+$, $[M+Na]^+$	(Song et al., 2022c)

Note: XGBoost: extreme gradient boosting; SVM: support vector machine; DNN: deep neural network; ANN: artificial neural

networks; PLS: partial least squares regression

There are three main factors affecting the accuracy of predicted CCS values: training set, molecular descriptors, and algorithms. Generally, the training set of CCS prediction model should be representative of the chemical characteristics for the intended applications. Zhou et al. (2020) excluded the lipid-like compounds from the training set and rebuilt the model. They found that the predicted CCS values of lipid-like compounds in testing set showed significantly larger errors after the exclusion of lipid from training set, while the prediction results of other super classes remain similar to that of original model. This confirmed that the chemical space of training set contributes a lot to the prediction accuracy. In addition, the training set of CCS prediction model may contain experimental CCS values gathered from different IMS devices, the discrepancies between different instrument types can affect the accuracy of the model. Hinnenkamp et al. (2018) compared the $^{TW}CCS_{N2}$ and $^{DT}CCS_{N2}$, finding that 7% $[M+H]^+$ ions and 13% $[M+Na]^+$ ions had CCS deviations higher than 2%, this result indicates that CCS database cannot be used without caring their types. However, in some cases, we have to say that the utilization of CCS values from different IMS types for model development is unavoidable, as the experimental CCS values of some kinds of compounds are only measured by a single type of IMS. The curation of different types of CCS records into a single dataset can provide high chemical diversities, which may compensate the CCS variations introduced. Recently, an improved CCS calibration approach has been proposed for the TWIMS system by Richardson and coworkers (Richardson et al., 2021), which has the potential to further improve the consistency between experimental $^{TW}CCS_{N2}$ and $^{DT}CCS_{N2}$ values, thus leading to a more accurate CCS prediction model.

Molecular descriptors used in the model can also affect the prediction accuracy. Currently, the molecular descriptors in the model are mainly calculated from neutral molecules. However, the CCS values are measured based on the mobility of ionized molecules, such as $[M+H]^+$ and $[M+Na]^+$ ions, this will inevitably introduce prediction

errors. The sodium ion has a higher atomic radius compared to a proton (Righetti et al., 2018), thus the conformational difference between sodium adducts and neutral molecules could be larger than the difference between protonated adducts and neutral molecules. This observation can support the findings that the predicted CCS values of sodium adducts always show larger errors than that of protonated adducts (Bijlsma et al., 2017; Song et al., 2022c).

Few researches studied the effect of algorithms on the accuracy of CCS prediction. In the work of Gonzales and coworkers (Gonzales et al., 2016), three algorithms: stepwise multiple linear regression (SMLR), principal components regression (PCR), and partial least squares regression (PLS), were used to develop the CCS prediction model; the result showed that PCR and PLS provided more accurate predictions than SMLR. In the study of Song et al. (2022c), support vector machine (SVM) was found to be able to provide more accurate prediction results than PLS. In addition, SVM also outperformed the least absolute shrinkage and selection operator (Lasso) and random forest (RF) regression in the study of Ross et al. (2020). SVM was also used for the development of AllCCS, CCSbase, MetCCS and LipidCCS (see Table I-4). The wide application of SVM is due to its easy configuration with few hyperparameters, as well as its ability to provide accurate and reproducible prediction results.

The current accuracy of predicted CCS values can reach around 5% or 6% deviations compared to measured values for many compounds. In the study of Bijlsma et al. (Bijlsma et al., 2017), over 95% of predicted CCS values of protonated molecules had prediction errors less than 6%. In another study of Song et al. (2022c), more than 92% of protonated molecules showed prediction errors less than 5%. The application of predicted CCS values in suspect and untargeted screenings allows to discard some false positives and improve the confidence level of identifications. However, the predicted CCS cannot be used to confirm the identification of unknown compounds, as many isomeric pairs have CCS differences less than 5% (Connolly et al., 2021). An improvement of IMS reproducibility can lead to more precise and accurate experimental

CCS values, thus can further improve the accuracy of predicted CCS values.

Section II: Objectives

1. General objectives

The main objective of this Thesis is to provide the analytical tools for identifying the unknown non-volatile compounds in untargeted specific migration analysis. To achieve the objective, instrumental devices such as ultrahigh performance liquid chromatography coupled to ion mobility and high-resolution spectrometry will be used for getting the experimental data and in silico prediction tools will be developed to built a useful system and a wide library able to help in the identification of migrants.

To comply with this goal, several specific objectives can be highlighted as follows:

2. Specific objectives

- To Build a comprehensive $^{TW}CCS_{N2}$ database for extractables and leachables found in FCMs in both positive and negative ion mode.

- To Compare the $^{TW}CCS_{N2}$ values of the three main adducts ($[M+H]^+$, $[M+Na]^+$, $[M-H]^-$) to previously published CCS values, in order to check the interlaboratory and interplatform reproducibility of CCS measurements.

- To Compare the experimental $^{TW}CCS_{N2}$ values in our database to the predicted CCS values from three prediction tools, in order to evaluate the applicability of CCS prediction tools in the field of FCMs.

- To develop a CCS prediction tool for chemicals in FCMs., studying the factors that can affect the prediction accuracy.

- To Collect more CCS values of plastic-related chemicals from and investigate the effect of algorithms and molecular descriptors on the CCS prediction.

- To characterizing the plastic-related chemicals in Ebro River water by using predicted CCS values.

- To Build a RT prediction model for FCCs in a LC-IM-QToF platform, prioritizing and identifying the non-volatile compounds in PA using RT and CCS prediction models.

- **To** Study the migration of PA oligomers from kitchenware utensils to food and quantifying the concentrations of migrated PA oligomers and evaluate their safety.

- **To** characterize the phenolic compounds in *Arctostaphylos uva-ursi* Leaves and evaluate the advantages and limitations of IMS-HRMS in the identification of phenolics.

Section III: Experimental part

Development of CCS database and in-silico prediction tools for identifying non-volatile compounds in FCMs

Chapter 1: *A Collision Cross Section Database for Extractables and Leachables from Food Contact Materials*

Chapter 2: *Prediction of Collision Cross Section Values: Application to Non-Intentionally Added Substance Identification in Food Contact Materials*

Chapter 3: *Prediction of Collision Cross Section Values for Extractables and Leachables from Plastic Products*

Chapter 4: *A Workflow for the Identification of Non-Volatile Migrates from Food Contact Materials Based on Ion Mobility−High Resolution Mass Spectrometry and In-Silico Prediction Tools*

Chapter 5: *The use of ion mobility time-of-flight mass spectrometry to assess the migration of polyamide 6 and polyamide 66 oligomers from kitchenware utensils to food*

The application of ion mobility in the characterization of phenolic compounds

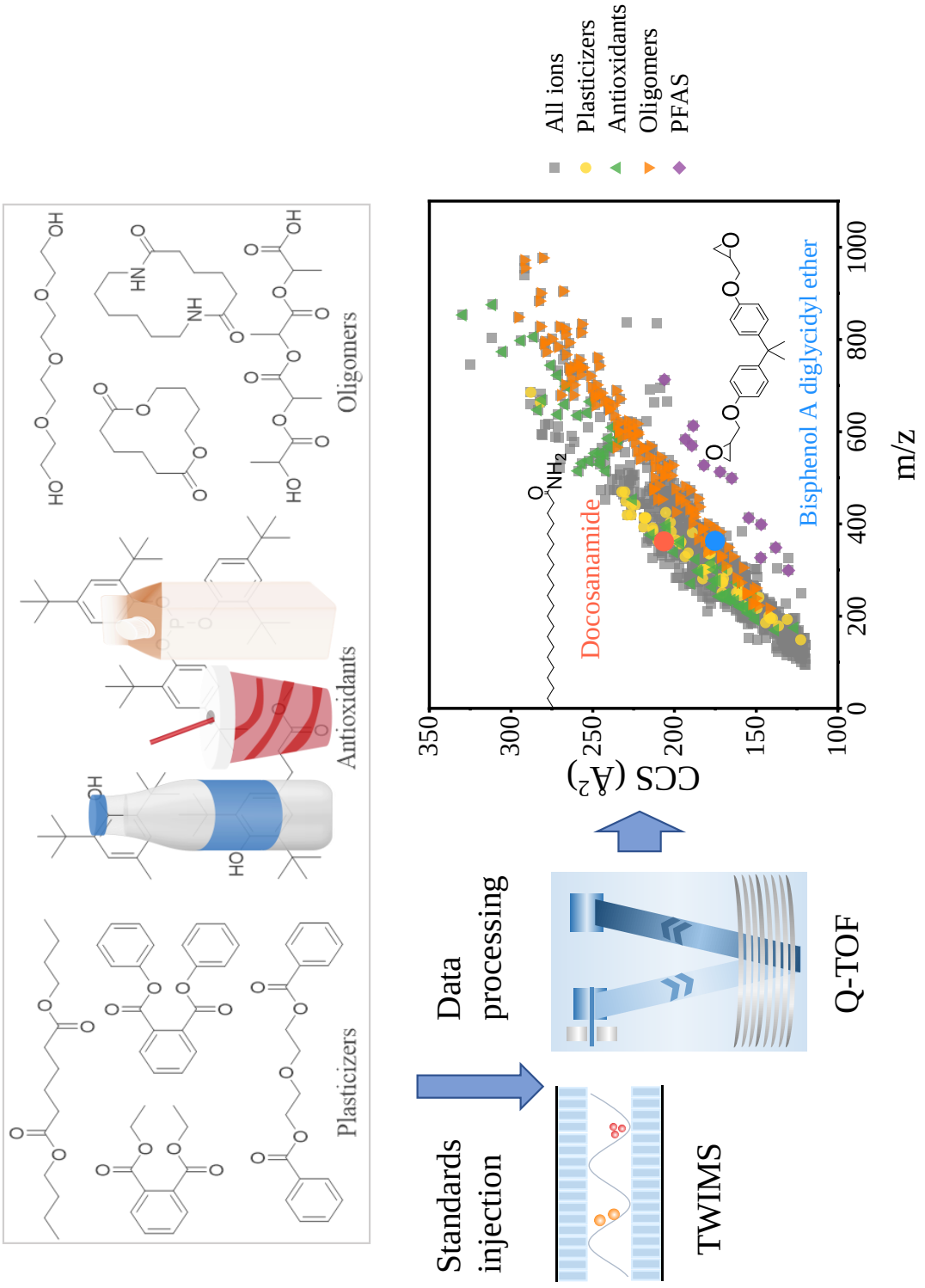
Chapter 6: *Discovery and Characterization of Phenolic Compounds in Bearberry (Arctostaphylos uva-ursi) Leaves Using Liquid Chromatography- Ion Mobility-High-Resolution Mass Spectrometry*

Chapter 1

A Collision Cross Section Database for Extractables and Leachables from Food Contact Materials

1. Abstract

The chemicals in food contact materials (FCMs) can migrate into food and endanger human health. In this study we developed a database of $^{TW}CCS_{N2}$ values for extractables and leachables from FCMs. The database contains a total of 1038 $^{TW}CCS_{N2}$ values from 675 standards including those commonly used additives and NIAS in FCMs. The $^{TW}CCS_{N2}$ values in the database were compared to previously published values, 85.7%, 87.7% and 64.9% $[M+H]^+$, $[M+Na]^+$ and $[M-H]^-$ adducts showed deviations <2%, with the presence of protomers, post-IMS dissociation of noncovalent clusters and inconsistent calibration possible sources of CCS deviations. Our experimental $^{TW}CCS_{N2}$ values were also compared to CCS values from three prediction tools. Of the three, CCSondemand gave the most accurate predictions. The $^{TW}CCS_{N2}$ database developed will aid the identification and differentiation of chemicals from FCMs in targeted and untargeted analysis.


Oligomers
Antioxidants
Plasticizers
Standards injection
TWIMS
Data processing
Q-TOF
CCS (Å²)
m/z
All ions
Plasticizers
Antioxidants
Oligomers
PFAS
Docosanamide
Bisphenol A diglycidyl ether

2. Introduction

Food contact materials (FCMs) are important sources of contaminations of the food. The chemical constituents of FCMs, termed food contact chemicals (FCCs), can be classified into two categories: intentionally added substances (IAS) and non-intentionally added substances (NIAS). IAS include known additives, including plasticizers, antioxidants, photoinitiators, lubricants and slip agents that are added to FCMs during processing in order to confer favorable characteristics and extend service life. NIAS can be broadly grouped into three categories: side products, breakdown products and contaminants. Side products form due to the incomplete polymerization of starting substances (Ubeda et al., 2018) or the interaction between migrants and food. (Canellas et al., 2021a). Breakdown products arise from the degradation of polymers and additives during manufacture and use (Liu and Mabury, 2018c; Yang et al., 2016). The origins of contaminants include the manufacturing process, shelf life and the recycling process (Su et al., 2021a; Su et al., 2021b). All these compounds can potentially migrate into food and pose a risk to the health of consumers (Biryol et al., 2017; Groh et al., 2021; Hahladakis et al., 2018).

It is challenging to achieve a full identification of FCCs in FCMs due to the high complexity of matrices. In the study by Zimmermann et al. (2021), only ~8% of detected features were identified by ultrahigh performance liquid chromatography coupled to a quadrupole-time-of-flight mass spectrometer (UPLC-QToF), indicating most of the chemicals from plastics remain unknown. Regulation (EC) No. 1935/2004 establishes that the substances in FCMs cannot migrate into food in quantities large enough to endanger human health (EC, 2004). Therefore, the FCCs in FCMs must be identified and quantified.

The coupling of ion mobility spectrometry (IMS) with HRMS provides a powerful tool for the identification and separation of small molecules commonly found in the food industry and environmental analyses, including mycotoxins (Righetti et al., 2018; Righetti et al., 2020). Pesticides (Bijlsma et al., 2017; Regueiro et al., 2017), drug and

drug-like compounds (Hines et al., 2017), phenolics (Song et al., 2021; Stander et al., 2017) and FCCs (Canellas et al., 2021b; Vera et al., 2019). It has been reported that some structural isomers (McCullagh et al., 2019b) and stereoisomers (Zheng et al., 2017b) can be separated by IMS. Collision cross section (CCS) is a parameter derived from the drift time (DT) using a power-law calibration for the traveling wave IMS (TWIMS) device. CCS measurement is consistent across different IMS platforms and laboratories (Hinnenkamp et al., 2018; Righetti et al., 2020; Stow et al., 2017). Hinnenkamp et al. (2018) compared the CCS values determined by TWIMS and drift tube IMS (DTIMS), finding that 93% of protonated adducts and 87% of sodiated adducts have deviations in the CCS values lower than 2%. The study of Righetti et al. (2020) indicated that the $^{TW}CCS_{N2}$ measurements of all ion species showed the deviations of less than 1.5% between two Vion platforms from different laboratories. Additionally, the deviation of $^{TW}CCS_{N2}$ values was within 2% for 96.4% of ions measured on Vion and Synapt platforms. The high reproducibility of CCS makes it a reliable parameter for inclusion in mass-spectral libraries. In addition to RT and fragments ion information, including CCS data in the identification process will improve confidence thereby reducing the number of tentative identifications.

Recently several open-source, experimental CCS databases have been constructed for mycotoxins (Righetti et al., 2018), steroids (Hernandez-Mesa et al., 2018), phenolics (Schroeder et al., 2019), pesticides (Bijlsma et al., 2017; Regueiro et al., 2016), drugs and drug-like compounds (Hines et al., 2017) and organic environmental micropollutants (Belova et al., 2021; Celma et al., 2020; Mullin et al., 2020). Additionally, some research groups have developed machine-learning-based tools to predict the CCS values of molecules. These include CCSondemand (Broeckling et al., 2021), AllCCS (Zhou et al., 2020), CCSbase (Ross et al., 2020), and DeepCCS (Plante et al., 2019).

The goal of this study was to build a comprehensive $^{TW}CCS_{N2}$ database for extractables and leachables found in FCMs in both positive and negative ion mode,

comprising of commonly used additives (e.g., plasticizers, antioxidants, photoinitiators and lubricants) and NIAS (degradation products of additives and oligomers). The $^{TW}CCS_{N2}$ values in our database were compared to previously published CCS measurements and reasons explaining high deviations, in some instances, are discussed. In addition, the experimental $^{TW}CCS_{N2}$ values in our database were compared to the predicted CCS values from three prediction tools, in order to evaluate the applicability of CCS prediction tools in the field of FCMs.

3. Materials and methods

3.1. Chemicals and Reagents.

Standards of commonly used additives in FCMs were purchased from Sigma-Aldrich Quimica S.A. (Madrid, Spain), Extrasynthese (Genay, France) and Cayman Chemical Company (Ann Arbor, Michigan, USA). Oligomers of adhesives, polyamide (PA) and polylactic acid (PLA) were isolated from associated polymers in our laboratory. The standard stock solutions at a concentration of 1000 mg kg^{-1} were prepared by dissolving 10 mg of standards with 10 g of methanol, using an electronic accurate balance from Mettler Toledo (XS205, 0.1 mg, Greifensee, Switzerland). If the standards were not dissolved in methanol, other solvents, such as ethanol, dichloromethane or dimethyl sulfoxide, were used. The measured $^{TW}CCS_{N2}$ values would not be affected by the solvents, as it is independent from sample matrices (Righetti et al., 2018). The working solutions at ~1 mg kg^{-1} were prepared by the dilution of 10 μL of stock solution with 10 mL of methanol. Each working solution contained a mixture of 8-10 analytes and all the mixtures were kept in dark at -20 ºC until analysis.

HPLC grade methanol (≥99.9%), ethanol (≥99.9%), dichloromethane (≥99.8%) and dimethyl sulfoxide (≥99.8%) were purchased from Scharlau Chemie S.A (Sentmenat, Spain). Ultrapure water was produced by a Millipore Milli-QPLUS 185 system (Madrid, Spain).

3.2. UPLC-IMS-QToF Analysis.

The working solutions at ~1 mg kg^{-1} were measured using an Acquity I-Class UPLC system coupled to a Vion IMS-QToF mass spectrometer (Waters, Manchester, UK). The chromatographic separation was performed using a CORTECS C18 column (2.1 × 100 mm, 1.6 µm particle size, 90 Å pore size) at a flow rate of 0.3 mL min^{-1}. Mobile phases were water (A) and methanol (B), both acidified with 0.1% of formic acid (v/v). The initial proportion of B was 5%, increased to 100% over 7 minutes, kept at 100% from 7 to 11 minutes, decreased to 5% over 0.1 minutes and re-conditioned until 13 minutes.

Data were acquired on the mass spectrometer in both positive and negative ion modes over the mass range of 50-1200 *m/z* with a scan time of 0.2 s. Electrospray ionization (ESI) conditions were as follows: capillary voltage, 1 kV; cone voltage, 30 V; source temperature, 120 ºC; desolvation temperature, 500 ºC; cone gas flow, 50 L h^{-1}; desolvation gas flow, 800 L h^{-1}. Data were acquired in high definition MS^E mode, with the instrument was switching between two collision energy states (low energy: 6 eV, high energy ramp: 20-40 eV) in order to obtain precursor and fragment ions within a single acquisition. Leucine-Enkephalin ($[M+H]^+$, *m/z* 556.2766 and $[M-H]^-$, *m/z* 554.2620) at a concentration of 100 ng/mL was infused at a rate of 15 µL/min for real-time mass correction. IM separations were performed with a travelling wave velocity of 250 m/s and IMS pulse height of 45 V, N_2 was used as the drift gas at a flow of 25 mL/min. The Vion platform works at a room temperature of 25 ºC.

The Major Mix IMS/ToF Calibration Kit (ref. 186008113) from Waters (Manchester, UK) was used for the CCS calibration. The calibration compounds and their CCS values in positive and negative ionization mode are shown in Table III-1.1 and Table III-1.2, respectively. In positive ion mode, the TWIMS platform was calibrated with polyalanine and nine drug-like compounds with a *m/z* range of 151.1−1154.6 Da and a CCS range of 130.4−333.6 $Å^2$. In negative ion mode, two fluoroalkanoic acids were added in the calibration mix, providing a *m/z* range of 151.1-

1167.0 Da and a CCS range of 130.1-322.4 Å^2. A quality control (QC) solution (Vion Test Mix, ref. 186008462) from Waters (Manchester, UK) was systematically injected before and after each batch of standard solutions to monitor the system stability. Detailed information about the nine compounds in QC solution are shown in Table III-1.3. The variations of the *m/z* and $^{TW}CCS_{N2}$ measurements for the QC solution were less than 5 ppm and 2%, respectively.

Table III-1.1. Calibration substances in Major Mix IMS/Tof Calibration Kit and their CCS values in positive mode.

Compound	Formula	m/z $[M+H]^+$	CCS (Å^2)
Acetaminophen	$C_8H_9NO_2$	152.0706	130.4
Reserpine fragment	$C_8H_{10}N_4O_2$	195.0877	138.2
Sulfaguanidine	$C_7H_{10}N_4O_2S$	215.0597	146.8
Sulfadimethoxine	$C_{12}H_{14}N_4O_4S$	311.0809	168.4
Val-Tyr-Val	$C_{19}H_{29}N_3O_5$	380.2180	191.7
Verapamil	$C_{27}H_{38}N_2O_4$	455.2904	208.8
Terfenadine	$C_{32}H_{41}NO_2$	472.3210	228.7
Polyalanine, n=7	$C_{21}H_{37}N_7O_8$	516.2776	211.0
Leucine Enkephalin	$C_{28}H_{37}N_5O_7$	556.2766	229.8
Polyalanine, n=8	$C_{24}H_{42}N_8O_9$	587.3148	228.0
Reserpine	$C_{33}H_{40}N_2O_9$	609.2807	252.3
Polyalanine, n=9	$C_{27}H_{47}N_9O_{10}$	658.3519	243.0
Polyalanine, n=10	$C_{30}H_{52}N_{10}O_{11}$	729.3890	256.0
Polyalanine, n=11	$C_{33}H_{57}N_{11}O_{12}$	800.4261	271.0
Polyalanine, n=12	$C_{36}H_{62}N_{12}O_{13}$	871.4632	282.0
Polyalanine, n=13	$C_{39}H_{67}N_{13}O_{14}$	942.5003	294.0
Polyalanine, n=14	$C_{42}H_{72}N_{14}O_{15}$	1013.5374	306.0
Polyalanine, n=15	$C_{45}H_{77}N_{15}O_{16}$	1084.5746	321.5
Polyalanine, n=16	$C_{48}H_{82}N_{16}O_{17}$	1155.6117	333.6
Ultramark 1621	$C_{20}H_{18}O_6N_3P_3F_{28}$	1022.0034	263.1
Ultramark 1621	$C_{22}H_{18}O_6N_3P_3F_{32}$	1121.9970	276.5

Table III-1.2. Calibration substances in Major Mix IMS/Tof Calibration Kit and their CCS values in negative mode.

Compound	Formula	m/z $[M-H]^-$	CCS (Å^2)
Acetaminophen	$C_8H_9NO_2$	150.0561	131.5

Theophylline	$C_7H_7N_4O_2$	179.0575	132.4
Sulfaguanidine	$C_7H_{10}N_4O_2S$	213.0452	145.2
Sulfadimethoxine	$C_{12}H_{14}N_4O_4S$	309.0663	170.1
Val-Tyr-Val	$C_{19}H_{29}N_3O_5$	378.2034	192.5
Leucine Enkephalin	$C_{28}H_{37}N_5O_7$	554.2620	225.3
Perfluoroheptanoic acid -CO2	C_6F_{13}	318.9798	130.1
Perfluorooctanoic acid -CO2	C_7F_{15}	368.9766	137.2
Polyalanine, n=8	$C_{24}H_{42}N_8O_9$	585.3002	227.7
Reserpine	$C_{33}H_{40}N_2O_9$	607.2661	265.2
Polyalanine, n=9	$C_{27}H_{47}N_9O_{10}$	656.3373	242.1
Polyalanine, n=10	$C_{30}H_{52}N_{10}O_{11}$	727.3744	255.9
Polyalanine, n=11	$C_{33}H_{57}N_{11}O_{12}$	798.4115	268.5
Polyalanine, n=12	$C_{36}H_{62}N_{12}O_{13}$	869.4487	280.2
Polyalanine, n=13	$C_{39}H_{67}N_{13}O_{14}$	940.4858	294.6
Polyalanine, n=14	$C_{42}H_{72}N_{14}O_{15}$	1011.5228	308.8
Polyalanine, n=15	$C_{45}H_{77}N_{15}O_{16}$	1082.5600	322.4
Ultramark 1621	$C_{23}H_{20}O_8N_3P_3F_{32}$	1165.9880	275.8

Table III-1.3. Molecular formula, monoisotopic mass, retention time (RT) and $^{DT}CCS_{N2}$ of nine compounds in quality control (QC) solution.

Compound	Molecular formula	Monoisotopic mass (Da)	RT (min)	CCS (Å^2) $[M+H]^+$	CCS (Å^2) $[M-H]^-$
Acetaminophen	$C_8H_9NO_2$	151.0633	1.96	130.4	131.5
Caffeine	$C_8H_{10}N_4O_2$	194.0804	2.76	138.2	-
Sulfaguanidine	$C_7H_{10}N_4O_2S$	214.0524	1.02	146.8	145.2
Sulfadimethoxine	$C_{12}H_{14}N_4O_4S$	310.0736	3.83	168.4	170.1
Val-tyr-val	$C_{19}H_{29}N_3O_5$	379.2107	2.87	191.7	192.5
Verapamil	$C_{27}H_{38}N_2O_4$	454.2832	4.57	208.8	-
Terfenadine	$C_{32}H_{41}NO_2$	471.3137	5.64	228.7	-
Leucine-enkephalin	$C_{28}H_{37}N_5O_7$	555.2693	3.97	229.8	225.3
Reserpine	$C_{33}H_{40}N_2O_9$	608.2734	4.92	252.3	265.2

Data acquisition and processing were performed on UNIFI v.1.9.4. (Waters Corp.). Only the $^{TW}CCS_{N2}$ values of singly charged ions were considered and included values for $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M-Na+2H]^+$ and $[M-Cl]^+$ in positive mode and $[M-H]^-$ and $[M+HCOO]^-$ in negative mode.

3.3. Precision of $^{TW}CCS_{N2}$ Measurement.

In order to validate the inter-day precision of $^{TW}CCS_{N2}$ measurements, a mixed standard solution at ~1 mg kg^{-1} containing 38 representative IAS and NIAS in FCMs was injected once a week over a period of two months. The mixed standard solution contained plasticizers, antioxidants, photoinitiators, bisphenols and common degradation products, such as 3,5-di-tert-butyl-4-hydroxybenzaldehyde. Detailed information on these compounds is provided in Table III-1.4.

Table III-1.4. Thirty-eight compounds for validating the inter-day precision of CCS measurement.

No.	Compounds	PubChem CID	ESI (+)	ESI (-)
1	Stearic acid	5281		•
2	Antioxidant 425	6928	•	•
3	N-(4-Hydroxyphenyl)Stearamide	7689	•	•
4	Antioxidant 2246	8398	•	•
5	4-Hydroxybenzophenone	14347	•	•
6	11-Aminoundecanoic acid	17083	•	•
7	Irganox 1024	61916	•	•
8	Irganox 1035	64883	•	•
9	12-Aminododecanoic acid	69661	•	•
10	MOPS	70807	•	•
11	Tioxolone	72139	•	•
12	3,5-Di-tert-butyl-4-hydroxybenzaldehyde	73219	•	•
13	3-(3,5-Di-tert-butyl-4-hydroxyphenyl)propionic acid	88389	•	•
14	Antioxidant 1790	93221	•	•
15	UV-234	112412	•	•
16	Ethyl 2-cyano-3,3-diphenylacrylate	243274	•	
17	3,5-Di-tert-butyl-4-hydroxyacetophenone	616296	•	•
18	Bisphenol A bis(3-chloro-2-hydroxypropyl) ether	3479589	•	•
19	Bisphenol F bis(2,3-dihydroxypropyl) ether	3928015	•	•
20	Bisphenol A (3-chloro-2-hydroxypropyl) (2,3-dihydroxypropyl) Ether	4166922	•	•
21	2,2'-Dihydroxy-4-methoxybenzophenone	8569	•	•
22	Erucamide	5365371	•	
23	Dibutyl phosphate	7881	•	•
24	Methylparaben	7456		•
25	Propylparaben	7175		•
26	Diethyl phthalate	6781	•	

No.	Compounds	PubChem CID	ESI (+)	ESI (-)
27	Tinuvin P	17113	•	
28	Dibutyl sebacate	7986	•	
29	Diphenyl phthalate	6778	•	
30	Tinuvin 327	77470	•	•
31	Tri(p-cresyl) phosphate	6529	•	
32	Uvitex OB	292429	•	
33	Cyasorb 2908	94623	•	•
34	Irganox 1076	16386	•	•
35	Irganox 245	91620	•	•
36	Irganox 1098	90004	•	•
37	Tinuvin 360	3571576	•	•
38	Irganox 1330	74370	•	•

3.4. Comparison with Published CCS Measurements.

The comparison between $^{TW}CCS_{N2}$ values in our database and those obtained from the literature is crucial to determine whether our database could be used across different laboratories and instrumental types. Hence, several CCS databases and publications were consulted for reference CCS values of some of the compounds considered in this study (Belova et al., 2021; Bijlsma et al., 2017; Celma et al., 2020; Corporation, 2020a, b, 2021; Gonzales et al., 2016; Hernandez-Mesa et al., 2018; Hines et al., 2017; Hinnenkamp et al., 2018; Mullin et al., 2020; Nichols et al., 2018; Poland et al., 2020; Regueiro et al., 2016; Righetti et al., 2020). The CCS deviations (ΔCCS%) were calculated using the $^{TW}CCS_{N2}$ values in our database as the reference values.

3.5. Evaluation of Public CCS Prediction Tools.

CCS values predicted by machine learning algorithms can be used when empirical CCS values are not available. To evaluate the accuracy of existing CCS prediction tools for FCCs, the $^{TW}CCS_{N2}$ values of the compounds in our database were compared against those generated by three CCS prediction tools: AllCCS (http://allccs.zhulab.cn/) proposed by Zhou et al. (2020), CCSbase (https://ccsbase.net/) from Libin Xu Lab (Ross et al., 2020), and CCSondemand (https://ccs.on-demand.waters.com/) from

Broeckling and co-workers (Broeckling et al., 2021).

4. Results and discussions

4.1. CCS deviations of QC compounds

The $^{TW}CCS_{N2}$ database was built over the period from November 2018 to July 2021. A total of 76 and 24 batches of QC solutions were injected in positive and negative mode, respectively, during the database creation. The comparison between reference and experimental CCS values for QC compounds is shown in Table III-1.5, and the distributions of their CCS deviations are shown in Figure III-1.1. It can be seen that for both ion modes the average CCS variation was less than 1.1% and the relative standard deviations (RSDs) ranged from 0.5% to 0.8%. These data indicate a high degree of accuracy and reproducibility of the $^{TW}CCS_{N2}$ measurements over the course of almost three years. Acetaminophen presented relatively high CCS deviations in both ion modes, which is possibly due to its low *m/z* and CCS values. This observation highlights the importance of adding more data points to calibration curve for below *m/z* values below 150.

Table III-1.5. Comparison between reference and experimental CCS values for quality control compounds for all batches of standards.

	Compounds	Adducts	m/z	CCS_{ref} ($Å^2$)	CCS_{exp} ± SD ($Å^2$)	RSD (%)	error (%)
Positive	Acetaminophen	$[M+H]^+$	152.0706	130.4	131.8 ± 1.1	0.8	1.1
(n = 76)	Caffeine	$[M+H]^+$	195.0877	138.2	137.7 ± 1.0	0.7	-0.4
	Sulfaguanidine	$[M+H]^+$	215.0597	146.8	146.5 ± 0.9	0.6	-0.2
	Sulfadimethoxine	$[M+H]^+$	311.0809	168.4	168.0 ± 1.0	0.6	-0.3
	Val-tyr-val	$[M+H]^+$	380.2180	191.7	193.6 ± 1.0	0.5	1.0
	Verapamil	$[M+H]^+$	455.2904	208.8	210.2 ± 1.2	0.6	0.7
	Terfenadine	$[M+H]^+$	472.3210	228.7	231.0 ± 1.5	0.6	1.0
	Leucine-enkephalin	$[M+H]^+$	556.2766	229.8	228.9 ± 1.3	0.6	-0.4
	Reserpine	$[M+H]^+$	609.2807	252.3	252.3 ± 1.6	0.6	-0.0
Negative	Acetaminophen	$[M-H]^-$	150.0561	131.5	130.2 ± 0.9	0.7	-1.0
(n = 24)	Sulfaguanidine	$[M-H]^-$	213.0452	145.2	144.1 ± 0.9	0.7	-0.8

Compounds	Adducts	m/z	CCS_{ref} ($Å^2$)	CCS_{exp} ± SD ($Å^2$)	RSD (%)	error (%)
Sulfadimethoxine	$[M-H]^-$	309.0663	170.1	170.4 ± 1.1	0.6	0.2
Val-tyr-val	$[M-H]^-$	378.2034	192.5	194.1 ± 1.2	0.6	0.9
Leucine-enkephalin	$[M-H]^-$	554.2620	225.3	223.5 ± 1.6	0.7	-0.8
Reserpine	$[M-H]^-$	607.2661	265.2	267.9 ± 1.5	0.5	1.0

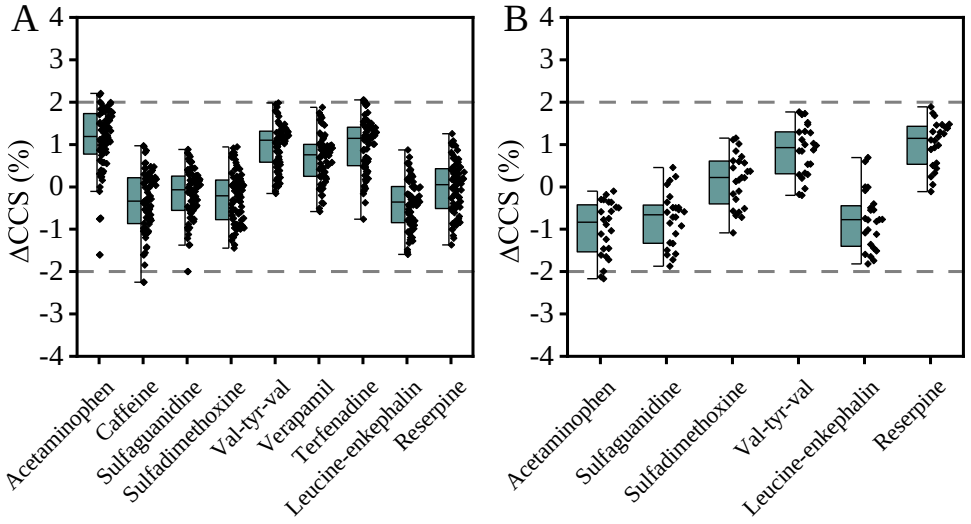


Figure III-1.1. CCS deviations of quality control compounds for all batches of standards, (A) positive ion mode and (B) negative ion mode.

4.2. $^{TW}CCS_{N2}$ Database Overview

This work presents a $^{TW}CCS_{N2}$ database with respect to extractables and leachables in FCMs, which consists of commonly used additives, degradation products of additives, oligomers and natural phenolic compounds from antioxidant active packaging. Detailed information about the chemicals in database, such as compound name, adduct, Monoisotopic Mass, Molecular Formula, Canonical SMILES, InChIKey and class can be seen in Supporting Information. Figure III-1.2A shows that a total of 1038 ions were detected for the 675 standards analyzed. The detected ions could be divided into two groups of 811 cations (446 $[M+H]^+$, 317 $[M+Na]^+$, 30 $[M+NH_4]^+$, 5 $[M-Na+2H]^+$ and 13 $[M-Cl]^+$) and 227 anions (190 $[M-H]^-$ and 37 $[M+HCOO]^-$). In positive ion mode, 580 compounds were detected, including the commonly used plasticizers, antioxidants, photoinitiators, primary aromatic amines, slip agents and

oligomers. These compounds contain either carbonyl oxygen, amine or ether oxygen in their structure. In negative ion mode, 205 compounds were detected, which included lubricants, hindered phenol antioxidants, bisphenols and perfluoroalkyl substances.

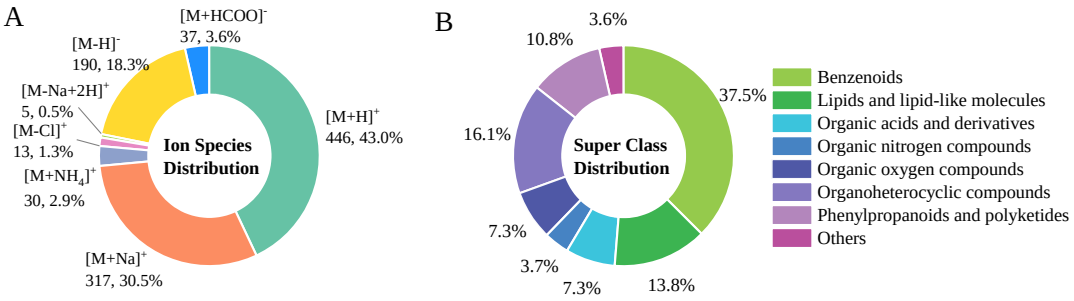


Figure III-1.2. (A) Distribution of the 1038 measured ions from positive and negative ionization modes; (B) Distribution of 675 detected compounds across super classes.

The super classes of the 675 standards analyzed were obtained using ClassyFire (Djoumbou Feunang et al., 2016) and the distribution of classes is shown in Figure III-1.2B. More compounds belong to the benzenoids super class than any other class in the database. This is unsurprising since commonly used additives in FCMs, such as antioxidants, biocides, bisphenols, nucleating agents, photoinitiators, phthalate-based plasticizers and UV absorbers belong to this super class. Lipid and lipid-like molecules and organoheterocyclic compounds also account for a large part of the database. The former contains lubricants, adipate-based and sebacate-based plasticizers, slip agents and fatty acid esters. The latter includes colorants, pesticides, drug-like compounds and UV absorbers.

A depiction of $^{TW}CCS_{N2}$ vs *m/z* for 1038 ions and the distribution of $^{TW}CCS_{N2}$ and *m/z* values are shown together in Figure III-1.3. The correlation between $^{TW}CCS_{N2}$ and *m/z* was described by power regression model, with R^2 of 0.882. $^{TW}CCS_{N2}$ values range from 119.6 Å^2 ($[M+H]^+$ of benzaldehyde) to 329.4 Å^2 ($[M+H]^+$ of 3,9-Bis(2,4-dicumylphenoxy)-2,4,8,10-tetraoxa-3,9-diphosphaspiro[5.5]undecane) and *m/z* values range from 94 Da ($[M+H]^+$ of aniline) to 977 Da ($[M+Na]^+$ of PLA 13). Figure III-1.3 shows that 95% of the measured $^{TW}CCS_{N2}$ values are accounted for in the *m/z* region

from 93 to 700 Da. $^{TW}CCS_{N2}$ values are mainly distributed in the range of 119-220 Å^2 which accounts for 83.3% of the measured $^{TW}CCS_{N2}$ values. Besides, 93.3% (968 out of 1038) of $^{TW}CCS_{N2}$ values located in the calibration range, the other 70 $^{TW}CCS_{N2}$ values were below the lowest CCS values of calibrates (130.4 Å^2 in positive mode and 130.1 Å^2 in negative mode).

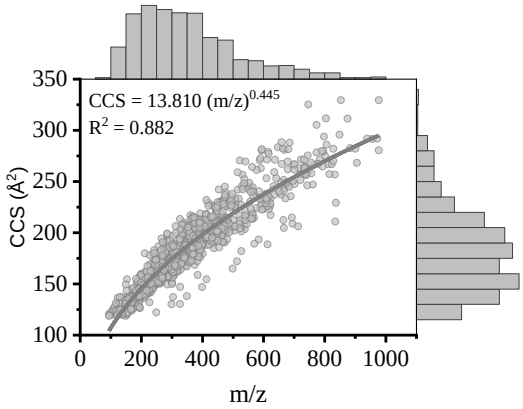


Figure III-1.3. Depiction of $^{TW}CCS_{N2}$ values vs *m/z* values for 1038 ions together with the distribution of $^{TW}CCS_{N2}$ and *m/z* values

4.3. CCS Distribution of Commonly Observed Additives and NIAS

Since CCS is a measurement related to the size, shape and charge of an ion (D'Atri et al., 2018), different relationships between CCS vs *m/z* have been observed for compounds presenting different structural characteristics (Belova et al., 2021; Hernandez-Mesa et al., 2018; Hines et al., 2017). Correlations between the CCS and *m/z* of the commonly used additives (plasticizers, antioxidants and photoinitiators) and oligomers studied here are shown in Figure III-1.4. Figure III-1.4A presents the $^{TW}CCS_{N2}$ vs *m/z* relationship of 57 ions from 49 plasticizers, with different colors denoting different types of plasticizers. It is evident that the trendline for adipates and sebacate-based plasticizers (10 $[M+Na]^+$ ions) have a steeper gradient than the trendlines for phthalates (6 $[M+H]^+$ and 24 $[M+Na]^+$ ions) and citrates (2 $[M+H]^+$ and 3 $[M+Na]^+$ ions). Adipates and sebacate-based plasticizers appear to have a more elongated structure due to their linear-chain molecules, which leads to a larger rotationally averaged collision area for a given *m/z*. The trendline for phthalates (i.e.

diesters of ortho-phthalic acid) has a shallower gradient. In general, the structures of this class of plasticizers contain both aryl and alkyl groups (e.g. dibutyl phthalate) and the compact aryl group will lead to a smaller $^{TW}CCS_{N2}$ value. This is supported by the lower $^{TW}CCS_{N2}$ values of benzyl butyl phthalate (BBP) and diphenyl phthalate (DPP) when compared to the $^{TW}CCS_{N2}$ values of other phthalates with a similar *m/z*. In the structures of these two compounds, one or two alkyl groups are replaced by aryl groups. Citrates have relatively lower $^{TW}CCS_{N2}$ values compared to phthalates, adipates and sebacate-based plasticizers with similar *m/z* values. This is may be due to the compact side chains in their structures, as demonstrated by Belova et al. (2021).

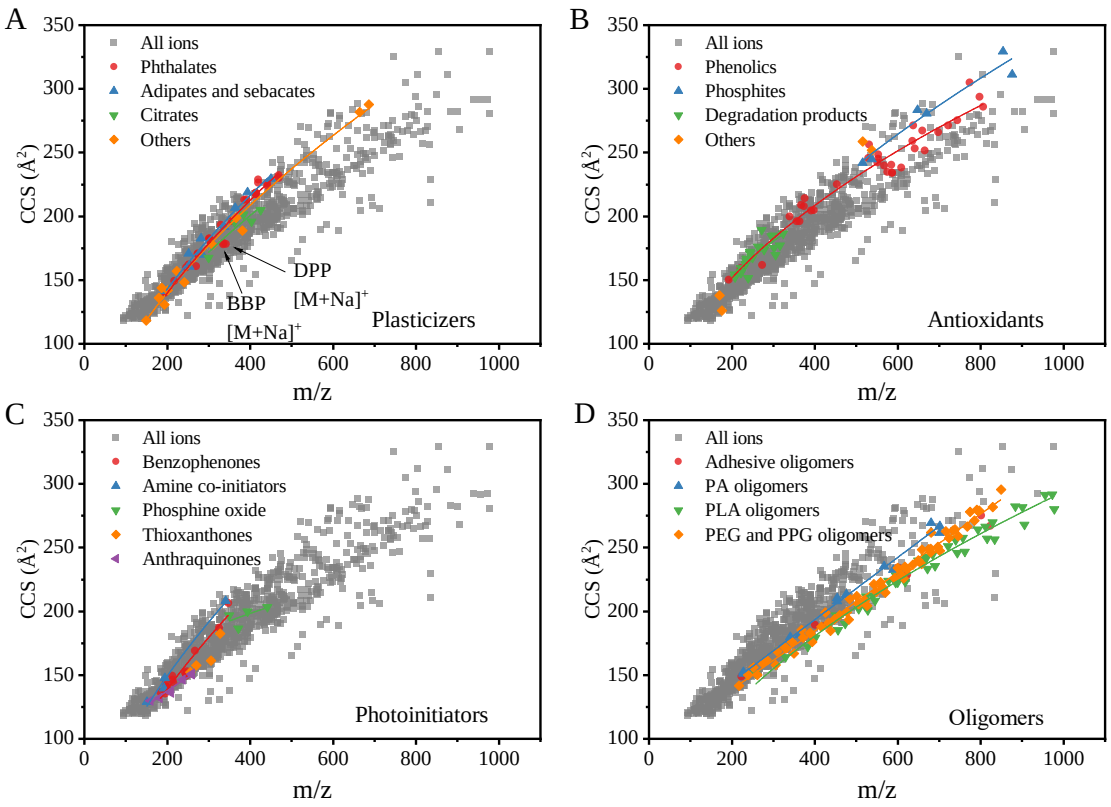


Figure III-1.4. Depiction of $^{TW}CCS_{N2}$ values vs *m/z* values for common additives and NIAS in food contact materials: (A) plasticizers; (B) antioxidants; (C) photoinitiators; (D) oligomers. BBP: Benzyl butyl phthalate, DPP: Diphenyl phthalate.

The presence of branched alkyl groups in phthalates produce various structural isomers. To study the effect of alkyl groups on the conformation of phthalates, the $^{TW}CCS_{N2}$ values of eight phthalates, with either linear or branched alkyl groups, were

measured in triplicate and the average $^{TW}CCS_{N2}$ values together with their standard deviations are presented in Table III-1.6. It can be seen that diisoalkyl phthalates have slightly lower $^{TW}CCS_{N2}$ values compared to the $^{TW}CCS_{N2}$ values for corresponding dialkyl phthalates. Compare, for example, dipropyl phthalate (171.79 ± 0.15 Å^2) and diisopropyl phthalate (170.65 ± 0.07 Å^2). This indicates that the branched alkyl group can lead to a slightly more compact molecule. The CCS deviations between diisoalkyl phthalates and dialkyl phthalates, though, were less than 1.5 Å^2 which is lower than variations in the $^{TW}CCS_{N2}$ values observed for other molecules obtained using TWIMS platforms (± 2%) (Hernandez-Mesa et al., 2020; Hernandez-Mesa et al., 2018; Righetti et al., 2020). As such, the definitive identification of such isomers may require a IMS device with higher resolving power and better reproducibility (providing CCS deviations <0.5%).

Table III-1.6. $^{TW}CCS_{N2}$ values of sodiated adducts of isomeric phthalate-based plasticizers (n=3).

Compounds	*m/z*	RT (min)	$^{TW}CCS_{N2}$ ± SD (Å^2)	RSD (%)
Dipropyl phthalate	273.1097	6.23	171.79 ± 0.15	0.09
Diisopropyl phthalate	273.1097	6.13	170.65 ± 0.07	0.04
Dibutyl phthalate	301.1410	6.88	183.61 ± 0.03	0.02
Diisobutyl phthalate	301.1410	6.80	182.13 ± 0.12	0.06
Dinonyl phthalate	441.2975	8.42	226.35 ± 0.13	0.06
Diisononyl phthalate	441.2975	8.37	225.13 ± 0.09	0.04
Didecyl phthalate	469.3288	9.01	233.92 ± 0.24	0.10
Diisodecyl phthalate	469.3288	8.65	232.47 ± 0.03	0.01

Diesters of isophthalic acid and terephthalic acid can also lead to the presence of isomers. The $^{TW}CCS_{N2}$ values of three phthalate-based plasticizers, bis(2-ethylhexyl) phthalate, bis(2-ethylhexyl) isophthalate and bis(2-ethylhexyl) terephthalate were measured as 218.42 ± 0.31 Å^2, 218.45 ± 0.14 Å^2 and 218.05 ± 0.20 Å^2, respectively. Their $^{TW}CCS_{N2}$ values were not significantly different, possibly because their molecules are flexible, and interact with the drift gas in a similar manner. This type of isomers

cannot be separated using TWIMS systems with Rp below 60 full width at half-maximum (FWHM) (Dodds et al., 2017).

$^{TW}CCS_{N2}$ values of other types of plasticizers (6 $[M+H]^+$ and 6 $[M+Na]^+$ ions), such as benzoates and organophosphates were also included in the database. Tris(2,4-di-tert-butylphenyl) phosphate has the highest $^{TW}CCS_{N2}$ value of all the plasticizers, with values of 282.1 $Å^2$ for $[M+H]^+$ and 287.9 $Å^2$ for $[M+Na]^+$. It should be mentioned that this compound is also an oxidation product of Irgafos 168 and can be used as a flame retardant.

A total of 67 $^{TW}CCS_{N2}$ values were obtained from 38 antioxidants and their degradation products. Two categories of antioxidants were included in dataset, hindered phenols (5 $[M+H]^+$, 14 $[M+Na]^+$, 12 $[M-H]^-$, 2 $[M+NH_4]^+$ and 1 $[M+HCOO]^-$ ions) and phosphites (3 $[M+H]^+$ and 3 $[M+Na]^+$ ions). The former category are primary antioxidants, which can eliminate free radicals, the latter category can decompose hydroperoxide, working as secondary antioxidants (Hu et al., 2021). The relationship between $^{TW}CCS_{N2}$ and *m/z* values of phenol antioxidants can be described by a power model with a determination coefficient $R^2 = 0.936$. The $^{TW}CCS_{N2}$ values of the phenols ranged from 150.3 to 305.2 $Å^2$ and those for the phosphites ranged from 241.9 to 329.4 $Å^2$ (see Figure III-1.4B).

Many degradation products can be generated by the oxidation of antioxidants and are an important set of NIAS in FCMs. A total of 23 $^{TW}CCS_{N2}$ values were obtained for degradation products, including 11 $[M+H]^+$, 7 $[M+Na]^+$, 4 $[M-H]^-$ and 1 $[M+HCOO]^-$. The relationship between their $^{TW}CCS_{N2}$ and *m/z* values are depicted in Figure III-1.4B. 2,6-di-tert-butyl-1,4-benzoquinone ($[M+H]^+$ 156.0 $Å^2$, $[M+Na]^+$ 171.2 $Å^2$) and 3,5-di-tert-butyl-4-hydroxybenzaldehyde ($[M-H]^-$ 162.7 $Å^2$, $[M+H]^+$ 164.9 $Å^2$) are degradation products of butylated hydroxytoluene (BHT) (Liu et al., 2015). 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionic acid ($[M+Na]^+$ 175.6 $Å^2$) can be produced from Irganox 245, Irganox 1076, Irganox 1035 and Irganox 1098. This compound can be further oxidated into 7,9-di-tert-butyl-1-oxaspiro(4,5)deca-6,9-diene-2,8-dione ($[M+H]^+$ 173.9 $Å^2$,

$[M+Na]^+$ 185.2 Å^2) (Liu and Mabury, 2021). Many of the $^{TW}CCS_{N2}$ values of degradation products are reported here for the first time, which contribute a lot to the application of IMS in the analysis of FCCs.

There are 33 $^{TW}CCS_{N2}$ values for 24 photoinitiators included in the $^{TW}CCS_{N2}$ database. Their $^{TW}CCS_{N2}$ values range from 129.2 ($[M+H]^+$ of 4-(dimethylamino)benzaldehyde) to 208.2 Å^2 ($[M+H]^+$ of 4-octadecylmorpholine). The photoinitiators are classified into benzophenones, amine co-initiators, phosphine oxides, thioxanthones and anthraquinones based on their structural characteristics. The relationship between the $^{TW}CCS_{N2}$ and *m/z* values for the photoinitiators is shown in Figure III-1.4C. The trendline for amine co-initiators (4 $[M+H]^+$ ions) has a slightly higher gradient than the other classes due to the high $^{TW}CCS_{N2}$ values of 4-octadecylmorpholine. This compound contains an octadecyl group which likely to increase the number of collisions of the molecule with the drift gas. Thioxanthones (3 $[M+H]^+$ and 1 $[M+Na]^+$ ions) and anthraquinones (6 $[M+H]^+$ ions) have slightly lower $^{TW}CCS_{N2}$ values than benzophenones, possibly due to the presence of additional rings in their molecular structures, as shown in Figure III-1.5. The trendline for phosphine oxide (1 $[M+H]^+$, 2 $[M+H]^+$ and 1 $[M+HCOO]^-$) has a relatively shallow gradient which is most likely due to the multiple phenyl groups in the structures of these molecules.

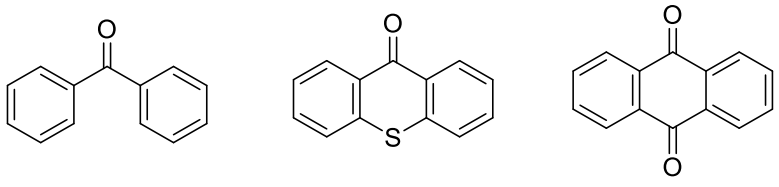


Figure III-1.5. Typical structures of benzophenone, thioxanthone and anthraquinone.

Oligomers are an important source of NIAS and 130 $^{TW}CCS_{N2}$ values from 56 oligomers of five types are included in the $^{TW}CCS_{N2}$ database. The relationship between $^{TW}CCS_{N2}$ and *m/z* values for the oligomers is shown in Figure III-1.4D. Adhesive oligomers are products of reactions between adipic acid and 1,4-butanediol. PA oligomers originate from two different polymers: PA6 (a polymer of caprolactam) and

PA66 (a polymer of 1,6-diaminohexane and adipic acid). The structure of PLA oligomers can be either linear or cyclic. The structures of polyethylene glycol (PEG) and polypropylene glycol (PPG) oligomers are similar so they are represented by the same color in Figure III-1.4D. For the PA (8 $[M+H]^+$ and 8 $[M+Na]^+$ ions) and the PEG and PPG oligomers (24 $[M+H]^+$, 23 $[M+Na]^+$ and 18 $[M+NH_4]^+$ ions), the relationship between $^{TW}CCS_{N2}$ and *m/z* values followed linear regression models, with R^2 values of 0.993 and 0.984, respectively. The $^{TW}CCS_{N2}$ and *m/z* relationship for PLA oligomers (13 $[M+H]^+$, 18 $[M+Na]^+$ and 11 $[M+NH_4]^+$ ions) followed a power model, with a R^2 value of 0.975. Adhesive, PA and PLA oligomers belong to the super class of phenylpropanoids and polyketides while PEG and PPG belong to the organic oxygen compounds class. 35 ions for the oligomers had $^{TW}CCS_{N2}$ values above 250 Å^2 and *m/z* values above 700 Da, which expanded the chemical space covered by the $^{TW}CCS_{N2}$ database.

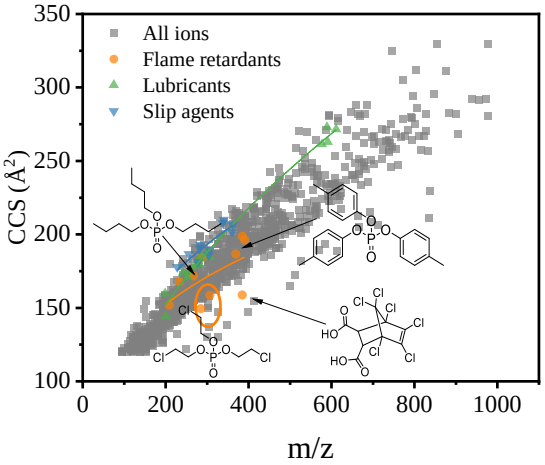


Figure III-1.6. CCS values vs *m/z* values for flame retardants, lubricants and slip agents.

Flame retardants, lubricants and slip agents are also commonly used additives in plastics and the $^{TW}CCS_{N2}$ values for those measured for this study are shown in Figure III-1.6. Lubricants mainly contain long-chain fatty acids and slip agents mainly contain long-chain fatty amides. As such it is understandable that these two types of additives have relatively high $^{TW}CCS_{N2}$ values. Plotting the $^{TW}CCS_{N2}$ values against the *m/z* values does not reveal any specific patterns for flame retardants. Dibutyl phosphate and

tributyl phosphate have high $^{TW}CCS_{N2}$ values with respect to *m/z* due to the presence of alkyl groups, tris(2-chloroethyl) phosphate and chlorendic acid have low $^{TW}CCS_{N2}$ values with respect to *m/z* due to the presence of chlorine, while tri-p-cresyl phosphate and octicizer reside between the two other groups. On plotting the $^{TW}CCS_{N2}$ values against the *m/z* values for halogenated compounds the resulting trendline tends to be different to that for compounds only containing C, H, O, N, S, P (Belova et al., 2021). To clearly show the effect of halogens on CCS values, the $^{TW}CCS_{N2}$ distributions of halogenated compounds are shown in Figure III-1.7. A total of 114 ions from 81 halogenated compounds were included in the $^{TW}CCS_{N2}$ database, and it is clear that their CCS values tend to be lower for a given *m/z* across the *m/z* range. PFAS and benzalkonium chloride are two types of surfactants used in plastic products. It should be mentioned that benzalkonium chlorides appear in the positive ion mode as benzalkonium cations, although they cannot be strictly classified in the halogenated compounds. However, the comparison of their CCS distribution with that of PFAS can clearly show the effect of halogens on CCS values. PFAS compounds contain carbon-fluorine bonds and their $^{TW}CCS_{N2}$ values are much lower in general than other compounds of similar *m/z* values. By contrast, $^{TW}CCS_{N2}$ values benzalkonium chloride tend to be high for a given *m/z*, since these compounds contain alkyl groups and will lose chloride in positive ion mode mass spectrometry.

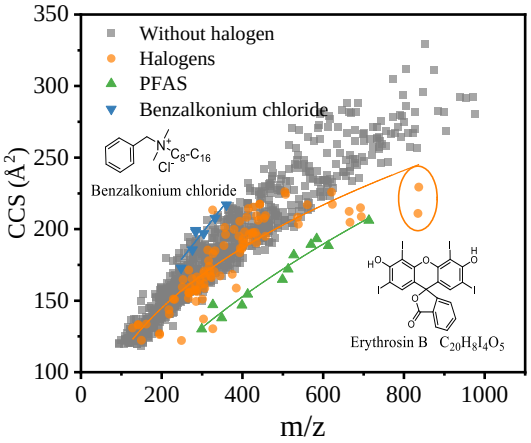


Figure III-1.7. Depiction of $^{TW}CCS_{N2}$ values vs *m/z* values for halogenated compounds.

The inter-day precision of $^{TW}CCS_{N2}$ measurements was monitored using 38 FCCs

over the course of two months within which the IMS cell was calibrated twice. 70 ions were detected including 15 $[M+H]^+$, 23 $[M+Na]^+$, 28 $[M-H]^-$ and 4 $[M+HCOO]^-$, and the distribution of the RSDs of the $^{TW}CCS_{N2}$ measurements of these ions is shown in Figure III-1.8. Excellent inter-day precision was obtained with all RSD values lower than 0.7%. 85.7% (60/70) of adducts had RSD values in the range of 0.3-0.5%. Similar inter-day precision of TWIMS platform has also been shown by Regueiro et al. (2016) with most RSDs in their work ranging from 0.3% to 0.5%.

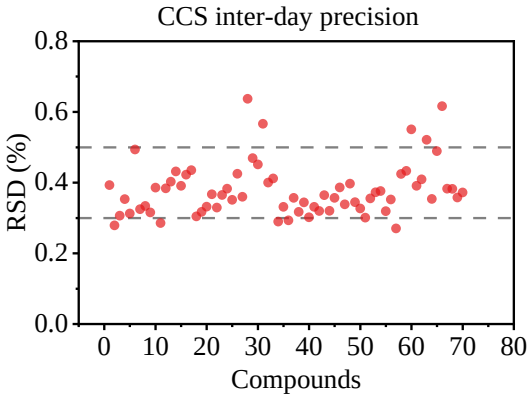


Figure III-1.8. Inter-day precision of CCS measurement over two months.

4.4. Comparison with Existing Literature CCS Values.

In order to check the accuracy of our $^{TW}CCS_{N2}$ data and ensure that it is independent of the IMS platform and laboratory used to acquire it, the $^{TW}CCS_{N2}$ values of the three main adducts ($[M+H]^+$, $[M+Na]^+$, $[M-H]^-$) were compared to previously published CCS values. CCS deviations (ΔCCS%) were calculated using the $^{TW}CCS_{N2}$ values in our database as the reference, and the results are shown in Figure III-1.9.

For some compounds, several CCS records can be found in different publications. A total of 300, 144 and 208 CCS measurements were found for 123 $[M+H]^+$, 71 $[M+Na]^+$ and 93 $[M-H]^-$ adducts, respectively. It is usual to use a tolerance of ±2.0% for CCS measurements in IMS analysis (Righetti et al., 2020). 85.7%, 87.7% and 64.9% of the $[M+H]^+$, $[M+Na]^+$ and $[M-H]^-$ adducts, respectively, showed CCS deviations less than 2%. There are several reasons which may lead to this threshold being exceeded.

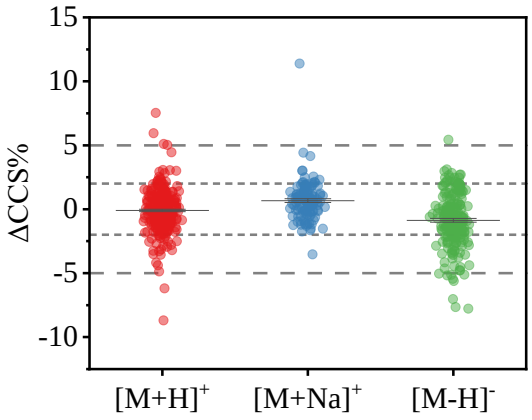


Figure III-1.9. Comparing $^{TW}CCS_{N2}$ values in the database with published CCS values.

(1) Protomers. The tendency for some compounds to form protomers can lead to different CCS values since different protonation sites can affect the shape and size of the molecules. As an example of this, ciprofloxacin has two competing protonation sites on its molecular structure, a carbonyl oxygen and an amine. Two different CCS values (173.3 $Å^2$ and 185.3 $Å^2$) were obtained for the $[M+H]^+$ adduct of ciprofloxacin by Hines et al. (Hines et al., 2017), each reflecting a different site of protonation. The $^{TW}CCS_{N2}$ value of the $[M+H]^+$ adduct of ciprofloxacin determined here was 184.8 $Å^2$. Only one CCS value was obtained here because ion mobility data was not sufficiently resolved, as shown on the mobility trace in Figure III-1.10.

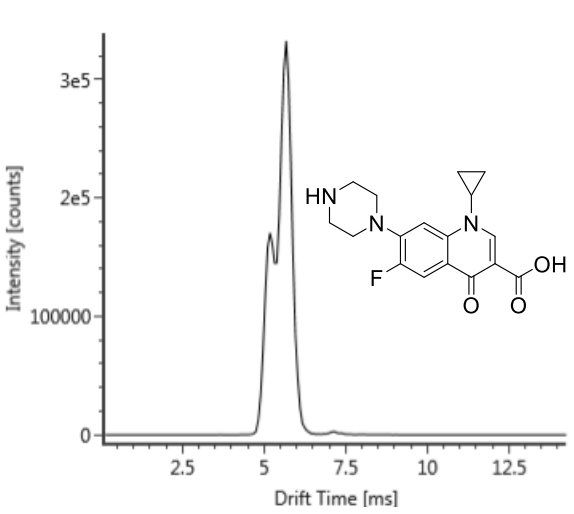


Figure III-1.10. Arrival time distribution (ATD) of ciprofloxacin.

Thus, a high CCS deviation of 6.2% (173.3 compared to 184.8 Å^2) was obtained on comparing the $^{TW}CCS_{N2}$ value determined in this work against previously published CCS values for ciprofloxacin. The presence of multiple protomers also explains the high CCS deviation of theobromine. Two CCS values for the $[M+H]^+$ adduct of theobromine, 131.1 Å^2 and 138.9 Å^2, were obtained in Nichols et al. (2018) (Nichols et al., 2018). The larger of these values had a deviation of 6.4% when compared to the value of 130.5 Å^2 measured here.

(2) Post-IMS dissociation of a noncovalent cluster. Occasionally, a noncovalent cluster can form in the ion source, prior to entering the travelling wave device, which can subsequently undergo a dissociation after drifting through the device. When this happens, elevated CCS values are generated since the noncovalent clusters are larger than the target ions they contain. As an example of this, the $[M+Na]^+$ adduct of chenodeoxycholic acid has a $^{TW}CCS_{N2}$ value of 196.9 Å^2. Two published CCS values for this compound are 202.8 Å^2 from (Poland et al., 2020) and 219.3 Å^2 from Metabolic Profiling CCS Library (Waters Corporation, 2020a). Multiple sites of protonation are not evident on the TWIMS platform. As such, the difference in the CCS values (196.9 compared to 219.3 Å^2, $\Delta CCS\% = 11.4\%$) may arise from the post-IMS dissociation of a noncovalent cluster.

It should be noted that, in some cases, the ion with highest abundance may not always yield the actual CCS value due to the presence of noncovalent clusters. The arrival time distribution (ATD) and mass spectrum of triclosan, a commonly used fungicide, are shown in Figure III-1.11. The ion with the highest abundance had a measured $^{TW}CCS_{N2}$ value of 177.5 Å^2, however, three CCS values ranging from 157.3 to 160.0 Å^2 were found for this compound in the literature (Belova et al., 2021; Celma et al., 2020; Waters Corporation, 2021). It can be seen from Figure III-1.11 that a small peak with the arrival time of 4.52 ms has a $^{TW}CCS_{N2}$ value of 157.4 Å^2, which is in good agreement with published values. A careful examination of the mass spectra or

comparing the experimental CCS values with those from literature can avoid this kind of discrepancy.

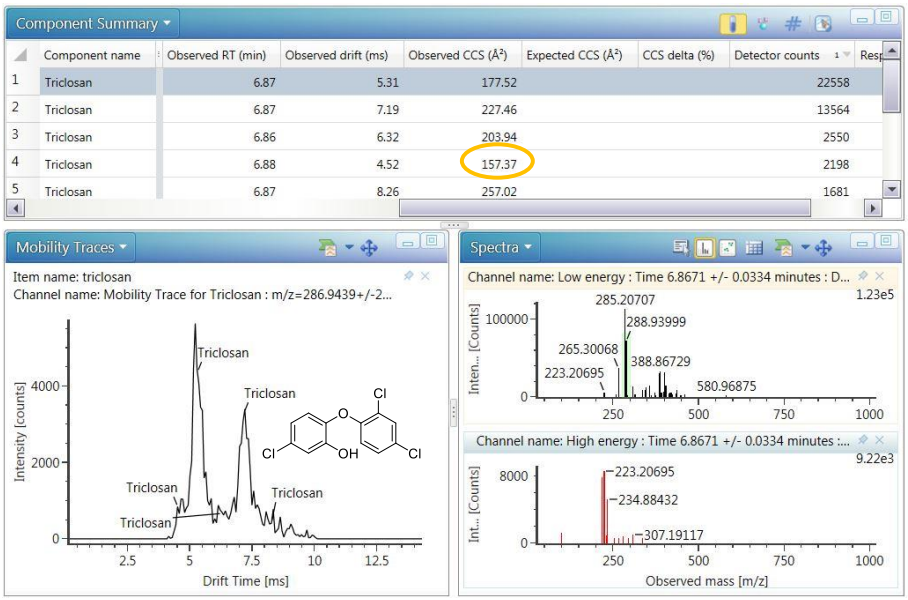


Figure III-1.11. Several CCS values of triclosan ($[M-H]^-$, m/z 286.9438) arising from post-IMS dissociation of noncovalent clusters.

(3) IMS calibration. $^{TW}CCS_{N2}$ values are obtained through appropriate calibration of the TWIMS platform and as such, using calibration compounds with similar structural characteristics as the analytes to be investigated leads to increased accuracy of the measured $^{TW}CCS_{N2}$ values (Hines et al., 2016). Therefore, performing the calibration using typical standards and over a limited data range may result in CCS deviations. For example, high CCS deviations were observed for the $[M-H]^-$ adduct of p-coumaric acid (129.9 Å^2) and ellagic acid (149.8 Å^2). In the case of p-coumaric acid, four published CCS values were found, three of which range from 128.9 to 132.6 Å^2 (Corporation, 2020a; Nichols et al., 2018; Zhou et al., 2016), while the fourth, in the study by Gonzales et al. (2016), has a value of 119.8 Å^2, deviating in the region of -7.8% from our data.

Three published CCS values were found for ellagic acid: 152.0 (Corporation, 2020b), 152.5 (Corporation, 2021) and 157.9 Å^2 (Gonzales et al., 2016), with the highest CCS deviation (5.4%) occurring, once again, between our data and Gonzales et

al. (2016). Gonzales et al. (2016) calibrated their TWIMS system using deprotonated polyalanine standards. In our work, two fluoroalkanoic acids and some drug-like compounds were added to the polyalanines to extend the range over which the calibration was valid (see Table III-1.2). These variations in the CCS measurements highlight the importance of establishing an appropriate CCS calibration strategy for the compounds to be analyzed on the TWIMS system. Recently, an improved CCS calibration approach has been proposed for TWIMS system by Richardson and coworkers (Richardson et al., 2021), which has the potential to further improve the accuracy the $^{TW}CCS_{N2}$ measurements.

(4) IMS reproducibility. 11% (33/300), 9.6% (11/114) and 26.9% (56/208) of the $[M+H]^+$, $[M+Na]^+$ and $[M-H]^-$ adducts respectively, had variations in their measured CCS values between 2% and 4%. This may be due to the reproducibility of the IMS measurements. A value of 2% is usually given for the variation in CCS measurements on TWIMS platforms. However, for some ions, the measurements may fall into the extreme ends of this ±2% tolerance, leading to elevated CCS discrepancies across different platforms and laboratories.

4.5. Comparison to Predicted CCS Values from Machine Learning Approaches.

When no reference standard is available, comparing experimental CCS values to theoretical predictions can increase the confidence of identifications and reduce false positives (Zhou et al., 2020). CCS values of 446 $[M+H]^+$, 317 $[M+Na]^+$ and 190 $[M-H]^-$ adducts were predicted using three publicly available CCS prediction tools (CCSondemand, AllCCS and CCSbase) and compared to our experimental values. The proportions of experimental $^{TW}CCS_{N2}$ records with relative deviations values less than 2%, 3% and 5% from predicted values were compared. Evaluating the predictive performance of the prediction tools enabled an assessment of their applicability to FCCs. The results of the comparison are presented in Figure III-1.12.

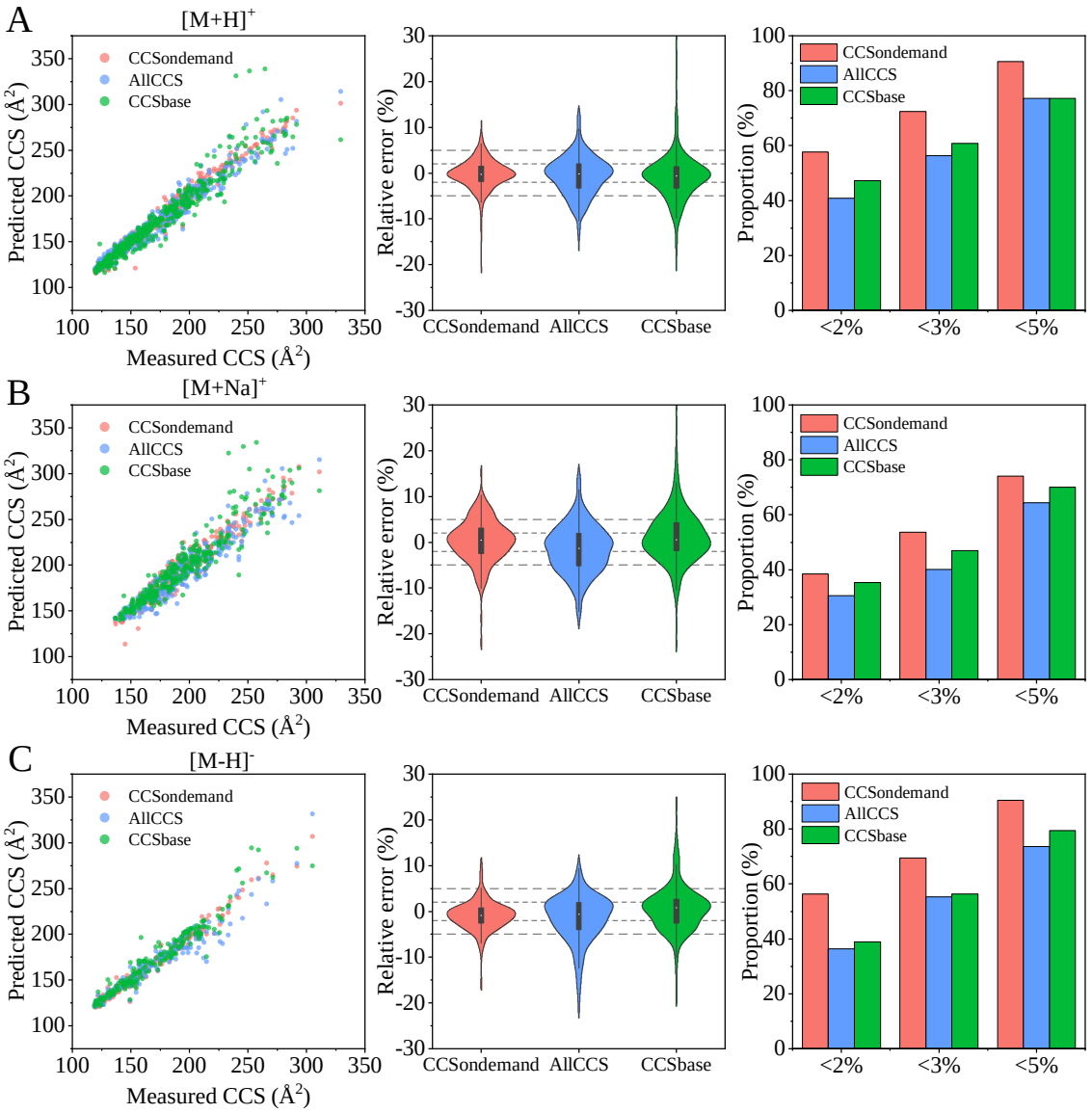


Figure III-1.12. Comparison between experimental and predicted CCS values, (A) $[M+H]^+$ adduct, (B) $[M+Na]^+$ adduct, (C) $[M-H]^-$ adduct.

CCSondemand was found to be the most accurate CCS prediction tool for FCCs, followed by CCSbase and AllCCS. For CCSondemand, 57.6% of the predictions for $[M+H]^+$ adducts and 56.3% of the predictions for $[M-H]^-$ adducts agreed with the measured $^{TW}CCS_{N2}$ values to within ±2%. On opening up the tolerance to 5%, 90.6% and 90.5% of the predicted CCS values for $[M+H]^+$ and $[M-H]^-$ adducts, respectively, agreed with the measured data. The training dataset for CCSondemand consisted of $^{TW}CCS_{N2}$ values measured on Vion or Synapt platforms calibrated with the Major Mix IMS/ToF Calibration Kit (Broeckling et al., 2021). The same calibration kit was used to

calibrate our instruments and this may explain why CCSondemand outperformed the other two tools in the prediction of CCS values of FCCs. Another reason for the more accurate prediction results provided by CCSondemand is that the training set of CCSondemand contains some experimental $^{TW}CCS_{N2}$ values of FCCs, as mentioned previously (Song et al., 2022c).

AllCCS and CCSbase predicted the CCS values of 40.8% and 47.3% of $[M+H]^+$ adducts to within 2% of the measured values, respectively. The CCS values predicted by AllCCS and CCSbase for some compounds, commonly detected in FCMs, had relatively high errors when compared to the measured data. These included the primary aromatic amines (PAAs), as mentioned in previous study (Song et al., 2022c), ultraviolet absorbers, chimassorb 81 (-6.6% and -5.6%, respectively) and UV-360 (-12.6% and -8.2%, respectively). Oligomers, such as PPG5-PPG11 showed variations between the AllCCS and CCSbase predictions and measured values in the range of 5.4% -12.7%. CCSbase predictions of the CCS values of cyclic PLA9 and cyclic PLA10 disagreed with the measured values by more than 30%. AllCCS predictions of the CCS values of antioxidants and their degradation products also showed elevated discrepancies to the measured data with Irgafos 168 (-11.6%), Irganox 1076 (-7.9%), 3,5-di-tert-butyl-4-hydroxybenzaldehyde (-9.1%), and 2,6-ditertbutyl-1,4-benzoquinone (-6.1%). Besides, it can be seen from Figure III-1.12 that AllCCS and CCSbase present a trend towards negative prediction errors.

A possible explanation of the high deviations in the CCS values of AllCCS and CCSbase is that the training datasets did not contain many FCCs or FCC-like compounds, Additionally, the training datasets for AllCCS and CCSbase contained CCS values originating from both drift tube and travelling wave devices, and discrepancies have been shown to occur in CCS values measured on different instrument types. Hinnenkamp et al. (2018) has compared the $^{TW}CCS_{N2}$ and drift tube collision cross section in nitrogen ($^{DT}CCS_{N2}$), finding that 7% protonated adducts and 13% sodium adducts have CCS deviations higher than 2%, this result indicates that CCS database

cannot be used without caring their types. As the DTIMS can determine the CCS directly, without the need of calibration (D'Atri et al., 2018), an improved CCS calibration approach of TWIMS may be able to increase the consistency between experimental $^{TW}CCS_{N2}$ and $^{DT}CCS_{N2}$ values, thus leading to a more accurate CCS prediction model.

Predicted CCS values for $[M+Na]^+$ adducts, for all three prediction tools, were relatively poor when compared to the measured values. CCSondemand once again provided the best results, but only 38.5% of the predicted CCS values for $[M+Na]^+$ agreed with the measure data to within 2%. This may due to there being less measurements for $[M+Na]^+$ adducts in the training datasets. Additionally, it is difficult to predicted the CCS values of sodiated molecules using the molecular descriptors from neutral molecules (Song et al., 2022c).

Considering the current accuracy of predicting CCS values, we believe that, at best, predicted CCS values can be used to help to eliminate false positives and to support tentative identifications. However, CCS prediction tools cannot be used to confirm the identification of an unknown compound. Connolly and coworkers have shown that predicted CCS values cannot accurately describe the difference of CCS values for isomeric glucuronide pairs (Connolly et al., 2021). Technological developments are on-going though, and as the accuracy and reproducibility of experimentally obtained CCS values improves, a similar improvement can be expected in the accuracy of predicted CCS values.

5. Conclusions

A database of $^{TW}CCS_{N2}$ values for extractable and leachable compounds from FCMs has been presented. This $^{TW}CCS_{N2}$ database contains both IAS and NIAS. Excellent inter-day precision of the measured values has been shown, with all RSD values lower than 0.7%, indicating good reproducibility and stability of measurements from the TWIMS system. The $^{TW}CCS_{N2}$ values in the database can serve as additional confirmation points for the identification of FCCs in targeted and untargeted screening

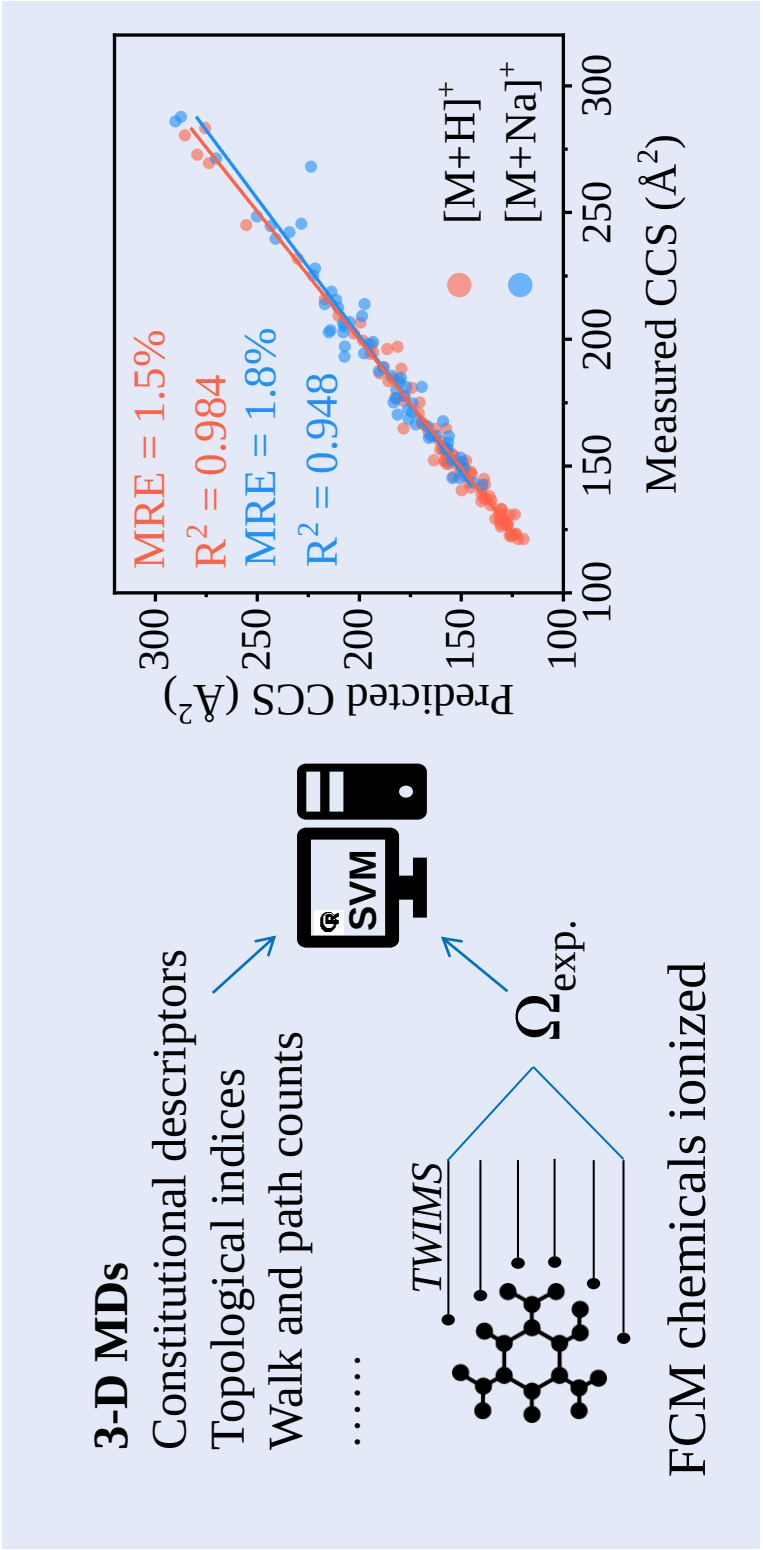
analyses. It has also been argued that CCSondemand is a promising tool for the prediction of CCS values of FCCs, the prediction performance of CCSondemand will be further improved by incorporating more high-quality CCS measurements in the training dataset.

Chapter 2

Prediction of Collision Cross Section Values: Application to Non-Intentionally Added Substance Identification in Food Contact Materials

1. Abstract

The synthetic chemicals in food contact materials can migrate into food and endanger human health. In this study, the $^{TW}CCS_{N2}$ values of more than 400 chemicals in food contact materials were experimentally derived by traveling wave ion mobility spectrometry (TWIMS). An SVM-based CCS prediction model was developed based on CCS values of food contact chemicals and a series of molecular descriptors. More than 92% of protonated and 81% of sodiated adducts showed relative deviation below 5%. Median relative errors (MRE) for protonated and sodiated molecules were 1.50% and 1.82%, respectively. The model was then applied to the structural annotation of oligomers migrating from polyamide adhesives. The identification confidence of eleven oligomers was improved by the direct comparison of experimental data with predicted CCS values. Finally, the challenges and opportunities of current machine learning models on CCS prediction were also discussed.


3-D MDs
Constitutional descriptors
Topological indices
Walk and path counts
……
SVM
FCM chemicals ionized
TWIMS
$\Omega_{exp.}$
Predicted CCS (Å^2)
Measured CCS (Å^2)
MRE = 1.5%
R^2 = 0.984
MRE = 1.8%
R^2 = 0.948
[M+H]$^+$
[M+Na]$^+$
100
150
200
250
300

2. Introduction

The food contact materials (FCMs) can provide a protection for food, but meanwhile it is also an important source of contaminations of food. In the manufacturing process of FCM, a range of synthetic additives (antioxidant, plasticizers, photoinitiators, lubricants, slip agents, etc.) are routinely employed to provide the material with desired mechanical and thermal properties. These compounds are intentionally added substances (IAS) and their specific migration limits (SMLs) are included in the positive list of Regulation (EU) No 10/2011 (EC, 2011). On the other hand, non-intentionally added substances (NIAS) are chemicals that are present in a FCM but have not been added for a technical reason during the manufacturing process, and origin from degradation of additives (e.g., 2,4-di-tert-butylphenol from Irgafos 168) (Yang et al., 2016), interactions between constituents (e.g., 1,6-dioxacyclododecane-7,12-dione from condensation reaction between 1,4-butanediol and adipic acid) (Canellas et al., 2015b), and impurities of raw materials (Nerin et al., 2013). Recycling can also introduce different kinds of NIAS due to the low efficiency of cleaning processes. Oligomers and degradation products can also be produced due to the high temperature and to the presence of oxygen in mechanical recycling (Song et al., 2019; Ubeda et al., 2018). Both IAS and NIAS can migrate through the packaging into food products and have the potential to endanger human health (Canellas et al., 2015a; Ubeda et al., 2020). The risk associated with the migration of NIAS from specific packaging materials have to be assessed. As the first step of risk assessment, the structural elucidation of such molecules is crucial for the correct quantification and the subsequent toxicological evaluation.

Compared to IAS, the identification of NIAS is much more challenging due to the complexity of composition of the final packaging material and limited ingredient information provided by manufacturers. Gas chromatography–mass spectrometry (GC-MS) (Osorio et al., 2019) and liquid chromatography–mass spectrometry (LC-MS) (Aznar et al., 2019) are widely used analytical techniques for the screening of volatile

and non-volatile NIAS. A high-resolution mass spectrometer (HRMS) operating in data independent acquisition (DIA) can provide accurate mass of precursor and product ions, thus deriving the elemental composition also based on isotopic pattern distributions. The chemical structure of unknowns can then be inferred from fragmentation studies, applying a combination of common rules. However, in this process, two main issues can be encountered. First, chromatographic co-elution exists, which makes it difficult to identify the actual precursor ion, especially where the number of adducts is limited due to different ionization efficiencies. Second, it is possible that two or more candidates conform to the exact mass and similar fragmentation pathway. In this case the experience and technical skillfulness of the analyst in the MS spectra interpretation are essential for reducing false detects and to bring confidence to the identification results, which ultimately rely on the confirmation with a pure standard. In this context, the availability of different separation techniques in combination with conventional LC-HRMS systems would be extremely beneficial.

Ion mobility spectrometry (IMS) is a gas-phase separation technique which enables the separation of ions by collisions with a buffer gas (usually nitrogen or helium) under a defined electric field profile and controlled gas pressure in a drift cell (D'Atri et al., 2018). The drift time of ions is associated with their size, shape and charge, which results in a partial orthogonality with MS separation (Zhang et al., 2018), besides, the drift time is generally in the range of milliseconds, which fits well between LC separation (in the range of several seconds) and MS detection (in the microseconds scale). The combination of ultra-high performance liquid chromatography with an ion mobility-mass spectrometer (UHPLC-IMS-MS) can provide a three-dimensional separation (retention time, drift time and *m/z*), thus increasing peak capacity compared to UHPLC-MS alone (Paglia et al., 2021). A few studies reported coelution of isomer pairs in conventional LC, which were then resolved by IMS (Lalli et al., 2013; McCullagh et al., 2019b). In recent years, UHPLC-IMS-MS has been widely used in the structural characterization of lipids (Hankin et al., 2016), glycans (Gray et al., 2016), and small molecules, such as pesticides (Bijlsma et al., 2017), steroids (Hernandez-

Mesa et al., 2018), phenolics (Gonzales et al., 2016; Song et al., 2021), and NIAS in food packaging (Canellas et al., 2021b; Vera et al., 2019).

Collision cross section (CCS) can be related to the mobility of ions and it is commonly recognized to represent the effective rotationally averaged collision area of the ions with neutral gas molecules, which is a physicochemical property of ions for a given compound. More precisely, CCS describes the momentum transfer between ions and drift gas particles. Therefore, it is considered as a structural property of ionized molecules, which depends on experimental conditions such as drift gas composition, temperature, and reduced field strength (E/N, where E represents the electric field and N is the gas number density) (Gabelica et al., 2019). However, unlike drift time, CCS values are not instrument-dependent, so they should be comparable across different instruments and laboratories operating under the same experimental conditions. CCS can then be treated as an additional structural descriptor obtained from IMS for confirmation of compound identification. A number of previous works has demonstrated a fairly good reproducibility of CCS values between different laboratories and platforms (Hinnenkamp et al., 2018; Stow et al., 2017). In recent years, several CCS databases have been generated from experimental measurements (Hernandez-Mesa et al., 2020; Hines et al., 2017; Nye et al., 2019; Paglia et al., 2015; Righetti et al., 2020; Zheng et al., 2017a), but many of them are still difficult to integrate into routine discovery analyses. In addition, unless costly and time-consuming chemical synthesis and purifications of suspect-compounds are addressed, the empirical CCS values of compounds cannot be obtained when their standards are not commercially available. In order to enhance the wider application of CCS in qualitative analysis, a number of efforts has been undertaken in the past few years for the prediction of a compound's CCS from its molecular descriptors (i.e., numeric values that provide a fingerprint of a compound's structural and physicochemical properties) by means of machine-learning tools (Bijlsma et al., 2017; Gonzales et al., 2016; Mollerup et al., 2018; Plante et al., 2019; Soper-Hopper et al., 2017; Zhou et al., 2016). Different algorithms, such as partial least squares regression (PLS-R), support vector machine (SVM) and artificial neural

network (ANN) have been applied to create predictive models for specific groups of analytes. The number of molecular descriptors (MDs) used to develop the predictive models varies from tens to thousands. As an alternative to MDs, Ross and co-workers used molecular quantum numbers (MQNs), which are obtained from analyzing compounds as a molecular graph (i.e. collections of nodes = atoms, and edges = bonds), claiming MQNs are invariant with respect to the software used to compute them (Ross et al., 2020). Plante and collaborators developed a convolutional neural network model (CNN) using simplified molecular-input line-entry systems (SMILESs) as input for CCS prediction, without the need for MDs (Plante et al., 2019). When no CCS database or commercial standards are available, the machine learning approach can become a potential alternative to predict and confirm CCS values.

In this study, a $^{TW}CCS_{N2}$ library was generated by measuring 488 standards available in our laboratory via UPLC-IMS-QToF. The majority of the measured compounds are commonly used chemicals in food packaging materials. The chemical structures of these compounds were then submitted to dedicated software to retrieve the physicochemical descriptors. The goal was to develop an *in-house* prediction model to predict $^{TW}CCS_{N2}$ values of food contact chemicals using MDs as input. After optimization and comparison with the currently available tools, the developed predictive model was implemented within our NIAS identification pipeline, and employed for the structural elucidation of unknown compounds migrating from packaging materials. Finally, we provide a discussion on the challenges and opportunities of existing machine learning CCS prediction tools.

3. Materials and methods

3.1. Chemicals and reagents.

A total of 488 standards, including the commonly used additives in food packaging, such as antioxidants, plasticizers, dyes, slip agents, UV-absorbers, lubricants,

as well as a large set of NIAS historically found from our previous studies (degradation products of hindered phenolic antioxidants, oligomers, by-reaction products, etc.) were included in the dataset. All standards were purchased from Sigma-Aldrich Quimica S.A. (Madrid, Spain), Extrasynthese (Genay, France) and Cayman chemical company (Ann Arbor, Michigan, USA). HPLC grade methanol (≥99.9%), ethanol (≥99.9%), dichloromethane (≥99.8%) dimethyl sulfoxide (≥99.8%) were purchased from Scharlau Chemie S.A (Sentmenat, Spain). Ultrapure water was produced by a Millipore Milli-QPLUS 185 system (Madrid, Spain). Formic acid was purchased from Waters (Milford, MA, USA). For building the CCS database, standard stock solutions (1000 mg kg-1) were prepared by dissolving 10 mg of standards in 10 g of methanol. Other solvents, such as ethanol, dichloromethane and dimethyl sulfoxide were used when the standards were not dissolved in methanol. The stock solutions were then diluted to create working solutions at ~1 mg kg^{-1}. Each working solution contained 8-10 analytes, avoiding isomers and co-eluting compounds in the same mixture. All standard solutions were stored in the dark at -20 ºC until analysis.

3.2. Measurements of Experimental CCS Values.

For the empirical measurements of $^{TW}CCS_{N2}$ values, An Acquity I-Class UPLC system coupled to a Vion IMS-QToF mass spectrometer (Waters, Manchester, UK) was used. UPLC separation was performed on a CORTECS C_{18} column (2.1 × 100 mm, 1.6 µm particle size, 90 Å pore size) at a flow rate of 0.3 mL min^{-1}. The column temperature was 40 ºC. The mobile phase was composed of water (A) and methanol (B), both with 0.1% of formic acid (v/v). The volumetric percentage of mobile phase B during the LC gradient was as follows: 0-7 min: 5-100%; 7-11 min: 100%; 11-11.10 min: 100-5%; 11.10-13 min 5%.

The Vion IMS-QToF (IMS resolution ~20 Ω/ΔΩ FWHM) consists of hybrids quadrupole orthogonal acceleration time-of-flight mass spectrometers (IMS-Q-oaToF), in which a stacked-ring ion guide, that is, the mobility cell, is positioned before the quadrupole mass filter. The system was operating in positive electrospray mode (ESI+).

The capillary voltage was 1 kV and sampling cone voltage was 30 V; the source temperature was 120 ºC; cone gas flow was 50 L h^{-1}, N_2 was used as desolvation gas with a flow rate of 800 L h^{-1} at 500 ºC. Mass and CCS calibration were performed in the range 50-1200 *m/z* and 130.4-372.6 $Å^2$, respectively, using the Major Mix IMS/ToF Calibration Kit (Waters Corp.). LockSpray containing Leucine-Enkephalin ($[M+H]^+$, *m/z* 556.2771) at a concentration of 100 ng mL^{-1} and infusion rate of 15 µL min^{-1} was used for real time mass correction. Raw data was acquired in high definition MS^E mode ($HDMS^E$), mass spectra were acquired with an acquisition rate of 0.2 s at two collision energy states (low energy = 6 eV, and high energy ramp = from 20 to 40 eV). Nitrogen was used as drift gas, argon was used as collision-induced dissociation (CID) gas. The ToF analyzer operated in sensitivity mode, and the ion mobility settings were as follows: IMS gas flow rate 25 mL min^{-1}, wave velocity 250 m s^{-1}, IMS pulse height 45 V. Data acquisition and processing were carried out on UNIFI v.1.9.4 software (Waters Corp.).

Prior to each analysis, an in-house made Test-Mix solution was injected as a system suitability test. The molecular formula, monoisotopic mass and expected CCS of nine compounds in Test-Mix are listed in Table III-1.3. The pass/fail criteria for mass and CCS accuracy were: mass error < 5 ppm and ΔCCS < 2%. All working solutions were injected in triplicate, with an injection volume of 5 and 10 µL, for a total of six technical replicates per each compound. This allowed an easier assignment of standard peaks and higher confidence in the experimental $^{TW}CCS_{N2}$ values, which were obtained by averaging independent measurements (n = 6).

3.3. Collision Cross Section Prediction.

Multivariate PLS is one of the most widely used machine-learning algorithms for both regression and classification purposes; its basic knowledge can be found in the literature (Geladi and Kowalski, 1986). SVM is a supervised learning algorithm that can be used for both classification and regression analysis, it has been used for CCS prediction in previous studies (Zhou et al., 2020; Zhou et al., 2016). Herein, both PLS and SVM models were developed between the physicochemical MDs of all the

compounds and their experimentally-derived $^{TW}CCS_{N2}$. MDs were obtained using alvaDesc software v.2.0.4 within Online Chemical database (OCHEM, **http://ochem.eu/home/show.do**), obtaining a total of 5666 molecular descriptors. The detailed list of the generated descriptors is reported in Table III-2.1.

Table III-2.1. Distribution of the molecular descriptors (MDs) calculated by alvaDesc software v.2.0.4.

Blocks	Number of variables	Blocks	Number of variables
Constitutional descriptor	50	Ring descriptors	35
Topological indices	79	Walk and path counts	46
Connectivity indices	37	Information indices	51
2D-matrix-based descriptors	608	2D autocorrelations	213
Burden eigenvalues	96	P-VSA-like descriptors	69
ETA indices	40	Edge adjacency indices	324
Geometrical descriptors	38	3D matrix-based descriptors	132
3D autocorrelations	80	RDF descriptors	210
3D-MoRSE descriptors	224	WHIM descriptors	114
GETAWAY descriptors	273	Randic molecular profiles	41
Functional group counts	154	Atom centered fragments	115
Atom-type E-state indices	346	Pharmacophore descriptors	165
2D Atom Pairs	1596	3D Atom Pairs	36
Charge descriptors	15	Molecular properties	27
Drug-like indices	30	CATS 3D	300
WHALES	33	MDE	19
Chirality	70		

The irrelevant descriptors were eliminated before the model building. The descriptors with constant values or with very few unique values relative to the number of samples contain few information, which were considered less important for the CCS prediction. These kinds of descriptors were removed by function of *nearZeroVar* in R package *caret*.

The dataset was randomly split into training and testing sets in a 3:1 ratio. By doing so, the prediction ability of a developed model can be assessed in an unbiased

manner. Descriptive statistics (i.e. mean, standard deviation, range, median) of $[M+H]^+$ and $[M+Na]^+$ adducts' CCS for both calibration and validation sets are summarized in Table III-2.2, Figure III-2.1 shows the distribution of data points in calibration and validation sets.

Table III-2.2. Descriptive statistical analysis of $^{TW}CCS_{N2}$ (Å^2) in training and testing sets.

Adducts	Data set	Min	Max	Mean	Median	Range	SD
[M+H]	Training	118.6	329.4	167.5	159.6	210.9	35.4
	Testing	121.3	283.4	161.2	152.2	162.2	36.0
[M+Na]	Training	136.8	311.4	191.7	186.9	174.6	32.3
	Testing	142.6	287.9	191.4	184.0	145.3	36.0

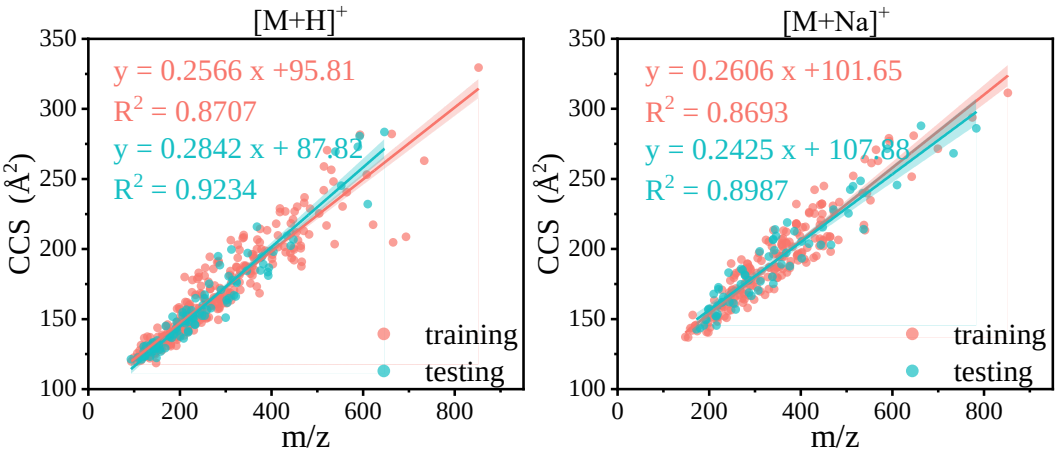


Figure III-2.1. Distribution of data points in training and testing sets for $[M+H]^+$ and $[M+Na]^+$.

Prior to modelling, natural logarithm transformation was applied to $^{TW}CCS_{N2}$ values to promote data normality. The MDs data (training set) were mean-centered and scaled to unit variance using the following equation:

$$z_i = \frac{x_i - \bar{m}_x}{s_x}$$

where z_i is the normalized data for the variable x of a particular molecule i; $\bar{m}_x$ and s_x are the mean and standard deviation of x. The $\bar{m}_x$ and s_x computed for the

training set were then used as normalization factors for the testing set. Both models were built on the preprocessed (training) data and optimized through 10-folded cross validation. The number of latent vectors of PLS was optimized based on the root mean squared error of cross validation (RMSECV) and prediction residuals, both statistically inspired modification of the partial least squares (SIMPLS) and kernel PLS were used to build the model. As for SVM, two hyper-parameters were optimized in order to get an accurate prediction: cost of constraints violation (C) and gamma (γ). Eight groups of C values (0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5)/N_{MD} (i.e. number of molecular descriptors) and nine γ values (2^0 to 2^8) formed 72 parameter combinations, The parameter combination providing the minimum RMSECV was used for further SVM model.

Sensitivity ratio (SR) is an embedded method within PLS-R for evaluating the contributions of variables for the model, which is defined as the ratio between the explained and the residual variance in the target-projected component (Rajalahti et al., 2009). F-test (99% confidence interval) criterion was used to define the boundary between highly important and less important variables, as proposed by Rajalahti et al. (2009). The important descriptors were then to build PLS and SVM models and their performances were compared with models built without feature selection.

Four CCS prediction models were developed for each adducts based on two algorithms (PLS and SVM) and two types of MDs (all MDS and important MDs selected by SR). The CCS values of testing set were predicted with four models obtained above. The prediction results of the model with better performance were then compared with the three main CCS prediction tools currently available, which use either MDs or MQNs: AllCCS from Zhu Lab (Zhou et al., 2020), CCSbase from Libin Xu Lab (Ross et al., 2020), and CCSondemand from Broeckling and co-workers (Broeckling et al., 2021).

All data processes and calculations were performed in R (version 4.0.5) using internal statistical functions and external packages (i.e., *pls* for PLS-based prediction,

e1071 for SVM-based model, *plsVarSel* for feature selection and *ggplot2* for plot creation) (Mevik and Wehrens, 2007).

3.4. Sample Preparation and Extraction.

The CCS predictive model was applied to the identification of NIAS in water-based adhesives, polyamide 6 (PA6) and polyamide 66 (PA66). Previous studies suggested that cyclic oligomers can be present in these types of materials (Abe et al., 2016; Canellas et al., 2015b). For the extraction of oligomers from adhesives, 5 g of sample was mixed with 50 mL of water, the mixture was centrifuged at 4000 g for 10 min, the supernatant was passed through a hydrophilic-lipophilic balance copolymer SPE (Oasis HLB cartridge, 6cc, Waters Corp.), previously activated with 10 mL of methanol and 10 mL of water. The oligomers were eluted with 50 mL of methanol and analysed via LC-IMS-HRMS. For the extraction of oligomers from PA6 and PA66, 10 g of pellets was extracted with 50 mL of methanol at 40 ºC overnight, the solution was filtered by 0.22-µm Nylon membrane filter and the filtrate was evaporated using a rotary evaporator. The residue was re-dissolved in 10 mL of 10% methanol in water (v/v). The reconstituted extract was cleaned-up on SPE and analyzed following the procedure described above.

As the commercial standards of these oligomers were not available, these were attempted to be produced at laboratory-scale to verify the identification. Briefly, one gram of adipic acid was mixed with one gram of 1,4-butanediol in a melting crucible with lid (40 mL), the mixture was heated at 135 ºC for 2 h, the obtained liquid was dissolved in methanol at a concentration of 10 mg kg^{-1} and then analyzed by LC-IMS-QToF under the conditions described in the experimental section.

4. Results and discussions

4.1. Mass-to-Charge and CCS Correlation

A total of 635 ions (i.e., 380 $[M+H]^+$ and 255 $[M+Na]^+$ adducts) were detected for the 488 analyzed standards, with $^{TW}CCS_{N2}$ values ranging from 118.6 to 329.4 Å^2, whose distribution is shown in Figure III-2.2. As expected, a significant correlation (R^2 = 0.880 and 0.878 for $[M+H]^+$ and $[M+Na]^+$, respectively) was found between the CCS and the respective ion *m/z*. Interestingly, lower R^2 were observed in the present work with respect to similar previous studies which focused on specific compound classes characterized by recurring sub-units/structures (e.g., phenolic compounds, peptides) (Gonzales et al., 2016; Mosier et al., 2002). In fact, the standard analyzed in this work contained several types of small molecules: carbonyls, organic acids, esters and amides; including alkyl and aryl moieties, typical of some classes of additives (see Figure III-2.2), the chemical classes of analyzed standards were obtained from ClassyFire (Djoumbou Feunang et al., 2016), and shown in Figure III-2.3. Benzenoids, lipids and lipid-like molecules and organoheterocyclic compounds seems to be the major classes, some additives: phthalate-based plasticizers, antioxidants, bisphenols, primary aromatic amines belong to benzenoids.

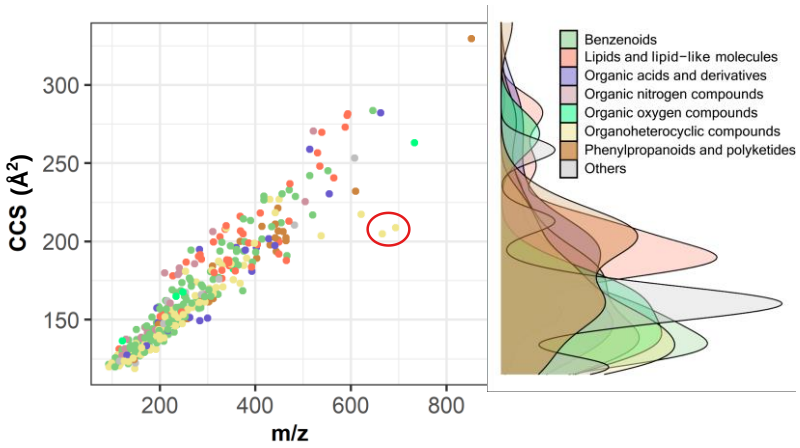


Figure III-2.2. Correlation between *m/z* and CCS of $[M+H]^+$ adducts and CCS density distributions according to the compounds' chemical class. The red circle highlights bromophenol blue and bromocresol green, that significantly deviate from the trend.

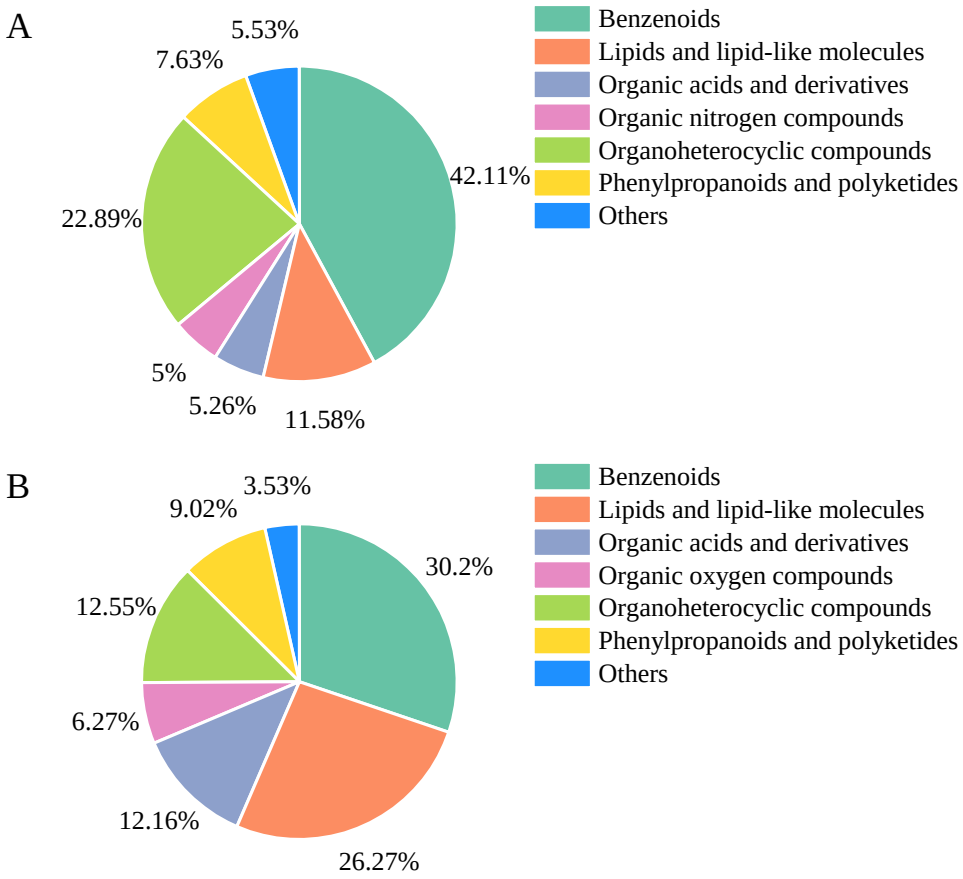


Figure III-2.3. Chemical classes of compounds for (A) $[M+H]^+$ and (B) $[M+Na]^+$.

Collision cross section and mass-to-charge ratio for both protonated and sodiated molecules presented 12% orthogonality (O) (Broeckling et al., 2021), calculated as follows:

$$O = (1 - R^2) * 100$$

where R^2 is the Pearson determination coefficient of the linear regression. This suggests that the inclusion of the CCS into a compound elucidation workflow for E&L testing could potentially increase peak capacity of >10% compared to retention time and accurate mass alone; this could ultimately increase the number of detected and identified analytes. Similar observations were made in metabolomics context by several authors (Blaženović et al., 2019; Nichols et al., 2018).

Molecular mass was not the only descriptor affecting the CCS values, two compounds significantly deviated from the mass/CCS trend. These were bromophenol

blue and bromocresol green, dyes used in the packaging industry and as pH indicators. In addition to C, H, O and N, these compounds include bromine (Br) within their structure, which could be ascribed to the observed negative deviation in the correlation plot.

For most of the analyzed molecules, $[M+Na]^+$ showed higher CCS values compared to $[M+H]^+$, as expected, due to the higher atomic radius and mass of Na over H. However, in some cases, the CCS values for $[M+H]^+$ were higher than those of $[M+Na]^+$. For example, Bis(2,4-dicumylphenoxy) pentaerythritol diphosphite, a common antioxidant used in food packaging, presented $^{TW}CCS_{N2}$ values of 329.4 and 311.4 $Å^2$ for its $[M+H]^+$ and $[M+Na]^+$ adducts, respectively. The sodium can be trapped in the core of the molecule and the proton might be protruding from one side of the molecule, thus resulting in the protonated adduct to be larger in size compared to the sodiated adduct.

4.2. Charge Isomers

In some cases, certain compounds can adopt multiple gas-phase conformers, resulting in multiple Gaussian-shaped arrival time distributions (ATD). In ESI+, this is commonly due to the presence of multiple equivalent protonation sites on the neutral molecule (giving rise to protomers), as well as multiple stable conformers from a single protonation site. If a charge isomer pair is sufficiently resolved in the IM dimension, the peak-detection algorithm will recognize two different components and will assign two discrete CCS values. The relationship between the charge location and the experimental CCS is logical, as the location of the charge affects the three-dimensional conformation of an ion, thus the CCS will be affected too. N-Ethyl-p-toluenesulfonamide, a commonly used plasticizer in polyamides and cellulose acetate materials, showed two $^{TW}CCS_{N2}$ values for its $[M+H]^+$ adduct. As shown in Figure III-2.4, protonation might occur on both O and N, leading to two different charge isomers, characterized by a double peak in the ATD of this compound, therefore leading to different CCS. Interestingly, by replacing methanol with acetonitrile as the organic

mobile phase, the formation of more compact conformation is favored. Warnke and co-workers, found that aprotic solvents can facilitate the protonation of amines, whereas methanol/water facilitate the protonation on carbonyl oxygen (Warnke et al., 2015). This led us to speculate that the first species (4.15 ms) corresponded to the protonation of the nitrogen atom, forming the quaternary ammonium cation; whilst the second species (4.83 ms) was represented by the protonation of the oxygen atom.

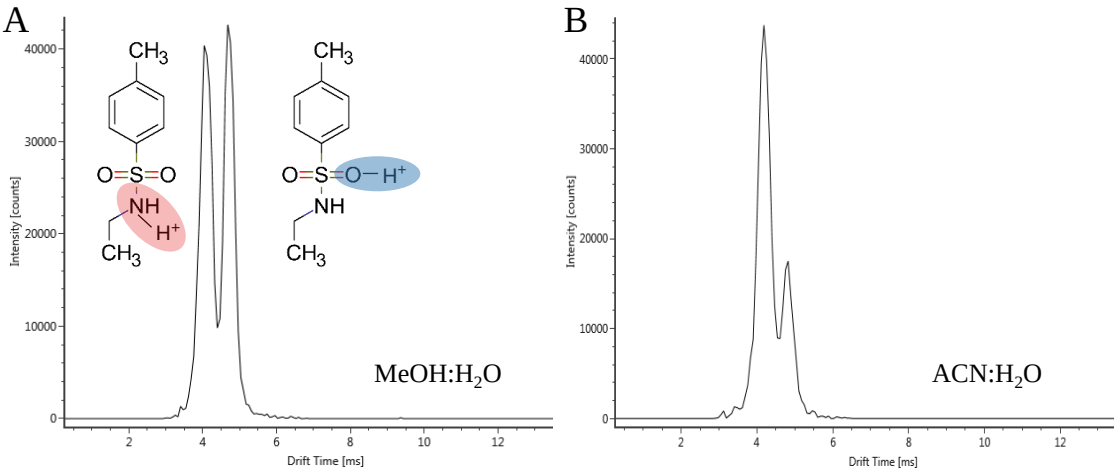


Figure III-2.4. Arrival time distribution (ATD) of N-Ethyl-p-toluenesulfonamide (m/z 199.0667) and its possible charge isomers. (A) in methanol/water mobile phase, (B) in acetonitrile/water mobile phase.

The relatively unpredictable formation of charge isomers and, generally speaking, conformers, represents a great challenge when attempting to create a CCS database and to apply prediction models. Essentially, the MDs for such isomers will likely be identical, regardless of the reference MD library of choice; thus, the prediction algorithm will be unable to generate multiple outputs for the isomeric adducts.

4.3. Dimeric Ionic Species.

In some instances, the presence of two (or more) features for the same precursor ion can be due to the formation of dimers, trimers or other noncovalent clusters in the ESI source, which will be subjected to a change in conformation or chemical reaction while traveling through the mobility cell (e.g. gas-phase collisional ion activation) or at

a later stage within the ion path (Ruotolo et al., 2008).

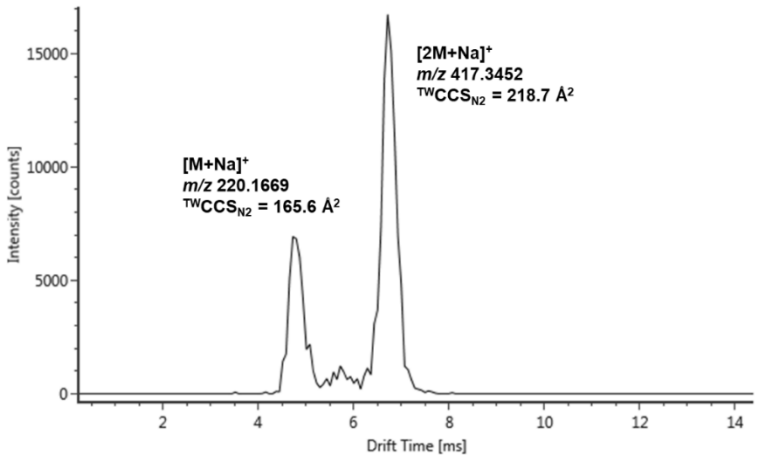


Figure III-2.5. Arrival time distribution of 12-aminododecanolactam.

For example, 12-Aminododecanolactam is a cyclic monomer of polyamide 12, the mobility trace of its sodiated precursor ions is shown in Figure III-2.5. Two distinct peaks were observed at 4.79 and 6.74 ms. Besides these two main peaks, a less intense peak at 5.78 ms was also discerned, corresponding to a CCS value of 191.3 $Å^2$. It can be speculated that the dimer precursor ion $[2M+Na]^+$ undergoes a fragmentation process to $[M+Na]^+$ during its permanence time in the drift region, leading to two different peaks in the ATD, with an additional peak which could correspond to a transition-state species or an artefact. Another example is tributyl phosphate, a commonly used defoamer and plasticizer, which formed $[2M+H]^+$ and $[2M+Na]^+$ in the ESI source. A portion of these ions could have fragmented to $[M+H]^+$ and $[M+Na]^+$ after passing through the drift cell or while transiting through the ion guides, thus leading to two distinct peaks in the mobility trace (Figure III-2.6).

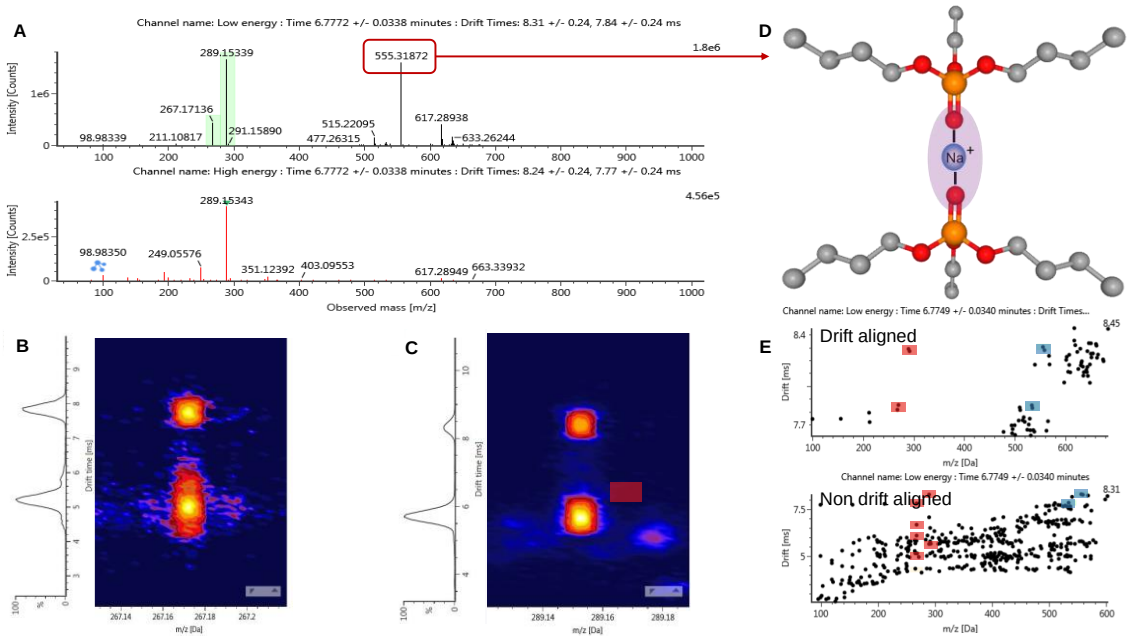


Figure III-2.6. (A) Mass spectra of tributyl phosphate; (B) mobility trace of $[M+H]^+$; (C) mobility trace of $[M+Na]^+$; (D) possible structure of the dimer $[2M+Na]^+$ of tributyl phosphate; (E) drift time of $[M+H]^+$, $[M+Na]^+$ and dimeric ions.

4.4. Develop and Optimize CCS Prediction Method.

After the elimination of non-informative descriptors, the final number of retained descriptors was n = 3036. Four models were developed for the $[M+H]^+$ and $[M+Na]^+$ adducts, respectively. Through the feature selection by SR, 1029 and 862 descriptors were selected for $[M+H]^+$ and $[M+Na]^+$ adducts, respectively.

The performances of the final PLS and SVM models are summarized in Table III-2.3. Overall remarkable prediction performance was achieved for $[M+H]^+$, regardless of the considered MDs set. Furthermore, it is worth stressing that any issue related to data overfitting can be diagnosed and excluded as the prediction ability was assessed by external validation. KernelPLS can deal with nonlinear behavior, For $[M+H]^+$, KernelPLS shows slight improvements in the prediction accuracy compared to SIMPLS, but is still less accurate than SVM. For $[M+Na]^+$, KernelPLS doesn't show significant difference with SIMPLS. Slightly better prediction accuracy was achieved by SVM for both $[M+H]^+$ and $[M+Na]^+$, 62.1% and 54.7% of compounds in $[M+H]^+$ and $[M+Na]^+$

were predicted with <2% errors. The prediction of CCS was less accurate along with the feature selection, as can be seen in Table III-2.3, in SVM-based models, the proportions of compounds with <2% predicted errors decreased from 62.1% to 53.7%, from 54.7% to 48.4%, for $[M+H]^+$ and $[M+Na]^+$, respectively. For these reasons, we refer to SVM and 3036 descriptors for the prediction of CCS of $[M+H]^+$ and $[M+Na]^+$.

Table III-2.3. Validation indices of PLS and SVM model using alvaDesc descriptors. R^2_P: external validation coefficient of determination; RMSEP: root mean square error of prediction; <2%, <3%, 5%: the proportions of molecules with predicted error less than 2%, 3% and 5%, respectively; MRE: median relative error.

Adducts	Descriptors	Algorithm	R^2_P	RMSEP	<2%	<3%	<5%	MRE
$[M+H]^+$	alvaDesc_3036	SIMPLS	0.980	5.3	58.9%	70.5%	93.7%	1.59%
		KernelPLS	0.983	4.7	57.9%	72.6%	93.7%	1.44%
		SVM	0.984	4.5	62.1%	74.7%	92.6%	1.50%
	alvaDesc_1029	SIMPLS	0.986	4.3	60.0%	71.6%	95.8%	1.81%
		SVM	0.987	4.2	53.7%	71.6%	96.8%	1.77%
$[M+Na]^+$	alvaDesc_3036	SIMPLS	0.949	8.0	45.3%	62.5%	82.8%	2.16%
		KernelPLS	0.949	8.1	45.3%	64.1%	81.3%	2.14%
		SVM	0.948	8.2	54.7%	71.9%	81.3%	1.82%
	alvaDesc_862	SIMPLS	0.922	10.0	40.6%	51.6%	76.6%	2.81%
		SVM	0.926	9.6	48.4%	64.0%	79.7%	2.18%

The first 25 important descriptors for the prediction of CCS are shown in Figure III-2.7. Hosoya-like index (log function) from Barysz matrix weighted by Sanderson electronegativity (Ho_Dz.e.), Hosoya-like index (log function) from Barysz matrix weighted by ionization potential (Ho_Dz.i.), McGowan volume (Vx), van der Waals volume from McGowan volume (VvdwMG), sum of atomic Van der Waals volumes (Sv) were the most important five descriptors for the prediction of CCS of $[M+H]^+$. Ho_Dz.e. was used to predict CCS of $[M-H]^-$ previously (Soper-Hopper et al., 2020). Other type of important MDs were 2D matrix-based descriptors, such as spectral moment of order 3 from Barysz matrix weighted by Sanderson electronegativity

(SM3_Dz.e.) and Hosoya-like index (log function) from topological distance matrix (Ho_D), these MDs were also used to predict CCS values (Soper-Hopper et al., 2017; Soper-Hopper et al., 2020). Sum of atomic polarizabilities (Sp) and Ghose-Crippen molar refractivity (AMR) were another two important MDs, which were used in CCS prediction in Zhou et al. (2016).

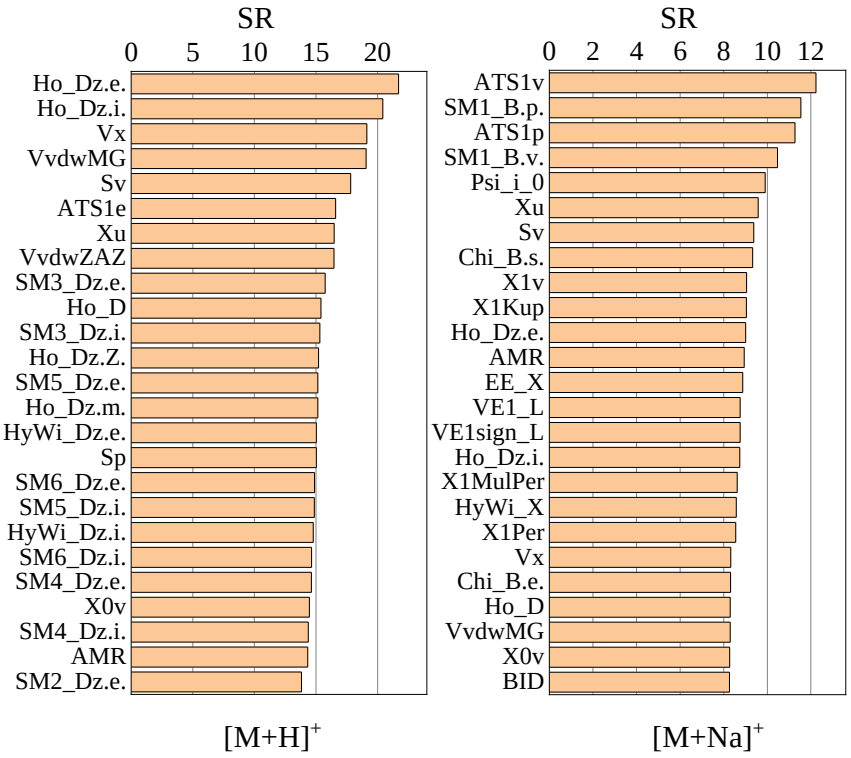


Figure III-2.7. SR values of first 25 important molecular descriptors in $[M+H]^+$ and $[M+Na]^+$ models.

Relative prediction residuals of validation set are shown in Figure III-2.8. When comparing the $[M+H]^+$ and $[M+Na]^+$ models, the former showed a better predictive performance, 92.6% (88/95) of protonated molecules showed prediction errors less than ±5%; for $[M+Na]^+$, only 81.3% (52/64) of molecules were predicted with ±5% error. This phenomenon can possibly be due to the fact that MDs were calculated on the neutral form of the molecules. The sodium ion has higher atomic radius compared with that of a proton, thus the descriptors of sodium adducts could differ significantly compared to the descriptors of the neutral molecules. This observation is in accordance with the findings from Bijlsma et al., where the author obtained lower prediction errors

for $[M+H]^+$ compared to $[M+Na]^+$ species (Bijlsma et al., 2017).

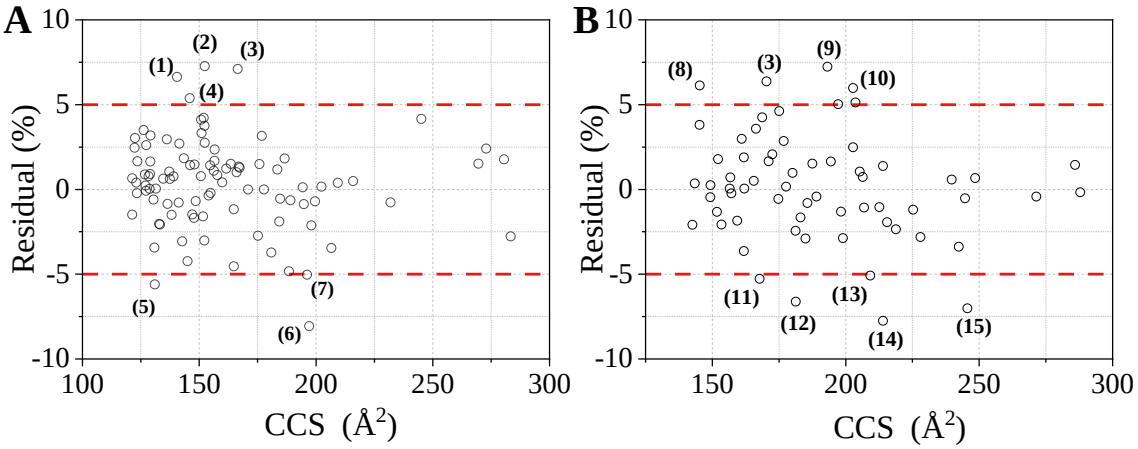


Figure III-2.8. Residuals% (percentage relative prediction error) of the external validation set for $[M+H]^+$ (A), and $[M+Na]^+$ (B) adducts: (1) 4,4'-Difluorobenzophenone; (2) 4-Aminophenyl sulfone; (3) Tebuconazole; (4) 2,2',6',2''-Terpyridine; (5) Phenyl isothiocyanate; (6) Diphenyl(2,4,6-trimethylbenzoyl)phosphine oxide; (7) 4-(Dodecyloxy)benzoic acid; (8) 1,4,7-trioxacyclotridecane-8,13-dione; (9) Dehydrocholic acid; (10) Tetracycline; (11) Dibutyl phosphate; (12) Triphenylphosphine oxide; (13) Corticosterone; (14) Testosterone propionate, (15) Rutin.

4.5. Outliers Detection.

The molecules with prediction errors higher than ±5% (outliers) were further investigated to try to understand the cause of poor prediction. The measured and predicted $^{TW}CCS_{N2}$ of sodiated (-)-Erythromycin were 268.2 and 223.6 Å^2, respectively (prediction deviation of -16.6%). The high prediction error of this compound could be a consequence of the fact that only three compounds with CCS values higher than 280 Å^2 were present in the calibration set. The limited training data points in the high-end CCS range could bias the prediction. In addition, some compounds containing halogens (fluorine, chlorine, bromine) also showed relatively high prediction errors. Tebuconazole, a commonly used triazole fungicide, its protonation ion had prediction error of 7.1% (measured 164.8 Å^2, predicted 178.3 Å^2). Bisacylphosphine oxides, a commonly used UV photoinitiator in packaging, containing phosphorus, also presented

a high prediction error of 5.1% (measured 203.6 Å^2, predicted 214.1 Å^2). The presence of these outliers may be due to the fact that most compounds in the data set prevalently contained C, H, O, N; only few compounds contained halogens and P. This highlights the importance of the chemical class when considering such tools. To further improve the model, the incorporation of more compounds with diverse chemical structures, especially the compounds with high molecular mass and with less common elements, such as halogens and phosphorus, should be considered.

4.6. Comparison of the Herein-Developed SVM Model with Existing CCS Predicting Tools.

CCSondemand is a recently developed CCS prediction tool, which is based on extreme gradient boosting (XGBoost) algorithm and 7325 $^{TW}CCS_{N2}$ data of different chemical classes (Broeckling et al., 2021). AllCCS is based on support vector regression algorithm and more than 5000 experimental CCS records (Zhou et al., 2020), CCSbase is a web interface which breaks down the chemical structural diversity by unsupervised clustering, followed by training of specific prediction models on each cluster (Ross et al., 2020). The comparison of CCS prediction of validation set between our SVM model with these three CCS predicting tools is illustrated in Table III-2.4 and Figure III-2.9.

Table III-2.4. Predictive performance indices of the tested models for the prediction of extractables and leachables compounds analyzed in this study.

Adducts	Models	R^2_p	RMSEP	<2%	<5%	MRE
M+H	SVM	0.984	4.5	62.1%	92.6%	1.50%
	CCSondemand	0.986	4.3	62.5%	92.6%	1.35%
	AllCCS	0.971	6.6	41.1%	83.2%	2.41%
	CCSbase	0.976	5.9	48.4%	81.1%	2.02%
M+Na	SVM	0.948	8.2	54.7%	81.3%	1.82%
	CCSondemand	0.949	8.9	34.4%	76.6%	1.82%
	AllCCS	0.922	13.3	25.0%	54.7%	4.37%
	CCSbase	0.940	8.8	37.5%	71.9%	2.42%

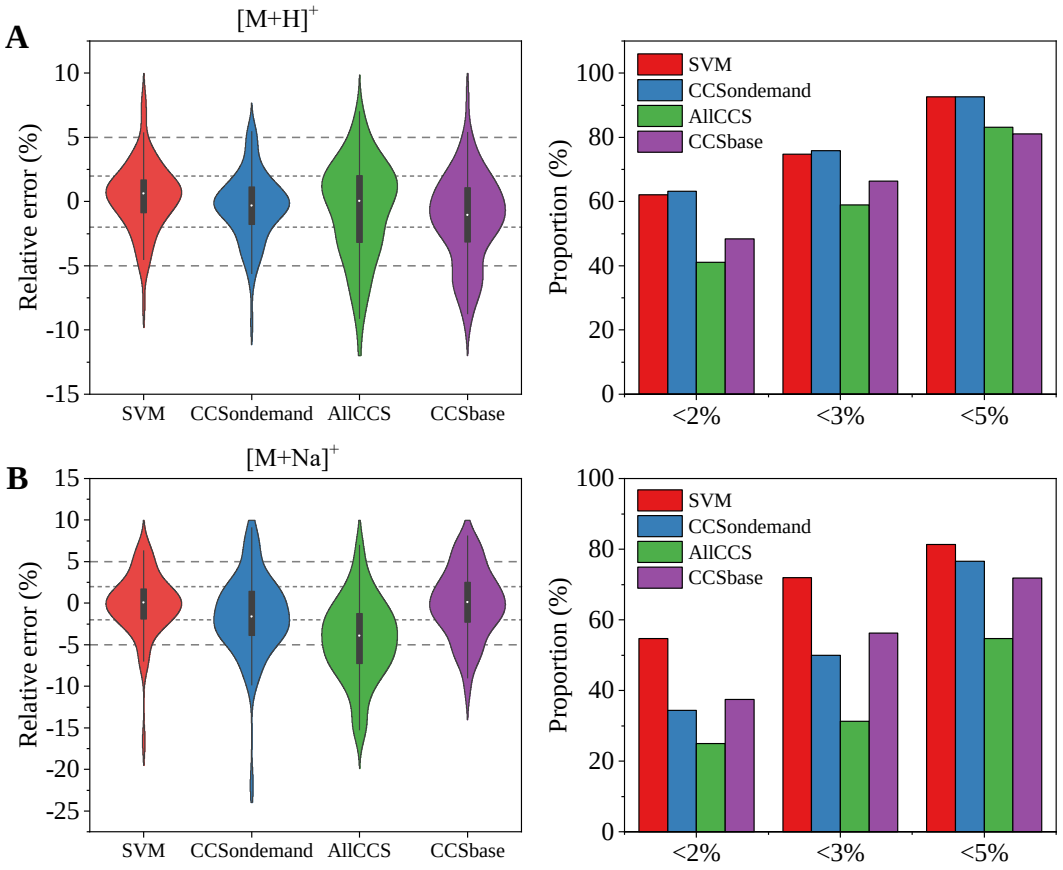


Figure III-2.9. Violin-plots and bar-plots showing the comparison between the developed PLS-R model with other predicting tools. $[M+H]^+$ (A), and $[M+Na]^+$ (B) adducts.

As can be seen, the herein developed model presented better predictive performance for both $[M+H]^+$ and $[M+Na]^+$ compared to other tools: 92.6% (88/95) of protonated molecules and 81.3% (52/64) sodiated molecules showed prediction error less than 5%. CCSondemand showed a comparable predictive ability for $[M+H]^+$ and a slightly worse predictive ability for $[M+Na]^+$. This is understandable because the dataset used for the development of CCSondemand includes a portion of experimental $^{TW}CCS_{N2}$ data of chemicals in food packaging. AllCCS and CCSbase showed less accurate predictions, where 83.2% and 81.1% of $[M+H]^+$ were predicted with <5% error, respectively. Some compounds typically used as food packaging additives were predicted with high errors: for 3,5-Di-Tert-butyl-4-hydroxybenzaldehyde, a degradation product of butylated hydroxytoluene (BHT) (Fries and Püttmann, 2002) with the measured $^{TW}CCS_{N2}$ of its $[M+H]^+$ is 164.9 $Å^2$, AllCCS gave a predicted CCS of 149.8

Å^2 (-9.1%) and CCSbase gave a predicted CCS of 157.1 Å^2 (-4.7%). Some primary aromatic amines also presented high prediction error by AllCCS and CCSbase, such as 4-aminobiphenyl (-5.8% and -6.4, respectively), benzidine (-6.3% and -8.0%, respectively) and 2,6-dimethylaniline (-6.0% and -5.6, respectively). Additionally, CCSbase gave a high prediction error for aniline (-5.2%), 4-chloroaniline (-8.0%), 4-chloro-2-methylaniline (-6.3%) and 5-chloro-2-methylaniline (-6.1%). The relatively less accurate CCS prediction of these kinds of compounds by AllCCS and CCSbase was possibly due to the fact that the molecules used for model training do not exhibit the similar structural characteristics with the chemicals in food packaging, the quality of prediction is notably affected by the types of molecules used for training (Zhou et al., 2020). Another thing needed to be mentioned is that even though the SVM model herein showed a better CCS predictive performance than AllCCS and CCSbase for the chemicals in food packaging, the more diverse chemical classes of AllCCS and CCSbase cannot be ignored, these two prediction tools can be applied to a wide variety of molecules.

4.7. Application of SVM to NIAS Identification.

The applicability of the developed CCS prediction model to the NIAS identification was further assessed by the analysis of a series of oligomers from adhesives and polyamides (PAs). Oligomers are molecules that consist of identical repeating units, which can be formed by the incomplete polymerization of monomers during polymer manufacturing and also due to the polymer degradation process (Ubeda et al., 2018). Based on previous knowledge of the composition of adhesives and PAs (Abe et al., 2016; Canellas et al., 2015b), 12 oligomers were tentatively identified through suspect screening. Four adhesive oligomers were derived from the reaction between adipic acid and 1,4-butanediol, five PA6 oligomers originated from the polymerization of caprolactam, and three PA66 oligomers were derived from the reaction of 1,6-diaminohexane and adipic acid. The detailed comparison between experimental and predicted CCS of oligomers is shown in Table III-2.5. For eleven

compounds the confidence of the structural elucidation process was improved by considering the predicted $^{TW}CCS_{N2}$ values within the workflow. The oligomers with low and high mass tend to present higher prediction errors. This also emphasized the importance of incorporating more high-mass and low-mass molecules in training set. The mass spectra and fragments assignment of 1,6,13,18,25,30-hexaoxacyclohexatriacontane-7,12,19,24,31, 36-hexone is shown in Figure III-2.10.

Table III-2.5. Experimental and predicted $^{TW}CCS_{N2}$ of tentatively identified oligomers from adhesive and polyamides (PAs).

	Compounds	MW	Adducts	Experimental $^{TW}CCS_{N2}$ (Å2)	Predicted $^{TW}CCS_{N2}$ (Å^2)	Errors (%)
Adhesive oligomers	1,6-Dioxacyclododecane-7,12-dione	200.1049	$[M+H]^+$	-	-	-
			$[M+Na]^+$	148.6	154.4	3.9
	1,6,13,18-Tetraoxacyclotetracosane-7,12,19,24-tetrone	400.2097	$[M+H]^+$	189.2	194.8	3.0
			$[M+Na]^+$	188.4	193.2	2.6
	1,6,13,18,25,30-Hexaoxacyclohexatriacontane-7,12,19,24,31,36-hexone	600.3146	$[M+H]^+$	234.2	234.9	0.3
			$[M+Na]^+$	228.5	234.8	2.8
	1,6,13,18,25,30,37,42-Octaoxacyclooctatetracontane-7,12,19,24,31,36,43,48-octaone	800.4194	$[M+H]^+$	275.1	262.9	**-4.4**
			$[M+Na]^+$	266.9	261.2	-2.1
PA 6 oligomers	PA6 dimer	226.1681	$[M+H]^+$	152.5	152.1	-0.2
			$[M+Na]^+$	155.5	163.0	**4.9**
	PA6 trimer	339.2522	$[M+H]^+$	180.0	177.9	-1.2
			$[M+Na]^+$	181.6	185.8	2.3
	PA6 tetramer	452.3363	$[M+H]^+$	208.0	208.1	0.1
			$[M+Na]^+$	212.7	214.6	0.9
	PA6 pentamer	565.4203	$[M+H]^+$	235.5	231.7	-1.6
			$[M+Na]^+$	232.6	234.8	1.0
	PA6 hexamer	678.5044	$[M+H]^+$	261.9	259.7	-0.8
			$[M+Na]^+$	261.5	259.7	-0.7
PA 66 oligomers	PA66 monomer	226.1681	$[M+H]^+$	151.3	150.6	-0.4
			$[M+Na]^+$	155.7	163.4	**5.0**
	PA66 dimer	452.3363	$[M+H]^+$	210.6	210.9	0.2
			$[M+Na]^+$	213.2	216.4	1.5
	PA66 trimer	678.5044	$[M+H]^+$	269.4	258.2	**-4.2**
			$[M+Na]^+$	266.9	262.8	-1.6

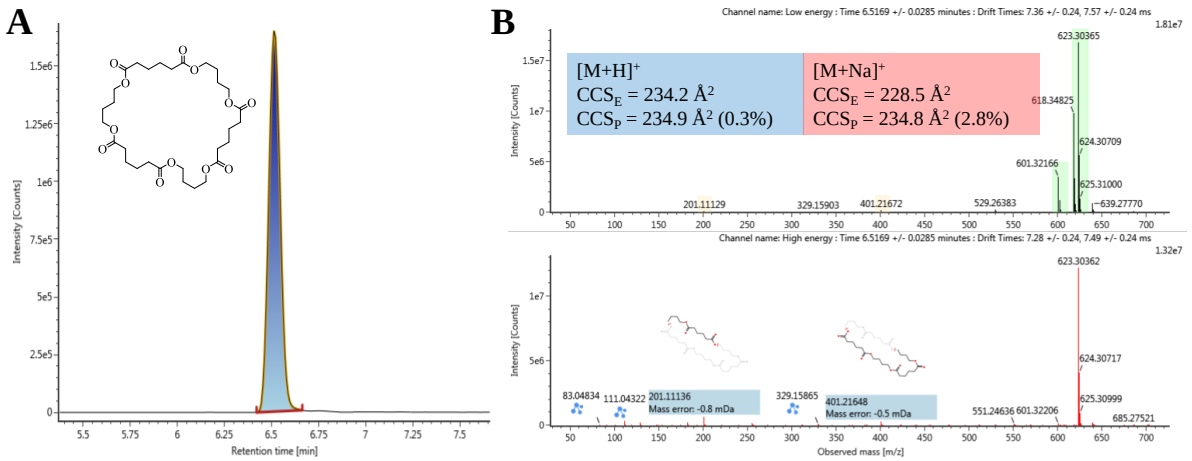


Figure III-2.10. (A) Chromatogram of 1,6,13,18,25,30-hexaoxacyclohexatriacontane-7,12,19,24,31,36-hexone; (B) Low and high-energy spectra, predicted CCS values, and fragments assignment.

Two cyclic oligomers were found by suspect screening of the reaction products between adipic acid and 1,4-butanediol, which showed the same RT and $^{TW}CCS_{N2}$ with the compounds identified in water-based adhesive: 1,6,13,18-Tetraoxacyclotetracosane-7,12,19,24-tetrone (5.83 min), $[M+H]^+$ (189.1 Å^2), $[M+Na]^+$ (188.0 Å^2) and $[M+NH_4]^+$ (190.9 Å^2); 1,6,13,18,25,30-hexaoxacyclohexatriacontane-7,12,19,24,31,36-hexone (6.47 min), $[M+H]^+$ (232.8 Å^2), $[M+Na]^+$ (228.7 Å^2) and $[M+NH_4]^+$ (238.3 Å^2). These data were in accordance with the prediction outcomes and further proved the reliability of identification.

4.8. Challenges and Opportunities of Existing Machine Learning CCS Prediction Models.

Charge isomers, dimers, chiral ions and IMS resolving power: In the previous section we have seen that small molecules can give rise to different charge isomers (e.g. protomers) and dimers. In both cases, multiple or distorted peaks in the ATD are obtained, which, when sufficiently resolved, are associated with multiple CCS values. Since current ML algorithms return a single CCS value for each compound as output data, these algorithms do not take into account the presence of charge isomers or chiral ions. This leads to potential incorrect predictions. In addition, conformers are often not

fully resolved due to the relatively low resolving power of existing IMS-MS systems (typically Rp <60 FWHM for linear temporally-dispersive IMS devices, such as TWIMS and DTIMS) (Dodds et al., 2017). Fortunately, technological development is on-going, and recent (or soon) commercially available platforms such as cyclic ion mobility (cIMS) (Giles et al., 2019) and structures for lossless ion manipulations (SLIM) (Webb et al., 2014) are expected to provide higher IMS resolving power, thus potentially better resolution of conformers.

The quality of input data contributes to a good prediction outcome. Perhaps we should dedicate more effort in the derivation of more accurate experimental CCS from instrumental analysis. So far, it is inappropriate to claim a prediction tool able to reach less than 2% CCS prediction relative error, as the current commercially available ion mobility platforms are set to produce CCS with deviations of ∼1-2% from standard values. For secondary IMS methods (i.e., TWIMS, TIMS, and DTIM operating via single-field method), the set of standards used as CCS calibrants should be specified. This is particularly important for TWIMS, as different compound classes used as calibrants can have an impact on the derivation of ^{TW}CCS (Gabelica and Marklund, 2018). Recently, Richardson and collaborators revisited the theory of T-wave IMS (Richardson et al., 2018) and proposed a more precise and robust calibration approach (Richardson et al., 2021), which will likely be adopted by next-generation TWIMS systems, and can further improve the experimental ^{TW}CCS values as input data.

MDs are mathematical representations of a compound calculated by well-specified algorithms which transform molecular structures into numbers (Soper-Hopper et al., 2020). MDs are used as X-block in SVM, and represent the second group of input data for model training in all MD-based machine learning algorithms. It is therefore crucial to obtain accurate MDs for reliable predictions. In the present work, as well as in most of the previously described studies, researchers make use of 2D-MDs calculated for the neutral form of the molecule. This tendency is not strictly correct, as the measured CCS is actually derived for the ionized form of compounds. Gonzales and co-

workers developed multiple ML prediction models for a group of deprotonated phenolic compounds (training n = 56, validation n = 16) using 3D-MDs after considering the proton removal from all possible titratable regions, followed by energy re-minimization, and considering the most stable conformers (Gonzales et al., 2016). The authors emphasized the ease of integration of their ML models in metabolite identification, compared to computational chemistry techniques (i.e. Mobcal). Yet, the generalization of Gonzales' method to a wider range of analytes and adduct types is not straightforward. When considering the 3D conformation of an ion, the first challenge is to assign the location of the charge. We have seen that not only the charge could reside on multiple discrete positions (i.e. charge isomers), but also some compounds can distribute the charge across the molecule (i.e. mesomeric effect due to the presence of conjugated bonds and aromatic structures). Furthermore, some compounds present dynamic conformations, which means that the transition from one energy-state to another could take place within the millisecond time frame, leading to splitting ATD peaks, sometimes ascribed to artifacts. Last but not least, also the ESI capillary temperature, voltage, and the source pressure can affect the internal energy distribution of electrosprayed ions, which in turn can affect the initial conformation of such species at the ionization stage (Gabelica et al., 2004). Factoring all these parameters into a prediction model becomes extremely complicated. A potential solution could be to integrate molecular modeling within the ML-prediction workflow in an automatic fashion, so that the user would be only requested to specify linear notation (e.g. SMILES) and adduct type into a script which automatically retrieves all possible ionic conformations, calculates 3D-MDs of the most stable ionic conformers, and uses such refined descriptors as input data for CCS prediction. The process of encoding refined ionic 3D-MDs as input features should be performed in a computationally cheap and easy-to-use manner, otherwise such prediction models would remain a tool for privileged users, not applicable to real-life identification workflows. Some authors used 3D-MDs of the neutral molecules, for example, Soper-Hopper and co-workers compared the prediction performance using 2D vs. 3D molecular descriptors (Soper-Hopper et al., 2020). They came to the conclusion that only in a few cases 3D models produced predictions better than 2D models,

obtaining a RMSE of 7.0 $Å^2$ (median error of 2%) using 2D-MDs. However, such performance could be further improved when considering 3D-MDs of the ionic species. Nevertheless, regardless of the discussion around 2D versus 3D, MDs mining remains highly customizable (i.e. different MD libraries and tools exist) and it is prone to user error. Thus, an efficient and standardized method for retrieving MDs should be pursued and agreed within the scientific community.

Model universality. Nowadays CCS prediction models tend to be built on a wider group of training data (e.g. Zhou et al. presented a model trained on more than 5000 experimental values) (Zhou et al., 2020), including a growing number of compounds and a mix of many different chemical classes. On the other hand, a different approach is to train ML algorithms on specific classes of compounds and to apply such prediction tools for specific applications. In the present work we demonstrated that the herein developed tool can outperform universal models for the prediction of chemicals in migration assessments from packaging materials. Nevertheless, the benefits of universal models should be acknowledged, as they can be used for all applications, regardless of the compound class.

Drift tube vs. traveling wave IMS. The most recent and comprehensive ML-based CCS prediction models also merged both ^{DT}CCS and ^{TW}CCS in both training and validation sets. This would further enhance the universality of such models. However, the fundamental difference of drift tube and traveling wave technology should not be neglected. Hinnenkamp et al. performed a study where the CCS of 124 different small molecules were measured on both DTIMS and TWIMS (Hinnenkamp et al., 2018). The authors found deviations <1% for most substances, but some compounds showed deviations up to 6.2%, which indicates that CCS databases cannot be used without care independently from the instrument type. Plante and co-workers noticed a decline of prediction performance of their CNN model based on a global testing set when considering only the Astarita dataset based on ^{TW}CCS (averaged R^2 from 0.97 to less than 0.9, and MRE from <2.6% to 5%) (Plante et al., 2019). The authors hypothesized

that a bias in measurement between data sets can be present.

5. Conclusions

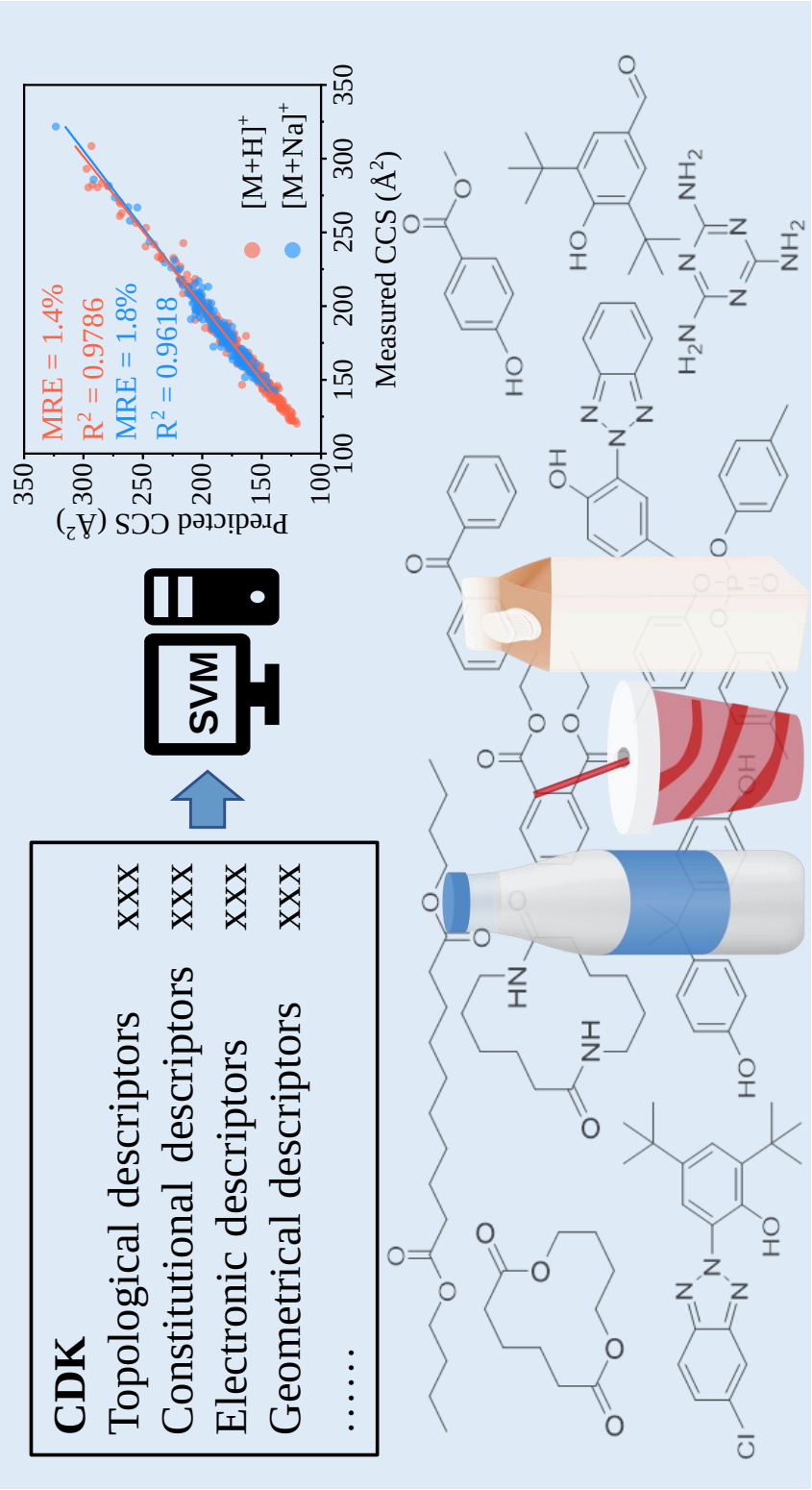
Unknown annotation is one of the major bottlenecks in untargeted E&L analysis. To accelerate the workflow from raw data processing to compound identification, multi-factor authentication with the integration of predicted CCS in combination with retention time, accurate mass and in-silico MS/MS tools, can facilitate this challenging task. In this study we developed a reliable $^{TW}CCS_{N2}$ prediction tool for chemicals in FCMs based on SVM. For more than 90% of protonated molecules the model accurately predicted CCS with relative errors below ±5%. The SVM model was successfully applied to the analysis of oligomers migrating from FCMs and adhesives, and it was integrated within our suspect and non-targeted analysis workflows for compound discovery and chemical migration assessment. The incorporation of a wider number of compounds in the training set, as well as the employment of a more accurate set of 3D-MDs based on energetically-minimized ion species could be explored to enhance model coverage and accuracy. Nevertheless, we believe that an automatic and universal approach for gathering the appropriate molecular descriptors from ionized species, also considering charge isomers, can be a game-changer in the prediction of CCS and it should be pursued in order to turn *in-house* prediction models into tools truly applicable in all laboratories.

Chapter 3

Prediction of Collision Cross Section Values for Extractables and Leachables from Plastic Products

1. Abstract

The use of ion mobility separation (IMS) in conjunction with high resolution mass spectrometry (HRMS) has proved to be a reliable and useful technique for the characterization of small molecules from plastic products. Collision cross section (CCS) values derived from IMS can be used as a structural descriptor to aid compound identification. One limitation of the application of IMS to the identification of chemicals from plastics is the lack of published empirical CCS values. As such, machine learning techniques can provide an alternative approach by generating predicted CCS values. Herein, experimental CCS values for over a thousand chemicals associated with plastics were collected from the literature and used to develop an accurate CCS prediction model for extractables and leachables from plastic products. The effect of different molecular descriptors and machine learning algorithms on the model performance were assessed. A support vector machine (SVM) model based on CDK descriptors, provided the most accurate prediction with 93.3% of CCS values for $[M+H]^+$ adducts and 95.0% of CCS values for $[M+Na]^+$ adducts in testing sets predicted with <5% error. Median relative errors (MRE) for the CCS values of the [M+H]+ and [M+Na]+ adducts were 1.42% and 1.76%, respectively. Subsequently, CCS values for the compounds in the Chemicals associated with Plastic Packaging Database (CPPdb) and the Food Contact Chemicals Database (FCCdb) were predicted using the SVM model developed herein. These values were integrated in our structural elucidation workflow and applied to the identification of plastic-related chemicals in river water. False positives were reduced and identification confidence level was improved by the incorporation of predicted CCS values in the suspect screening workflow.


CDK
Topological descriptors XXX
Constitutional descriptors XXX
Electronic descriptors XXX
Geometrical descriptors XXX
……
SVM
Predicted CCS (Å²)
Measured CCS (Å²)
MRE = 1.4%
R² = 0.9786
MRE = 1.8%
R² = 0.9618
[M+H]⁺
[M+Na]⁺
100
150
200
250
300
350

2. Introduction

Plastics play an important role in our daily life, as they are used in a variety of materials, including packaging, building and construction materials, transportation, and electrical and electronic components (Ilyas et al., 2018). It has been reported that up to 2015, approximate 6300 million metric tons of plastic waste was generated of which only 9% was recycled. The remaining plastic waste was either incinerated, accumulated in landfills or disposed of in natural environments (Geyer et al., 2017). _ENREF_3The impact of plastic packaging on the environment and, subsequently, human health is of great concern due to the release of microplastics (He et al., 2021; Hernandez et al., 2019; Prata et al., 2020), and low molecular weight (MW) chemicals (Biryol et al., 2017; Hahladakis et al., 2018; Liu and Mabury, 2020; Su et al., 2021a). During the production of plastics, a variety of additives are incorporated into the polymeric formulations to enhance favorable characteristics and extend service life. Commonly used additives include plasticizers, flame retardants, lubricants, antioxidants and UV stabilizers (Hahladakis et al., 2018). Such additives have been detected in indoor dust (Liu and Mabury, 2019a; Liu et al., 2019), airborne particulate matters (Liu et al., 2020; Liu et al., 2021; Shi et al., 2020), waste water (Gonzalez-Marino et al., 2021), soils (Gong et al., 2021), rivers and oceans (Bolivar-Subirats et al., 2020; Bolivar-Subirats et al., 2021; Schmidt et al., 2019). Plastic products have become an important source of contaminants in aquatic and terrestrial environments.

In addition to the known substances included during the production of plastic materials, non-intentionally added substances (NIAS) can also occur. Typical NIAS include impurities, oligomers and degradation products of material components (Nerin et al., 2013), For example, organophosphate esters can result from the oxidation of organophosphite antioxidants in plastics, and have been detected in indoor dust (Liu and Mabury, 2018b, 2019b). If plastic products are made from recycled plastics, NIAS can also include contaminants resulting from the previous use of the material or from the recycling process itself (Cecon et al., 2021). In recent years, the presence of

perfluoroalkyl substances (PFAS) in plastic products has also attracted the attention of food safety and environmental authorities (Curtzwiler et al., 2021; Dodds et al., 2021; Luo et al., 2020b).

The complete structural elucidation of extractables and leachables from plastics is a challenging process due to the complexity of the matrix. In recent years, ion mobility separation (IMS) coupled to high resolution mass spectrometry (HRMS) has emerged as a promising tool for analyzing complex samples (Belova et al., 2021; Canellas et al., 2019; Celma et al., 2020; Fabregat-Safont et al., 2021; Vera et al., 2019). IMS can separate molecules based on their shape, size and charge (D'Atri et al., 2018). Collision cross-section (CCS), derived from IMS, is a physico-chemical property of an ion and is related to the chemical structure and three-dimensional conformation of the molecule (D'Atri et al., 2018). In addition, since CCS measurements are independent from chromatographic and mass spectrometric conditions, as well as sample matrix (Paglia and Astarita, 2017), they can be used as an additional molecular identifier to increase specificity and identification confidence. Celma et al. (2021) showed that CCS of imazalil was not affected by sample matrix, while the retention time (RT) deviations ranged from 0.14 min to 0.30 min, the consistent CCS values provided an extra point for unknown identification. In addition, incorporation the CCS values into the annotation process can help to reduce false positive identifications (Zhou et al., 2020) and enable structural isomers to be separated and identified (McCullagh et al., 2019b; Song et al., 2021).

Experimental CCS values of reference standards are often measured in order to confirm compound identification by comparing them to CCS values of candidate compounds in qualitative analyses. Although public CCS databases of pesticides (Bijlsma et al., 2017; Regueiro et al., 2016), drugs (Hines et al., 2017), steroids (Hernandez-Mesa et al., 2018), mycotoxins (Righetti et al., 2020) and chemicals in plastic food packaging (Song et al., 2022c) have been established, there remain many compounds that are not included in such libraries. As a matter of fact, many

experimental CCS values of chemicals in plastics are not available due to the unavailability or high price of commercial standards. In this case, theoretical CCS values can be alternatives to be used for suspect and untargeted screening analysis. Several public CCS machine learning prediction tools have appeared in recent years, such as MetCCS (Zhou et al., 2016), AllCCS (Zhou et al., 2020), CCSondemand (Broeckling et al., 2021), CCSbase (Ross et al., 2020) and DeepCCS (Plante et al., 2019). Some laboratories have also developed their own CCS prediction tools for specific classes of compounds, such as pesticides (Bijlsma et al., 2017), phenolics (Gonzales et al., 2016) and drugs (Mollerup et al., 2018). Many CCS values, belonging to different chemical classes, can provide a high structural diversity, and as such, the developed model can provide satisfactory prediction results for diverse chemical classes. At the time of writing, there are 3539, 7325, 7405 and 2439 CCS values in the datasets used by AllCCS, CCSondemand, CCSbase and DeepCCS, respectively.

In a previous study (Song et al., 2022c), 635 CCS values derived from 488 standards associated with plastic packaging, were used to develop a support vector machine (SVM) model to predict CCS values. The CCS values of 92.6% of protonated molecules were predicted with an error of less than 5%. The CCS values of some halogenated compounds were inaccurately predicted due to the lack of halogenated compounds in the training set. Consequently, in this study additional experimental CCS values of molecules related to plastics have been collected from literature, with the aim of achieving more accurate CCS prediction for chemicals found in plastics. The effect of different molecular descriptors and algorithms on the accuracy of the CCS prediction were also explored. Following optimization and external validation, the model was used to predict CCS values of molecules in two plastic-related databases: The Chemicals associated with Plastic Packaging Database (CPPdb) (Groh et al., 2019) and the Food Contact Chemicals Database (FCCdb) (Groh et al., 2021). FCCdb also contains many plastic-related chemicals since approximately 37% of food contact materials (FCMs) are made from plastics (Muncke, 2016). The two databases were subsequently converted into screening libraries, containing the predicted CCS values, and used for

the suspect screening of plastic-related chemicals in river water.

3. Materials and methods

3.1. CCS data collection and processing

A total of 2145 experimental traveling wave CCS ($^{TW}CCS_{N2}$) and drift tube CCS ($^{DT}CCS_{N2}$) values were collected from seven recent publications (Belova et al., 2021; Bijlsma et al., 2017; Celma et al., 2020; Hinnenkamp et al., 2018; Mullin et al., 2020; Regueiro et al., 2016; Song et al., 2022c), of which 1425 and 720 CCS values were for $[M+H]^+$ and $[M+Na]^+$ ions, respectively (Table III-3.1). The CCS values in the publication of Song and co-workers (Song et al., 2022c) were experimentally measured by injecting standards of chemicals associated with plastic food packaging. Four of the publications (Bijlsma et al., 2017; Celma et al., 2020; Hinnenkamp et al., 2018; Regueiro et al., 2016) include CCS values mainly for pesticides and pharmaceuticals found in environmental studies. The CCS values in these four databases were used in this study because pesticides are an important type of NIAS in plastic materials, especially those made from the recycled plastics (Su et al., 2021b). Additionally, many pesticides contain halogens in their structure, as such, the predictions of CCS values for halogenated compounds will be more accurate by including the pesticides in the CCS dataset. The last two publications (Belova et al., 2021; Mullin et al., 2020) mainly contain CCS values for organophosphorus flame retardants, compounds with a phosphate structure, which are common additives used in plastic materials. Since only three organophosphorus flame retardants were included in the previous self-built CCS database (Song et al., 2022c), the addition of the CCS values from these two publications significantly expanded the chemical diversity of the current study.

Table III-3.1. Experimental CCS values retrieved from scientific literatures.

Compound type	Technology	$[M+H]^+$	$[M+Na]^+$	Reference
Chemicals in food plastic packaging,	$^{TW}CCS_{N2}$	401	272	(Song et al.,

Compound type	Technology	$[M+H]^+$	$[M+Na]^+$	Reference
including antioxidant, plasticizers, UV absorbers, lubricants, and NIAS				2022c)
Pesticides	$^{TW}CCS_{N2}$	205	0	(Bijlsma et al., 2017)
Organic environmental pollutants, including illicit drugs, hormones, mycotoxins, new psychoactive substances, pesticides, and pharmaceuticals	$^{TW}CCS_{N2}$	460	243	(Celma et al., 2020)
Pesticides	$^{TW}CCS_{N2}$	177	34	(Regueiro et al., 2016)
Pesticides and pharmaceuticals	$^{TW}CCS_{N2}$	91	88	(Hinnenkamp et al., 2018)
Contaminants of emerging concern in human matrix, including bisphenols, plasticizers, organophosphate flame retardants and triazoles	$^{DT}CCS_{N2}$	55	81	(Belova et al., 2021)
Pollutants in indoor dust: flame retardants, pesticides	$^{TW}CCS_{N2}$	36	2	(Mullin et al., 2020)

CCS values for some compounds appeared in more than one publication. In such cases, the CCS data were rationalized as follows:

(1) Chemical information retrieval: Information including the compound identifier (CID), monoisotopic mass, molecular formula, canonical SMILES and InChIKey of each CCS record was retrieved from PubChem using the R package *webchem* (Szöcs et al., 2020).

(2) Calculation of median CCS values for duplicated records: In the cases where different names were used for the same compound in the different publications, the InChIKey was used as a unique identifier. The median and relative standard deviation (RSD) of multiple CCS values were calculated, and the median CCS values were used in the model.

A total of 1721 CCS values were retained after the consolidation of duplicate records, which included 1076 CCS values for $[M+H]^+$ ions and 645 CCS values for $[M+Na]^+$ ions. In consolidated data, the CCS values of 248 $[M+H]^+$ ions (23.0%) and

72 $[M+Na]^+$ ions (11.2%) were median values of multiple CCS records. The CCS dataset rationalization was performed using the R package *tidyverse* (Wickham et al., 2019) and the chemical class of each compound contributing to the model was obtained from ClassyFire (Djoumbou Feunang et al., 2016).

3.2. Calculation and selection of molecular descriptors

Molecular descriptors (MDs) play a crucial role in the prediction of CCS values. In this work, three types of molecular descriptors were calculated using OCHEM (Sushko et al., 2011) and ChemDes (Dong et al., 2015). The first descriptor dataset was calculated using alvaDesc v.2.0.4 within OCHEM, which contains 5666 descriptors including constitutional, topological, charge, and geometrical descriptors. The second descriptor dataset was calculated using CDK v2.3 from OCHEM, which contains 256 constitutional, topological, geometric, electronic and hybrid descriptors, the CDK descriptors were used for the prediction of CCS values in the AllCCS webserver (Zhou et al., 2020). The third descriptor dataset contains 196 RDKit descriptors calculated using ChemDes. RDKit descriptors were used in the development of the CCSondemand prediction tool) (Broeckling et al., 2021).

Descriptors that have a constant value or very few unique values relative to the number of samples have variance values equal or close to zero. Such descriptors contain little information and were considered less important for the model and excluded from the dataset. Correlation coefficients (r) between individual MDs and CCS were subsequently calculated and only the MDs for which $r > 0.6$ were retained. The remaining descriptors were auto-scaled to normalize the effect of magnitude. The alvaDesc MDs were further rationalized by considering Extreme Gradient Boosting (XGBoost) importance. In XGBoost, the contribution of each variable to model is calculated with respect to the number of times the variable is selected for splitting, weighted by the squared improvement to the model as a result of each split. The variable importance is then averaged across all the decision trees within the model (Elith et al., 2008). In this study, the alvaDesc MDs accounting for 99% and 95% of the total

XGBoost importance were retained.

3.3. Development of the CCS prediction model

For both $[M+H]^+$ and $[M+Na]^+$ ions, the data were randomly divided into training and testing sets in the ratio of 7:3. The training set was used for the calibration and optimization of the model and testing set was used for the external validation. The comparison of CCS prediction accuracy between various models (models developed with different algorithms and MDs in this study as well as public CCS prediction tools) was based on the testing set data. The R code for model building was provided in GitHub (https://github.com/songxuechao/plasticCCS).

In addition to the CCS data and descriptors, the machine learning algorithm employed was another important factor affecting the predictive performance of model. In this study two algorithms that are often used for CCS prediction were compared: XGBoost and SVM. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient and flexible (Chen and Guestrin, 2016), and was used to develop CCSondemand (Broeckling et al., 2021). XGBoost model tuning consisted of 576 combinations of five important model parameters: eta (0.01, 0.05, 0.1, 0.3), max_depth (3, 5, 7), min_child_weight (1, 3, 5), subsample (0.6, 0.7, 0.8, 0.9), and colsample_bytree (0.6, 0.7, 0.8, 0.9). All combinations were evaluated using the training dataset by a 10-fold cross validation. The optimal value of the *nrounds* parameter which controls the maximum number of iterations was returned using the minimized root mean square error of cross validation (RMSECV). Finally, the XGBoost model was built using the training dataset with the optimized combination of parameters using the R package *xgboost*. The importance of MDs in the model was also calculated.

SVM is also a commonly used machine learning algorithm and has previously been used for the prediction of CCS values (Zhou et al., 2016; Zhou et al., 2017). In this study, SVM with the radial basis function (RBF) kernel was used to build the model. Two important hyper-parameters were optimized in order to get accurate predictions:

cost of constraints violation (C) and gamma (γ). The C parameter trades off the predictive performance of training set against the model's margin, while the γ parameter defines how far the influence of a single training example reaches. Eight groups of C values (0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5)/N_{MD} (i.e. number of molecular descriptors) and nine γ values (2^0 to 2^8) formed 72 parameter combinations, which were then evaluated using 10-fold cross validation on the training set. The parameter combination providing the minimum RMSECV was used in the SVM model using the R package *e1071*.

The performance of the models was assessed by comparing the following parameters: the coefficient of determination of the prediction (R^2_p), the root mean square error of the prediction (RMSEP), the median relative error (MRE) and the percentage of molecules with relative deviations from experimental CCS values of less than 2%, 3% and 5%.

The prediction performance of our model was compared to three publicly available CCS prediction tools: CCSondemand (https://ccs.on-demand.waters.com) from Broeckling, C. and co-workers (Broeckling et al., 2021), AllCCS (http://allccs.zhulab.cn) from Zhu lab(Zhou et al., 2020) and CCSbase from Xu lab (http://ccsbase.net) (Ross et al., 2020).

3.4. The prediction of CCS values for compounds in CPPdb and FCCdb

The CPPdb consists of 4283 substances associated with plastic food packaging. The dataset was rationalized by removing the metals and salts together with any substances with same InChIKey (replicates). Finally, only substances with a neutral mass between 50-1200 were retained. After following this procedure, 2883 substances from the CPPdb were retained. The FCCdb dataset was also rationalized using the procedure described above, leading to 6508 substances retained in dataset. The CCS values of the compounds retained from the databases were then predicted using the model that yielded the best performance in this study. Meanwhile, the chemical space

covered by CPPdb, FCCdb and our collected molecules was compared.

3.5. Application of predicted CCS values to the analysis of plastic-related chemicals in Ebro River water

Two liters of surface water were sampled from the Ebro River near the urban areas of Zaragoza, Spain. The river water was stored in an amber glass bottle and treated on the day of collection, using the previously developed procedures (Fabregat-Safont et al., 2021). All the water sample was filtered by 0.7 glass fiber filter, 100 mL of aliquots were passed through the solid phase extraction (SPE) (Oasis HLB cartridge, 6cc/200mg, Waters Corp.), previously conditioned with 10 mL of methanol and 10 mL of water. The SPE cartridge dried for 10 mins and was eluted with 12 mL of methanol, the filtrate was evaporated to dryness at 45 ºC under a gentle stream of N_2, the residue was redissolved in 1.5 mL of methanol and analyzed by Vion IMS-QTof (Waters, Manchester, UK), this treatment was performed in triplicates.

The final samples were analyzed using a Vion IMS-QTof mass spectrometer. The chromatographic separation was performed using a CORTECS C18 column (2.1 × 100 mm, 1.6 μm particle size, 90 Å pore size) at a flow rate of 0.3 mL min-1. Mobile phases were water (A) and methanol (B), both acidified with 0.1% of formic acid (v/v). The initial proportion of B was 5%, increased to 100% over 7 minutes, kept at 100% from 7 to 11 minutes, decreased to 5% over 0.1 minutes and re-conditioned until 13 minutes. Data were acquired on the mass spectrometer in positive mode over the mass range of 50-1000 m/z with a scan time of 0.2 s. Electrospray ionization (ESI) conditions were as follows: capillary voltage, 1 kV; cone voltage, 30 V; source temperature, 120 ºC; desolvation temperature, 500 ºC; cone gas flow, 50 L h^{-1}; desolvation gas flow, 800 L h-1. Data were acquired in high definition MS^E mode, with the instrument was switching between two collision energy states (low energy: 6 eV, high energy ramp: 20-40 eV) in order to obtain precursor and fragment ions within a single acquisition. Leucine-Enkephalin ($[M+H]^+$, m/z 556.2766) at a concentration of 100 ng/mL was infused at a rate of 15 μL/min for real-time mass correction. IM separations were

performed with a travelling wave velocity of 250 m/s and IMS pulse height of 45 V, N_2 was used as the drift gas at a flow of 25 mL/min. The Vion platform works at a room temperature of 25 ºC.

The features (m/z_RT_CCS pairs) obtained from Vion IMS-QTof, were then screened against two plastic-related databases, CPPdb (2883 compounds) and FCCdb (6508 compounds), containing *m/z* values, adducts and predicted CCS values. The *m/z* deviations of the measured values were less than 5 ppm, as for CCS deviation, the filter setting was based on its prediction accuracy.

4. Results

4.1. CCS dataset

A total of 1076 and 645 CCS values were collated for $[M+H]^+$ and $[M+Na]^+$ adducts, respectively. CCS values ranged from 118.6 to 332.2 $Å^2$ for the $[M+H]^+$ data and from 134.7 to 321.9 $Å^2$ for the $[M+Na]^+$ data. Using ClassyFire (Djoumbou Feunang et al., 2016), the compounds were categorized into 10 super classes for the $[M+H]^+$ adduct and 11 super classes for the $[M+Na]^+$ adduct. The principal super classes assigned were benzenoids, organoheterocyclic compounds, lipids and lipid-like molecules, organic acids and derivatives (Figure III-3.1). Benzenoids include compounds commonly detected in plastics such as phthalate-based plasticizers, antioxidants, bisphenols, primary aromatic amines, and pesticides.

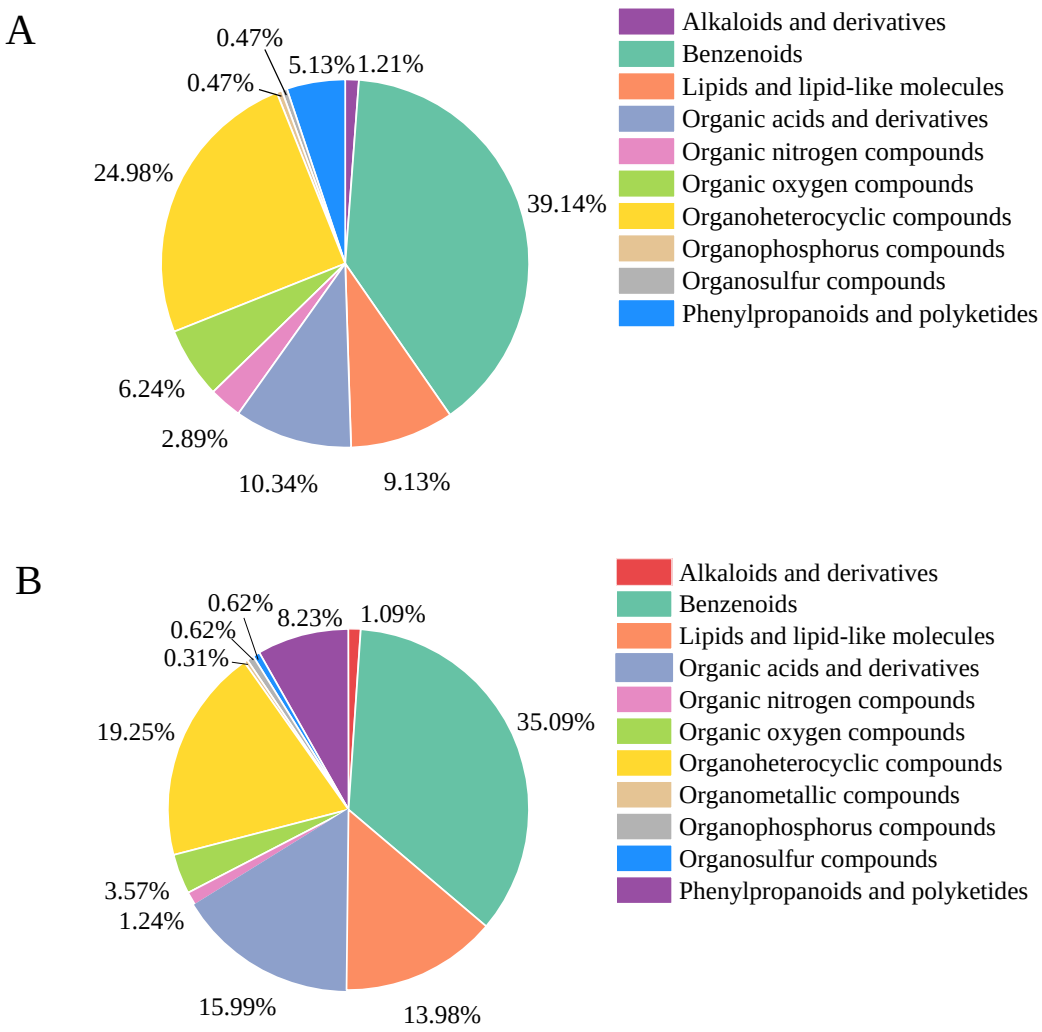


Figure III-3.1. Chemical classes of 1076 compounds in $[M+H]^+$ (A) and 645 compounds in $[M+Na]^+$ (B).

248 and 72 duplicate CCS values were found for $[M+H]^+$ and $[M+Na]^+$ adducts, respectively, across the seven publications, and the RSDs of the measurements are shown in Figure III-3.2. The RSD variation is less than 2% for 89.1% (221/248) of the $[M+H]^+$ adducts of the molecules, and 95.8% (69/72) of the $[M+Na]^+$ adducts. Consequently, there are 27 and 3 CCS values with RSDs higher than 2% for the $[M+H]^+$ and $[M+Na]^+$ adducts, respectively. The majority of CCS value with RSDs greater than 2% were obtained from the publications of Bijlsma et al. (2017), Celma et al. (2020), and Regueiro et al. (2016). It appears that pesticide and drug-like compounds are more likely to produce a high variation of CCS values. Such compounds include

picoxystrobin, acetopromazine, prochloraz and oxadixyl, with the variation of the CCS measurements for the last two compounds being more than 20 $Å^2$. The limit of CCS reproducibility, presence of protomers and inconsistent CCS calibration across different instrument systems are three possible sources of deviations in CCS measurements. A more detailed explanation is given in following section:

(1) Studies have shown that while the deviations of CCS values measured on traveling wave IMS (TWIMS) devices from different laboratories are generally less than 2%, in some cases the deviations can be at either extreme of the acceptable range leading to a higher RSD value.

(2) Presence of protomers. Isobaric protomers have been identified for some compounds due to molecules having multiple protonation sites (Hinnenkamp et al., 2018; McCullagh et al., 2019a). If a charged isomer pair is sufficiently resolved by IMS, different CCS values will be assigned to each isomer. Most pesticides contain multiple protonation sites in their structure, the presence of amine and carbonyl oxygen groups, for example. Two CCS values, 179.6 $Å^2$ and 191.21 $Å^2$, have been reported for marbofloxacin (Bijlsma et al., 2017; Celma et al., 2020), both protomers were detected by a cyclic ion mobility systems in McCullagh et al. (2019a). Warnke et al. (2015) has shown that the use of methanol can favor the protonation of carbonyl oxygen while acetonitrile can favor the protonation of the amine. In their work Regueiro et al. (2016) used acetonitrile and water as the mobile phase, which might explain the different CCS values of fenpyroximate (205.3 $Å^2$ versus 215.9 $Å^2$), picoxystrobin (194.5 $Å^2$ versus 177.5 $Å^2$) and oxadixyl (178.3 $Å^2$ versus 158.8 $Å^2$).

(3) Inconsistent CCS calibration across different instrument systems. The CCS values of 16 compounds measured by Bijlsma et al. (2017) were lower than those measured by Celma et al. (2020) and Regueiro et al. (2016). This consistent difference might imply that the TWIMS was calibrated using a different set of standards or by considering a different set of reference points.

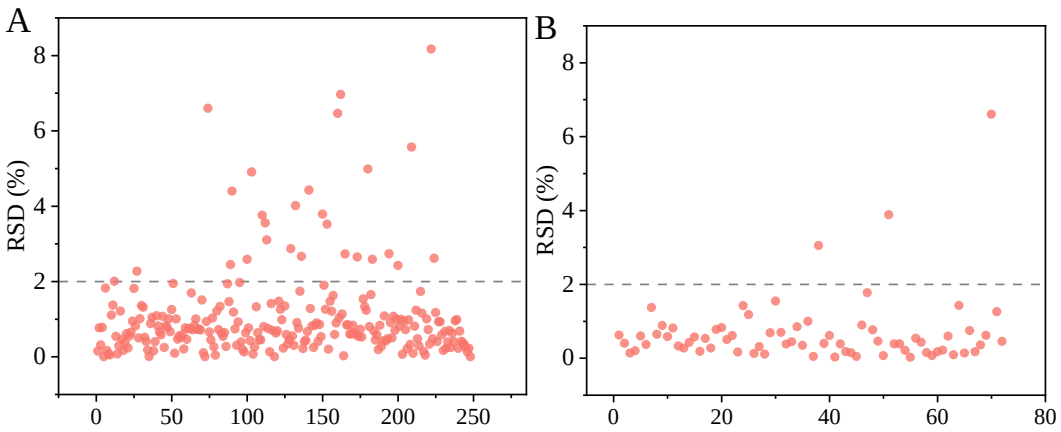


Figure III-3.2. Relative standard deviation (RSD) of collision cross section (CCS) values of identical molecules obtained from different instrument platforms and different laboratories, (A) $[M+H]^+$ and (B) $[M+Na]^+$.

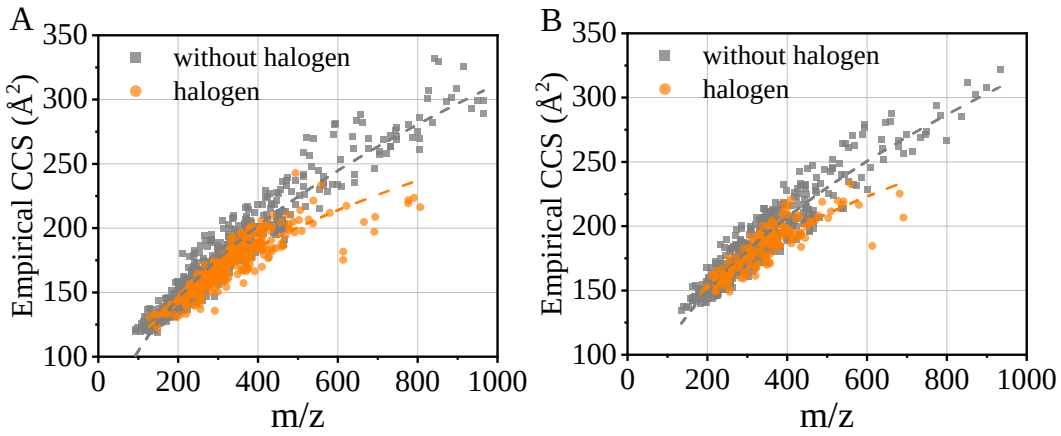


Figure III-3.3. Empirical CCS vs *m/z* for (**A**) $[M+H]^+$ and (**B**) $[M+Na]^+$.

CCS is a value related to the size, shape and charge of a molecule and understandably, CCS is also strongly correlated with the *m/z* value of a compound (Belova et al., 2021; Hernandez-Mesa et al., 2018; Mullin et al., 2020; Vera et al., 2019). The correlation between *m/z* and the CCS value of the compounds considered in this study is shown in Figure III-3.3. In general, the relationship between *m/z* and CCS can be described by a power regression model. The inclusion of more halogenated compounds in this study (a total of 302 and 149 halogenated molecules were included for $[M+H]^+$ and $[M+Na]^+$ adducts, respectively), highlighted a distinct difference in their *m/z* and CCS relationship when compared to the relationship for non-halogenated

compounds. The halogenated compounds tended to have smaller CCS values for a given *m/z*. It is believed that halogens have a lower atomic radius per atomic mass unit in comparison to other elements, such as C, H, O, N.

As CCS is related to the size, shape, and charge of gas-phase ions, it is understandable that CCS is highly correlated to *m/z* values. However, distinct correlations between CCS and *m/z* have been observed for the compounds that possess different structural characteristics (Belova et al., 2021; Hines et al., 2017). Belova et al. showed that the plasticizers, organophosphate flame retardants and per- and polyfluoroalkyl substances (PFAS) present different CCS versus *m/z* trendlines (Belova et al., 2021). PFAS contain carbon-fluorine bonds, and their CCS values are much lower in general than other compounds of similar *m/z* values. Hines et al. present that most 1440 CCS values of drugs and drug-like compounds fall within ±10% threshold, the compounds with *m/z* 300-350 possess CCS values ranging from 150 to 210 $Å^2$ (Hines et al., 2017). These studies proved that besides the mass of molecules, the molecular shape and compactness can also affect the CCS values. The different CCS values of molecules at a given *m/z* value indicate that CCS can provide partially orthogonal molecular information for features in targeted and untargeted screening analysis.

Some CCS values collated in this study, were measured using drift tube IMS (DTIMS) (Belova et al., 2021) and deviations between $^{TM}CCS_{N2}$ and $^{DT}CCS_{N2}$ have previously been observed (Hinnenkamp et al., 2018). Since, accurate CCS values are fundamental to obtain a reliable CCS prediction model, the $^{DT}CCS_{N2}$ values were compared to $^{TW}CCS_{N2}$ values available in the literature (Tables III-3.2 and III-3.3). 16 $^{TW}CCS_{N2}$ values were found in literature that could be directly compared to $^{DT}CCS_{N2}$ values, and most of these values were for compounds in the types of plasticizers and organophosphorus flame retardants. Table III-3.2 shows that 81.3% of the values agree to within 2% and the deviations ranged from 0.11% (for atrazine) to 2.88% (for tri-n-butyl phosphate) with an average of 1.15%. In the case of the $[M+Na]^+$ adduct, 75.0% of the values agree to within 2% and the deviations ranged from 0.15% (for di-n-butyl

phosphate) to 4.23% (for mono(2-ethylhexyl) adipate), with an average of 1.32%. The median of the $^{TM}CCS_{N2}$ and $^{DT}CCS_{N2}$ values was used when building the model to reduce any outlier measurements arising from the use of different IMS technologies.

Table III-3.2. Comparison between $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ for $[M+H]^+$ adducts.

No.	Name	CID	$^{DT}CCS_{N2}$	$^{TW}CCS_{N2}$	Differences (%)	Remarks
1	3,5-ditert-butyl-4-hydroxybenzaldehyde	73219	165.21	164.86	0.21	Degradation product of BHT
2	Atrazine	2256	149.53	149.70	0.11	Herbicide
3	Diazinon	3017	173.15	173.38	0.13	Insecticide
4	Benzotriazole	7220	122.42	123.49	0.87	UV absorbent
5	Di(2-ethylhexyl) phthalate	8343	211.00	213.33	1.10	Plasticizer
6	Diisononyl phthalate	590836	220.60	226.37	2.62	Plasticizer
7	Isodecyl diphenyl phosphate	34697	200.20	202.40	1.10	Plasticizer
8	Tributyl acetyl citrate	6505	199.82	195.99	1.92	Plasticizer
9	Antiblaze V6	92310	211.37	212.10	0.35	Flame retardant
10	Tri-n-butyl phosphate	31357	166.73	171.53	2.88	Flame retardant
11	Triphenyl phosphate	8289	174.74	170.00	2.71	Flame retardant
12	Tri-p-tolyl phosphate	6529	190.02	187.00	1.59	Flame retardant
13	Tris(1,3-dichloro-2-propyl) phosphate	26177	178.56	176.50	1.15	Flame retardant
14	Tris(2-butoxyethyl) phosphate	6540	196.44	198.30	0.95	Flame retardant
15	Tris(2-chloroethyl) phosphate	8295	151.31	150.40	0.60	Flame retardant
16	Tris(2-chloroisopropyl) phosphate	26176	161.66	161.87	0.13	Flame retardant

Note: $^{DT}CCS_{N2}$ values are from (Belova et al., 2021), $^{TW}CCS_{N2}$ values are from (Song et al., 2022c).

Table III-3.3. Comparison between $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ for $[M+Na]^+$ adducts.

No.	Name	CID	$^{DT}CCS_{N2}$	$^{TW}CCS_{N2}$	Differences (%)	Remarks
1	Di(2-ethylhexyl) adipate	7641	218.46	218.84	0.18	Plasticizer
2	Di(2-ethylhexyl) phthalate	8343	215.33	217.47	0.99	Plasticizer
3	Di(2-ethylhexyl) terephthalate	22932	215.81	217.69	0.87	Plasticizer
4	Dibutyl sebacate	7986	193.48	191.61	0.97	Plasticizer
5	Diisodecyl phthalate	33599	226.42	232.18	2.54	Plasticizer
6	Diisononyl phthalate	590836	220.94	225.45	2.04	Plasticizer
7	Dimethyl sebacate	7829	159.71	160.41	0.44	Plasticizer
8	Mono(2-ethylhexyl) adipate	20342	177.32	169.83	4.23	Plasticizer
9	Mono(2-ethylhexyl) phthalate	20393	182.27	181.25	0.56	Plasticizer
10	Tributyl acetyl citrate	6505	205.77	205.28	0.24	Plasticizer
11	Diphenyl phthalate	6778	181.27	178.28	1.65	Plasticizer

No.	Name	CID	$^{DT}CCS_{N2}$	$^{TW}CCS_{N2}$	Differences (%)	Remarks
12	2-Ethylhexyl diphenyl phosphate	14716	202.70	198.80	1.92	Flame retardant
13	Di-n-butyl phosphate	7881	167.54	167.79	0.15	Flame retardant
14	Tri-n-butyl phosphate	31357	184.54	184.16	0.21	Flame retardant
15	Tris(2-chloroethyl) phosphate	8295	161.39	157.89	2.17	Flame retardant
16	Tri-p-tolyl phosphate	6529	200.02	196.04	1.99	Flame retardant

Note: $^{DT}CCS_{N2}$ are from (Belova et al., 2021), $^{TW}CCS_{N2}$ are from (Song et al., 2022c).

4.2. Selection and weighting of molecular descriptors

The selection of molecular descriptors can reduce training time, simplify the prediction model and avoid overfitting, however, it is possible that meaningful information can be also lost leading to a decrease in accuracy. For this reason, it is necessary to achieve a balance between the simplicity and accuracy of model.

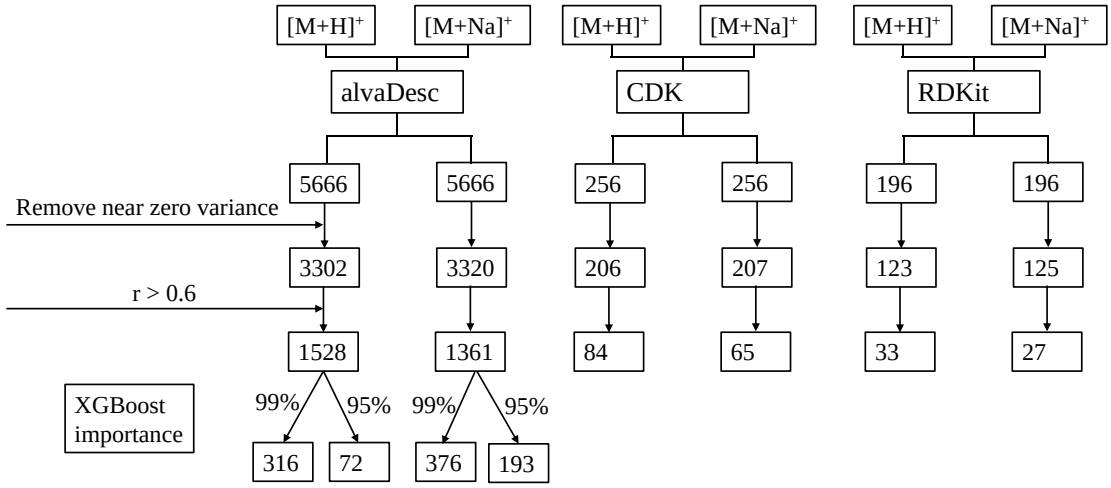


Figure III-3.4. The number of molecular descriptors retained after each step of variable selection.

The numbers of MDs retained after each step of variable selection are shown in Figure III-3.4, and the comparison of model performance before and after variable selection is presented in Tables III-3.4-3.6. For alvaDesc MDs, the first 316 and 72 descriptors accounted for 99% and 95% of total importance for $[M+H]^+$ adducts. When the number of MDs was decreased from 1528 to 72, both the SVM and XGBoost model showed a slight decrease in performance. The R^2_P of the SVM model decreased from

0.9802 to 0.9737, RMSEP increased from 4.47 to 5.43 and MRE increased from 1.50% to 1.52%. Considering the model was significantly simplified and the performance was still acceptable, the 72 most important alvaDesc MDs were selected for the $[M+H]^+$ adduct data. In the case of the $[M+Na]^+$ adduct CCS predictions, the models based on first 193 MDs showed comparable performance with the models built on 1361 MDs. Therefore the 193 most significant MDs were selected for $[M+Na]^+$ adduct data.

On determining the descriptors using CDK and RDKit, after the elimination of MDs that show low correlation with CCS ($r < 0.6$), 84 and 65 CDK descriptors, 33 and 27 RDKit descriptors were retained for $[M+H]^+$ adducts and $[M+Na]^+$ adducts, respectively; they were not filtered further. Table III-3.5 shows that 84 CDK MDs can provide accurate prediction results for $[M+H]^+$ adducts. A remarkable reduction in the performance of the model was observed for $[M+Na]^+$ adducts, when the number of MDs were reduced from 207 to 65. Therefore, 84 and 207 CDK MDs were selected for the $[M+H]^+$ and $[M+Na]^+$ adducts, respectively. In the case of RDKit, 33 and 125 MDs were retained for $[M+H]^+$ and $[M+Na]^+$ adducts based on the performance of the model (Table III-3.6).

Table III-3.4. Optimization of alvaDesc descriptors.

Adducts	Descriptor	Algorithm	R^2_P	RMSEP	<2%	<3%	<5%	MRE
$[M+H]^+$	alva_1528	SVM	0.9802	4.71	65.3	81.2	94.8	1.50
		XGBoost	0.9724	5.56	60.5	76.3	90.6	1.48
	alva_316	SVM	0.9779	5.00	62.3	79.3	93.6	1.54
		XGBoost	0.9734	5.46	63.2	79.6	92.4	1.46
	alva_72	SVM	0.9737	5.43	61.7	79.0	91.8	1.52
		XGBoost	0.9727	5.53	61.4	75.7	90.6	1.44
$[M+Na]^+$	alva_1361	SVM	0.9629	5.44	53.0	72.9	92.8	1.86
		XGBoost	0.9511	6.20	51.4	67.4	86.2	1.95
	alva_376	SVM	0.9589	5.70	50.3	70.2	91.7	1.94
		XGBoost	0.9611	5.57	54.1	71.8	90.1	1.75
	alva_193	SVM	0.9570	5.73	54.1	67.4	90.1	1.81
		XGBoost	0.9593	5.76	52.5	72.9	89.0	1.88

Table III-3.5. Optimization of CDK descriptors.

Adducts	Descriptor	Algorithm	R^2_P	RMSEP	<2%	<3%	<5%	MRE
$[M+H]^+$	CDK_206	SVM	0.9778	4.99	60.5	78.1	92.7	1.56
		XGBoost	0.9775	5.02	64.4	77.2	93.6	1.46
	CDK_84	SVM	0.9786	4.90	64.7	82.7	93.3	1.42
		XGBoost	0.9765	5.14	59.6	78.7	94.2	1.61
$[M+Na]^+$	**CDK_207**	SVM	0.9618	5.53	58.0	74.6	95.0	1.76
		XGBoost	0.9555	5.95	53.0	68.5	90.1	1.81
	CDK_65	SVM	0.9484	6.35	47.5	63.0	88.4	2.21
		XGBoost	0.9481	6.45	46.4	63.0	85.6	2.22

Table III-3.6. Optimization of RDKit descriptors.

Adducts	Descriptor	Algorithm	R^2_P	RMSEP	<2%	<3%	<5%	MRE
$[M+H]^+$	RDKit_123	SVM	0.9787	4.90	62.0	78.7	94.5	1.44
		XGBoost	0.9744	5.36	59.9	74.5	92.4	1.61
	RDKit_33	SVM	0.9772	5.09	63.8	79.6	93.0	1.46
		XGBoost	0.9700	5.80	58.1	74.2	90.3	1.58
$[M+Na]^+$	**RDKit_125**	SVM	0.9511	6.18	49.2	72.9	90.1	2.01
		XGBoost	0.9577	5.82	53.6	69.1	87.8	1.81
	RDKit_27	SVM	0.9389	6.96	43.1	61.9	85.1	2.35
		XGBoost	0.9564	5.89	51.4	66.3	89.0	1.97

4.3. Model performance

After dividing the collated CCS values into a training data set and a testing data set, 329 and 181 CCS values were included in the testing set for $[M+H]^+$ and $[M+Na]^+$ adducts, respectively. For each adduct, six CCS prediction models were developed based on the combinations of two algorithms (XGBoost and SVM) and three types of molecular descriptors (alvaDesc, CDK and RDKit). The distribution of prediction errors and model parameters for each model are shown in Figure III-3.5 and Table III-3.7, respectively. In the case of the $[M+H]^+$ adducts, more than 90% of molecules showed prediction errors within 5% for all six models. The SVM-based model in conjunction with the CDK descriptors provided the best predictive performance. R^2_P and MRE were

0.9786 and 1.42%, respectively, and more than 93% and 64% of molecules had prediction errors of less than 5% and 2%, respectively. This model also provided a better predictive performance for the $[M+Na]^+$ adduct with more than 95% and 58% molecules having prediction errors of less than 5% and 2%, respectively. The results also show that the model for $[M+H]^+$ adducts should use a different set of descriptors to those used for the model for $[M+Na]^+$ adducts, implying that a unique CCS prediction model should be developed for each adduct. This is highlighted by the studies of Bijlsma et al. (2017), in which a single set of descriptors was used for the CCS prediction of all positive ions and demonstrated that the CCS values predicted for $[M+Na]^+$ adducts were less accurate in general.

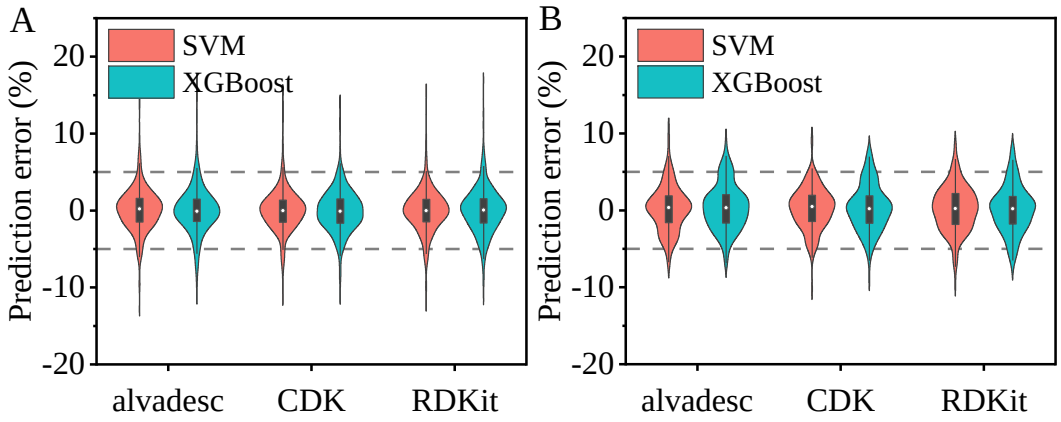


Figure III-3.5. Violin-plot illustrating the prediction errors of the SVM and XGBoost models using different sets of descriptors: (**A**) $[M+H]^+$; (**B**) $[M+Na]^+$.

Table III-3.7. Performance of the models developed using different descriptors and algorithms.

Adducts	Descriptor	Algorithm	R^2_P	RMSEP	<2%	<3%	<5%	MRE(%)
$[M+H]^+$	alvaDesc	SVM	0.9737	5.43	61.7	79.0	91.8	1.52
		XGBoost	0.9727	5.53	61.7	75.7	90.6	1.44
	CDK	SVM	**0.9786**	**4.90**	**64.7**	**82.7**	**93.3**	**1.42**
		XGBoost	0.9765	5.14	59.6	78.7	94.2	1.61
	RDKit	SVM	0.9772	5.09	63.8	79.6	93.0	1.46
		XGBoost	0.9700	5.80	58.1	74.2	90.3	1.58
$[M+Na]^+$	alvaDesc	SVM	0.9570	5.83	54.1	67.4	90.1	1.81

Adducts	Descriptor	Algorithm	R^2_P	RMSEP	<2%	<3%	<5%	MRE(%)
		XGBoost	0.9593	5.76	52.5	72.9	89.0	1.88
	CDK	SVM	**0.9618**	**5.53**	**58.0**	**74.6**	**95.0**	**1.76**
		XGBoost	0.9555	5.95	53.0	68.5	90.1	1.81
	RDKit	SVM	0.9511	6.18	49.2	72.9	90.1	2.01
		XGBoost	0.9577	5.82	53.6	69.1	87.8	1.81

In comparison to our previous study (Song et al., 2022c), a more accurate prediction of CCS values for $[M+Na]^+$ adducts is achieved here. The value of RMSEP decreased from 8.2 to 5.5 $Å^2$, the percentage of molecules with prediction errors less than 5% increased from 81.3% to 95.0% and those with prediction errors less than 2% increased from 54.7% to 58%. Even though there is a dramatic improvement, the prediction of CCS values for $[M+Na]^+$ adducts was still less accurate than that for $[M+H]^+$ adducts. It is believed that the main reason for this is that the molecular descriptors are calculated from neutral molecules and a sodium adduct can lead to a more diverse range of molecular conformations in 3D space compared to protonation (Bijlsma et al., 2017; Hernandez-Mesa et al., 2018; Righetti et al., 2018). One way to improve the accuracy of CCS predictions would be to determine the descriptors for ionized, rather than neutral molecules. However, such an approach is more complicated and computationally expensive, in addition, a conformational analysis is always required before the calculation of the descriptors (Boschmans et al., 2016).

The CCS prediction for halogenated molecules was also more accurate using the current SVM model compared to our previous study (Song et al., 2022c). 95.3% (81 out of 85 molecules) and 65.9% (56 out of 85 molecules) of protonated halogenated molecules had prediction errors of less than 5% and 2%, respectively. This compares to our previous study (Song et al., 2022c) for which the percentages were only 86.7% and 40%, respectively. This significant improvement could be due to the additional halogenated molecules in the training set, which supports previous observations that structure similarity between predictions and the training set significantly affect the accuracy of CCS predictions (Zhou et al., 2020). To further validate this conjecture, we

excluded the 217 halogenated molecules from the training set for $[M+H]^+$ adducts, leaving 530 non-halogenated molecules to rebuild SVM model for the prediction of CCS values for molecules in the testing set. A comparison of CCS prediction results, with and without halogenated compounds in the training set, is shown in Figure III-3.6. It is evident that upon excluding halogenated compounds from the training set, the prediction errors for the 244 non-halogenated compounds in the test data are similar to those generated when the halogens were included in the training data. However, the predicted CCS values of 85 halogenated compounds in the test data has significantly larger errors when the halogens were excluded from the training data: MRE increased from 1.46% to 1.87% and the proportion of halogenated compounds with prediction errors <2% decreased from 65.9% to 54.1%. This confirms that the chemical diversity of training set is an important factor which affects the prediction accuracy for the test data.

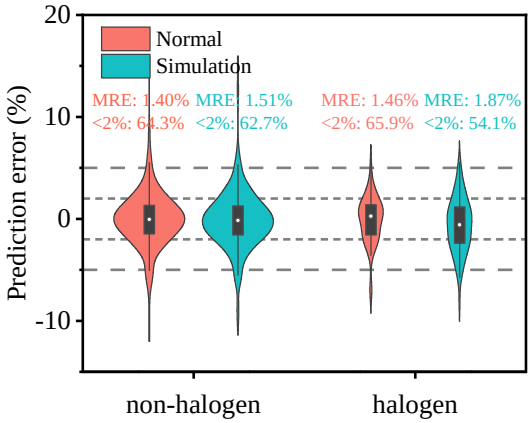


Figure III-3.6. Comparison of the CCS prediction accuracy including and excluding halogenated compounds from the training set for $[M+H]^+$ adducts. "Normal" represents the SVM model built with all 747 compounds in the training set; "Simulation" represents the SVM model built with only the 530 non-halogenated compounds in the training set; "non-halogen" and "halogen" represent 244 non-halogenated compounds and 85 halogenated compounds in testing set, respectively.

The protonated molecules for which the prediction error in the CCS value was

greater than 5% were further investigated. The presence of protomers can lead to high CCS prediction errors. Among the 22 protonated molecules for which the errors in the predicted values were more than 5%, there were some pesticide and drug compounds that had structures which exhibited multiple possible protonation sites. For example, two different CCS values (160.5 $Å^2$ and 176.2 $Å^2$) have been reported for acetopromazine in previous studies (Bijlsma et al., 2017; Celma et al., 2020), and protonation can occur on the carbonyl oxygen or the aminic group for this molecule, the predicted CCS value of 179.8 $Å^2$ matched well with the CCS value of the more extended protomer. The high prediction error (8.9%) for glycocholic acid (measured 187.6 $Å^2$/predicted 207.8 $Å^2$) may also be due to the presence of multiple protomers. A $^{TW}CCS_{N2}$ value of 205.2 $Å^2$ for protonated glycocholic acid, was obtained in Hines et al. (2017) which matched well with our predicted CCS value.

Similar behavior was also observed in the work of Zhou et al. (2016), in which the predicted CCS value (172.3 $Å^2$) for S-methyl-5'-thioadenosine matched well with CCS of the more extended protomer (170.9 $Å^2$) compared to the CCS of the more compact protomer (162.5 $Å^2$). One possible explanation for this phenomenon is that for the compact protomers, the protons are trapped in the core of the molecules and do not increase the overall size of the molecule. Thus, slightly lower experimental CCS values are obtained which can lead to relatively higher prediction errors. The less accurate CCS prediction for doramectin (measured 308.7 $Å^2$/predicted 284.7 $Å^2$), N,N'-Ethylenebis(stearamide) (measured 280.5 $Å^2$/predicted 296.6 $Å^2$), may be attributed to the low number of similar chemical structures in the training set.

Through the comparison of the six models and the comparison with our previous study (Song et al., 2022c), the SVM model based on CDK MDs provided the most accurate predictions. the chemical diversity of training set seems to be a more crucial factor for CCS prediction than descriptors and algorithm. The possibility of multiple protomers is another important factor affecting the accuracy since only one predicted CCS value can currently be determined for a given adduct by machine learning models.

Besides, we opted to use SVM due to its easy configuration with few hyperparameters, as well as its ability to provide reproducible prediction results.

4.4. Comparison between the SVM model and public CCS prediction tools

The outcomes from the SVM model based on CDK MDs were compared to those from three publicly available CCS prediction tools: CCSondemand, AllCCS and CCSbase. The distributions of the prediction errors for all models are illustrated in Figure III-3.7, and the corresponding MRE for each chemical class is shown in Figure III-3.8.

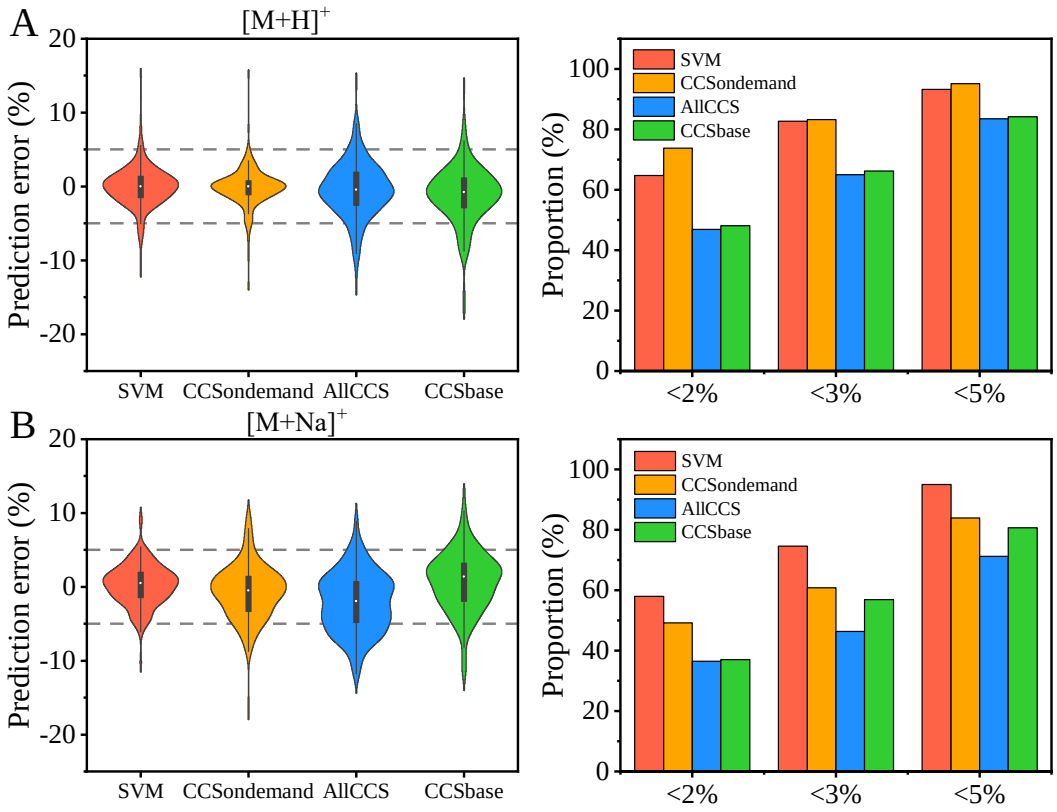


Figure III-3.7. Violin-plot and bar-plot showing the comparison of the CCS predictions of the SVM model to other CCS prediction tools: (A) $[M+H]^+$; (B) $[M+Na]^+$.

The CCS values of 65% and 74% of protonated molecules were predicted with an error of less than 2% by SVM and CCSondemand, respectively. More than 93% of protonated molecules has prediction errors less than 5% for both models.

CCSondemand was trained by approximately 7325 experimental $^{TW}CCS_{N2}$ values obtained from 3775 compounds (Broeckling et al., 2021). The training dataset contains CCS values of chemicals found in plastic food packaging and pesticides, so when the CCS values of such molecules are predicted by CCSondemand, one would expect smaller prediction errors. The predictive capabilities of AllCCS and CCSbase were not as good as those for SVM and CCSondemand for the compounds considered in this study. This is possibly due to the dissimilarity of the structures of chemicals in plastics and the molecules used in the training sets of AllCCS and CCSbase.

The results for $[M+Na]^+$ adducts showed that the SVM model gave more accurate predictions than the other tools. The enhanced performance of SVM is possibly due to the higher number of MDs used in this model (n = 207), as only 15 MDs were used in AllCCS (Zhou et al., 2020). The heat-map in Figure III-3.8 shows that SVM provides more accurate CCS predictions for the $[M+H]^+$ adduct of lipid and lipid-like molecules, organic nitrogen compounds and organic oxygen compounds, when compared against the other models. The prediction of CCS values by the SVM model for benzenoids, which are common class of molecules found in plastics, is also acceptable, with a MRE of 1.6%. The SVM also outperforms the other tools in the prediction of CCS values for the $[M+Na]^+$ adduct for benzenoids, organic acids and derivatives and organoheterocyclic compounds, all of which have an MRE of less than 2%. The relatively high prediction errors for $[M+Na]^+$ adducts of organic nitrogen compounds and organic oxygen compounds may be due to the low numbers of experimental CCS values available in these two super classes. Organic nitrogen compounds account for only 1.2% (8/645) of the CCS values available for $[M+Na]^+$ adducts and organic oxygen compounds for 3.6% (23/645). The same reason can be used to explain the high MRE values for $[M+H]^+$ adducts of organic nitrogen compounds and organosulfur compounds, which only account for 2.9% and 0.5% of the experimental CCS value available for $[M+H]^+$ adducts. This highlights the need for collecting more experimental CCS values from a range of compound classes.

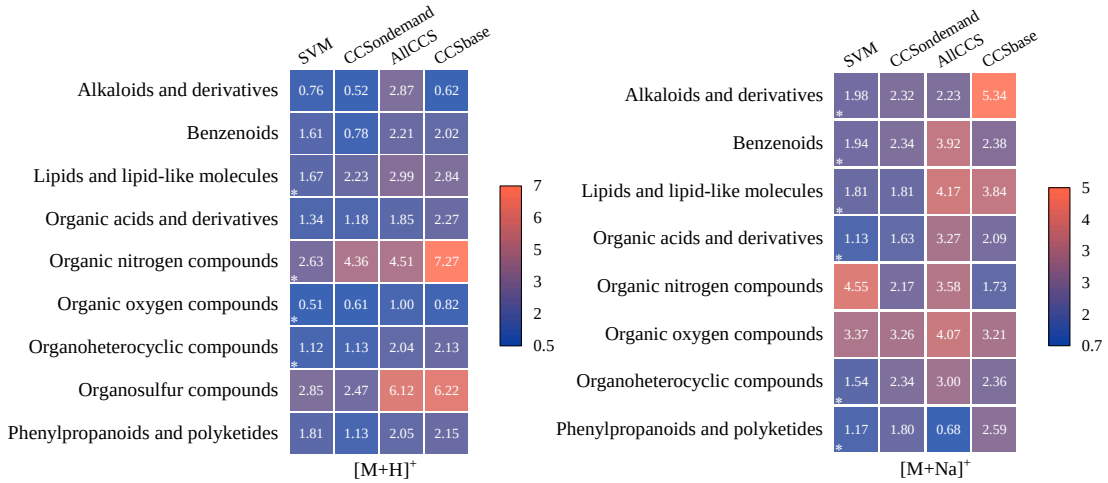


Figure III-3.8. Heat-map displaying median relative errors (MRE) of different chemical super classes obtained from the model presented here and other CCS prediction tools.

Table III-3.8. Comparison of the prediction results obtained from SVM models using 84 and 207 descriptors in our study, SVM models using 15 selected descriptors in AllCCS, and AllCCS.

Adducts	Models	R^2_P	RMSEP	<2%	<3%	<5%	MRE (%)
$[M+H]^+$	SVM_84	**0.9786**	**4.90**	**64.7**	**82.7**	**93.3**	**1.42**
	SVM_15	0.9763	5.17	61.4	76.3	92.7	1.57
	AllCCS	0.9609	6.92	47.0	64.9	83.5	2.16
$[M+Na]^+$	SVM_207	**0.9618**	**5.53**	**58.0**	**74.6**	**95.0**	**1.76**
	SVM_15	0.9430	6.68	49.2	64.6	82.9	2.10
	AllCCS	0.9185	9.21	36.5	46.4	71.3	3.29

The AllCCS tool is also based on the SVM algorithm and CDK MDs, however, there are two main differences between AllCCS and our model: the training data and the number of MDs. In order to investigate which factor leads to the significantly different prediction results between AllCCS and our model, we built a SVM model based on our CCS training data and the 15 MDs used for AllCCS, and compared their prediction results for the testing set to those obtained from our original model and AllCCS (Table III-3.8). For both $[M+H]^+$ and $[M+Na]^+$ adducts, less accurate prediction results were obtained from SVM models based on 15 MDs than with our original SVM model. MRE values increased from 1.4% to 1.6%, for $[M+H]^+$ adducts and 1.8% to 2.1%

for $[M+Na]^+$ adducts. The results from the SVM model based on 15 MDs and the AllCCS tool using the same MDs but different training data, can be seen in Table III-3.8. AllCCS shows significantly larger prediction errors, with MRE values of 2.2% for $[M+H]^+$ adducts and 3.3% for $[M+Na]^+$ adducts. These results show that the data used to train the model have a greater effect on the prediction accuracy of the model than the MDs.

These results show that, when compared to other available prediction tools, the SVM model based on the CDK MDs can improve the prediction of CCS values, especially for sodiated molecules. The CCS values for the $[M+H]^+$ and $[M+Na]^+$ adducts of the molecules in CPPdb and FCCdb were subsequently predicted by the SVM model developed here. The two databases were then transformed into screening libraries, which were used for the suspect screening of plastic-related chemicals in Ebro River water.

4.5. Plastic-related chemicals tentatively identified in Ebro River water

Approximately 95% of predicted CCS values (93.3% for $[M+H]^+$ adducts and 95.0% for $[M+Na]^+$ adducts) are within 5% deviation with respect to experimental values. Thus, the tolerance for CCS deviations was set as 5% in the suspect screening of plastic-related chemicals in Ebro River water. Two main aspects of using predicted CCS values in the identification of unknowns were investigated: reducing the number of false-positives and increasing the confidence level of identified compounds. The river water samples were screened against 9391 compounds in CPPdb and FCCdb to search for plastic-related chemicals. The number of candidates with and without the confirmation of CCS values was compared. The addition of the CCS filter decreased the number of candidate compounds from 376 to 204 (45.7%).

Table III-3.9. The plastic-related chemicals tentatively identified in Ebro River water

RT (min)	Measured CCS (Å^2)	Predicted CCS (Å^2)	m/z	adducts	Molecular formula	Mass error	Name	Use
9.42	288.28	283.54	685.4361	$[M+Na]^+$	$C_{42}H_{63}O_4P$	0.7	Tris(2,4-ditert-butylphenyl)phosphate	plasticizers, flame retardents, degradation products of Irgafos 168
9.87	251.06	249.38	537.3945	$[M+H]^+$	$C_{30}H_{58}O_4S$	-0.5	Dilauryl thiodipropionate	antioxidants
9.42	237.54	232.86	495.2663	$[M+Na]^+$	$C_{30}H_{39}O_4P$	0.9	Phenol, isobutylenated, phosphate (3:1)	flame retardants
8.16	218.05	214.09	393.2974	$[M+H]^+$	$C_{22}H_{42}O_4$	-0.3	Diisooctyl adipate	plasticizers
7.63	215.16	207.91	580.9145	$[M+Na]^+$	$C_{13}H_{24}Cl_6O_8P_2$	-0.9	Antiblaze V6	flame retardants
8.12	209.78	201.74	397.329	$[M+H]^+$	$C_{22}H_{46}O_4$	0.5	Ceteth-3	surfactants
8.13	215.95	218.46	413.2663	$[M+Na]^+$	$C_{24}H_{38}O_4$	0.5	Bis(2-ethylhexyl) phthalate	plasticizers
6.52	225.68	231.21	575.3763	$[M+Na]^+$	$C_{28}H_{56}O_{10}$	-0.5	Octaethylene Glycol Laurate	surfactants
7.41	209.62	212.91	409.2925	$[M+Na]^+$	$C_{22}H_{42}O_5$	0.1	Diethylene glycol dipelargonate	surfactants
6.2	201.57	199.35	437.1932	$[M+Na]^+$	$C_{24}H_{30}O_6$	-0.7	2,6-Bis(4-ethylphenyl)perhydro-1,3,5,7-tetraoxanaphth-4-ylethane-1,2-diol	nucleating agent
9.78	261.77	262.83	559.5172	$[M+Na]^+$	$C_{34}H_{68}N_2O_2$	-0.2	Hexadecanamide, N,N'-1,2-ethanediylbis-	lubricants
8.28	245.04	253.22	551.3921	$[M+H]^+$	$C_{30}H_{56}O_7$	0.6	Tris(2-ethylhexyl) 2-hydroxypropane-1,2,3-tricarboxylate	plasticizers
8.33	205.73	208.01	338.3422	$[M+Na]^+$	$C_{22}H_{43}NO$	1.5	Erucamide	slip agents
6.7	176.8	174.64	281.1722	$[M+Na]^+$	$C_{14}H_{26}O_4$	-0.6	Dibutyl adipate	plasticizers
9.42	193.9	192.24	383.1409	$[M+H]^+$	$C_{22}H_{23}O_4P$	0.5	Tert-Butylphenyl Diphenyl Phosphate	flame retardants
8.02	192.73	189.46	284.2947	$[M+Na]^+$	$C_{18}H_{37}NO$	-0.4	Octadecanamide	slip agents
8.83	263.39	256.47	769.1285	$[M+Na]^+$	$C_{37}H_{34}N_2Na_2O_9S_3$	-4.3	Brilliant Blue FCF	dyes
9.29	277.36	273.16	647.5582	$[M+H]^+$	$C_{39}H_{76}O_5$	-0.4	Distearin	antistatic agents
9.04	256.66	258.29	753.1286	$[M+H]^+$	$C_{36}H_{32}Cl_4N_6O_4$	-3.4	Pigment Yellow 81	dyes
8.21	271.09	271.97	825.2532	$[M+Na]^+$	$C_{45}H_{44}N_3NaO_7S_2$	2.4	Acid Blue 83	dyes
9.33	247.09	253.62	535.4336	$[M+Na]^+$	$C_{31}H_{60}O_5$	0.6	1,2-Dimyristoyl-rac-glycerol	antistatic agents

RT (min)	Measured CCS (Å^2)	Predicted CCS (Å^2)	m/z	adducts	Molecular formula	Mass error	Name	Use
9.1	236.2	233.50	497.2786	$[M+H]^+$	$C_{28}H_{43}O_4P$	-0.9	Phosphoric acid, hexadecyl diphenyl ester	flame retardants
9.1	232.44	229.68	475.2967	$[M+Na]^+$	$C_{28}H_{43}O_4P$	-0.9	Phosphoric acid, hexadecyl diphenyl ester	
8.61	207.24	206.33	340.3574	$[M+Na]^+$	$C_{22}H_{45}NO$	0.1	Docosanamide	slip agents
10.66	263.09	265.58	591.4954	$[M+Na]^+$	$C_{35}H_{68}O_5$	-0.8	Glyceryl 1,3-dipalmitate	antistatic agents
8.4	222.92	227.78	441.2984	$[M+H]^+$	$C_{26}H_{42}O_4$	1.9	Dinonyl phthalate	plasticizers
8.22	283.76	289.70	806.5096	$[M+H]^+$	$C_{48}H_{69}N_3O_6$	2.1	1,3,5-Tris(3,5-di-tert-butyl-4-hydroxybenzyl)-1,3,5-triazinane-2,4,6-trione	antioxidants
7.72	184.99	179.77	256.2636	$[M+Na]^+$	$C_{16}H_{33}NO$	0.4	Hexadecanamide	slip agents
8.76	276.77	274.74	701.5541	$[M+Na]^+$	$C_{42}H_{78}O_4S$	4.0	Dioleyl thiodipropionate	degradation of antioxidants
7.13	205.59	207.66	425.2152	$[M+Na]^+$	$C_{20}H_{34}O_8$	0.4	Acetyl tributyl citrate	plasticizers
7.53	217.6	228.80	575.1009	$[M+H]^+$	$C_{30}H_{24}O_8P_2$	-1.8	Resorcinol bis(diphenyl phosphate)	flame retardants
6.56	163.07	162.40	235.1691	$[M+Na]^+$	$C_{15}H_{22}O_2$	-0.5	3-Phenylpropyl Hexanoate	flavouring
7.59	178.69	182.35	302.2444	$[M+Na]^+$	$C_{18}H_{33}NO$	-3.6	Linoleamide	slip agents
7.55	250.39	250.87	693.1805	$[M+H]^+$	$C_{39}H_{34}O_8P_2$	0.5	Bisphenol A bis(diphenyl phosphate)	flame retardants
8.81	306.3	305.79	925.2544	$[M+H]^+$	$C_{50}H_{42}Cl_2N_6O_8$	3.3	C.I.Pigment red 221	dyes
7.16	189.34	188.34	320.2563	$[M+Na]^+$	$C_{18}H_{35}NO_2$	0.8	(R-(Z))-12-Hydroxy-9-octadecenamide	slip agents
8.3	200.18	198.11	312.3261	$[M+H]^+$	$C_{20}H_{41}NO$	0	Icosanamide	slip agents
8.71	209.65	203.63	379.1615	$[M+Na]^+$	$C_{20}H_{24}N_2O_4$	-3.5	2H-Azepin-2-one, 1,1'-(1,3-phenylenedicarbonyl)bis[hexahydro-	intermediates
6.9	220.47	221.52	577.1338	$[M+H]^+$	$C_{30}H_{24}O_{12}$	-0.4	Ethylene Terephthalate Cyclic Trimer	oligomers
8.41	234.13	240.67	493.3495	$[M+Na]^+$	$C_{27}H_{50}O_6$	-1	Tricaprylin	antistatic agents
10.34	247.87	244.75	553.459	$[M+H]^+$	$C_{35}H_{62}O_3$	-0.3	Irganox 1076	antioxidants
8.36	246.3	247.89	569.4381	$[M+Na]^+$	$C_{40}H_{56}O_2$	5	Lutein	personal care
8.46	286.34	274.33	701.4303	$[M+H]^+$	$C_{42}H_{63}O_5P$	-0.3	(4,6-Ditert-butyl-2,3-dihydroxyphenyl) bis(2,4-ditert-butylphenyl) phosphite	NIAS, degradation of Irgafos 168

RT (min)	Measured CCS (Å^2)	Predicted CCS (Å^2)	m/z	adducts	Molecular formula	Mass error	Name	Use
8.72	203.85	200.41	355.2846	$[M+Na]^+$	$C_{21}H_{38}O_4$	1.0	Methyl acetyl ricinoleate	plasticizers
8.19	258.3	251.31	587.5027	$[M+Na]^+$	$C_{36}H_{68}O_4$	2.9	Fatty acids, C18-unsatd., dimers	adhesives and sealant chemicals
5.15	185.56	185.99	399.1776	$[M+H]^+$	$C_{21}H_{28}O_6$	-0.6	Bisphenol A bis(2,3-dihydroxypropyl) ether	NIAS, degradation product of bisphenol A diglycidyl ether
8.19	273.52	266.29	815.0535	$[M+Na]^+$	$C_{40}H_{24}Cl_4N_6O_4$	3.6	Pigment Red 166	dyes
8.61	198.87	196.98	351.2568	$[M+H]^+$	$C_{18}H_{38}O_4S$	1.3	2-Hydroxy-1-octadecanesulfonic acid	
3.42	144.5	147.64	239.0888	$[M+Na]^+$	$C_{10}H_{16}O_5$	-0.9	1,4,7-trioxacyclotridecane-8,13-dione	NIAS, neoformed, reaction product from adipate plasticizer/adipate acid
7.24	187.88	187.32	341.2302	$[M+Na]^+$	$C_{17}H_{34}O_5$	1.0	Dodecanoic acid, 3-hydroxy-2,2-bis(hydroxymethyl)propyl ester	
3.08	176.51	177.93	393.2094	$[M+Na]^+$	$C_{16}H_{34}O_9$	-0.4	Octaethylene Glycol (PEG8)	oligomers
9.79	196.88	199.83	310.3106	$[M+H]^+$	$C_{20}H_{39}NO$	0.7	Cis-11-Eicosenamide	slip agents
7.7	176.2	169.70	267.1634	$[M+Na]^+$	$C_{12}H_{26}O_4S$	3.5	Dodecyl sulfate	adhesive, pesticides, surfactants
2.61	157.77	158.64	305.1568	$[M+Na]^+$	$C_{12}H_{26}O_7$	-1.0	hexaethylene glycol (PEG6)	oligomers
7.07	199.38	201.04	395.2771	$[M+Na]^+$	$C_{21}H_{40}O_5$	0.7	Glyceryl Monoricinoleate	surfactants
9	244.86	234.22	605.4233	$[M+H]^+$	$C_{30}H_{62}O_{10}$	-0.4	Polidocanol	surfactants
9.51	166.61	168.10	249.1849	$[M+H]^+$	$C_{16}H_{24}O_2$	-0.2	Octyl phenylacetate	flavouring
6.79	180.26	178.88	301.141	$[M+Na]^+$	$C_{16}H_{22}O_4$	-0.1	Diisobutyl phthalate	plasticizers
8.4	265.74	255.57	769.1324	$[M+Na]^+$	$C_{37}H_{34}N_2Na_2O_9S_3$	0.7	Light Green SF Yellowish	dyes
7.59	170.27	173.81	263.2368	$[M+Na]^+$	$C_{18}H_{30}O$	-0.6	Tributylphenol	
6.15	201.91	196.44	389.2198	$[M+H]^+$	$C_{23}H_{30}N_2O_2$	-0.3	2-Benzyl-2-dimethylamino-1-(4-morpholinophenyl)-1-	paint additives and

RT (min)	Measured CCS (Å^2)	Predicted CCS (Å^2)	m/z	adducts	Molecular formula	Mass error	Name	Use
							butanone	coating additives
9.11	199.54	193.04	363.1726	$[M+Na]^+$	$C_{20}H_{27}O_4P$	1.6	Octyl Diphenyl Phosphate	flame retardants
9.83	224.91	233.35	495.3809	$[M+Na]^+$	$C_{31}H_{52}O_3$	0.1	Alpha-tocopherol acetate	antioxidants
9.42	152.53	150.44	251.0468	$[M+H]^+$	$C_{12}H_{11}O_4P$	0.1	Diphenyl phosphate	paints and coatings
7.03	191.21	186.15	283.2627	$[M+Na]^+$	$C_{18}H_{34}O_2$	-1.5	Tetradecyl methacrylate	adhesives and sealant chemicals
8.16	182.71	181.52	297.137	$[M+H]^+$	$C_{19}H_{18}N_2$	2.8	1,4-Benzenediamine, N-(2-methylphenyl)-N'-phenyl-	
9.58	248.15	259.64	718.9976	$[M+Na]^+$	$C_{26}H_{17}ClN_7Na_3O_{10}S_3$	2.2	C.I. Reactive Red 24	dyes
7.79	188.22	190.61	282.279	$[M+H]^+$	$C_{18}H_{35}NO$	-0.4	Oleamide	slip agents
9.33	204.92	203.12	341.3053	$[M+H]^+$	$C_{21}H_{40}O_3$	0.8	Propylene Glycol 1-Oleate	antistatic agents
7.29	186.08	185.31	321.2395	$[M+Na]^+$	$C_{18}H_{34}O_3$	-1.5	Ricinoleic acid	lubricants
9.28	269.42	272.23	675.5178	$[M+H]^+$	$C_{39}H_{72}O_7$	1.2	Bis[(R)-12-hydroxyoleic] acid, diester with glycerol	antistatic agents
6.88	197.56	206.41	380.2102	$[M+H]^+$	$C_{24}H_{27}N_3$	1.4	Solvent violet 8	dyes
8.83	245.23	252.96	605.4256	$[M+H]^+$	$C_{32}H_{60}O_{10}$	-0.5	Polyoxyethylene sorbitan monooleate	antistatic agents
8.23	198.87	195.70	311.2945	$[M+Na]^+$	$C_{20}H_{38}O_2$	0.2	Hexadecyl methacrylate	lubricants, paints and coatings, adhesives and sealants
9.55	279	282.16	661.5369	$[M+Na]^+$	$C_{39}H_{74}O_6$	-1.3	Trilaurin	antistatic agents
7.39	240.58	231.26	595.3814	$[M+Na]^+$	$C_{31}H_{56}O_9$	-0.4	23-(Nonylphenoxy)-3,6,9,12,15,18,21-heptaoxatricosan-1-ol	surfactants
9.34	256.28	245.92	597.3946	$[M+H]^+$	$C_{40}H_{52}O_4$	1.3	Astaxanthin	dyes, colorant
3.05	146.82	145.01	215.1255	$[M+H]^+$	$C_9H_{20}O_4$	0.5	Tripropylene glycol	oligomers
9.56	190.55	196.15	337.2351	$[M+Na]^+$	$C_{18}H_{34}O_4$	0.4	3,5,5-Trimethylhexanoyl peroxide	initiator
6.94	199.78	202.68	421.2314	$[M+Na]^+$	$C_{18}H_{39}O_7P$	-2.8	Tris(2-Butoxyethyl) Phosphate	flame retardants
8.83	192.7	192.48	299.2951	$[M+H]^+$	$C_{19}H_{38}O_2$	2.3	Methyl stearate	lubricants

RT (min)	Measured CCS (Å^2)	Predicted CCS (Å^2)	m/z	adducts	Molecular formula	Mass error	Name	Use
8.24	208.51	205.12	395.3125	$[M+Na]^+$	$C_{22}H_{44}O_4$	-1.8	Diethylene glycol monostearate	surfactants
6.97	208.45	211.50	425.2873	$[M+Na]^+$	$C_{22}H_{42}O_6$	-0.3	Triethylene glycol dicaprylate	antistatic agents
5.37	124.33	128.95	163.0389	$[M+Na]^+$	$C_9H_6O_3$	-0.7	Umbelliferone	
7.5	250.07	251.29	714.964	$[M+H]^+$	$C_{24}H_{15}ClN_7Na_3O_{10}S_3$	2.3	Reactive Orange 13	dyes
6.46	177.73	185.23	359.2404	$[M+Na]^+$	$C_{17}H_{36}O_6$	-0.2	5,8,11,13,16,19-Hexaoxatricosane	surfactants
8.88	222.53	233.02	449.3612	$[M+Na]^+$	$C_{26}H_{50}O_4$	2.4	Glycol dilaurate	antistatic agents
2.41	164.79	159.01	335.1674	$[M+H]^+$	$C_{13}H_{28}O_8$	-0.6	1,3-Propanediol, 2,2-bis(hydroxymethyl)-, polymer with oxirane	
8.09	204.72	202.31	381.2973	$[M+Na]^+$	$C_{21}H_{42}O_4$	-0.6	Glyceryl monostearate	antistatic agents
9.48	268.71	272.20	587.5489	$[M+Na]^+$	$C_{38}H_{70}N_2O_2$	-3.6	N-(2-Oleamidoethyl)-linoleamide	
8.59	228.12	236.10	469.3302	$[M+Na]^+$	$C_{28}H_{46}O_4$	2.9	Didecyl phthalate	plasticizers
6.8	180.9	180.43	301.1409	$[M+Na]^+$	$C_{16}H_{22}O_4$	-0.4	Dibutyl phthalate	plasticizers
8.41	218.66	227.94	425.3608	$[M+Na]^+$	$C_{26}H_{48}O_4$	-4.1	Bis(7-methyloctyl) Cyclohexane-1,2-dicarboxylate	plasticizers
11.72	270.88	269.78	619.5266	$[M+H]^+$	$C_{37}H_{72}O_5$	-0.9	Glycerides, C16-18 mono-and di-	antistatic agents

A total of 98 plastic-related chemicals were tentatively identified in the Ebro River surface water samples from the CPPdb and FCCdb databases, of which 26 compounds were confirmed using reference standards. The tentatively identified compounds consisted of 12 plasticizers, 10 flame retardants, 6 antioxidants, 9 slip agents, 10 dyes and 26 surfactants (including glycol and glycerol derivatives). NIAS were also detected in Ebro River water, including ethylene terephthalate cyclic trimer, a common oligomer of polyethylene terephthalate (PET) (Ubeda et al., 2018), and bisphenol A bis(2,3-dihydroxypropyl) ether, a hydrolysis product of bisphenol A diglycidyl ether (Gallart-Ayala et al., 2011). Detailed information about the identified compounds is available in Table III-3.9.

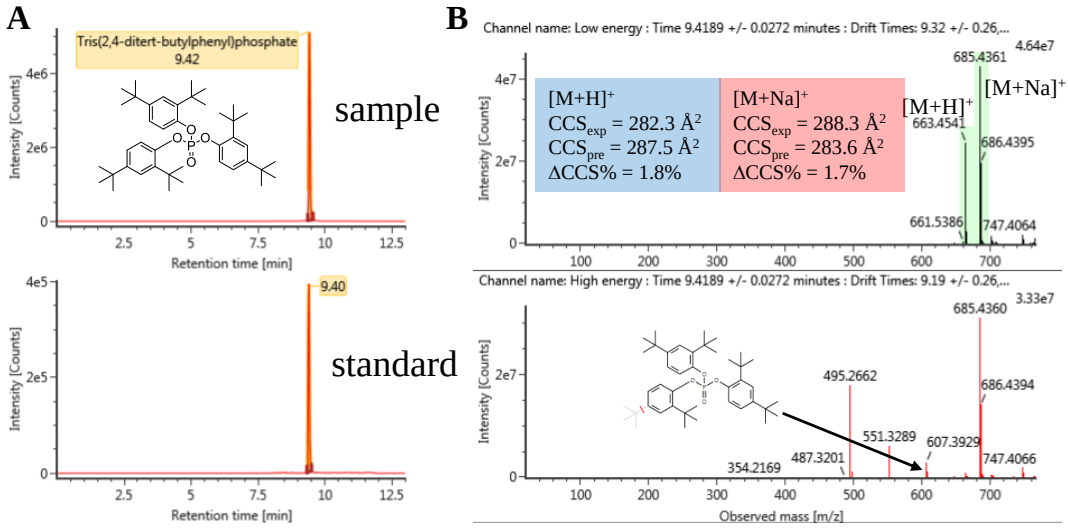


Figure III-3.9. Identification of tris(2,4-ditert-butylphenyl)phosphate. (A) extracted ion chromatograms from sample and standard, (B) low and high energy spectra, fragment assignment, comparison between experimental and predicted CCS values.

The most abundant compound detected in the Ebro River water samples was tris(2,4-ditert-butylphenyl)phosphate. This is a degradation product of Irgafos 168 (a commonly used phosphite antioxidant in plastics) (Yang et al., 2016). Previous studies have shown that tris(2,4-ditert-butylphenyl)phosphate was an abundant contaminant in indoor dust (Liu and Mabury, 2018b) and fine particulate matter (Shi et al., 2020). The predicted CCS values for tris(2,4-ditert-butylphenyl)phosphate had deviations less than

2% versus the experimental values (Figure III-3.9). For this compound, the predicted CCS values provided by AllCCS are 245.5 and 246.6 Å^2 for $[M+H]^+$ and $[M+Na]^+$ ions, respectively, which are significantly different with the experimental values. This highlights the necessity of developing a specialized CCS prediction tool for FCCs.

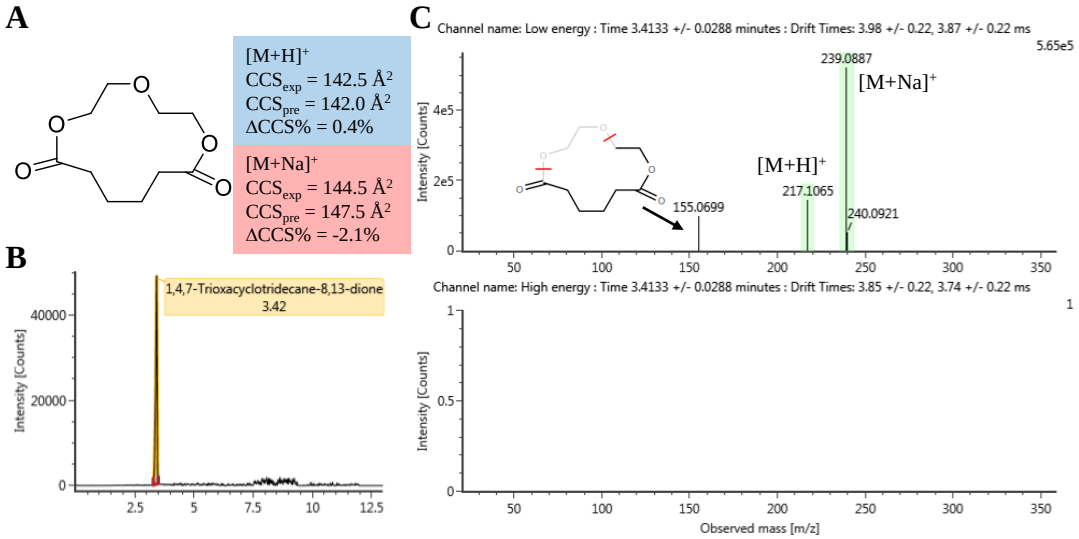


Figure III-3.10. Identification of 1,4,7-trioxacyclotridecane-8,13-dione. (A) molecular structure and predicted CCS values; (B) extracted ion chromatogram from sample; (C) low and high energy spectra, fragment assignment.

The benefit of predicted CCS values in identification of unknowns is more relevant either when the analyte is at low concentration levels or the reference standard is not available. 1,4,7-trioxacyclotridecane-8,13-dione is a reaction product from adipate plasticizer/adipate acid and ethylene glycol and its molecular structure and mass spectra are shown in Figure III-3.10. The figure shows the isotopic pattern for the $[M+H]^+$ adduct was indistinct and no fragment ions were observed in the high energy spectrum for the compound, possibly due to the low concentration and ineffectual assignment of fragment ions to the respective precursor ions. A fragment ion at *m/z* value 155.0699 was observed in low energy spectrum, which corresponds to the loss of ethylene glycol unit. CCS deviation for the $[M+H]^+$ and $[M+Na]^+$ adducts of 1,4,7-trioxacyclotridecane-8,13-dione were 0.4% and -2.1%, respectively. In this case, even

though no abundant fragmentation information was obtained for 1,4,7-trioxacyclotridecane-8,13-dione, the combination of RT, *m/z* and CCS contribute to a reliable identification.

In some cases, even when the analyte is at a high concentration in the sample, fragments ions may still not be assigned in the high-energy mass spectrum, as a result of rigid structures of less labile molecules. An example of this is given in Figure III-3.11 in which the mass spectra of Antiblaze V6, a flame retardant in plastics, are shown. The fragment ions observed in high energy spectrum are in low abundance and substructures of the parent molecule could not be assigned. The predicted CCS value (207.9 Å^2) had a 3.5% deviation when compared to the experimental value (215.2 Å^2). Additionally, an experimental CCS value (211.4 Å^2) for Antiblaze V6 was found in the literature (Belova et al., 2021), and has a deviation of -1.8% from our experimental value. It is not possible to confirm the identification of this compound due to the lack of a reference standard, however, the comparison between predicted and experimental CCS values, the *m/z* values and the characteristic chlorine isotopic pattern provide high confidence for the assignment.

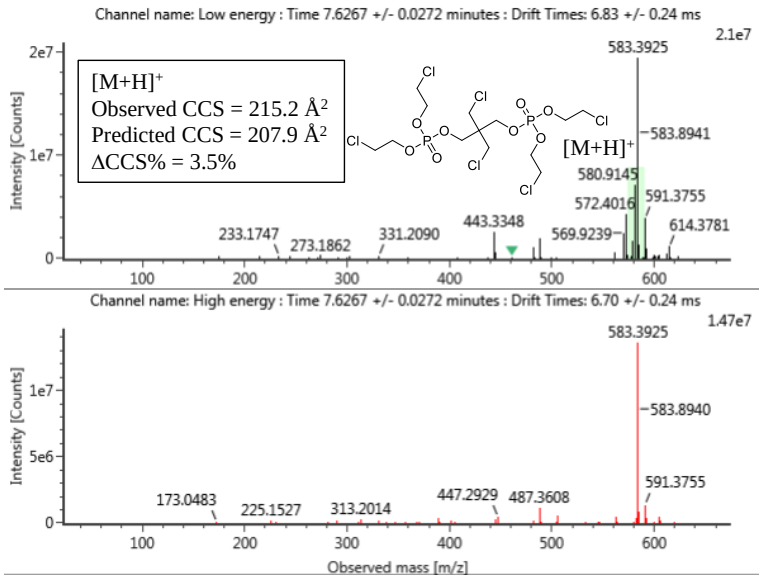


Figure III-3.11. Mass spectra and CCS values of Antiblaze V6.

False positive assignments have been observed for which the tolerances of *m/z*

error < 5 ppm and CCS deviation < 5% are satisfied. For example, the ion with *m/z* 327.0785 and CCS 169.9 $Å^2$ at a retention time of 9.42 min is a good match for triphenyl phosphate. However, the reference standard was detected with a retention time of 6.50 min, thereby showing this assignment to be a false positive. The addition of retention time predictions may be able to eliminate this kind of false positives, as shown by previous studies (Bijlsma et al., 2019; Bonini et al., 2020).

5. Discussions

5.1. Suitability of combining both $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ values in model

CCS values measured using DTIMS can differ from those measured using traveling wave IMS (TWIMS) platforms (Hinnenkamp et al., 2018). Therefore, the suitability of combining both $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ values in the CCS prediction model was investigated. 16 compounds have both $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ values for both $[M+H]^+$ and $[M+Na]^+$ adducts (Tables III-3.2 and III-3.3). Measurements of $^{DT}CCS_{N2}$ alone are present for 39 $[M+H]^+$ adducts and 65 $[M+Na]^+$ adducts. Of these, 27 $^{DT}CCS_{N2}$ values for $[M+H]^+$ adducts and 51 $^{DT}CCS_{N2}$ values for $[M+Na]^+$ adducts are present in the training data. The $^{DT}CCS_{N2}$ values were removed from the training data, the SVM models were rebuilt and their performance was compared to the original SVM models (Table III-3.10). The prediction accuracy of the CCS values for the $[M+H]^+$ adducts remained similar to the original results, however, the predicted CCS values for the $[M+Na]^+$ adducts were less accurate. RMSEP increased from 5.5 $Å^2$ to 5.8 $Å^2$, and the proportion of compounds with prediction errors was <2% decreased from 58.0% to 55.8%. The reduction in the prediction accuracy upon removing the $^{DT}CCS_{N2}$ values from the training data is probably due to the reduction in the diversity of chemical structures. The $^{DT}CCS_{N2}$ values were mainly for organophosphate flame retardants and phthalate monoesters, both of which are additives commonly used in plastics (Liu et al., 2019; Liu et al., 2021).

It should be noted that in this study, the difference between $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ values are relatively small. Higher CCS deviations were observed between $^{TW}CCS_{N2}$ values from different laboratories than between $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ values (Tables S2-S5). Based on these observations, we decided to use both $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ values in the training data of the model.

Table III-3.10. Comparison of SVM models before and after excluding 27 and 51 $^{DT}CCS_{N2}$ values from the training set of $[M+H]^+$ and $[M+Na]^+$.

Adducts	Training set	R^2_P	RMSEP	<2%	<3%	<5%	MRE (%)
$[M+H]^+$	747	**0.9786**	**4.90**	**64.7**	**82.7**	**93.3**	**1.42**
	720	0.9785	4.91	64.1	82.1	93.3	1.40
$[M+Na]^+$	464	**0.9617**	**5.53**	**58.0**	**74.6**	**95.0**	**1.76**
	413	0.9572	5.80	55.8	72.9	91.7	1.78

5.2. Weighting and collinearity of CDK MDs

The important CDK MDs for the prediction of the CCS values are shown in Figure III-3.12. The Atomic and Bond Contributions of Van der Waals volume (VABC) was found to be the most important CDK descriptor for the prediction of CCS values. CCS was strongly correlated with VABC, with the linear regression line having an R^2 value of 0.9533. VABC is a molecular volume property, which is estimated by the atomic contributions and the number of atoms, bonds, and rings (Zhao et al., 2003). Molecular connectivity chi indices: VP.0, VP.1, SP.1 and SP.2 were also found to be influential CDK descriptors, each chi index is a sum of weighted subgraphs in which the weights are functions of the molecular connectivity delta values (Devillers and Balaban, 1999). Molecular connectivity chi indices is one type of topological indices, which were previously used to predict the retention time of small molecules (Aćimović et al., 2020). Atom molar refractivity (AMR) is a constitutive-additive property, as described by Ghose and Crippen (Ghose and Crippen, 1987), it represents the volume of molecules (Padron et al., 2002). Atomic polarizability (apol) calculates the sum of the atomic

polarizabilities (including implicit hydrogens) and bpol calculates the sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule. AMR and apol were used to predict CCS values by Zhou et al. (2016). The Weiner path number (WPATH) is a topological descriptor, which is calculated as the sum of the lengths of the shortest paths between all pairs of vertices in the chemical graphs (Rouvray and King, 2002). Eccentric connectivity index (ECCEN) is a topological descriptor, it considers the eccentricity and valency of each vertex involved in a molecular graph. Its calculation can be performed from the distance matrix of a hydrogen-suppressed molecular graph after the vertices have been numbered arbitrarily (Sharma et al., 1997). The first kappa shape index (Kier1) is a also a topological property, it characterize the molecular shape (Hall and Kier, 2007), kappa index was used for predicting the CCS values in AllCCS (Zhou et al., 2020). The nAtom and MW are constitutional descriptors, which measure the total number of atoms and molecular weight of molecules. The analysis of MDs weights substantiates the correlation of the CCS of a molecule to the size and shape of that ionized molecule.

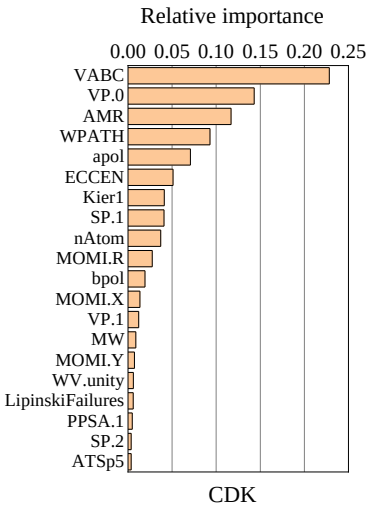


Figure III-3.12. Relative importance of the 20 most influential CDK descriptors for the prediction of CCS vlaues of $[M+H]^+$ adducts in XGBoost model.

The effect of collinearity between CDK MDs was investigated by building models that omitted highly correlated MDs. The variance inflation factor (VIF) is a measure of the correlation between MDs with higher values indicating greater correlation. Two

models were built to study collinearity of MDs, one for which MDs with a VIF value greater than 50 were excluded and one for which MDs with a VIF value greater than 20 were excluded. A comparison of the predictive performance between the two new models and the original model is shown in Figure III-3.13. The figure shows that for $[M+H]^+$ adducts, 33 MDs with a VIF value below 50 were retained and 24 MDs with VIF value below 20 were retained. The reduction in the number of the MDs slightly decreased the prediction accuracy of both the SVM and XGBoost models. Similar behavior was observed when the same procedure was applied to the models for the $[M+Na]^+$ adduct. Since the models built with 84 and 207 CDK MDs for the $[M+H]^+$ and $[M+Na]^+$ adducts respectively, provide more accurate predictions, and the complexity of the model was still deemed to be acceptable, the number of MDs were not reduced in the final models.

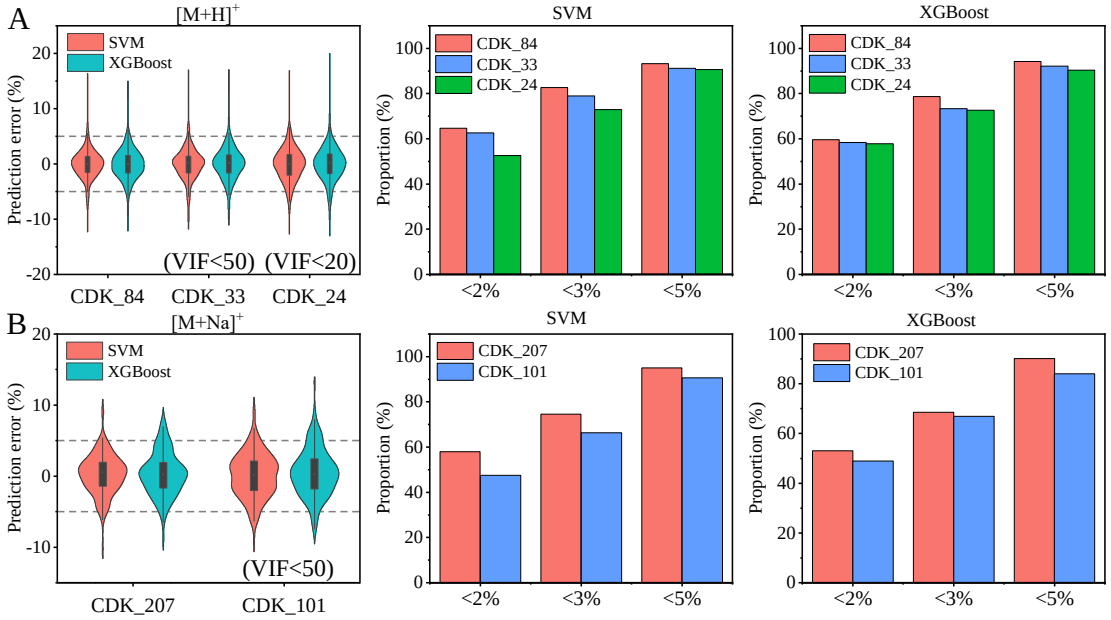


Figure III-3.13. Comparison of CCS prediction accuracy between before and after excluding highly correlated descriptors, (A) $[M+H]^+$, CDK_84, CDK_33 and CDK_24 represent models based on 84, 33 (VIF<50) and 24 (VIF<20) CDK descriptors; (B) $[M+Na]^+$, CDK_207 and CDK_101 represent models based on 207 and 101 (VIF<50) CDK descriptors.

5.3. Approaches to improve the prediction accuracy.

There are several ways in which the accuracy of CCS predictions could

potentially be improved. Firstly, more experimental CCS values can be collected for the training set to increase the chemical diversity and universality of the model. A total of 17 and 15 chemical super classes are considered in CCSondemand and AllCCS respectively, while in this study only 10 and 11 super classes are covered by 1076 and 645 CCS values for $[M+H]^+$ and $[M+Na]^+$ adducts, respectively. Figure III-3.14 compares the chemical space of FCCdb, CPPdb and our collected CCS records. As can be seen, a large proportion of molecules in CPPdb and FCCdb was covered by the chemical space of our collected CCS records. However, there are still many molecules which are out of this chemical space, we carefully examined these kinds of molecules in order to find their structural characteristics. Table III-3.11 presents the 50 compounds in CPPdb and FCCdb that were not covered by the chemical space of our collected CCS records. Generally, many of these molecules have relatively high molecular weights (MW), their MW range from 650 to 1200 Da. However, most compounds in our CCS dataset have MW ranging from 150 to 600 Da (see Figure III-3.3).

In addition to their high MW, these compounds appear to have linear-chain molecular structures, such as polyoxyethylene (23) lauryl ether and cetyl poly(oxyethylene) ether, these compounds are alkyl PEG ethers, which are normally used as non-ionic surfactants in plastics (Cowan-Ellsberry et al., 2014). Alkyl glycerol esters, such as glyceryl tribehenate, glyceryl tri(12-acetoxystearate), glycerol trimyristate and glycerol dibehenate, were also included, these compounds are normally used as antistatic agents (Hu et al., 2021). Other compounds include emulsifier in plastics, such as sorbitan trioleate, sorbitan dioleate and sorbitan trihexadecanoate; plasticizers: tristearyl citrate, distearyl citrate, dioctyldodecyl adipate; antioxidant: distearyl thiodipropionate, trioctadecyl phosphite; organophosphate flame retardant: tristearyl phosphate. These compounds contain long alkyl chains in their structures and can undergo more collisions with drift gas when passing through the drift cell. The comparison of the chemical space between our collected CCS dataset and CPPdb, FCCdb highlights that more compounds with high molecular mass and linear-chain structure should be incorporated into CCS dataset.

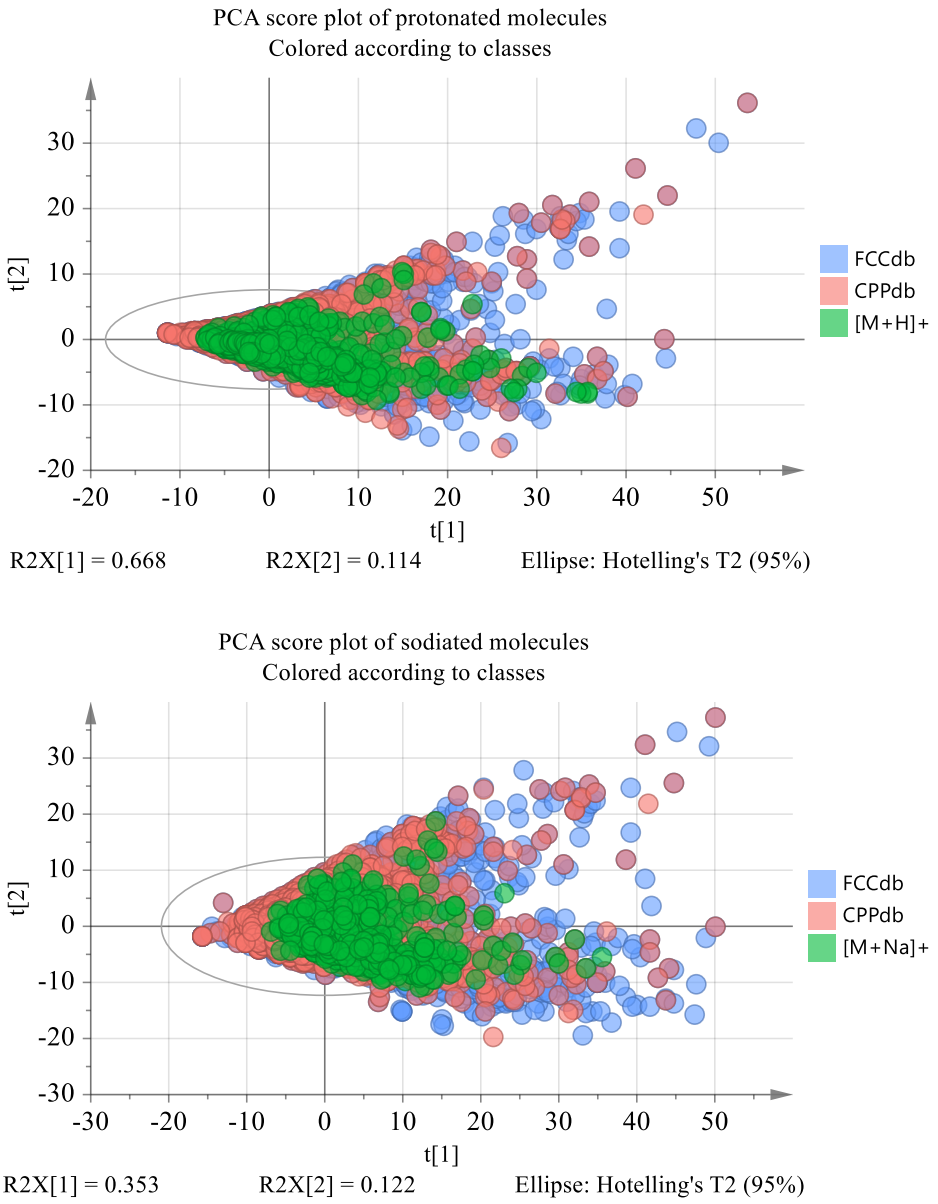


Figure III-3.14. Comparison of the chemical space of FCCdb, CPPdb and our collected CCS records.

Table III-3.11. The 50 molecules in CPPdb and FCCdb that are not covered by the chemical space of the collected CCS records.

Name	PubChem CID	Monoisotopic Mass	InChIKey
Polyoxyethylene (23) lauryl ether	2724258	1198.8013	IEQAICDLOKRSRL-UHFFFAOYSA-N
Pentaerythritol tetraoleate	6436503	1193.0548	QTIMEBJTEBWHOB-PMDAXIHYSA-N
Cetyl poly(oxyethylene) ether	2724259	1122.7853	NLMKTBGFQGKQEV-UHFFFAOYSA-N
2,2-bis[[3-(dodecylthio)-1-	122423	1160.8179	VSVVZZQIUJXYQA-UHFFFAOYSA-N

Name	PubChem CID	Monoisotopic Mass	InChIKey
oxopropoxy]methyl]propane-1,3-diyl bis[3-(dodecylthio)propionate]			
Glyceryl tribehenate	62726	1059.0180	DMBUODUULYCPAK-UHFFFAOYSA-N
Homopolymer of glyceryl triester with 12-glycidyl-9-octadecenoic acid	58604493	1100.8467	ZFJYZDDXGKWNCH-UHFFFAOYSA-N
Glyceryl tri(12-acetoxystearate)	6451270	1064.8467	FNOXLRARSOMOQK-UHFFFAOYSA-N
Tris(dinonylphenyl) phosphite	74003	1066.9210	WRSPWQHUHVRNFV-UHFFFAOYSA-N
Pentaerythritol tetrakis(3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionate)	64819	1176.7841	BGYHLZZASRKEJE-UHFFFAOYSA-N
Starch, hydrogen phosphate, 2-hydroxypropyl ether	24847848	1198.4140	DVROLKBAWTYHHD-UHFFFAOYSA-N
Octadecanoic acid, 12-hydroxy-, polymer with alpha-hydro-omega-hydroxypoly(oxy-1,2-ethanediyl)	121596032	948.6233	GJHBWXDMLVECGP-UHFFFAOYSA-N
Sorbitol trioleate	129772621	1022.7997	IEFFZOKIZPKMOG-RJUOWQTLSA-N
2-Heptadecyl-4,4'-bis(methylene stearate)-1,3-oxazoline	95104	901.8462	BPUYDDKNZNJELI-UHFFFAOYSA-N
Sorbitan tristearate	15181202	962.8514	IJCWFDPJFXGQBN-RYNSOKOISA-N
Tristearyl citrate	24493	948.8721	UKBHVNMEMHTWQO-UHFFFAOYSA-N
Castor oil, hydrogenated	25100	938.8150	WCOXQTXVACYMLM-UHFFFAOYSA-N
Pentaerythritol tristearate	14252002	934.8565	FWCDLNRNBHJDQB-UHFFFAOYSA-N
Sorbitan triisostearate	171343	962.8514	QWSHIYVIOOXKLL-LLPUSWRMSA-N
Sorbitan trioleate	9920343	956.8044	PRXRUNOAOLTIEF-ADSICKODSA-N
Trimethylolpropane trioleate	6436686	926.8302	BTGGRPUPMPLZNT-PGEUSFDPSA-N
Distarch glycerol	24832114	1176.4895	IQZVGYOIHLNAKB-UHFFFAOYSA-N
dioctyldodecyl adipate	3020369	706.6839	WLFITRMCTPBSQS-UHFFFAOYSA-N
Sorbitan dioleate	22833309	692.5591	TTZKGYULRVDFJJ-GIVMLJSASA-N
Distearyl citrate	11643318	696.5904	PGGUBOZIFQYBOV-UHFFFAOYSA-N
Glycerol trimyristate	11148	722.6424	DUXYWXYOBMKGIN-UHFFFAOYSA-N
Glycerol dibehenate	9831860	736.6945	GNWCZBXSKIIURR-UHFFFAOYSA-N
Tetrakis(2,4-di-tert-butyl-5-methylphenyl) [1,1'-biphenyl]-2,3-diylbis(phosphonite)	22672285	1090.7097	XMKVUPOWKZQBDQ-UHFFFAOYSA-N
Hexanoic acid, 3,5,5-trimethyl-, 1,1-[oxybis[2,2-bis[[(3,5,5-trimethyl-1-oxohexyl)oxy]methyl]-3,1-propanediyl]] ester	90684455	1094.8572	GJIDQGCYINNRBJ-UHFFFAOYSA-N
Hydroxylated lecithin	57508518	821.6146	XSEOYPMPHHCUBN-FGYWBSQSSA-N
alpha-D-Glucopyranoside, beta-D-fructofuranosyl, dioctadecanoate	5360827	874.6381	MZNXRHOLDWQYRX-CBKJUIDTSA-N

Name	PubChem CID	Monoisotopic Mass	InChIKey
diester of 3-(laurylthio)propionic acid with 4,4'-[thiobis(2-tert-butyl-5-methylphenol)]	105368	870.5688	MILWQXYAFKWZBP-UHFFFAOYSA-N
Glycerol tripalmitate	11147	806.7363	PVNIQBQSYATKKL-UHFFFAOYSA-N
1-Octadecanaminium, N-ethyl-N,N-dioctadecyl-, ethyl sulfate	106121	927.9016	MGUHJFIMPPINPG-UHFFFAOYSA-M
N,N'-[ethylenebis(iminoethylene)]bisbehenamide	44151217	790.8003	SFBNWBOCWQUMEJ-UHFFFAOYSA-N
Glycerol trioleate	5497163	884.7833	PHYFQTYBJUILEZ-IUPFWZBJSA-N
Tristearyl phosphate	74962	854.8220	FDGZUBKNYGBWHI-UHFFFAOYSA-N
1-(Hexadecanoyloxy)-3-(octadecanoyloxy)propan-2-yl octadec-9-enoate	53422263	860.7833	QXPXMOHHFYONAC-UHFFFAOYSA-N
Sorbitan, trihexadecanoate	171319	878.7575	NVANJYGRGNEULT-BDZGGURLSA-N
Trioctadecyl phosphite	248442	838.8271	CNUJLMSKURPSHE-UHFFFAOYSA-N
Pyrrolo[3,4-c]pyrrole-1,4-dione, 2,5-dihydro-3,6-bis[4-(octadecylthio)phenyl]-1	135565960	856.5974	OECIMFUOKDGJQO-UHFFFAOYSA-N
Fatty acids, C18-unsatd., trimers	6437702	800.4077	CFQZKFWQLAHGSL-FNTYJUCDSA-N
Sucrose octabenzoate	25113553	1174.3259	AKIVKIDZMLQJCH-KWOGCLBWSA-N
beta-Cyclodextrin	444041	1134.3698	WHGYBXFWUBPSRW-FOUAGVGXSA-N
2,2-Bis(((2-cyano-3,3-diphenylacryloyl)oxy)methyl)propane-1,3-diyl bis(2-cyano-3,3-diphenylacrylate)	16134382	1060.3472	CVSXFBFIOUYODT-UHFFFAOYSA-N
Distearyl thiodipropionate	12738	682.5934	PWWSSIYVTQUJQQ-UHFFFAOYSA-N
Glycerol tristearate	11146	890.8302	DCXXMTOCNZCJGO-UHFFFAOYSA-N
Triisodecyl tridecyl trimellitic ester	3085422	672.5329	YNKHAYUWCVQHBA-UHFFFAOYSA-N
1,2,3-Propanetriol, homopolymer, (Z)-9-octadecenoate	9963243	1022.6237	NPTLAYTZMHJJDP-KTKRTIGZSA-N
Distearyl adipate	70706	650.6213	GYFBKUFUJKHFLZ-UHFFFAOYSA-N
Pentaerythritol decanoate, Decanoic acid, 2,2-bis[[(1-oxodecyl)oxy]methyl]-1,3-propanediylester	83733	752.6166	MXNODNKXIIQMMI-UHFFFAOYSA-N

Secondly, MDs based on ionized molecules could improve predictions, especially for $[M+Na]^+$ adducts. There is a much bigger difference between the structural conformation of sodiated and neutral molecules than there is between protonated and

neutral molecules (Righetti et al., 2018). This makes it difficult to obtain accurate predicted CCS values for sodiated molecules when the descriptors are derived from neutral molecules. Taking into account that deriving MDs from ionized molecules is time-consuming and complex, and as such, MDs derived from neutral molecules are probably sufficiently accurate for $[M+H]^+$ adducts. In the study by Gonzales et al. (2016), MDs of deprotonated phenolics were determined for a CCS prediction model and 92.8% (52/56) of molecules were predicted within 5% of their measured values. In the present study a similar proportion (93.3%) of protonated molecules were predicted with an error less than 5%, highlighting that MDs determined from neutral molecules are sufficient for the accurate prediction of CCS values of protonated molecules.

Thirdly, improving the reproducibility of commercially available IMS devices such as TWIMS and DTIMS will lead to more precise and accurate CCS measurements, which, when used as inputs to prediction models, will improve the performance of the models. At the time of writing, commercial IMS devices have relatively low reproducibility, which makes it impractical to adopt an accuracy threshold lower a than 2% when matching measured CCS values to library values within a suspect screening workflow (Hernandez-Mesa et al., 2020; Hernandez-Mesa et al., 2018; Righetti et al., 2020).

5.4. Current limitations and future prospects

In this study, the CCS prediction models were only built for positive ions. This is understandable to some extent, as most additives in plastic products, such as plasticizers, antioxidant, flame retardants, photoinitiators, slip agents are detected in positive ion mode (Nerin et al., 2013). In some cases, compounds can be only detected, or show a higher response, in negative ion mode. Such compounds include lubricants (lauric acid, oleic acid) and surfactants (perfluorooctanesulfonic acid, perfluorobutanesulfonic acid), which were detected in the Ebro River water samples using our in-house plastic additives library. Therefore, a CCS prediction model for negative ions needs to be developed in the future. Additionally, the CCS prediction models developed herein are

only available to a small set of privileged users and work needs to be undertaken to develop them into an open-access tool.

Many emerging contaminants associated with plastics, such as tricaprin and polyethylene glycol (PEG) and polypropylene glycol (PPG) oligomers do not exist in the CPPdb and FCCdb databases. These compounds were detected at high abundance in the Ebro River water samples using our in-house plastic additives library. With the rapid growth of newly reported plastic-related chemicals, the CPPdb and FCCdb databases need to be continuously expanded and updated. The construction of an integrated plastic-related database containing name, adducts, *m/z* values, predicted CCS values and predicted retention times will facilitate the identification of extractables and leachables from plastics in HRMS-based screening strategies.

6. Conclusion

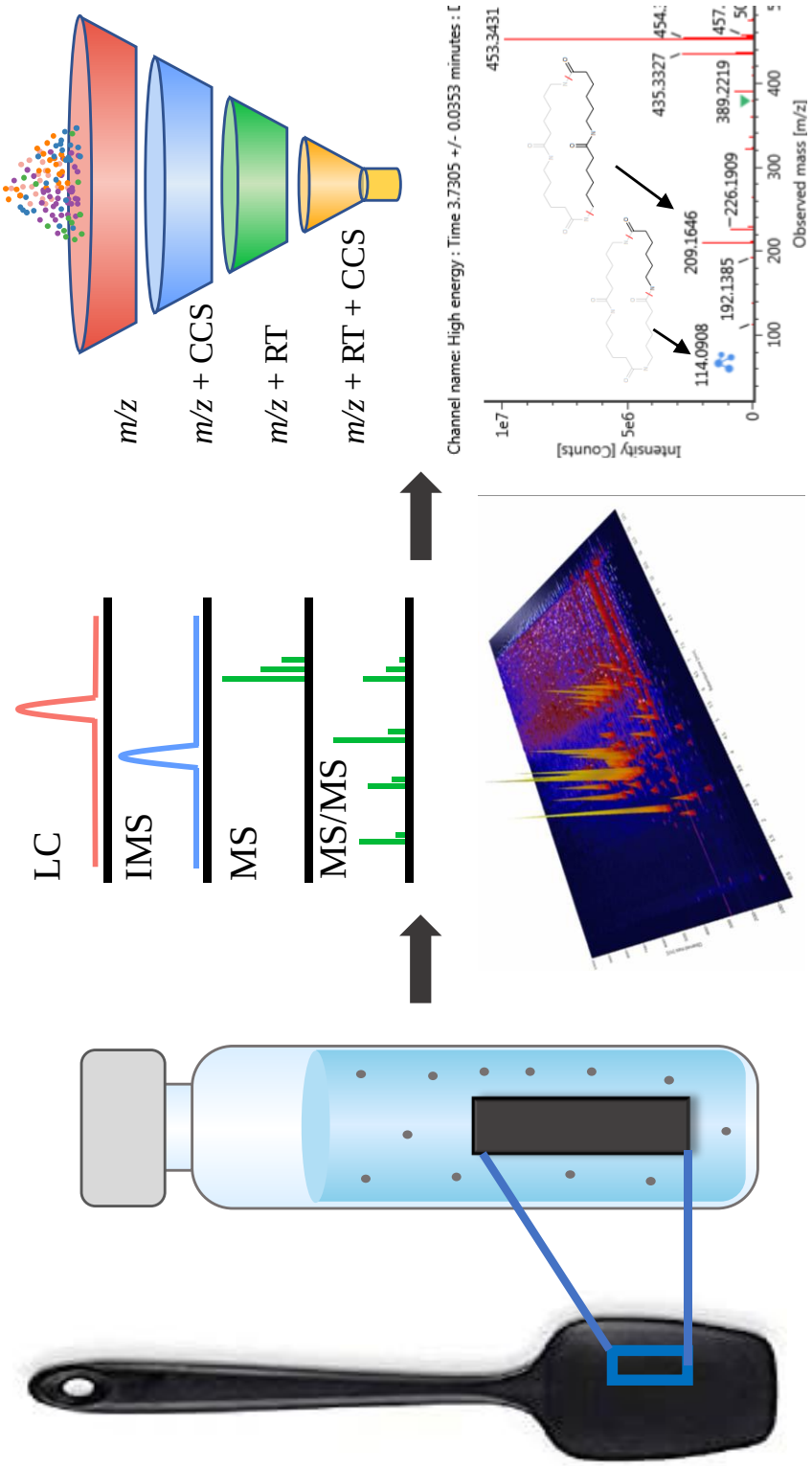
In summary, the SVM model based on CDK descriptors presented here, provided more accurate CCS predictions than the XGBoost algorithm and other descriptors. The CCS values of 93.3% $[M+H]^+$ adducts and 95.0% $[M+Na]^+$ adducts were predicted within 5% of their measured values. It has been shown that the chemical diversity of training set appears to have more influence on the predictive performance than alternative algorithms and MDs investigated here. Indeed, CCS predictions for halogenated compounds were more accurate following the incorporation of more CCS records of halogenated compounds in the training set. Increasing the number of experimental CCS values and improving the reproducibility of CCS measurements seem to be two feasible ways to further increase the performance of prediction models. In future work, a CCS prediction model for negative ions will be developed and work towards making all models open-access will be undertaken.

Chapter 4

A Workflow for the Identification of Non-Volatile Migrates from Food Contact Materials Based on Ion Mobility−High Resolution Mass Spectrometry and In-Silico Prediction Tools

1. Abstract

The identification of compounds migrating from food contact materials (FCMs) is challenging due to the complex matrices and unavailability of commercial standards. The use of machine learning-based prediction tools could be of great help for this purpose. This study mainly established a workflow for identifying the migrates from FCMs based on liquid chromatography-ion mobility-high resolution mass spectrometry and in-silico tools for the prediction of retention time (RT) and collision cross section (CCS). The applicability of this workflow was evaluated by the screening the chemicals migrated from polyamide (PA) spatulas. Approximately 75% and 29% of candidates was reduced by applying the predicted RT and CCS filters, respectively. A total of 95 compounds were identified in PA spatulas, of which 54 compounds were confirmed by reference standards. The building of FCMs-related database with predicted RT and CCS values have shown to be a promising approach to aid the identification of chemicals in FCMs.


m/z
m/z + CCS
m/z + RT
m/z + RT + CCS
LC
IMS
MS
MS/MS
Channel name: High energy : Time 3.7305 +/- 0.0353 minutes : [
Intensity [Counts]
Observed mass [m/z]
453.3431
435.3327
389.2219
209.1646
192.1385
114.0908
-226.1909

2. Introduction

Foods and beverages are in contact with food contact materials (FCMs) during the production, transport, storage and serving. The chemicals in FCMs, termed food contact chemicals (FCCs), can migrate into the contacted food and pose health concerns for consumers (Muncke et al., 2020). It was estimated that the food contamination with organic chemicals from FCMs may be 100 times higher than that from pesticides and other environmental pollutions (Grob et al., 2006). Plastic polymers are the predominant materials among other types of FCMs, such as paper and board, glass, metal and biomaterials, multilayers, etc. (Nerin et al., 2022). In the production of plastics, different types of additives are deliberately added into the polymers in order to enhance the desirable performances of the final product. The most commonly used additives in plastic packaging are plasticizers, antioxidants, stabilizers, flame retardants, slip agents and lubricants. These additives are not chemically bound to the polymers and can easily migrate into food (Hahladakis et al., 2018).

Non-intentionally added substances (NIAS) are also included in FCMs as chemical contaminants, generally coming from degradation products of additives or polymer itself, side products, reaction products and contaminants. The degradation products of additives have been widely studied by different groups (Liu and Mabury, 2018c; Reingruber and Buchberger, 2010; Shi et al., 2020; Yang et al., 2016). Oligomers originated from the incomplete polymerization or degradation of polymer chains (Hoppe et al., 2016). which are commonly present in any type of polymer. The oligomers from polylactic acid (PLA) (Aznar et al., 2019), polyethylene terephthalate (PET) (Alberto Lopes et al., 2021; Ubeda et al., 2018), and polyurethane (PU) adhesive (Ubeda et al., 2020) have been studied.

The complexity of plastic matrices and incomplete information about the ingredients make it challenging to achieve a full identification of FCCs coming from FCMs. Currently, high resolution mass spectrometry (HRMS), such as time-of-flight (TOF), has been commonly used for the identification of non-volatile FCCs (Canellas

et al., 2015b; Dreolin et al., 2019; Osorio et al., 2020; Vera et al., 2018), where the elemental composition of features can be inferred from the precursor ion and isotopic pattern distribution benefit from the high resolving power of HRMS. However, when searching for compounds with a given molecular formula in public scientific databases, like ChemSpider or PubChem, an extensive number of potential structures can be found. In this case, the experience of analysts and their technical skillfulness in the MS spectral interpretation are essential for discarding unlike entries and reducing false positives.

In recent years, the hyphenation of ultrahigh performance liquid chromatography with ion mobility spectroscopy (IMS) and a quadrupole time-of-flight mass spectrometry (UPLC-IMS-QTof) has been widely used for targeted and untargeted screening of complex samples (Celma et al., 2020; Hernandez-Mesa et al., 2018; Hines et al., 2017; Righetti et al., 2018; Song et al., 2021). Collision cross section (CCS), derived from IMS, is a stable parameter related to the size, shape and charge of ions (D'Atri et al., 2018). Excellent inter-day precision of CCS measurements have been observed in the study of Regueiro et al. (2016) and Song et al. (2022b) with relative standard deviations (RSDs) values below 1% for all pesticides and FCCs, and most of them ranging from 0.3% to 0.5%. Righetti et al. (2020) has proved that the CCS values of mycotoxins measured from two IMS platforms located in two different laboratories have deviations within 1.5%. Because of the high reproducibility of CCS measurements, these could be used as an extra molecular identifier in addition to retention time (RT), accurate mass, isotopic pattern and fragment ions (Righetti et al., 2018).

One limitation in the identification of FCCs originating from FCMs, is that many of these compounds are not commercially available, thus the tentative identification cannot be confirmed by comparing the RT, MS spectra and CCS values of unknowns with those of reference standards. To deal with this situation, some research groups have developed RT and CCS prediction models using machine learning approaches. Bonini and coworkers developed a R package: *Retip*, which can provide the predicted RT values for metabolomics studies, and they found that 68% of candidate structures were

filtered out with the addition of RT match (Bonini et al., 2020). Some groups have developed CCS prediction tools, such as AllCCS (Zhou et al., 2020), CCSondemand (Broeckling et al., 2021), CCSbase (Ross et al., 2020), and DeepCCS (Plante et al., 2019). Zhou et al. (2020) proved that ~75% of annotated candidates were reduced through addition of CCS match. In addition to the significant reduction of candidates brought by RT and CCS match, the comparison between experimental RT and CCS values with the predicted ones can increase the identification confidence, especially in absence of reference standards.

There are two databases related to FCMs: Chemicals associated with Plastic Packaging Database (CPPdb) and Food Contact Chemicals Database (FCCdb). CPPdb consists of 4283 substances associated with plastic food packaging, and FCCdb consists of 12285 substances that could possibly be used worldwide to make FCMs. The search of compounds with formulas proposed by HRMS in these two databases rather than ChemSpider or PubChem may reduce the list of potential candidates.

In this work, we firstly developed RT and CCS prediction models based on machine learning approaches and experimental values. Then the RT and CCS values of substances in CPPdb and FCCdb were predicted, these two databases were then transformed into screening libraries and integrated in our structural elucidation workflow. The applicability of these two databases to the identification of FCCs was evaluated by the migration test of PA spatulas, where the chemicals migrated from PA spatulas were screened against self-built plastic additives database, CPPdb and FCCdb. Some approaches aiding the identification of FCCs were also discussed.

3. Materials and methods

3.1. Chemicals and reagents

A total of 675 standards consisting of antioxidants, plasticizers, photoinitiator, UV absorbers, slip agent, lubricants, degradation products of additives and oligomers of

polyethylene glycol (PEG), polypropylene glycol (PPG) of were purchased from Sigma-Aldrich Quimica S.A. (Madrid, Spain), Cymit Química S.L (Barcelona, Spain), Extrasynthese (Genay, France) and Cayman chemical company (Ann Arbor, Michigan, USA). Oligomers of adhesives, polyamide (PA), PLA were isolated from corresponding polymers using the procedures in Canellas et al. (2021b), HPLC grade ethanol (≥99.9%) and methanol (≥99.9%), dichloromethane (≥99.8%), acetone (≥99.8%), dimethyl sulfoxide (≥99.8%) were purchased from Scharlau Chemie S.A (Sentmenat, Spain). Deionized water was obtained from a Millipore Milli-QPLUS 185 system (Madrid, Spain).

Standard stock solutions of all standards were prepared in methanol at a concentration of 1000 mg/kg. For compounds not fully soluble in methanol, alternative solvents (ethanol, dichloromethane, acetone and dimethyl sulfoxide) were used. Working solutions consisting of 8-10 analytes at ~1 mg/kg were prepared from the dilution of stock solutions. All solutions were kept at -20 ºC until analysis.

3.2. Instrumentation

The standard solutions were analyzed by an Acquity I-Class UPLC system coupled to a Vion IMS-QTof mass spectrometer with an electrospray ionization (ESI) source (Waters, Manchester, UK). Compounds were separated by a CORTECS C18 UPLC column (2.1 × 100 mm, 1.6 µm particle size, 90 Å pore size) at a flow rate of 0.3 mL/min. Gradient elution was performed using water (A) and methanol (B) as mobile phases, both with 0.1% (v/v) formic acid. The initial percentage of B was 5%, which linearly increased to 100% over 7 min, followed by a 4 min isocratic period, then returned to initial conditions from 11 to 11.1 min and re-equilibrated at 5% B from 11.1 to 13 min. The sample and column temperature were 10 and 40 °C, respectively. Injection volume was 5 µL.

ESI was performed in both positive and negative ionization modes. Capillary voltage was 1 kV and cone voltage was 30 V, the source temperature was set to 120 °C

and desolvation temperature was set to 500 °C. Cone gas flow rate was 50 L/h, and desolvation gas flow was 800 L/h. Data were acquired in high definition MS^E mode ($HDMS^E$) over the *m/z* range of 50-1000. The instrument switched between two collision energies (low energy: 6 eV, high energy: ramp 20−40 eV) to obtain precursor and fragment ions in a single run. Leucine-Enkephalin ($[M+H]^+$: *m/z* 556.2766; $[M-H]^-$: *m/z* 554.2620) at a concentration of 100 ng/mL was infused at a rate of 15 µL/min for real-time mass correction. The IMS gas flow rate was 25 mL/min with a wave velocity of 250 m/s and an IMS pulse height of 45 V. The Major Mix IMS/Tof Calibration Kit (ref. 186008113) from Waters (Manchester, UK) was used for CCS calibration. Data acquisition and processing were carried out on UNIFI v.1.9.4 Scientific Information System (Waters, Manchester, UK).

3.3. RT prediction

The RT prediction was performed in R package *Retip* (Bonini et al., 2020). Three algorithms: Extreme Gradient Boosting (XGBoost), random forest (RF), Bayesian regularized neural network (BRNN) were used for model building and their prediction performances were compared.

Briefly, SMILES and InChIKey of each compounds was retrieved from PubChem using the R package *webchem* (Szöcs et al., 2020). Then CDK chemical descriptors of each compound were calculated, the returned data contained 667 compounds as the descriptors of 8 sodium-containing compounds were not successfully calculated. The descriptors with zero or low variance were eliminated, 134 descriptors were retained in the data. The dataset was then randomly split into training and testing set in the ratio of 8:2; 535 and 132 RT records were included in training and testing set, respectively. The model was built using the training set and then it was externally validated by the testing set.

3.4. CCS prediction

The prediction of CCS values of chemicals associated with plastic products was

described in a previous study (Song et al., 2022a), where the CCS prediction models were built by support vector machine regression (SVM) based on 1076 CCS values of $[M+H]^+$ and 645 CCS values of $[M+Na]^+$.

3.5. Prediction of RT and CCS values of substances in CPPdb and FCCdb

The cleaning-up of CPPdb and FCCdb databases was performed in the previous study (Song et al., 2022a), where metals and salts together with any substances with the same InChIKey (replicates) were removed. Only compounds with a molecular weight between 50-1200 Da were retained, and a total of 2883 and 6508 substances were retained in CPPdb and FCCdb, respectively. Subsequently, RT and CCS values for the compounds in CPPdb and FCCdb were predicted using the models described above. These two databases were then transformed into screening libraries and were used for suspect screening.

3.6. Migration test

PA spatulas were purchased from a local market, the migration test conditions were described ina previous study (Canellas et al., 2021b). Briefly, 95% (v/v) ethanol was used as food simulants, the spatula was cut into 1 × 5 cm pieces and placed into glass vials filled with 41.6 g of simulant and the vials were then placed in an oven at 60 °C for 2.5 h. Migration test was conducted in triplicate and 95% ethanol without spatula was used as blank.

3.7. Identification workflow

The identification process of the migrants from PA spatula is shown in Figure III-4.1, which generally consists of two stages. In the first stage, 675 standards, which included commonly used additives and NIAS in FCMs, were analyzed and their *m/z* values, adducts, RT, CCS values and fragment ions were added to an in-house database. The migration samples were screened against the in-house library with the criteria of *m/z* error < 5 ppm, RT error < 0.1 min, CCS delta < 2% and at least one fragment ion

matches.

The unidentified features (*m/z*_RT_CCS) were then screened against two plastic packaging-related databases: CPPdb (2883 compounds) and FCCdb (6508 compounds), which contain the information of *m/z* values, adducts, prediction RT and CCS values. The filter setting was based on their prediction accuracy and is shown in Figure III-4.1. After the tentative assignment, commercial standards of proposed compounds were purchased to further confirm the identification, when available.

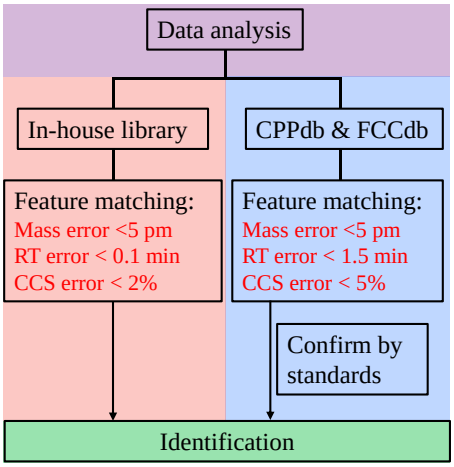


Figure III-4.1. The workflow of identification of migrates from PA spatula.

4. Results and discussion

4.1. RT and CCS prediction

After splitting the dataset into training and testing set, 132 RT records were included in the testing set, herein, three algorithms were used to build RT prediction models. The performance comparison is shown in Figure III-4.2, where it can be seen that RF and XGBoost present similar prediction results, and both outperformed BRNN. Considering that the configuration of RF algorithm involves fewer hyperparameters and needs less effort for tuning, RF was used for developing the RT prediction model and then predicting the RT values of molecules in CPPdb and FCCdb. As shown in Figure III-4.2, RF-based model has prediction errors within 1.22 min for testing set when

considering 95% confidence interval, so when the unknown features were screened against CPPdb and FCCdb, the RT filter was set as 1.5 min. Herein a wider acceptance error window of RT was used in order to avoid the elimination of true positives.

In the CCS prediction model, 93.3% of $[M+H]^+$ adducts and 95.0% of $[M+Na]^+$ adducts in testing sets were predicted with <5% error (Song et al., 2022a), thus the CCS filter of CPPdb and FCCdb library searching was set as 5%.

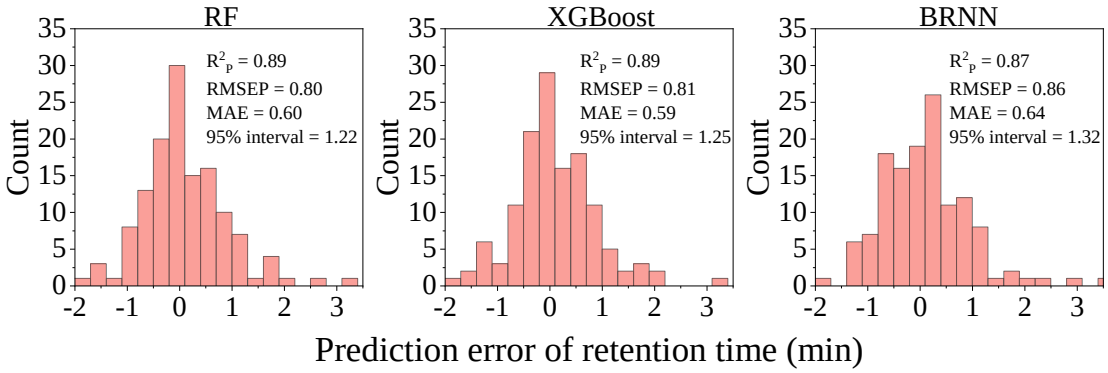


Figure III-4.2. Histograms showing prediction errors of retention time using the models of random forest (RF), extreme Gradient Boosting for tree algorithms (XGBoost) and Bayesian Regularized Neural Network (BRNN). R^2_p: external validation coefficient of determination; RMSEP: root mean square error of prediction; MAE: mean absolute error; 95% interval: prediction errors in 95% confidence interval.

4.2. Identification of migrating compounds by self-built library

A total of 51 compounds were identified by searching against the self-built library, which included common additives: plasticizers, antioxidant, slip agents, lubricants. Some NIAS (oligomers and degradation products of additives) were also found in the PA spatulas. Detailed identification results are shown in Table III-4.1.

Table III-4.1. The identified compounds migrated from polyamide (PA) spatula into 95% ethanol by self-built plastic additives library.

No.	RT_{exp} (min)	ΔRT (min)	CCS_{exp} ($Å^2$)	ΔCCS (%)	Observed *m/z*	*m/z* error (ppm)	Adducts	Molecular formula	Candidate name	Remarks
1	7.17	0.04	178.4	0.43	309.2036	0.0	$[M+Na]^+$	$C_{16}H_{30}O_4$	2,2,4-Trimethyl-1,3-Pentanediol Diisobutyrate	Plasticizer
2	7.74	0.09	191.4	-0.10	337.2351	0.4	$[M+Na]^+$	$C_{18}H_{34}O_4$	Dibutyl sebacate	Plasticizer
3	7.95	0.05	211.0	0.01	363.2502	-1.2	$[M+Na]^+$	$C_{20}H_{36}O_4$	Bis(2-ethylhexyl) maleate	Plasticizer
4	8.20	0.07	219.0	0.35	413.2660	-0.5	$[M+Na]^+$	$C_{24}H_{38}O_4$	Dioctyl phthalate/Diisooctyl phthalate/Bis(2-ethylhexyl) phthalate	Plasticizer
5	8.22	0.07	221.3	1.08	393.2973	-0.6	$[M+Na]^+$	$C_{22}H_{42}O_4$	Bis(2-ethylhexyl) adipate	Plasticizer
6	8.49	0.10	226.2	0	441.2975	0.0	$[M+Na]^+$	$C_{26}H_{42}O_4$	Dinonyl phthalate/Diisononyl phthalate	Plasticizer
7	6.60	0.03	176.8	0.71	277.1808	-0.6	$[M-H]^-$	$C_{17}H_{26}O_3$	3-(3,5-di-tert-butyl-4-hydroxyphenyl)propanoic acid	Degradation products
8	6.60	0.06	164.7	1.21	233.1544	-1.2	$[M-H]^-$	$C_{15}H_{22}O_2$	3,5-ditert-butyl-4-hydroxybenzaldehyde	Degradation products
9	6.66	0.04	169.6	0.44	247.1703	-0.3	$[M-H]^-$	$C_{16}H_{24}O_2$	3,5-di-tert-butyl-4-hydroxyacetophenone	Degradation products
10	7.16	-0.09	241.5	-0.46	551.3845	-1.6	$[M-H]^-$	$C_{34}H_{52}N_2O_4$	Irganox 1024	Antioxidant
11	7.47	0.09	273.9	0.95	637.4941	0.4	$[M+H]^+$	$C_{40}H_{64}N_2O_4$	Irganox 1098	Antioxidant
12	7.78	0.07	187.8	0.68	256.2635	0.1	$[M+H]^+$	$C_{16}H_{33}NO$	Hexadecanamide	Slip agent
13	7.87	0.04	191.0	0.29	282.2794	1.0	$[M+H]^+$	$C_{18}H_{35}NO$	Oleamide	Slip agent
14	7.93	-0.07	195.7	0.48	284.2947	-0.3	$[M+H]^+$	$C_{18}H_{37}NO$	Octadecanamide	Slip agent
15	8.41	0.10	202.6	0.42	360.3239	0.5	$[M+Na]^+$	$C_{22}H_{43}NO$	Erucamide	Slip agent
16	7.80	0.01	209.0	1.00	392.3121	-3.5	$[M+Na]^+$	$C_{22}H_{43}NO_3$	N,N-Diethanololeamide	Antistatic agent
17	8.03	0.08	201.7	0.56	365.2661	-0.5	$[M+Na]^+$	$C_{20}H_{38}O_4$	Glycol ricinoleate	Antistatic agent
18	8.13	0.06	172.8	0.30	317.2458	2.1	$[M+Na]^+$	$C_{19}H_{34}O_2$	9,12-Octadecadienoic acid, methyl ester	Antistatic agent
19	8.14	0.07	206.1	1.15	381.2974	-0.5	$[M+Na]^+$	$C_{21}H_{42}O_4$	Glyceryl monostearate	Antistatic agent
20	7.36	0.07	161.4	1.04	199.1702	-0.6	$[M-H]^-$	$C_{12}H_{24}O_2$	Lauric acid	Lubricant
21	8.08	0.08	176.3	0.68	255.2329	-0.1	$[M-H]^-$	$C_{16}H_{32}O_2$	Palmitic acid	Lubricant
22	8.38	0.10	184.5	-0.80	283.2643	0.2	$[M-H]^-$	$C_{18}H_{36}O_2$	Stearic acid	Lubricant
23	11.73	0.01	280.9	0.68	615.5811	1.9	$[M+Na]^+$	$C_{38}H_{76}N_2O_2$	N,N'-Ethylenebis(stearamide)	Lubricant
24	3.17	0.04	181.5	-0.09	362.2415	0.3	$[M+Na]^+$	$C_{18}H_{33}N_3O_3$	1,8,15-Triazacyclohenicosane-2,9,16-trione	PA6 trimer

No.	RT_{exp} (min)	ΔRT (min)	CCS_{exp} ($Å^2$)	ΔCCS (%)	Observed *m/z*	*m/z* error (ppm)	Adducts	Molecular formula	Candidate name	Remarks
25	3.67	0	210.9	1.21	453.3432	-0.8	$[M+H]^+$	$C_{24}H_{44}N_4O_4$	1,8,15,22-Tetraazacyclooctacosane-2,9,16,23-tetrone	PA6 tetramer
26	4.10	0.04	233.7	0.19	588.4090	-0.9	$[M+Na]^+$	$C_{30}H_{55}N_5O_5$	1,8,15,22,29-Pentazacyclopentatriacontane-2,9,16,23,30-pentone	PA6 pentamer
27	4.43	0.08	265.5	1.21	701.4935	-0.1	$[M+Na]^+$	$C_{36}H_{66}N_6O_6$	1,8,15,22,29,36-Hexazacyclodotetracontane-2,9,16,23,30,37-hexone	PA6 hexamer
28	2.70	0.09	152.0	0.49	227.1753	-0.5	$[M+H]^+$	$C_{12}H_{22}N_2O_2$	1,8-Diazacyclotetradecane-2,7-dione	PA66 monomer
29	3.92	0.04	211.0	0.19	453.3432	-0.8	$[M+H]^+$	$C_{24}H_{44}N_4O_4$	1,8,15,22-Tetraazacyclooctacosane-2,7,16,21-tetrone	PA66 dimer
30	4.61	0.06	266.2	-0.28	701.4939	0.4	$[M+Na]^+$	$C_{36}H_{66}N_6O_6$	1,8,15,22,29,36-Hexaazacyclodotetracontane-2,7,16,21,30,35-hexone	PA66 trimer
31	2.32	0.03	150.9	0.47	261.1306	-1.1	$[M+Na]^+$	$C_{10}H_{22}O_6$	PEG5	PEG oligomer
32	2.61	0.01	160.1	1.25	305.1569	-0.6	$[M+Na]^+$	$C_{12}H_{26}O_7$	PEG6	PEG oligomer
33	2.87	0.02	167.0	1.08	349.1833	0.0	$[M+Na]^+$	$C_{14}H_{30}O_8$	PEG7	PEG oligomer
34	3.08	0	178.6	1.40	393.2095	0.0	$[M+Na]^+$	$C_{16}H_{34}O_9$	PEG8	PEG oligomer
35	3.28	0.01	188.4	0.23	437.2355	-0.6	$[M+Na]^+$	$C_{18}H_{38}O_{10}$	PEG9	PEG oligomer
36	3.44	0	197.3	0.16	481.2616	-0.7	$[M+Na]^+$	$C_{20}H_{42}O_{11}$	PEG10	PEG oligomer
37	3.59	0	207.3	1.37	525.2877	-0.8	$[M+Na]^+$	$C_{22}H_{46}O_{12}$	PEG11	PEG oligomer
38	3.72	0	220.5	1.77	569.3147	0.7	$[M+Na]^+$	$C_{24}H_{50}O_{13}$	PEG12	PEG oligomer
39	4.72	-0.03	173.8	1.48	331.2088	-0.8	$[M+Na]^+$	$C_{15}H_{32}O_6$	PPG5	PPG oligomer
40	5.24	-0.03	185.9	0.15	389.2504	-1.6	$[M+Na]^+$	$C_{18}H_{38}O_7$	PPG6	PPG oligomer
41	5.66	-0.03	198.5	-0.34	447.2927	-0.4	$[M+Na]^+$	$C_{21}H_{44}O_8$	PPG7	PPG oligomer
42	5.98	-0.05	211.1	0.60	505.3343	-0.8	$[M+Na]^+$	$C_{24}H_{50}O_9$	PPG8	PPG oligomer
43	6.28	-0.02	224.6	1.49	563.3756	-1.7	$[M+Na]^+$	$C_{27}H_{56}O_{10}$	PPG9	PPG oligomer
44	6.51	-0.03	238.2	1.73	621.4179	-0.8	$[M+Na]^+$	$C_{30}H_{62}O_{11}$	PPG10	PPG oligomer
45	6.71	-0.03	253.2	1.65	679.4592	-1.6	$[M+Na]^+$	$C_{33}H_{68}O_{12}$	PPG11	PPG oligomer
46	5.29	0.05	141.1	0.61	179.0711	-1.3	$[M-H]^-$	$C_{10}H_{12}O_3$	Propylparaben	Biocide
47	5.58	0.08	161.9	-1.05	242.1761	-0.2	$[M+HCOO]^-$	$C_{12}H_{23}NO$	12-Aminododecanolactam	Monomer
48	5.97	-0.1	184.2	-0.02	355.1457	4.6	$[M+H]^+$	$C_{23}H_{18}N_2O_2$	2-Diphenylacetyl-1,3-indandione-1-hydrazone	-
49	7.29	0.06	203.4	-0.62	379.1700	-0.5	$[M+H]^+$	$C_{21}H_{28}Cl_2N_2$	4,4'-Methylenebis(3-Chloro-2,6-Diethylaniline)	Curing agent

No.	RT_{exp} (min)	ΔRT (min)	CCS_{exp} ($Å^2$)	ΔCCS (%)	Observed *m/z*	*m/z* error (ppm)	Adducts	Molecular formula	Candidate name	Remarks
50	7.48	0.07	233.5	0.06	507.2709	-1.6	$[M+Na]^+$	$C_{29}H_{40}O_6$	1,2,3-Trideoxy-4,6:5,7-bis-O-((4-propylphenyl)methylene)-nonitol	Nucleating agent
51	8.87	0.10	228.9	1.02	431.1788	0.0	$[M+H]^+$	$C_{26}H_{26}N_2O_2S$	2,5-Bis(5-Tert-Butyl-2-Benzoxazolyl)thiophene	Brightener

Table III-4.2. Compounds migrating from polyamide (PA) spatula using CPPdb and FCCdb databases.

No.	RT_{exp} (min)	ΔRT (min)	CCS_{exp} ($Å^2$)	ΔCCS (%)	Observed *m/z*	*m/z* error (ppm)	Adducts	Molecular formula	Compound name	PubChem CID	Remarks
1	0.84	-0.82	141.8	-0.78	217.1042	-0.6	$[M+H]^+$	$C_6H_{12}N_6O_3$	Trimethylolmelamine	70549	FCCdb
2	2.03	-0.68	166.6	0.76	245.1861	0.6	$[M+H]^+$	$C_{12}H_{24}N_2O_3$	6-(6-Aminohexanamido)hexanoic acid	895	CPPdb
3	2.26	-0.59	137.5	-3.36	190.1077	1.9	$[M+H]^+$	$C_8H_{15}NO_4$	Diethyl iminodiacetate	80502	FCCdb
4	2.68	0.59	150.9	-0.22	209.1647	-0.6	$[M+H]^+$	$C_{12}H_{20}N_2O$	Bis[(dimethylamino)methyl]phenol	3018067	FCCdb
5	3.67	0.05	149.1	-0.10	210.1488	-0.1	$[M+H]^+$	$C_{12}H_{19}NO_2$	N,N-Bis(2-hydroxypropyl)aniline	76498	FCCdb
6	4.45	0.47	137.0	-0.48	172.1117	-2.4	$[M+H]^+$	$C_{12}H_{13}N$	(S)-(-)-1-(1-Naphthyl)ethylamine	66325	FCCdb
7	4.61	0.83	160.1	1.85	228.1594	-0.1	$[M+Na]^+$	$C_{12}H_{21}NO_3$	4-[(2-Ethylhexyl)amino]-4-oxoisocrotonic acid	6913277	FCCdb
8	4.75	1.29	183.8	-3.88	399.2596	1.8	$[M+H]^+$	$C_{18}H_{38}O_9$	Oxirane, 2-methyl-, polymer with oxirane, ether with 1,2,3-propanetriol (3:1)	86278135	FCCdb
9	5.12	-1.23	199.4	-4.43	426.2681	-5.0	$[M+H]^+$	$C_{18}H_{43}NO_6Si_2$	Bis(3-(triethoxysilyl)propyl)amine	83535	FCCdb
10	5.6	-0.66	186.1	-2.88	369.2246	-0.4	$[M+Na]^+$	$C_{18}H_{34}O_6$	Sorbitan laurate	347468	FCCdb
11	5.72	-0.31	156.6	0.06	225.1485	-0.2	$[M+H]^+$	$C_{13}H_{20}O_3$	Isoamyl 4-(2-furan)butyrate	68968644	FCCdb
12	5.84	-0.91	169.1	-2.07	291.1566	-0.3	$[M+Na]^+$	$C_{15}H_{24}O_4$	1,9-Nonanediol diacrylate	9795378	FCCdb
13	5.95	-0.07	153.3	1.62	209.1538	1.1	$[M+H]^+$	$C_{13}H_{20}O_2$	(2-(1-Propoxyethoxy)ethyl)benzene	61403	FCCdb
14	6.07	0.59	151.9	0.86	211.1693	0.1	$[M+H]^+$	$C_{13}H_{22}O_2$	2-Bornyl propionate	89306	FCCdb
15	6.13	-0.73	177.1	-3.50	300.1934	0.1	$[M+Na]^+$	$C_{17}H_{27}NO_2$	2-Ethylhexyl 4-(dimethylamino)benzoate	30541	CPPdb/ FCCdb
16	6.25	-0.12	187.9	0.22	367.1875	-1.4	$[M+Na]^+$	$C_{21}H_{28}O_4$	Bisphenol A bis(2-hydroxypropyl) ether	8306	FCCdb
17	6.3	-0.38	189.1	-3.75	369.2239	-2.4	$[M+Na]^+$	$C_{18}H_{34}O_6$	Bis(2-butoxyethyl) adipate	8837	CPPdb

No.	RT_{exp} (min)	ΔRT (min)	CCS_{exp} ($Å^2$)	ΔCCS (%)	Observed *m/z*	*m/z* error (ppm)	Adducts	Molecular formula	Compound name	PubChem CID	Remarks
18	6.35	-1.11	224.2	2.86	509.3344	4.6	$[M+H]^+$	$C_{25}H_{48}O_{10}$	Pentanedioic acid, bis(2-(2-(2-butoxyethoxy)ethoxy)ethyl) ester	103407	CPPdb
19	6.40	-0.07	171.9	-0.08	327.0781	0.0	$[M+H]^+$	$C_{18}H_{15}O_4P$	Triphenyl phosphate	8289	CPPdb/FCCdb
20	6.62	-0.73	204.8	3.63	439.2672	1.2	$[M+Na]^+$	$C_{22}H_{40}O_7$	2-(2-(2-(Dodecyloxy)ethoxy)ethoxy)ethyl hydrogen maleate	6437561	FCCdb
21	6.72	0.46	181.5	-3.24	341.1278	-1.8	$[M+H]^+$	$C_{22}H_{16}N_2O_2$	3,10-Dimethylquinolino[2,3-B]Acridine-7,14(5H,12H)-Dione	11382230	CPPdb
22	6.87	-0.08	234.6	-2.31	553.3969	4.1	$[M+H]^+$	$C_{28}H_{56}O_{10}$	Octaethylene Glycol Laurate	10187601	FCCdb
23	6.87	0.01	245.4	-4.10	531.4149	3.2	$[M+Na]^+$	$C_{30}H_{56}N_2O_4$	Bis(1,2,2,6,6-Pentamethyl-4-Piperidyl) Sebacate	586744	CPPdb/FCCdb
24	7.27	1.27	202.9	2.03	465.2582	-1.2	$[M+Na]^+$	$C_{20}H_{43}O_8P$	3,6,9,12-Tetraoxatetracosan-1-ol, dihydrogen phosphate	170762	CPPdb/FCCdb
25	7.34	-0.27	194.3	-3.89	337.2134	-1.2	$[M+Na]^+$	$C_{21}H_{30}O_2$	Methyl dehydroabietate	14697	CPPdb
26	7.46	-0.51	204.3	2.93	375.2504	-0.4	$[M+Na]^+$	$C_{21}H_{36}O_4$	5-Carboxy-4-hexyl-2-cyclohexene-1-octanoic acid	105841	FCCdb
27	7.46	0.53	170.6	2.91	259.1691	-0.5	$[M+H]^+$	$C_{17}H_{22}O_2$	Geranyl benzoate	5353011	FCCdb
28	7.59	0.86	250.7	3.49	605.4233		$[M+Na]^+$	$C_{30}H_{62}O_{10}$	Polidocanol	656641	CPPdb/FCCdb
29	7.66	-0.39	196.8	0.74	349.2729	4.6	$[M+Na]^+$	$C_{20}H_{38}O_3$	Glycol oleate	5364420	CPPdb/FCCdb
30	7.82	-0.15	216.3	2.97	395.277	0.5	$[M+Na]^+$	$C_{21}H_{40}O_5$	Decanoic acid, ester with 1,2,3-propanetriol octanoate	19026760	FCCdb
31	7.83	0.78	176.0	2.53	273.1845	-1.3	$[M+H]^+$	$C_{18}H_{24}O_2$	Geranyl Phenylacetate	5366044	FCCdb
32	7.86	0.31	198.0	1.85	353.2659	-1.0	$[M+Na]^+$	$C_{19}H_{38}O_4$	Methyl 9,10-dihydroxyoctadecanoate	66194	CPPdb/FCCdb
33	8.16	0.39	212.2	3.44	397.3285	-0.9	$[M+Na]^+$	$C_{22}H_{46}O_4$	ceteth-3	4639427	FCCdb
34	8.2	1.02	229.3	-0.20	475.2374	-1.3	$[M+H]^+$	$C_{32}H_{30}N_2O_2$	9,10-Anthracenedione, 1,4-bis[(2-ethyl-6-methylphenyl)amino]-	6451771	FCCdb
35	8.24	0.06	203.4	0.77	312.3259	-0.6	$[M+Na]^+$	$C_{20}H_{41}NO$	icosanamide	3016647	CPPdb
36	8.26	-0.35	314.2	-0.83	844.6006	-3.7	$[M+Na]^+$	$C_{44}H_{88}NO_{10}P$	Hydroxylated lecithin	57508518	FCCdb
37	8.32	-0.21	237.0	-0.14	493.3503	0.7	$[M+Na]^+$	$C_{27}H_{50}O_6$	Glyceryl trioctanoate	10850	FCCdb
38	8.61	0.64	211.5	3.13	381.3339	-0.1	$[M+Na]^+$	$C_{22}H_{46}O_3$	Arosurf	10991978	FCCdb
39	8.74	1.05	203.0	1.72	351.2557	-1.9	$[M+H]^+$	$C_{18}H_{38}O_4S$	2-Hydroxy-1-Octadecanesulfonic Acid	414070	FCCdb
40	9.02	-0.73	277.0	0.00	619.527	-0.3	$[M+Na]^+$	$C_{37}H_{72}O_5$	Glycerides, C16-18 Mono- And Di-	3086206	CPPdb/FCCdb
41	9.02	-0.74	249.9	-1.01	533.4541	0.1	$[M+Na]^+$	$C_{32}H_{62}O_4$	Ditridecyl adipate	85653	CPPdb
42	9.21	-0.25	249.8	-1.41	547.3992	-0.2	$[M+H]^+$	$C_{33}H_{54}O_6$	Tris(2-ethylhexyl) trimellitate	18725	CPPdb/FCCdb

No.	RT_{exp} (min)	ΔRT (min)	CCS_{exp} ($Å^2$)	ΔCCS (%)	Observed *m/z*	*m/z* error (ppm)	Adducts	Molecular formula	Compound name	PubChem CID	Remarks
43	10.11	0.04	264.5	0.60	559.5177	0.7	$[M+Na]^+$	$C_{34}H_{68}N_2O_2$	Hexadecanamide, N,N'-1,2-ethanediylbis-	79654	CPPdb/FCCdb
44	11	0.83	272.1	-0.91	587.5487	0.2	$[M+Na]^+$	$C_{36}H_{72}N_2O_2$	Ethylene-N-palmitamide-N'-stearamide	101432497	CPPdb/FCCdb

Among these additives, six plasticizers were found in PA spatula, including the structures of fatty acid ester and phthalates. It should be noted that the identities of phthalate-based plasticizers were not confirmed due to the presence of structural isomers. For example, three isomers (dioctyl phthalate, diisooctyl phthalate, bis(2-ethylhexyl) phthalate) showed good matches with the feature of 413.2660_8.20 min_219.0 $Å^2$, their corresponding RT and CCS values in library were 8.13 min_218.3 $Å^2$, 8.10 min_217.9 $Å^2$, 8.19 min_217.5 $Å^2$. All these three compounds met the criteria of RT error < 0.1 min and CCS error < 2%. To differentiate these isomers, IMS with higher resolving power and better reproducibility may be required (ΔCCS <0.5%).

Antioxidants are other commonly used additives in plastics to prevent the oxidation of polymers during the process of processing, storage and usage; two 3-(3,5-Di-tert-butyl-4-hydroxyphenyl)propionate antioxidants: Irganox 1024 and Irganox 1098 were found in PA spatula. Besides, three degradation products: 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propanoic acid, 3,5-ditert-butyl-4-hydroxybenzaldehyde, 3,5-di-tert-butyl-4-hydroxyacetophenone were identified, which can originate from the oxidation of Irganox 1024 and Irganox 1098. These degradation products were also found in other plastic products (Felix et al., 2012; Vera et al., 2018).

Fatty amides like oleamide, erucamide and fatty acids like palmitic acid, stearic acid were identified in PA spatula, which were commonly used as slip agents and lubricants in plastics (Vera et al., 2019). Slip agents like oleamide and erucamide, could form microcrystalline structure on the surface of films and reduce the friction coefficient between films and equipment (Getachew et al., 2017). Some antistatic agents: N,N-diethanololeamide and glycerin derivatives, such as glyceryl monostearate were also detected in PA spatula.

The most abundant group of compounds detected in spatula was PA oligomers, including four PA6 oligomers and three PA66 oligomers. PA6 is a polymer of ε-caprolactam and PA66 is a polymer of 1,6-hexanediamine and adipic acid. The PA6 tetramer 1,8,15,22-tetraazacyclooctacosane-2,9,16,23-tetrone and PA66 dimer 1,8,15,22-tetraazacyclooctacosane-2,7,16,21-tetrone showed the same molecular mass

and similar CCS values, one obvious difference is their RT values, PA6 tetramer was eluted earlier than PA66 dimer. We compared the high energy spectra of PA6 tetramer and PA66 dimer to explore the different fragment ions between these two compounds. As can be seen in Figure III-4.3, Some common fragment ions at *m/z* 209.1646, 226.1909 and 435.3327 were observed for both PA6 tetramer and PA66 dimer. PA6 tetramer presented a characteristic fragment ion with *m/z* 114.0908 which corresponds to the monomer of PA6, i.e., caprolactam. PA66 dimer presented characteristic fragment ions with *m/z* 100.1118, 111.0436, 128.0703, 182.1535, 326.2800, the fragment ion 100.1118 can be obtained from 1,6-hexanediamine losing NH_2, fragment 111.0436 can be from adipic acid losing two hydroxyl groups. These fragmentation information are valuable to distinguish between PA6 and PA66 oligomers.

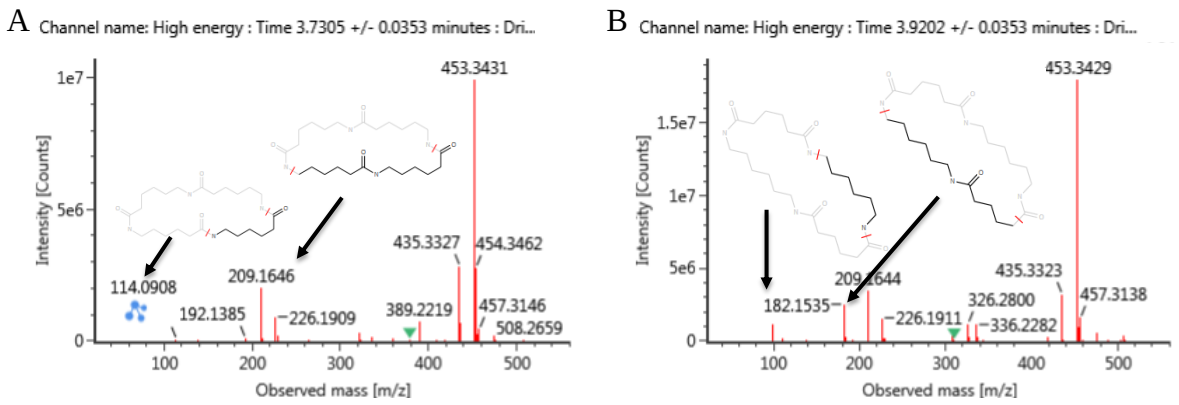


Figure III-4.3. High energy spectra of PA6 tetramer (A) and PA66 dimer (B).

Eight PEG oligomers and seven PPG oligomers were found in PA spatula, these oligomers are highly viscous and always blended with lower viscosity plasticizers in plastics. These oligomers were also detected in PA kitchenware (Hu et al., 2021). It has been reported that the grafting of PEG on polyvinyl chloride (PVC) surface reduced the diffusion of bis(2-ethylhexyl) phthalate from PVC matrix, possibly due to the high hydrophilicity of PEG (Rahman and Brazel, 2004). The CCS values versus *m/z* values of PEG and PPG oligomers is shown in Figure III-4.4, where there is a significant trend of linear correlation between their CCS and *m/z* values, with R^2 of 0.9907. PEG and PPG oligomers tend to form compact structures by intramolecular hydrogen bonds (Kozlowska et al., 2016), thus, their CCS values are relatively lower than that of other

compounds with similar *m/z* values. Figure III-4.5 shows the mass spectra of PPG5 with or without drift time (DT) alignment. As precursor and fragment ions always share the same DT, the alignment based on both RT and DT enables to eliminate many interfering ions and obtain cleaner mass spectra. This is a big advantage of using IMS-HRMS in targeted and suspect screening analysis.

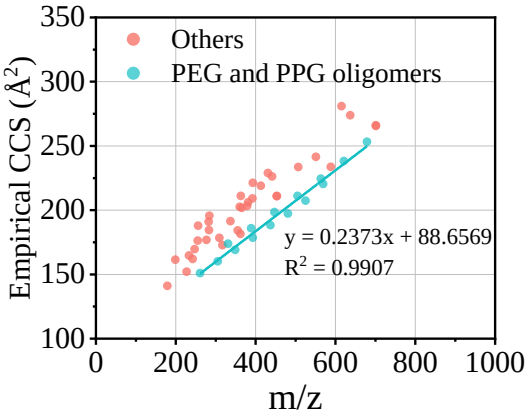


Figure III-4.4. Depiction of $^{TW}CCS_{N2}$ values vs *m/z* values for PEG and PPG oligomers.

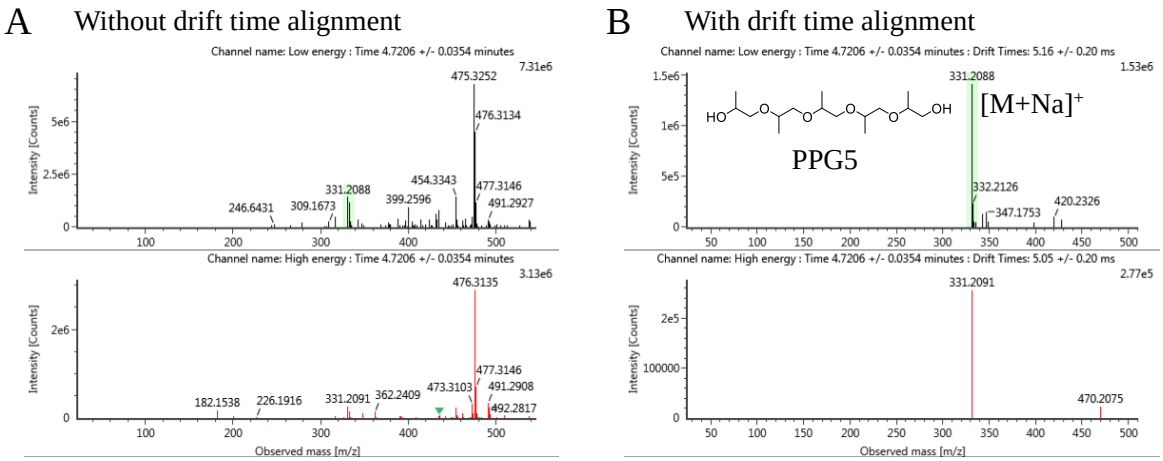


Figure III-4.5. Comparison of mass spectra of PPG5 ([M+Na]$^+$, *m/z* 331.2088) without drift time alignment (A) and with drift time alignment (B).

Some features present no fragment ions in their high energy spectra, for example: PPG5, as shown in Figure III-4.5(B). This could be due to the low concentrations of analytes or the analytes are not easily fragmentated. The use of IMS gives higher confidence to the identification of compounds without fragment ions, as it provides an additional dimension of separation, with three identification points *(m/z*, RT and CCS).

Despite these benefits of IMS in targeted and suspect screening analysis, the crucial information provided by LC cannot be ignored. For example, the isomeric pairs PA6 tetramer (CCS = 210.91 Å^2, RT = 3.73 min) and PA66 dimer (CCS = 210.98 Å^2, RT = 3.92 min) have the similar experimental CCS values, and their complete identification was achieved by the different RT values. The power of LC-IMS-HRMS originates from the multi-dimensional structural information that this technique provides.

4.3. Identification of migrating compounds using CPPdb and FCCdb database

Firstly, we studied the effect of using predicted RT and CCS filter on the number of candidates in the suspect screening. The features from PA migration samples were screened against 9391 compounds in CPPdb and FCCdb, and the number of candidates retained by applying different filters (*m/z*, RT and CCS) is shown in Figure III-4.6.

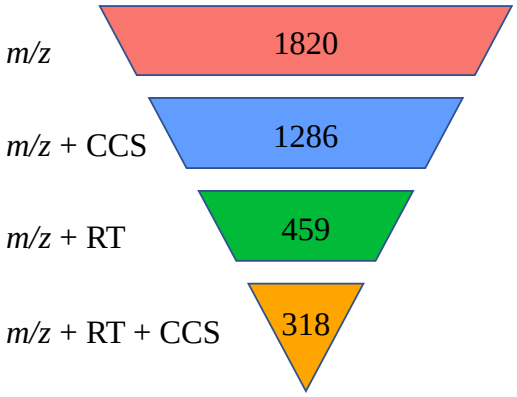


Figure III-4.6. The number of candidates retained by applying different filters. *m/z* filter is 5 ppm, CCS filter is 5% and RT filter is 1.5 min.

It is obvious that compared with CCS filter, the RT filter can lead to a higher reduction of false positives, approximately 75% (from 1820 to 459) of candidate structures were further excluded as likely false positives by the inclusion a RT filter of 1.5 min. Similar reduction of candidates was also observed in Bonini et al. (Bonini et al., 2020), an average of 68% of all candidate were eliminated by using a RT filter of 1 min. The use of CCS filter alone resulted in a reduction of 29% (from 1820 to 1286) of candidates, this is in accordance with the study of Bijlsma et al. (Bijlsma et al., 2019), in which around 9% to 39% of candidates were excluded by applying a CCS filter of

6%. The lower reduction of candidates by applying CCS filter is understandable, as CCS is highly correlated to the molecular weight (MW) (Hernandez-Mesa et al., 2018; Hines et al., 2017; Song et al., 2022b), however, RT is mainly determined by octanol/water partition coefficient (Bonini et al., 2020), and has less correlation to MW (see Figure III-4.7), which can provide more complementary structural information for features.

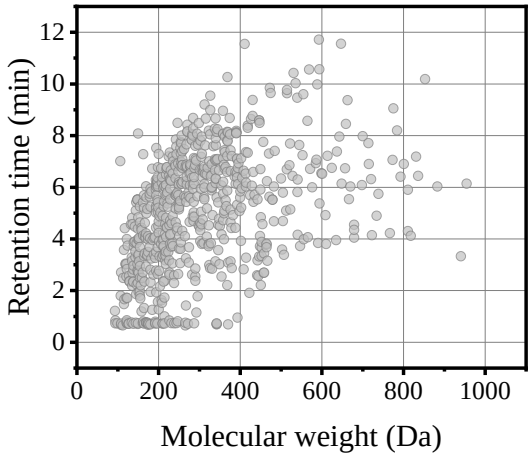


Figure III-4.7. Retention time (RT) versus molecular weight (Da) for 667 compounds in the database.

A total of 44 compounds were tentatively identified based on the match of RT, CCS and *m/z* of features with that of molecules in CPPdb and FCCdb databases, the detailed information including RT, CCS, adducts, molecular formula, mass error are summarized in Table III-4.2.

Three alkyl PEG ethers were identified, including polidocanol, ceteth-3 (emulsifying), arosurf, This kind of compounds was also called alkyl ethoxylates, which are commonly used as non-ionic surfactants in plastics (Cowan-Ellsberry et al., 2014). Tisler and coworkers also found the alkyl PEG ethers in migration experiments from reusable sport polyethylene (PE) water bottles (Tisler and Christensen, 2022). The ethoxylate (EO) group in chain is able to increase the polarity of molecule, thus, the higher number of EO groups in alkyl PEG ethers leads to an earlier chromatographic elution in reverse phase LC. This was proven by the elution order of polidocanol, ceteth-3 and arosurf, which contain 9, 3 and 2 EO groups in their molecules, respectively, and

their corresponding RT were 7.59, 8.16 and 8.61 mins. Three derivatives of alkyl PEG ethers were identified: 2-(2-(2-(Dodecyloxy)ethoxy)ethoxy)ethyl hydrogen maleate can originate from the reaction between triethylene glycol monododecyl ether and maleic acid; 3,6,9,12-tetraoxatetracosan-1-ol, dihydrogen phosphate can originate from the esterification between tetraethylene glycol monododecyl ether and phosphoric acid; octaethylene glycol laurate can originate from the reaction between octaethylene glycol and lauric acid, these three derivatives can also be used as surfactants.

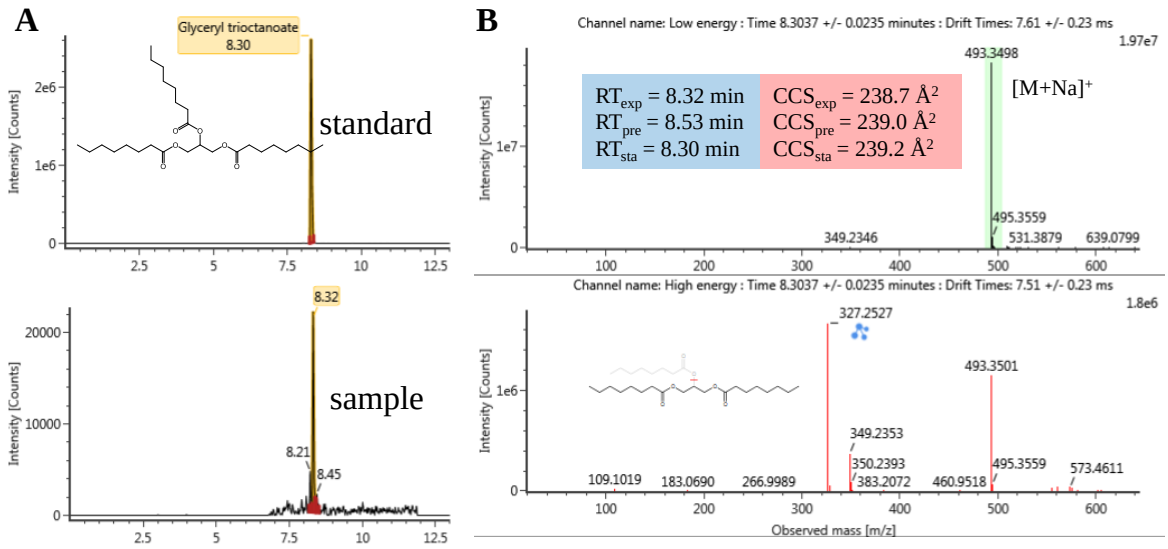


Figure III-4.8. Identification of glyceryl trioctanoate. (A) extracted ion chromatograms for its $[M+Na]^+$ adduct (*m/z* 493.3498), (B) low and high energy spectra, fragment assignment, comparison between experimental and predicted RT and CCS values.

Five antistatic agents were tentatively identified, including one sorbitan fatty acid ester (sorbitan laurate), one glycol derivative (glycol oleate) and three glycerol derivatives (decanoic acid, ester with 1,2,3-propanetriol octanoate; glyceryl trioctanoate; glycerides, C16-18 mono- and di-.) Antistatic agent is also one type of surfactants, the polyol amine (N,N-Diethanololeamide), sorbitan laurate, glycol and glycerol derivatives have moisturizing properties, which are capable of forming a water film on the surface of plastics to prevent static electricity. As an example, the identification of glyceryl trioctanoate was further confirmed by the reference standard and its extracted ion chromatogram, comparison between experimental and predicted RT and CCS values, mass spectra and assignment of fragments are shown in Figure III-4.8. The fragment

ion with *m/z* value of 327.2527 $[M - C_8H_{15}O_2]^+$ can be derived from the loss of a octanoate anion, the fragment ion with *m/z* value of 349.2353 had a deviation of 21.9826 with 327.2527, corresponding to $[M + Na - C_8H_{15}O_2]^+$. For this compound, the predicted RT and CCS values showed a good match with experimental values, with ΔRT = 0.21 min and ΔCCS = 0.3 Å^2. The confidence of identification was improved by the integration of predicted RT and CCS values.

Four plasticizers were tentatively identified by CPPdb and FCCdb, including bis(2-butoxyethyl) adipate, pentanedioic acid, bis(2-(2-(2-butoxyethoxy)ethoxy)ethyl) ester, ditridecyl adipate and tris(2-ethylhexyl) trimellitate. The identity of tris(2-ethylhexyl) trimellitate was confirmed by comparison to the reference standard, this compound was also previously detected in PA kitchenware (Hu et al., 2021). Two lubricants: hexadecanamide, N,N'-1,2-ethanediylbis- and ethylene-N-palmitamide-N'-stearamide were found in the migrates from PA spatula, the former has been detected in PE films by Vera and coworkers (Vera et al., 2019).

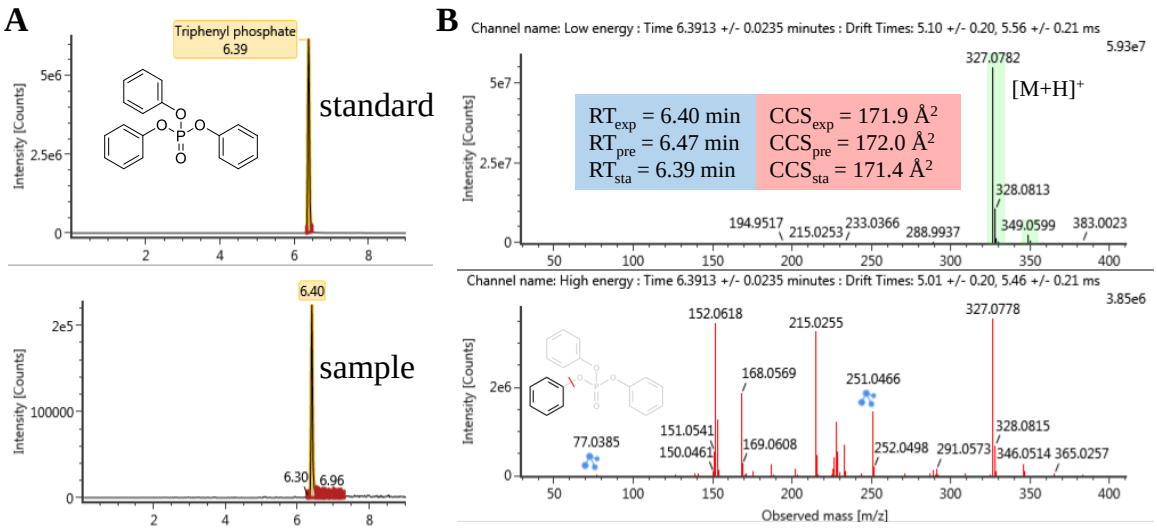


Figure III-4.9. Identification of triphenyl phosphate. (A) extracted ion chromatograms from standard and sample injection, (B) low and high energy mass spectra, fragment assignment, comparison between experimental and predicted RT and CCS values for its $[M+H]^+$ adduct.

One organophosphate flame retardant: triphenyl phosphate, was identified and further confirmed by a reference standard. Its chromatograms, mass spectra, fragment ions, predicted RT and CCS values are shown in Figure III-4.9. The fragment ion with

m/z value of 251.0466 indicated the loss of phenyl group, fragment ion 77.0385 corresponded to the retained benzene. The fragment ions with *m/z* values of 152.0618, 168.0569 and 215.0255 were also observed in mass spectra from MassBank of North America (MoNA). The experimental and predicted CCS values of its $[M+H]^+$ adduct showed a good match, with ΔCCS% = 0.1%, whist the predicted CCS value for $[M+Na]^+$ of triphenyl phosphate was not very accurate, with experimental and predicted CCS values of 181.8 $Å^2$ and 187.8 $Å^2$, respectively, the CCS deviation was 3.3%, which was still less than 5%.

2-Ethylhexyl 4-(dimethylamino)benzoate was identified and it can be used as photoinitiator. One dimer: 6-(6-Aminohexanamido)hexanoic acid was identified and it can be derived from the polymerization of two 6-aminohexanoic acid. It should be noted that 6-aminohexanoic acid can be used to produce caprolactam. One fatty acid ester: methyl 9,10-dihydroxyoctadecanoate may be from the oxidation of methyl oleate.

As limited by the accuracy of predicted RT and CCS values, some false positive candidates were also found in the identification, the feature of 553.3973_6.85_234.4 and 665.3831_7.06_260.8 showed a good match with Irganox 1024 and Irganox 1035, respectively, using the filtering criteria of RT error < 1.5 min and ΔCCS% < 5%, however, the subsequent injection of standards denied these identifications. This means that although the compounds in Table S1 met all the identification criteria, the definite identities still need to be confirmed by reference standards. The addition of predicted RT and CCS values into identification process in this study can improve the confidence but do not unequivocally confirm the identities.

4.4. Approaches aiding the identification of FCCs

The use of CPPdb and FCCdb in untargeted screening of FCCs can significantly reduce the number of candidates. For example, bisphenol A bis(2-hydroxypropyl) ether was identified by searching against CPPdb and FCCdb, which is a common bisphenol derivative in plastic products. However, by searching its molecular formula in PubChem, 3247 compounds were found and the first listed compounds were steroids; these substances are unlikely to appear in plastic packaging. In fact, we can confidently state

that in the identification process of FCCs, the use of plastic-related or FCM-related databases can reduce the false candidates and improve the identification confidence.

One limitation still exists when using CPPdb and FCCdb database. Although CPPdb and FCCdb currently contain a total of 4283 and 12285 substances, respectively, there are still some emerging substances associated with FCMs that are not included in these two databases, such as PA, PEG and PPG oligomers. With the rapid growth of reported new FCCs, these two databases need to be continually expanded and updated.

The improvement of IMS resolving power and CCS reproducibility would also help the identification of FCCs, indeed some structural isomers of phthalate-based plasticizers (dioctyl phthalate, diisooctyl phthalate, bis(2-ethylhexyl) phthalate) could be well differentiated with a CCS error window of 0.5% or lower. The current TWIMS system was found to operate at a resolving power of 40-50 (Dodds et al., 2017), efficient separations of these structural isomers on the basis of IMS may require a resolving power of ≥200. Recently, it has been reported that one state-of-the-art ion mobility platform: structures for lossless ion manipulations (SLIM), was able to provide a resolving power >300, with low CCS measurement deviations <0.5% (May et al., 2021). This technique is promising to differentiate the structural isomers of plasticizers.

5. Conclusions

A total of 51 migrates were identified from PA spatula by screening against self-built plastic additives library. Furthermore, 44 migrates from PA spatula were identified by screening against two public FCC-related databases (CPPdb and FCCdb), of which 3 tentative identifications were further confirmed by reference standards. The most abundant compounds migrated from PA spatula were PA6 and PA66 oligomers, some other additives, like plasticizers, slip agents and antistatic agents were also found in the migrates. Regarding the predicted RT and CCS values from machine learning approaches, they can be used to reduce false candidates and to add some identification confidence, the accuracy of prediction can be further improved by the incorporation of more RT and CCS records in training set, and more exhaustive characterization of the molecular performance by descriptors.

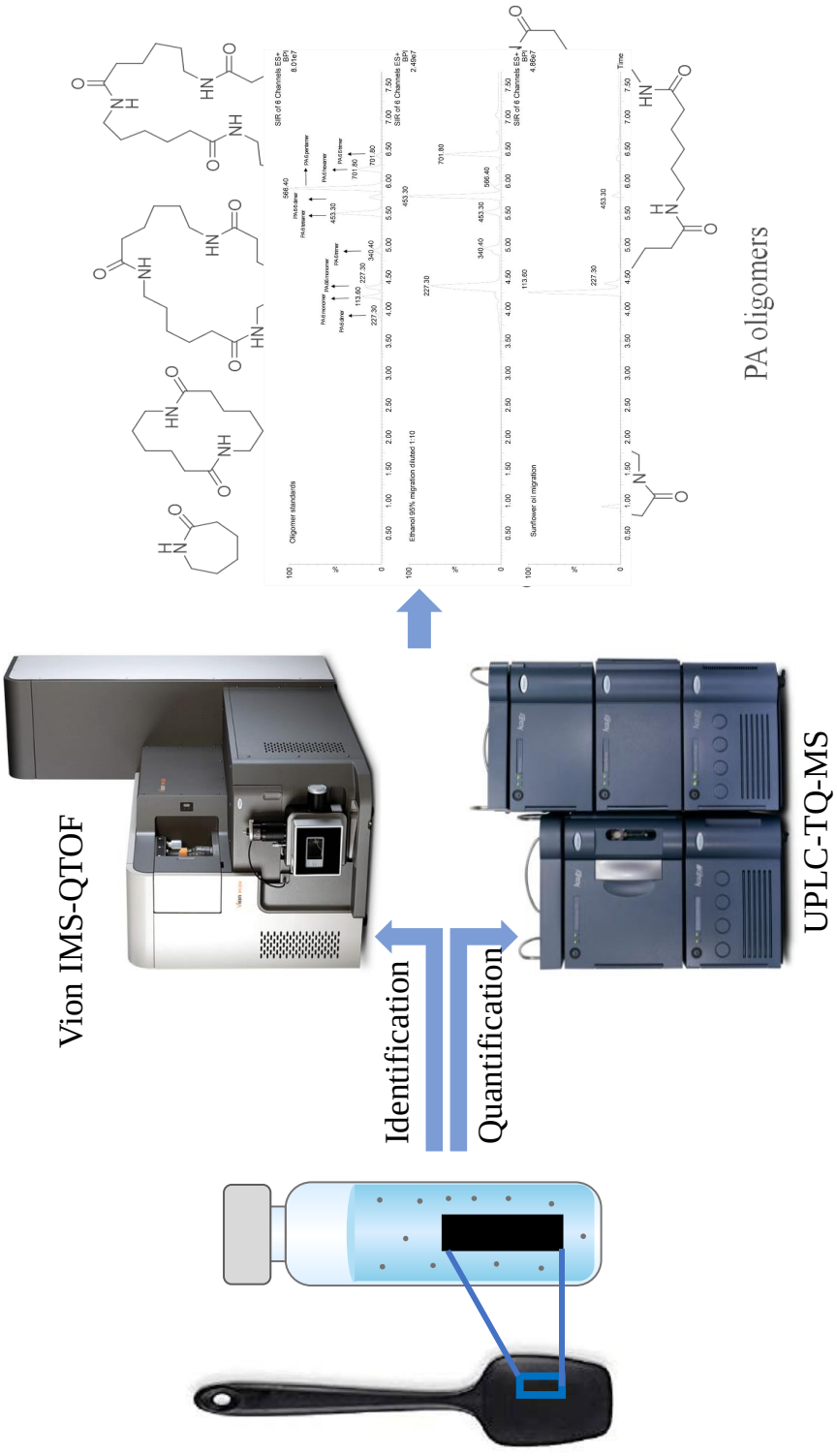
The building of an integrated FCM-related database is a promising approach for the identification of FCCs, a lot of repetitive identification work can be avoided by the direct screening against developed database. Besides, the CPPdb and FCCdb databases need to be continually expanded and updated with the rapid growing of emerging FCCs found in FCMs.

Chapter 5

The use of ion mobility time-of-flight mass spectrometry to assess the migration of polyamide 6 and polyamide 66 oligomers from kitchenware utensils to food

1. Abstract

Oligomers, are, in general, unknown components of the polymer. These oligomers can migrate from the polymer into the food and become a non-intentionally added substance to the food. In this work, ion mobility time-of flight mass spectrometry has been used to identify oligomers migrating from kitchenware. The structure elucidation of oligomers from polyamide 6 and polyamide 66 was achieved through the analysis of accurate *m/z* values of adducts and collision cross section values of precursor ions together with high-energy fragmentation patterns. Additionally, a method to extract oligomers from sunflower oil, cooked beans, soup and whole milk has been developed. Extraction recoveries ranged from 87 to 102% and limits of detection were from 0.03 to 0.11 mg/kg. It was observed that the migration from kitchenware to real food was below the specified migration limit of 5 mg/kg. However, this limit was exceeded for food simulants, which therefore overestimated the oligomer migration.


PA oligomers
UPLC-TQ-MS
Identification
Quantification
Vion IMS-QTOF

2. Introduction

A food contact material is any material or article that is intended to come into contact with food as defined in regulation 1935/2004 (Commission, 2004). The safety of food contact materials must be evaluated since compounds can migrate from the materials into food. The materials must be manufactured in compliance with European regulations so that any potential transfer to foods does not raise safety concerns, change the composition of the food in an unacceptable way or have adverse effects on the taste or odour of food (Commission, 2004).

Kitchenware has been broadly studied in the context of intentionally added substances proven to migrate to food or food simulants (García Ibarra et al., 2016; Gelbke et al., 2019). According to the European Commission though, the migration of non-intentionally added substances (NIAS) should also be investigated (Commission, 2011). The study of NIAS is complex due to the fact that they can occur as impurities, reaction products, breakdown products, degradation products or oligomers (Nerin et al., 2013). As such, accurate mass spectrometry-based techniques are required to detect and identify potential NIAS, especially in the case of oligomers (Canellas et al., 2014; Hoppe et al., 2018a, b; Kappenstein et al., 2018; Rusko et al., 2020; Tsochatzis et al., 2020; Ubeda et al., 2018). However, even using such high-resolution techniques, it is not always possible to identify every NIAS present in a sample due to the complexity of the mass spectral data and the interpretation thereof. To address this, ion-mobility quadrupole time-of-flight mass spectrometry, coupled to the ultra-high performance liquid chromatography (UPLC-IM-Q/TOF) has been used for this work. This technique produces cleaner, less complex spectra, since the alignment of precursors and fragment ions is based on both retention time and drift time. Moreover, in addition to accurate m/z values of precursor and fragmentation ions, ion-mobility mass spectrometry enables the determination of a Collision Cross Section (CCS) value, which reflects the size and shapes of each molecule. Therefore, the use of this technique can provide more information than traditional accurate mass spectrometry such as traditional time-of-flight or Orbitrap techniques. UPLC-IM-Q/TOF analysis has previously been used to successfully identify NIAS migrating from several different food contact materials

(Canellas et al., 2019; Canellas et al., 2020; Vera et al., 2019).

Once the NIAS have been identified, a risk assessment must be performed to ensure that there is no risk to human health due to their migration into food. Food simulants are used as substitutes for food due to the simplification of the chemical analysis they provide. Food simulants vary in terms of their chemical properties in order to simulate different food types, for example hydrophilic, lipophilic or amphiphillic. In general, the use of food simulants provides a good approximation to the actual migration into foods. However, for some compounds, the migration to simulants can overestimate or underestimate the extent of migration to food due to the fact that temperatures, time and the matrix affect the partition and diffusion coefficients responsible for migration mechanism (Elizalde et al., 2020; Martínez-López et al., 2018; Stärker and Welle, 2019). Although migration studies from food contact material to foodstuff require further analytical developments, they can still be used to confirm the compliance of a product for intended use, and determine the suitability of a given simulant for a migrant or a group of migrants.

In the present study, identification and quantification of oligomers migrating from kitchenware is studied using ion mobility-quadrupole time of flight mass spectrometry coupled to the ultra-high pressure liquid chromatography. The individual oligomers were isolated and purified to be used as pure standards to confirm the identification and to quantify the migration. Migration to food simulants and to several foods has been studied, and the risk associated with the migration to both food simulants and foodstuffs has been assessed.

3. Materials and methods

3.1. Reagents

Ethanol and methanol of high performance liquid chromatography (HPLC) grade were supplied by Scharlau Chemie S.A (Sentmenat, Spain). Glacial acetic acid was supplied by Scharlau Chemie S.A (Sentmenat, Spain). Caprolactam was purchased from Sigma-Aldrich Química S.A (Madrid, Spain). Oligomers from polyamide 6 (PA6) and

polyamide 66 (PA 66) were isolated from PA6 pellets and PA 66 pellets, respectively, which were purchased from Sigma-Aldrich Química S.A (Madrid, Spain). 6 g and 200 mg OASIS HBL cartridges were supplied by Waters (Spain).

3.2. Oligomer isolation

The oligomers were isolated following the method developed by Abe et al. (Abe et al., 2016) with some minor modifications. 10 g of pure PA 6 and PA66 were extracted with methanol (50 mL) at 40 ºC overnight, and the residue was subsequently removed by filtration. The filtrates and washes were evaporated using a rotary evaporator and the residue was dissolved in 10% methanol (v/v). The solution was passed through 6 g OASIS HLB solid phase extraction cartridges. 50 mL of 10, 20, 30, 40, and 50% methanol (v/v) were then passed through the cartridge and each collected separately. The monomers and oligomers in each fraction were determined using ultra-high performance liquid chromatography and triple quadrupole mass spectrometry (UPLC-TQ-MS) and the solvent was evaporated. The residue remaining from each fraction was re-dissolved in ethanol 10% and passed through a 200 mg OASIS HLB cartridge. Subsequently, 50 mL of 10, 20, 30, 40, and 50% methanol (v/v) was passed through a 200 mg OASIS HLB cartridges and 10 mL of each sample was collected. Isolated monomers and oligomers in each fraction were determined by UPLC-TQ-MS. These steps were repeated for each fraction until pure PA cyclic monomers and oligomers were obtained. The concentration of each oligomer was then calculated using a gravimetric method.

To separate caprolactam and the PA6 dimer, caprolactam was hydrolyzed following the method developed by Abe. et al. (Abe et al., 2016).

3.3. Migration samples

Four kitchenware utensils, each belonging to a different brand, were purchased from a local market; two spatulas, which were labelled as spatula 1 and spatula 2 and two ladles, labelled as ladle 1 and ladle 2 were tested.

3.4. Food used for migration studies

Cans of beans, “Fabada Litoral”, (8.3% fat content), chicken soup, “Caldo casero de pollo Gallina Blanca”, (0.3% w/w of fat content), whole milk, “Leche Pascual”, (3% w/w of fat content) and sunflower oil, “Koipesol”, (100% w/w of fat content) were purchased from a local market.

3.5. Oligomer extraction from food matrices

Two methods were followed for the extraction of oligomers from sunflower oil and food.

The first method was a slight variant of that developed by Heimrich 2015 (Heimrich et al., 2015). The method consisted of taking an aliquot of 50 g of homogenized food, transferring it into a separation funnel and adding 50 mL of hexane to dissolve the fat content. Cyclic oligomers were extracted three times using 25 mL acetonitrile/water 1:1 (v/v). The combined acetonitrile fractions were washed with another fraction of 50 mL hexane and, together with the extract, evaporated in a rotary evaporator and subsequently re-dissolved in 50 mL of 10% methanol (v/v). This solution was then passed through 200 mg OASIS HLB solid phase extraction (SPE) cartridges which had been activated with methanol and water. 30 mL of methanol was required to ensure the elution of all oligomers. Recovery was determined by spiking the oil and food with the oligomers previously isolated.

For the second procedure, an aliquot of 5 g of homogenized food was collected in a separation funnel. 50 mL of hexane was added to dissolve the fat and subsequently 50 mL of methanol was added. The resulting solution was shaken until it separated into two phases. The methanol phase was collected, evaporated in a rotary evaporator and then dissolved in 50 mL of 10% methanol (v/v). The solution was passed through the 200 mg OASIS HLB SPE cartridges that had previously been activated with methanol and water. 30 mL of methanol was required to ensure the elution of all oligomers. Recovery was determined by spiking the oil and the food with the mixture of oligomers previously isolated. Comparing the results from the two extraction methods, the second

method was selected for the extraction of oligomers from beans, chicken soup and whole milk samples.

3.6. Migration assays

Migration assays were performed following the guidelines set out in the BfR document (2019, 2019). In the document the migration value is determined to be 5 mg/kg of food/day (group migration value). This is based on the available toxicological data for the group of PA 6 oligomers with n = 2 to 8 (dimer to octamer) and PA 6,6 oligomers with n = 1 to 4 (monomer to tetramer). Based on the finding in the BfR report, the 5 mg/kg value is considered as the SML. The test conditions used were 30 min or 2 h at 100 ºC which agree with those published by (Beldi et al., 2021). Therefore, to compare the migration results presented here with the SML established in the BfR report, the samples were analysed as follows. The samples were dispensed into DURAN® glass beakers together with 1L of simulant solution at a temperature of 100 ºC for 0.5 h. An area of 12 dm^2 of the spatulas and 9 dm^2 of ladles were immersed in the simulants. The same sample size was used for the migration assays of beans, soup, milk and sunflower oil. The temperature was monitored in water and food experiments using two thermometers inside the Duran beaker. In the analysis of other simulants, the solution was preheated to the temperature used in the test and dispensed into beaker with the sample inside a preheated oven. A laboratory oven with an internal thermometer and accurate temperature control was used. When using simulants containing ethanol, evaporation occurs during the migration test and if the test is performed in a glass beaker. Therefore, for the simulants containing ethanol 10%, ethanol 50% and ethanol 95% (v/v), the samples were cut into 1 × 5 cm pieces and placed into pressure resistant tubes filled with 41.6 g of simulant, thereby ensuring the same ratio of spatula to simulant (area/w) to that in the beakers. Preheated ethanol 10% (v/v) was placed into the pressure resistant tubes, the spatula pieces were introduced and the tube was closed and placed in the oven at a temperature of 100 ºC for 0.5 h. The same procedure was followed in ethanol 50% and ethanol 95% (v/v) with the exception that the oven temperature was held at 60 ºC for 2.5 h. Three replicates of each test were carried out and blanks were also prepared for both the simulants and the food.

3.7. Ultra-high-pressure-liquid chromatography coupled to an ion mobility-quadrupole time of flight analyzer (UPLC-IM-Q/TOF)

Screening analyses were carried out using an Acquity™ UPLC chromatography system coupled to an electrospray interface (ESI) and VION® ion mobility-quadrupole time of flight (IMS-Q-TOF) mass spectrometer, from Waters (Manchester, UK). A UPLC BEH C18 column of 1.7 μm particle size (2.1 × 100 mm) was used with a flow rate of 0.3 mL/min and a column temperature of 35 ºC. The mobile phases were water (phase A) and methanol (phase B), both with 0.1% formic acid. The gradient used was 95% A to 100% B after 13 min, with 2 min of re-equilibration to initial conditions. The volume of sample injected was 5 μL. The electrospray interface (ESI) was used in positive ionization, sensitivity mode with a capillary voltage of 3 kV and a sampling cone of 30 V. The temperatures used were 120 ºC and 500 ºC for source block and desolvation gas, respectively, and the desolvation gas flow rate was 800 L h^{-1}. The system was calibrated and data were acquired in the range 50-1000 *m/z*. Leucine-Enkephalin $[M + H]^+$, *m/z* 556.2765, was used as the lock-mass compound for real time mass correction. A collision energy ramp of 20 to 40 V was applied with argon used as the collision gas. Nitrogen was used as the mobility gas. The acquisition was set to high definition mass spectrometry (HDMSE) mode, with a 0.1 s scan time. The ion mobility resolving power was ~ 20 Ω/ΔΩ FWHM. Data acquisition and processing were carried out using UNIFI v.1.9.4 software.

Table III-5.1. Compounds detected via UPLC-IM-Q/TOF in simulant 95% ethanol after exposure, retention time (RT), measured m/z, theoretical m/z, mass error, elemental composition and CCS detected. UPLC-TQ cone voltages (V), limits of detection and recoveries of the polyamide oligomers (percentage relative standard deviation in brackets).

Compounds	RT (min)	Experimental m/z	Theoretical m/z	Mass error (mDa)	Elemental composition	CCS ($Å^2$)	Cone voltage (V)	LOD (mg/kg)	Recovery sunflower oil	Recovery beans	Recovery chicken soup	Recovery whole milk
PA6 monomer	2.50	113.0841	113.0841	0.0	$C_6H_{12}NO$	127.88	30	0.031	98 ± 3.0	78 ± 5.5	96 ± 3.5	88 ± 5.6
PA6 dimer	2.29	227.1753	227.1760	-0.7	$C_{12}H_{23}N_2O_2$	155.44	30	0.062	100 ± 5.6	95 ± 7.6	106 ± 2.6	90 ± 5.2
PA6 trimer	3.10	340.2593	340.2600	-0.7	$C_{18}H_{34}N_3O_3$	182.05	30	0.110	99 ± 7.1	85 ± 8.1	102 ± 5.1	85 ± 4.8
PA6 tetramer	3.66	453.3436	453.3441	-0.5	$C_{24}H_{45}N_4O_4$	209.71	30	0.082	87 ± 2.6	80 ± 9.2	98 ± 1.6	81 ± 8.6
PA6 pentamer	4.03	566.4269	566.4281	-1.2	$C_{30}H_{56}N_5O_5$	237.61	40	0.054	80 ± 5.0	79 ± 10.3	91 ± 2.2	93 ± 4.8
PA6 hexamer	4.33	701.4928	701.4942	-1.4	$C_{36}H_{66}N_6O_6Na$	265.41	40	0.110	78 ± 11.0	73 ±15.0	83 ± 5.0	87 ± 2.6
PA66 monomer	2.61	227.1753	227.1760	-0.7	$C_{12}H_{23}N_2O_2$	153.61	30	0.054	97 ± 5.3	98 ± 7.9	89 ± 6.9	91 ± 8.1
PA66 dimer	3.88	453.3437	453.3441	-0.4	$C_{24}H_{45}N_4O_4$	214.07	30	0.071	102 ± 5.8	95 ±8.7	100 ± 8.0	100 ± 6.3
PA66 trimer	4.55	701.4942	701.4942	-0.7	$C_{36}H_{66}N_6O_6Na$	270.69	40	0.130	89 ± 8.0	81 ± 10.0	99 ± 7.0	98 ± 1.9

3.8. Ultra-high-pressure-liquid chromatography coupled to tandem quadrupole mass spectrometer (UPLC-TQ-MS)

Quantitation analyses were carried out using a AcquityTM UPLC chromatography system coupled to an electrospray interface (ESI) and TQ detector, supplied by Waters (Manchester, UK). Same UPLC method/ conditions were used as in section 3.7. The electrospray interface (ESI) was used in positive ionization mode with a capillary voltage of 3 kV and an optimized sampling cone voltage for each single compound (Table III-5.1). The temperatures used were 120 ºC and 450 ºC for source and desolvation gas, respectively, and the desolvation gas flow was 600 L h^{-1}. Selected Ion Recording (SIR) mode was used for monitoring the precursor ions of the target analytes. MassLynx v.4.1 software was used for data acquisition and processing

4. Results and discussion

4.1. Identification of migrants

The kitchenware samples were purchased in a supermarket, therefore the additives used in the production of the polymer of which the kitchenware was composed were unknown.

UPLC-IM-Q/TOF was selected to identify the compounds migrating from the kitchenware to the food simulants. In addition to mass accuracy and fragment ion information, this technique also provides separation by means of a traveling wave ion mobility spectrometry (TWIMS). The ions, which are subjected to a constant electric field while traveling through a buffer gas, are separated based on their shape and size. The time ions take to traverse the drift cell is called the ion-mobility “drift time”. By applying a calibration, a collision cross section (CCS) value can be derived from the drift time of each compound. The CCS value is related to the three-dimensional conformation of the of the chemical structure compound.

Since, $HDMS^E$ was used for this analysis, cleaner spectra were obtained as the alignment of precursors and fragment ions is based on both retention time and drift time.

The spectra are drift time-aligned, fragments can be distinguished from background ions and can be easily assigned to the precursor ion, thus bringing additional confidence in the identification of unknowns.

The workflow for the identification of the compounds begins with molecular ion of the unknown compound in the measured low energy spectrum. The elements considered for the derivation of the elemental formula were: carbon, oxygen, hydrogen, nitrogen, chlorine, bromine, fluorine, sulfur, phosphorous, silicon and sodium. Since they are the most common elements in the molecular formulae of plastic additives. Sodium was included specifically because the compounds have a high tendency to form sodium adducts with the mobile phase. Two criteria were used to establish the elemental formula for an unknown: (1) the i-Fit, which is a measure of the goodness-of-fit of the theoretical isotope pattern of a particular elemental composition to the peaks in the measured spectrum, and (2) the mass tolerance, which was set at 3 mDa. Once the molecular formulae of each accurate mass had been determined, it was necessary to use a database of chemical compounds (http://www.chemspider.com/) to obtain a list of candidate compounds for the unknowns. An in-silico fragmentation algorithm was applied to the structure of each candidate compound to produce theoretical fragments. The mass of each theoretical fragment was then compared to masses of the ions in the measured high-energy spectrum for the corresponding unknown compound. Therefore, assignment of a compound to an unknown migrant was made on the basis of the accurate mass of the precursor and fragment ions and additionally on the accuracy of the measured isotope pattern as compared to the theoretical isotope pattern. In all cases, the mass error between measured and theoretical values for the compounds identified by UPLC-IM-Q/TOF was<3 mDa.

This workflow was applied to identify migrants from spatula 1 to the food simulants used in the migration studies. Table III-5.1 shows the compounds identified, their retention times, CCS values and measured m/z values, together with the theoretical m/z and elemental composition of the proposed candidate molecule.

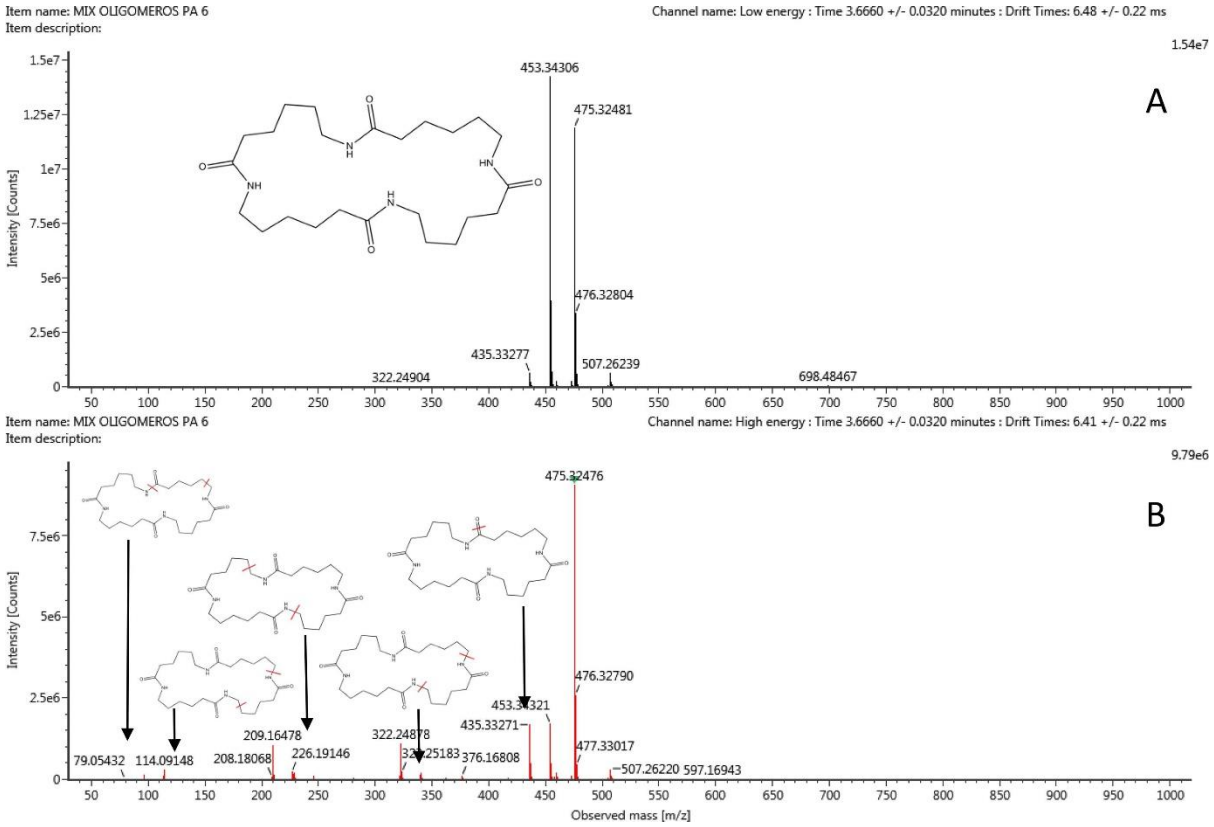


Figure III-5.1. Low-energy (A) and high energy drift time-aligned spectrum (B) of the PA6 tetramer obtained by UPLC-IM-Q/TOF.

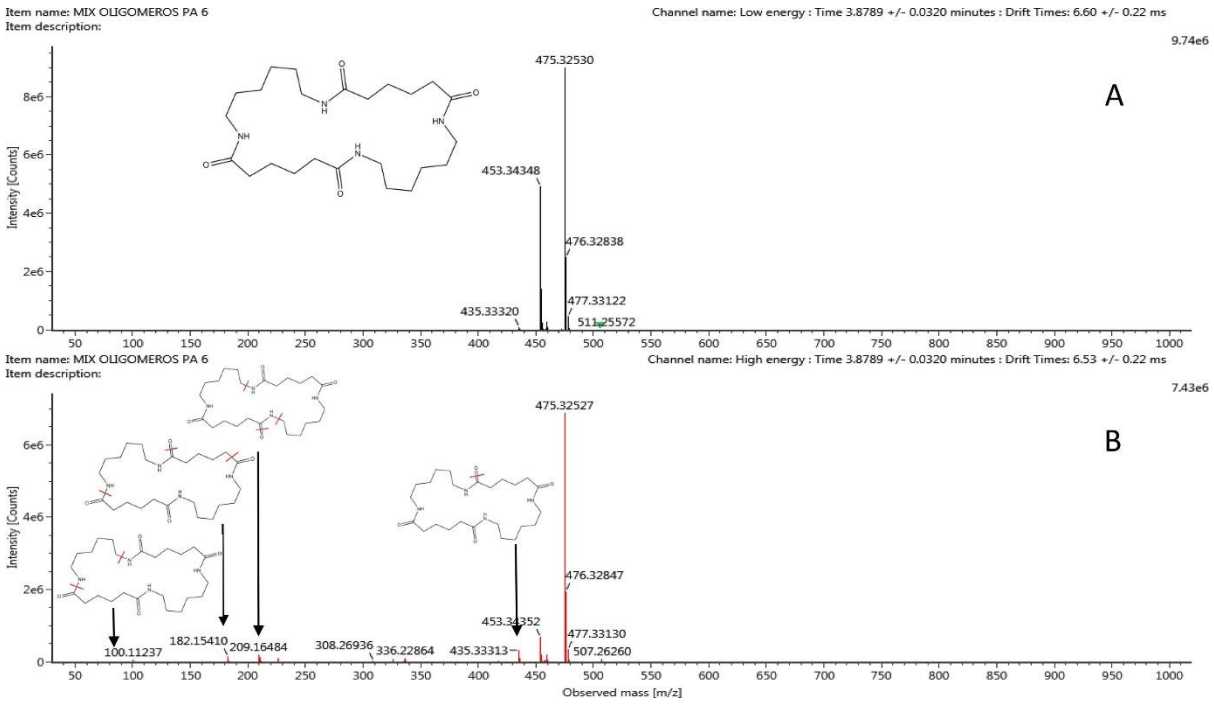


Figure III-5.2. Low-energy (A) and high energy drift time-aligned spectrum (B) of the PA66 dimer obtained by UPLC-IM-Q/TOF.

Figures III-5.1 and III-5.2 show the fragment assignment for the candidates proposed for the accurate mass *m/z* 453.344 and an elemental composition of $C_{24}H_{45}N_4O_4$. In the absence of fragmentation data, it is not possible to distinguish between the candidate compounds since they have the same elemental composition and similar CCS values (Table III-5.1). CCS has been useful to find out that not only the *m/z* of the pairs candidates was equal but also to know that the shape was equal too. Therefore, the molecules must be similar in shape and it could lead to discard other candidates in the identification process. CCS is a parameter related with the molecular shape and size, therefore it can be complementary to accurate mass and fragmentation patterns. The candidate compounds are PA6 tetramer and the PA66 dimer. The fragments measured in the high energy function for each compound are clearly different, though, and thereby enables the identification through the agreement between the experimental fragments and the predicted fragments for each candidate compound. The measured fragments 79.054, 114.091, 226.191 *m/z* correspond to theoretical fragments of the PA6 tetramer molecule and the measured fragments 100.112, 182.154 and 209.165 *m/z* correspond to the theoretical fragments of the PA66 dimer molecule. The accurate mass fragment assignment was performed for all compounds observed in the sample, leading to the identification of the oligomers shown in Table 1. Some fragments were found on several PA6 and PA66 oligomers since part of their molecular structure is common in all of them.

Table III-5.1 shows, that two candidate compounds, with the same elemental composition, were found for three of the unknown migrants. As such a definitive assignment could not be made for these migrants based on the *m/z* of the precursor ion alone. In all cases though, a definitive assignment was made by comparing the theoretical fragmentation patterns for each candidate compound with ions in the high energy data. The ability to differentiate the compounds using their fragmentation pattern was the key to the structural elucidation in all cases and was simplified by having cleaner high energy spectra afforded by ion mobility. Moreover, in order to confirm the identification, the self-generated standards (section 3.2.) were injected. Retention time and mass fragmentation of the standards match with the substances found in the

migration solutions.

All unknown migrants were identified as oligomers of PA6 and PA66 and therefore, the material of which the spatula is made is a mixture of PA6 and PA66. Since, oligomers are produced as by-products of the polymerization processes of PA and remain in the PA-based materials, they can be considered as NIAS.

According to Article 19 of Regulation (EU) No 10/2011 (Commission, 2011), no migration limits for these compounds have been established. Nevertheless, a specific migration limit (SML) of 15 mg/kg in food is specified for the PA6 monomer (caprolactam).

Additionally, the BfR document on PA oligomers (2019/BfR, 2019) states that a group migration value of <5 mg/kg in food is considered toxicologically acceptable. Therefore, migration studies were carried out to establish whether the migration from these kitchenware utensils exceeds this limit.

4.2. Oligomer extraction from food matrices

Upon identifying the migrants from the spatula, a quantitation method was developed to determine the extent of the migration. Commercial standards of the oligomers are not available and therefore they must be isolated from the polymer. The purity of each isolated monomer and oligomer was determined from their mass chromatogram peak area on the total ion chromatograms (TIC) obtained by UPLC-IM-Q/TOF (SCAN) mode ranging from *m/z* 50-1200. The peak purities were calculated by dividing the peak area of each isolated monomer or oligomer by the total peak area on the TIC. Purities were determined as follows; PA6 monomer 95%, PA6 dimer 90%, PA6 trimer 82%, PA6 tetramer 87%, PA6 heptamer 84%, PP6 hexamer 86%, PA66 monomer 91%, PA66 dimer 83% and PA66 trimer 82%. The purity of each standard was considered in all calculations. Once the oligomers were obtained, they were used to confirm the identity of the migrants by their analysis by UPLC-IM-Q/TOF. They were used for identification confirmation and quantification purposes. Moreover, CCS values were determined with the standard solutions of theses pure oligomers obtained. CCS

values were constant in the range ± 5%. It is remarkable, that a library of compounds for UPLC-MS-IMS/Q-TOF was being built by the injection of standards of these NIAS and collection of their CCS data. The migrants found in this study have been included and will be used to facilitate future screening analysis. CCS values were determined with the standard solutions of the pure oligomers previously obtained and described in this paper. CCS values were constant in the range ± 5%. A UPLC-TQ-MS method was developed to perform the quantitation of the oligomers. Each single oligomer was infused directly through the ESI the at a concentration of 10 mg/kg in order to determine the capillary voltage, desolvation nitrogen flow and cone voltages optimal for that oligomer. An efficient separation of all the PA oligomers was achieved in a single run using the UPLC BEH C18 column of 1.7 μm particle size column. Table III-5.1 shows the optimum cone voltage obtained for each compound.

Calibration curves were generated using the isolated oligomers and a commercial standard of caprolactam. Regression coefficients were at least 0.999 for all of the calibration curves. Limits of detection were calculated as follows: (LOD) = 3SD of blank/slope of calibrators, and are shown in Table III-5.1. Similar LOD were obtained for the molecules with the same mass but different structure. Moreover, the method was less sensitive for the larger oligomers than for the smaller ones. Good recoveries were found for sunflower oil ranging from 78 to 102%, chicken soup ranging from 83 to 106% and for milk ranging from 88 to 100%. The recoveries found for the extraction from the beans were lower due to the fact that it is the most complex matrix; a heterogeneous mixture of solid and liquid content, with beans and sauce (73 to 95%). The RSD % for the food matrices ranged from 1.6% to 15% depending on the food matrix and the oligomer. The worst RSD % was found for beans ranging (5.5 to 15%). This again could be due to the fact that this is the most complex matrix of those studied. Although, PA oligomer analytical methods with high sensitivity have been presented recently (Kappenstein et al., 2018; Song et al., 2018), no methods for the extraction of the oligomers from very complex food matrices have been developed. In this work a methodology for the extraction of the oligomers migrating from kitchenware has been optimized.

A mixture of the standards at a concentration of 1 mg/L was spiked into the sunflower oil in order to calculate recoveries from both methods previously described. Recoveries were calculated as follows: Recovery (%) = (Concentration found on the sunflower oil/ concentration spiked) × 100%.

Two methods were investigated for the extraction of the compounds from the food matrices. After the samples had been extracted from the food SPE was applied, using SPE cartridges as that for the isolation of the oligomers, to clean the matrix for the extraction and concentration of the oligomers. The methanol phase resulting from the extraction was evaporated in a rotary evaporator and dissolved in 50 mL of 10% methanol (v/v). Ten fractions, each of 5 mL, of the methanol solution were eluted through the SPE cartridge. Each fraction was collected and analyzed by UPLC-TQ/MS in order to determine the quantity of methanol required to elute the oligomers. Six fractions of 5 mL (30 mL) were needed to completely elute the oligomers.

For the first method, described in section 3.5, analysis by UPLC-TQ-MS revealed recoveries ranging from 78 to 103% for sunflower oil. The Table III-5.1 also indicates recoveries ranging from 78 to 102% for the second extraction methodology which is also described in section 2.5. The latter extraction method was deemed to be the most desirable due to the fact that the extraction process is cleaner and only one extraction is required. Therefore, it was more efficient than the first extraction method for this class of compounds. Subsequently, the food products selected for the study were spiked, extracted, analyzed by UPLC-TQ-MS and recoveries were obtained. Recoveries ranged from 78 to 98% for beans, 83–106% for chicken soup and 88-100% for whole milk. Employing a two steps sample preparation, with a liquid–liquid extraction followed by a SPE extraction, good recoveries were obtained.

The limits of detection obtained using this technique ranged from 0.031 to 0.110 mg/kg. This was considered sufficiently sensitive since the SML of the sum of the oligomers is 5 mg/kg.

Table III-5.2. Concentrations of PA oligomers found after the migration assays from spatula 1 in the food and food simulants (mg/kg)

Compounds	10% ethanol	3% acetic acid	50% ethanol	95% ethanol	Sunflower oil	Beans	Chicken soup	Whole milk
PA6 monomer	0.28 ± 0.02	0.37 ± 0.10	0.61 ± 0.13	0.73 ± 0.21	3.2 ± 0.3	0.32 ± 0.02	0.15 ± 0.01	0.42 ± 0.01
PA6 dimer	0.081 ± 0.001	0.11 ± 0.04	0.098 ± 0.001	0.16 ± 0.05	< 0.062	< 0.062	0.082 ± 0.013	< 0.062
PA6 trimer	0.160 ± 0.001	0.19 ± 0.03	0.230 ± 0.003	0.31 ± 0.03	< 0.110	0.24 ± 0.02	0.24 ± 0.02	0.54 ± 0.01
PA6 tetramer	< 0.082	< 0.082	< 0.082	0.28 ± 0.07	< 0.082	< 0.082	< 0.082	< 0.082
PA6 pentamer	< 0.054	0.071 ± 0.001	< 0.054	0.08 ± 0.02	< 0.052	< 0.054	1.2 ± 0.3	< 0.054
PA6 hexamer	< 0.110	< 0.110	< 0.110	0.16 ± 0.03	< 0.110	< 0.110	< 0.110	< 0.110
PA66 monomer	4.8 ± 0.4	5.5 ± 1.2	7.0 ± 0.2	12 ± 1	0.95 ± 0.31	2.2 ± 0.2	2.6 ± 0.4	2.3 ± 0.4
PA66 dimer	0.74 ± 0.03	0.78 ± 0.18	0.87 ± 0.02	1.1 ± 0.1	0.42 ± 0.02	0.31 ± 0.01	0.11 ± 0.01	0.32 ± 0.02
PA66 trimer	1.1 ± 0.1	1.1 ± 0.4	1.5 ± 1.0	2.2 ± 0.2	< 0.120	< 0.120	< 0.120	< 0.120
Sum of oligomers	7.2	8.1	10.3	16.9	4.6	3.1	4.4	3.6

4.3. Migration study

Table III-5.2 shows the migration results from spatula 1 to each of the simulants and food products studied. Results for the migration to the food simulants show higher values were obtained for ethanol 95%, with the sum of oligomers (without caprolactam) close to 16 mg/kg. The main contributor to the total migration of the oligomers was PA 66 monomer in ethanol 95% (migration concentration of 12 ± 1 mg/kg). It was also observed that the migration of low molecular weight PA 66 oligomers was higher than that for heavier PA 66 oligomers. Similar findings were obtained by Kappenstein et al.(Kappenstein et al., 2018).

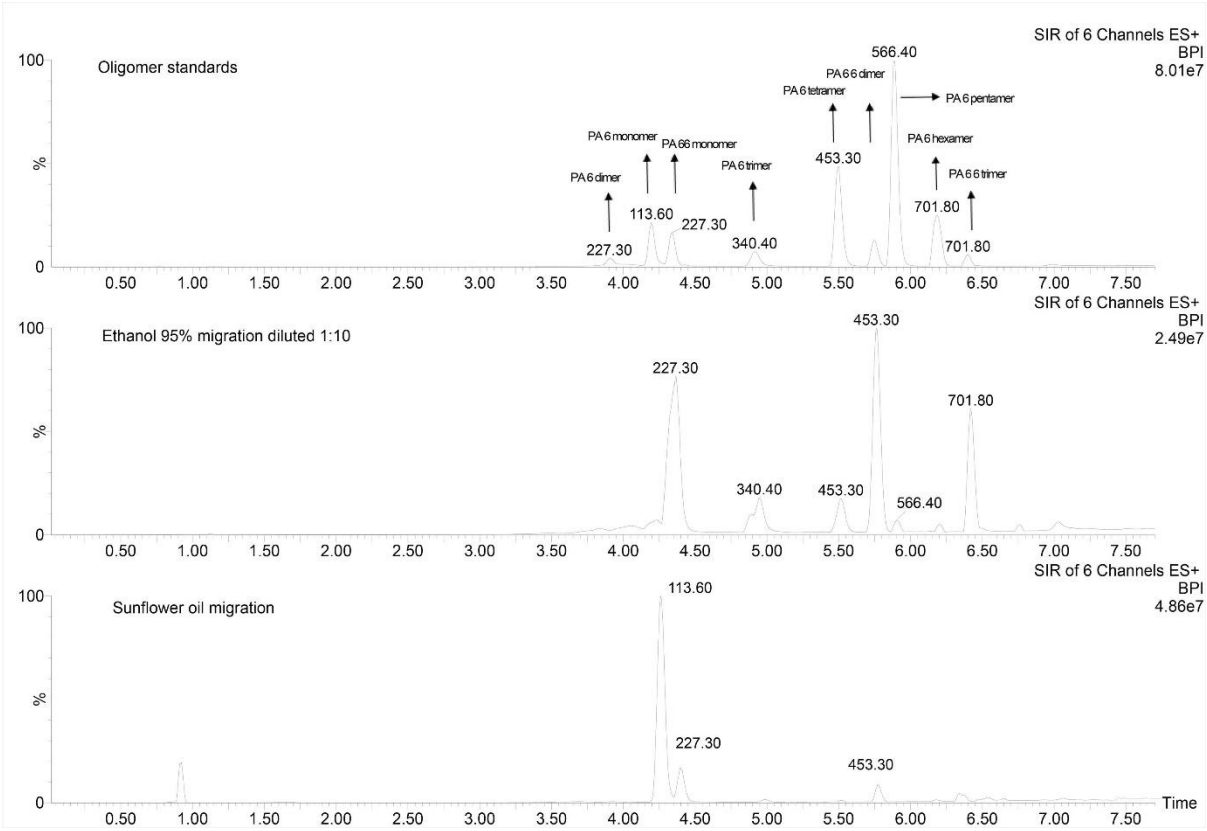


Figure III-5.3. Combined SIR chromatograms of the PA oligomers. From top to bottom: standard mix, migration to ethanol 95%, and migration to sunflower oil

Ethanol 95% is the food simulant usually used as substitute for the fat in sunflower oil. Therefore, the migration values obtained for ethanol 95% were compared to those for sunflower oil. Figure III-5.3 shows the UPLC-TQ-MS base peak ion (BPI)

chromatograms which combines the nine SIR channels monitored for the extract obtained from the migration to sunflower oil and ethanol 95%, together with the BPI chromatogram of a standard mix of PA 6 and PA 66 oligomers. The values obtained for the ethanol 95% simulant and the sunflower oil were very different. The migration of PA 6 dimer, PA 6 trimer, PA 6 tetramer, PA 6 pentamer, PA 6 hexamer and PA 66 trimer was below the limit of detection in sunflower oil. Additionally, the migration of PA66 monomer and PA 66 dimer was much lower in sunflower oil than in ethanol 95% (12 ± 1 and 1.1 ± 0.1 mg/kg respectively for ethanol 95%; and 0.95 ± 0.31 and 0.42 ± 0.02 mg/kg respectively for sunflower oil). The total concentration of the oligomers migrating to ethanol 95% was over the established limit, whereas the migration concentration was below this limit for sunflower oil. Therefore, in general, ethanol 95% overestimates the migration of PA6 and PA 66 oligomers to sunflower oil. This agrees with the work published by Heimrich et al. in which ethanol 95% was found to overestimate the migration when compared to the migration in oil (Heimrich et al., 2015), In this work, the main contribution to the migration to sunflower oil was caprolactam, which at 3.2 ± 0.3 mg/kg, and below the SML of 15 mg/kg, exceeded the migration of the compound in ethanol 95% (0.73 ± 0.21 mg/kg). This is in contrast with Abe et al. and Heimrich et al. (Abe et al., 2016; Heimrich et al., 2015), who found that the concentration in the simulant was higher than that in the oil.

The sum of the migration values obtained for the oligomers in each of the ethanol 10%, ethanol 50%, ethanol 95% and acetic acid 3% (v/v) simulants exceeded the established migration limit. The compound that contributes most to this migration is the PA 66 monomer (4.8 ± 0.4, 7.0 ± 0.2, 12 ± 1, and 5.5 ± 1.2 mg/kg respectively). In contrast to this, the migration of the PA 6 monomer, also known as caprolactam, was below the SML (15 mg/kg) for all simulants studied. Since the results of migration to ethanol 95% were inconsistent with those for the migration to sunflower oil, the migration of oligomers to the three food products was performed in order to establish migration values of the oligomers to real food. Cooked beans, milk and chicken soup were selected for the study since they are common foods (high consumption rate) and usually transferred with a spatula or a spoon when heated. The sum of the oligomers in

the three foods studied was below the established limit of 5 mg/kg (in the absence of caprolactam which has a SML 15 mg/kg). The simulants representative of beans were ethanol 10% and sunflower oil. The measured migration of the oligomers to the beans was closed to 3 mg/kg. Whilst the total oligomer migration to ethanol 10% was found to be 7 mg/kg, therefore overestimating the migration to the beans. Consequently, the oligomer migration from the spatula to the ethanol 10% as simulant exceeded the SML, whereas the migration to the real food was below the SML. In this work, it has been observed that the PA 66 monomer migrates to ethanol far more readily than it does to food with no ethanol but a high of water content. It is important to remark that, no swelling effect was observed, the spatulas maintained their structure after the migration process. Therefore, the ethanol effect on migration it is not due to swelling. This behaviour for ethanol has been shown previously, for example by Abe et al., for the migration of PA oligomers from ladles in which the migration of PA 66 monomer was an order of magnitude higher in ethanol 20% than in water (Abe et al., 2016; Heimrich et al., 2015).

The simulants used for chicken soup were ethanol 10%, acetic acid 3% and sunflower oil (with a correction factor of 1/3 according to Regulation (EU) No 10/2011 (Commission, 2011). The total oligomer migration to ethanol 10% (7 mg/kg), and to acetic acid 3% (8 mg/kg) both exceeded the SML and overestimated the migration to the chicken soup (4.3 mg/kg). Migration of caprolactam to chicken soup (0.15 mg/ kg) was below migration of this monomer to sunflower oil (3.2 mg/kg), and this again may be attributed to the lower amount of fat in the soup. Additionally, the migration of caprolactam to the soup was less than that found for the beans which have a higher level of fat (8.3% fat). Based on these findings, it is suggested that fat content is the most likely factor contributing to the migration of caprolactam. Moreover, the sum of oligomer migration of soup and oil is nearly the same but the migration of the single oligomers varies. The migration of PA66 monomer was much lower in sunflower oil than in the soup. As it was explained before, PA 66 had a very low tendency to migrate to sunflower oil when compared with food with a high content of water or ethanol-based simulants.

Finally, whole milk was studied, with the simulant for the whole milk being ethanol 50%. The simulant overestimated the oligomer migration (9.8 mg/kg) to that for the whole milk (3% fat content) which was found to be 3.3 mg/kg and below the SML. As previously demonstrated, ethanol tends to extract PA oligomers more readily than the food it is used to simulate. Therefore, its utility as a simulant for food for these compounds should be considered carefully.

In all cases, the oligomer migration from the spatula to the simulants were above the SML, however, after studying the migration to real foods, the migration level was found to be below the SML.

Table III-5.3. Concentrations of PA oligomers (mg/kg) found after the migration assays from spatula 1, 2 and ladle 1 and 2.

Compounds	Bean spatula 1	Bean spatula 2	Bean ladle 1	Bean ladle 2
PA6 monomer	0.32 ± 0.01	< 0.031	< 0.031	< 0.031
PA6 dimer	0.041 ± 0.001	< 0.062	< 0.062	< 0.062
PA6 trimer	0.24 ± 0.02	< 0.110	< 0.110	< 0.110
PA6 tetramer	< 0.082	< 0.082	< 0.082	< 0.082
PA6 pentamer	< 0.054	< 0.054	< 0.054	< 0.054
PA6 hexamer	< 0.110	< 0.110	< 0.110	< 0.110
PA66 monomer	2.2 ± 0.2	1.7 ± 0.4	2.1 ± 0.5	3.2 ± 0.3
PA66 dimer	0.31 ± 0.01	2.1 ± 0.3	1.9 ± 0.3	2.5 ± 0.4
PA66 trimer	< LOD	0.11 ± 0.25	0.091 ± 0.030	0.31 ± 0.02

To extend the sample set, three additional kitchenware utensils were purchased and the oligomer migration from each utensil to beans was measured (Table III-5.3). The aim of this study was to compare the migration from different kitchenware utensils to real food, thus improving the coverage of items that are available to consumers. It was found that the monomer and oligomers of PA6 did not migrate from spatula 2 or either of the ladles. It is possible that this is because they may only be made of PA66. PA66 monomer was the main migrant from the four utensils studied (2.2 mg/kg, 1.7 mg/kg, 2.1 mg/kg and 3.2 mg/kg for spatula 1, spatula 2, ladle 1 and ladle 2 respectively). Migration of PA 66 dimer was also observed and found to be higher from spatula 2 and ladles 1 and 2 (2.1, 1.9 and 2.5 mg/kg, respectively) than that from spatula 1 (0.31

mg/kg). The total migration of oligomers from ladle 2 was the only value found to exceed the recommended limit of 5 mg/kg.

5. Conclusions

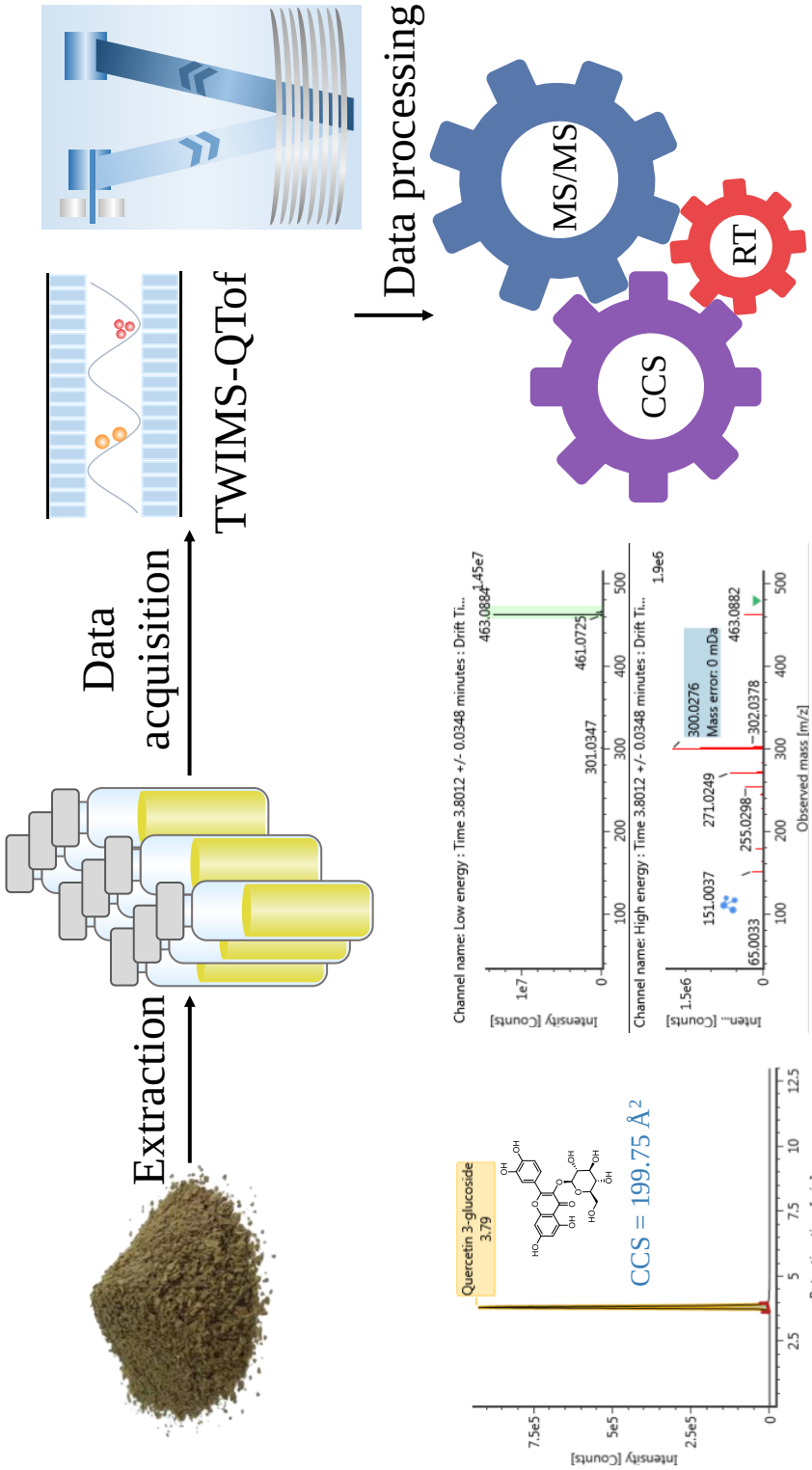
The use of hyphenated traveling wave ion mobility-high resolution mass spectrometry system has been successfully used to identify oligomers in migration extracts. The technique was able to differentiate oligomers with the same elemental composition and similar CCS values through unique fragmentation patterns. Migration of the identified compounds from different kitchenware utensils to a range of food products and simulants was performed. It was observed that the migration of PA6 and PA 66 oligomers to ethanol 95%, commonly used as substitute for sunflower oil, overestimated the migration when compared to migration to sunflower oil itself. Moreover, an extraction method of oligomers from cooked beans, soup and whole milk was developed, and the migration to simulants and real foods was studied. Migration values of the sum of oligomers obtained in the simulants were higher than those obtained for real food. Finally, ethanol generally overestimated the migration and therefore careful consideration should be given to using ethanol as a food simulant for this family of compounds.

Chapter 6

Discovery and Characterization of Phenolic Compounds in Bearberry (Arctostaphylos uva-ursi) Leaves using Liquid Chromatography Ion Mobility High-Resolution Mass Spectrometry

1. Abstract

The characterization and quantification of phenolic compounds in bearberry leaves was performed using a hyphenated ion mobility (IMS) quadrupole time-of-flight mass spectrometer. A higher identification confidence level was obtained by comparing measured collision cross section ($^{TW}CCS_{N2}$) with predicted values using a machine learning algorithm. A total of 88 compounds were identified, including 14 arbutin derivatives, 33 hydrolysable tannins, 6 flavanols, 26 flavonols and 9 saccharide derivatives and glycosidic compounds. Those most reliably reproduced in all samples were quantified against respective standards. Arbutin (47-107 mg/g), 1,2,3,4,6-pentagalloylglucose (6.6-12.9 mg/g) and quercetin 3-galactoside/quercetin 3-glucoside (2.7-5.7 mg/g) were the most abundant phenolic components in the leaves. Quinic acid and ellagic acid were also detected at relatively high concentrations. The antioxidant activity of the most abundant compounds was evaluated. A critical view of the advantages and limitations of traveling wave IMS and CCS for the discovery of natural products is given.


Extraction
Data acquisition
TWIMS-QTof
Data processing
MS/MS
RT
CCS
Quercetin 3-glucoside
3.79
CCS = 199.75 Å²
Retention time [min]
Observed mass [m/z]
Channel name: Low energy : Time 3.8012 +/- 0.0348 minutes : Drift Ti...
Channel name: High energy : Time 3.8012 +/- 0.0348 minutes : Drift Ti...
Mass error: 0 mDa

2. Introduction

Arctostaphylos uva-ursi, also known as bearberry, is a wild shrub in the genus Arctostaphylos, belonging to the Ehododendron family. This plant is mainly found in Europe, Asia and North America (Naczk et al., 2011). Over recent decades, many studies have shown that *Arctostaphylos uva-ursi* extracts possess remarkable antioxidant, antimicrobial and anti-inflammatory properties (Krzywicki, 1979; Mazza et al., 1999; Zhang et al., 2013). *Arctostaphylos uva-ursi* leaf extracts have been used as a skin-whitening agent, as well as antioxidant agents in food and food packaging formulations (Carpenter et al., 2007; Mohd Azman et al., 2016; Wrona et al., 2019). The health benefits can be attributed to bioactive components, which include arbutin, flavonols, flavanols and phenolic acids. The phenolic compounds in bearberry leaves have been tentatively characterized in previous studies (Kurkin et al., 2018; N. and V., 2013; Panusa et al., 2015; Wrona et al., 2019), in which the techniques of photodiode array (PDA) and time-of-flight mass spectrometry (Tof/MS) were used, and compound identification was mainly based on retention time (RT), *m/z* and fragmentation pathway analysis. The presence of isomers and lack of reference standards limited the complete elucidation of such components and many remain unconfirmed. Furthermore, chromatographic co-elution of structurally-similar compounds yield complex mass spectra which are extremely challenging and time-consuming to interpret. The inclusion of an additional dimension of separation can aid in increasing peak capacity and improving the identification confidence when analyzing such complex matrices (McCullagh et al., 2018; Paglia et al., 2021).

In recent years, ultrahigh performance liquid chromatography coupled to a ion mobility spectroscopy (IMS) and a quadrupole-time-of-flight mass spectrometer (UPLC-IMS-QTof) has been shown to be a powerful technique for the untargeted screening of complex samples (Bauer et al., 2018; Canellas et al., 2019; Claassen et al., 2019; Dodds et al., 2020; Qi et al., 2020). The ionized molecules pass through a drift cell under an electric field and collide with a buffer gas of neutral molecules (normally nitrogen or helium). The arrival time of the ionized molecules is related to their size, charge and shape (D'Atri et al., 2018), and as such, IMS potentially allows the

separation of isobaric and isomeric compounds that co-elute in liquid chromatography. Following ion-mobility separation, the ions undergo a collision-induced dissociation (CID) which ensures that a precursor ion and its corresponding product ions have the same arrival time. The alignment of precursor and fragment ions based in both RT and arrival time dimensions dramatically reduces the number of ions that do not belong to the candidate of interest (e.g., ions belonging to matrix interferences or co-eluting compounds) and provide much cleaner mass spectra.

Unlike the direct determination of rotationally-averaged collision cross section (CCS) in drift tube IMS (DTIMS), CCS values are derived from the arrival time of ionized species via a power-law calibration in traveling wave IMS (TWIMS), the calibration was performed by using a set of reference compounds (always polyalanine) with known CCS values. CCS represents a unique physicochemical parameter of a molecule, which is associated with the size, shape and charge of ions. Unlike arrival time, CCS is a stable parameter, which has been proven to be transferrable between different systems, different laboratories, and different experimental conditions, provided that the same calibration procedure is applied (Hinnenkamp et al., 2018; Righetti et al., 2020; Stow et al., 2017). Therefore, CCS can be treated as an additional structural descriptor for compound identification (Zhou et al., 2020).

In this work, we explore the potential benefits of including CCS values within the compound elucidation process, in order to assess the capabilities of a hyphenated technique for high throughput phenolic compound discovery. The reference CCS values were obtained either from existing literature or from open-source machine-learning prediction models. Phenolic compounds, confirmed using reference standards, were quantified to evaluate their distributions in nine bearberry leaves from different geographic origins. The antioxidant activity of these compounds was also evaluated to assess the contribution to the total antioxidant capacity of bearberry leaf derivatives.

3. Material and methods

3.1. Plants samples

Nine *Arctostaphylos uva-ursi* leaf samples were collected from three provinces of Spain. Detailed information of locations (latitude and altitude) and sampling season have been described in previous work (Wrona et al., 2019). A representative sample from each location was prepared by mixing all the leaves collected from that specific sampling point. After collection, the leaves were dried in an oven at 60 ºC for 24 h, ground into powder and stored in a refrigerator at 4 ºC until further analysis.

3.2. Chemicals

Arbutin (≥98%, CAS: 497-76-7), gallic acid (CAS: 149-91-7), ellagic acid (CAS: 476-66-4), ferulic acid (≥98%, CAS: 1135-24-6), p-coumaric acid (≥98%, CAS: 501-98-4), chlorogenic acid (≥95%, CAS: 327-97-9), (-)-quinic acid (≥98%, CAS: 77-95-2), caffeic acid (≥98%, CAS: 331-39-5), carnosol (≥98%, CAS:5957-80-2), rosmarinic acid (≥97%, 20283-92-5), sinapinic acid (≥98%, CAS: 530-59-6), hydroquinone (≥99%, CAS: 123-31-9), propyl gallate (CAS: 121-79-9), morin (≥90%, CAS: 480-16-0), oleuropein (≥98%, 32619-42-4), glycitein 7-glucoside (CAS: 40246-10-4), myricetin (≥96%, CAS: 529-44-2), quercetin (≥98%, CAS: 117-39-5), kaempferol (≥97%, CAS: 520-18-3), silibinin (≥98%, CAS: 22888-70-6), quercetin 3-rhamnoside (CAS: 522-12-3), quercetin 3-glucoside (CAS: 482-35-9), quercetin 3-galactoside (≥97%, CAS: 482-36-0), myricetin 3-rhamnoside (≥99%, CAS:17912-87-7), luteolin 7-glucoside (≥98%, CAS: 5373-11-5), (-)-epicatechin (>97%, CAS: 490-46-0), (+)-catechin (≥ 99%, CAS:154-23-4), (-)-epigallocatechin gallate (≥95%, CAS: 989-51-5), D-(+)-Glucose (≥99%, CAS: 50-99-7), (+)-lactose(≥99%, CAS: 63-42-3), D-(+)-trehalose (≥99%, CAS: 99-20-7) and formic acid (≥98.0%, CAS: 64-18-6) were obtained from Sigma-Aldrich Química S.A. (Madrid, Spain). (-)-epicatechin gallate (CAS: 1257-08-5), 1,2,3,4,6-penta-O-galloyl-β-D-glucose (CAS: 14937-32-7) were obtained from Cayman chemical company (Ann Arbor, Michigan, USA). HPLC grade ethanol (≥99.9%) and methanol (≥99.9%) were purchased from Scharlau Chemie S.A (Sentmenat, Spain). Deionized water used for all the solutions and dilutions was obtained from a Millipore Milli-QPLUS 185 system (Madrid, Spain).

Standard stock solutions of all compounds were prepared in methanol at a

concentration of 1000 mg/kg, with the exception of ellagic acid, which was dissolved in dimethyl sulfoxide. Mixed standard working solutions, of different concentrations, were produced by diluting the stock solutions with the appropriate amount of 70% ethanol (*v/v*). All the standard solutions were prepared under gravimetric control to accurately maintain the concentrations, and kept at -20 °C until analysis.

3.3. Extraction of *Arctostaphylos uva-ursi* leaves

The method used to extract the phenolic compounds from *Arctostaphylos uva-ursi* leaves, was based on that described in a previous study (Wrona et al., 2019) with some modifications. Briefly, 0.1±0.01 g of leaf powder was placed in a 20 mL glass vial, 10 g of 70% ethanol (*v/v*) was added and the vial was agitated in an ultrasound-assisted bath for 30 min at 100 W. The sample was then centrifuged for 20 min at 4000 g_n. A portion of the supernatant was filtered through a 0.22 μm nylon membrane directly into an LC vial for analysis. The remainder was stored at 4 °C.

3.4. UPLC-IMS-QTof analysis

An Acquity I-Class UPLC system (with a Flow Through Needle sample manager) coupled to a Vion IMS-QTof mass spectrometer with an electrospray ionization (ESI) source (Waters, Manchester, UK) was employed for all analyses in this study. The chromatographic separation was carried out on a CORTECS C18 UPLC column (2.1 × 100 mm, 1.6 μm particle size, 90 Å pore size) at a flow rate of 0.3 mL/min. Mobile phase consisted of water (A) and methanol (B), both with 0.1% formic acid (*v/v*). The LC gradient followed a ramp from 5 to 100% B for the first 7 minutes, was isocratic at 100% B from 7-11 minutes, returned to the initial conditions (5% B) from 11-11.1 minutes and then re-equilibrated at 5% B from 11.1-13 minutes. The injection volume was 10 μL and the sample and column temperature were maintained at 10 °C and 40 °C, respectively.

The mass spectrometer was operated in negative ion mode over the mass range 50 -1000 *m/z* with a scan time of 0.2 s. Capillary and cone voltages were -1 kV and 30 V, respectively. The source temperature was set at 120 °C and the cone gas flow rate was

50 L/h. N_2 was used as the desolvation gas at a flow rate of 800 L/min and temperature of 500 °C. Data were acquired in high definition MS^E mode ($HDMS^E$), in which the instrument switches between two energy states (low energy: 6 eV, high energy: ramp 20-40 eV) to obtain precursor and product ions information in separate spectra but within a single acquisition. Since the ion mobility cell is positioned before the collision chamber in the Vion IMS-QTof, the precursor ion and its corresponding fragments have the same arrival time. Leucine-Enkephalin ($[M-H]^-$ ion, *m/z* 554.2620) at a concentration of 100 ng/mL and infused at rate of 15 µL/min through a LockSpray channel, was used for real-time mass correction. The IMS gas flow rate was 25 mL/min with a wave velocity of 250 m/s and IMS pulse height of 45 V. Data acquisition and processing were carried out on UNIFI v.1.9.4 Scientific Information System (Waters, Manchester, UK).

Major Mix IMS/Tof Calibration Kit (ref. 186008113, Waters, Wilmslow, UK) was used for CCS calibration of Vion instruments. The detailed information about the calibration substances and their CCS values is shown in Table III-1.2. A Test-Mix solution was injected after each batch of 10 standard injections to monitor the stability of the system. Table III-1.3 shows *m/z* and CCS values for each compound in the text mix. For stability it was ensured that the variation in the *m/z* and CCS measurements were less than 5 ppm and less than 2%, respectively.

3.5. Identification workflow

Thirty-three standards, which included phenolic acids, flavonols, flavanols were analyzed using the conditions above mentioned. The *m/z* values, adducts, RT, CCS values and fragment ions were added to in an in-house database. The leaf samples were subsequently screened against the in-house library created from the phenolic standards. The criteria use to match a target to a component in the data were: RT < 0.1 min, mass error < 5 ppm, CCS delta < 2%, and for at least one fragment ion to be found.

In addition to the targeted screen, a visual inspection of the data was performed using the base peak intensity (BPI) chromatograms and a binary comparison between a blank and a sample within UNIFI to determine unidentified peaks unique to the sample.

The Multivariate Analysis tool in UNIFI was also used to determine whether there were any low-intensity compounds, unique to the samples, that were masked by the background present in full scan chromatograms. A list of markers unique to the samples, was generated with each marker uniquely defined by a *m/z* value, an RT and an arrival time. The markers were subsequently analysed using the Discovery Tool in UNIFI, which links together Elemental Composition calculations, on-line ChemSpider database searching and Fragment Match, in which substructures generated from the structure of a candidate compound are matched to ions in the high-energy data.

Elemental compositions were derived for the precursor ion of each marker using the elements C, H, O since most phenolic derivatives consist of just these three elements. Two criteria were used when calculating the elemental formula mass error (< 5 ppm) and i-Fit confidence, which is a measure of how well the theoretical isotope pattern of a proposed composition matches the measured data. Once an elemental composition for each marker had been derived, searches for compounds with that formula were performed in the ChemSpider database, automatically by UNIFI, and in PubChem manually. Structures of the compounds returned from the database searches were automatically submitted to Fragment Match which matched substructures to ions in the high energy data of each marker, providing added confidence in the compound assignment. Additionally, the MS/MS spectral databases METLIN and MoNA (Massbank of North America) were used for the interpretation of some compounds.

A total of 333 CCS values of the deprotonated ion of the phenolic compounds were mined from a range of publications (Corporation, 2020b; Gonzales et al., 2016; Masike et al., 2020; Schroeder et al., 2019; Stander et al., 2017; Venter et al., 2019) and compared to the experimental CCS measured using the Vion IMS-QTof in our lab (with a ΔCCS% threshold of 2%). For the compounds for which a CCS value was not available in the literature, a theoretical CCS was determined by AllCCS (Zhou et al., 2020) (http://allccs.zhulab.cn/) which was then compared to the measured CCS value. A tolerance of 5% between the measured and predicted CCS values was considered to be acceptable.

All of the identified compounds were assigned an identification confidence level proposed by Celma et al. (2020) and based on the experimental data (RT, CCS value, *m/z* value and fragments) acquired during the analysis. The five confidence levels are:

- Level 1, empirical data of unknowns fully match that of a reference standard.
- Level 2, an exact structure is proposed, but not cross-checked against its respective chemical standard.
- Level 3, different compounds are compatible with empirical data.
- Level 4, only molecular formula can be determined and no fragmentation information is available.
- Level 5, only exact mass is obtained.

3.6. Quantification of phenolic compounds in bearberry leaves

Fourteen phenolic compounds were quantified using an external standard calibration, where the standards at different concentrations were injected under the same conditions as the samples by means of a matrix-matched calibration curve. Phenolic compounds always show precursor ion $[2M-H]^-$ when high concentrations are injected, as such we used the sum of the responses of $[M-H]^-$ and $[2M-H]^-$ to derive the calibration curve.

The limit of detection (LOD) and limit of quantification (LOQ) of the analytes were calculated by calibration curve procedure using the slope and standard deviation of regression.

3.7. DPPH radical scavenging capacity

The IC_{50} of fourteen standards was determined using a DPPH assay, for which a series of standard solutions at different concentrations (200-1000 mg/kg for arbutin, 10-200 mg/kg for other standards) were prepared in 70% ethanol (*v/v*). Exactly 100 µL of each standard solution was mixed with 3.5 mL of the DPPH solution at a concentration

of 30 μg/g, and placed in the dark at room temperature for 30 min. The absorbance at 515 nm was measured with a UV-Vis spectrophotometer (UV-1700, Shimadzu, Japan) using 10 mm quartz cuvettes. A solution of 100 μL of 70% (*v/v*) ethanol and 3.5 mL of DPPH was used as a control blank. Inhibition percentage was calculated using the following equation:

$$I\% = \frac{A_0 - A_s}{A_0} \times 100\%$$

in which *I%* is the inhibition rate (%), A_0 is the absorbance of the control blank, and A_s is the absorbance of each sample. IC_{50} (mg/kg), the concentration at which 50% of inhibition rate is obtained, was calculated using calibration curves by plotting *I%* versus concentrations.

4. Results and discussion

Table III-6.1 shows the 88 phenolic compounds identified in bearberry leaves together with their RT, CCS value, observed *m/z*, fragment ions, molecular formula, mass error and confidence level. Fourteen arbutin derivatives, thirty-three hydrolysable tannins, six flavanols, twenty-six flavonols and nine other compounds were identified, 17 these compounds were confirmed using reference standards. Gallotannins and quercetin glycosides were the dominant compounds detected in the bearberry leaves while myricetin glycosides gave a low response. Phenolic acids, such as coumaric acid, ferulic acid, caffeic acid, chlorogenic acid were not detected in the bearberry leaves. This has also been shown to be the case in previous studies (Panusa et al., 2015; Saleem et al., 2010).

Table III-6.1. Phenolic compounds identified in bearberry leaves in negative ion mode. Retention time (RT), collision cross section (CCS).

No.	RT (min)	Measured CCS (Å^2)	Reference CCS (Å^2)	*m/z* Observed	Fragments	Molecular formula	Candidate name	Mass error (ppm)	Confidence level
1	0.74	133.67	133.22	191.0562	127.0401, 108.0215, 85.0295	$C_7H_{12}O_6$	Quinic acid	0.42	1
2	0.84	161.42		295.1031	-	$C_{11}H_{20}O_9$	Unknown	-1.24	4
3	0.84	172.71	168	331.0667	271.0456, 211.0244, 169.0139, 125.0242	$C_{13}H_{16}O_{10}$	Galloyl glucose	-1.18	3
4	0.84	206.62		493.1194	323.0988, 161.0457	$C_{19}H_{26}O_{15}$	Galloyl di-hexoside	-0.98	3
5	0.85	194.11	191	483.0774	331.0669, 169.0139	$C_{20}H_{20}O_{14}$	Digalloyl glucose	-1.19	3
6	0.86	172.13	172.74	271.0819	108.0215	$C_{12}H_{16}O_7$	Arbutin	-1.62	1
7	0.95	124.04	123.83	169.014	125.0242	$C_7H_6O_5$	Gallic acid	-1.32	1
8	1.21	188.06		389.1084	347.0979	$C_{16}H_{22}O_{11}$	Unknown	-1.40	4
9	1.24	173.49	168	331.0665	271.0451, 211.0244, 169.0138, 125.0242	$C_{13}H_{16}O_{10}$	Galloyl glucose	-1.75	3
10	1.36	170.53	168	331.0669	169.0142, 125.0242	$C_{13}H_{16}O_{10}$	Galloyl glucose	-0.63	3
11	1.52	170.88	168	331.0669	271.0461, 211.0247, 169.0142	$C_{13}H_{16}O_{10}$	Galloyl glucose	-0.63	3
12	1.57	172.33		325.0563	169.0141, 125.0242	$C_{14}H_{14}O_9$	Galloylshikimic acid	-0.59	3
13	1.67	193.78	191	483.0776	331.0669	$C_{20}H_{20}O_{14}$	Digalloyl glucose	-0.89	3
14	1.69	160.88		295.1031	-	$C_{11}H_{20}O_9$	Unknown	-1.03	4
15	1.78	170.17		315.0718	152.0113, 108.0216	$C_{13}H_{16}O_9$	Unknown	-0.10	4
16	1.79	203.65		423.0928	313.0561, 169.0140, 125.0241	$C_{19}H_{20}O_{11}$	Galloyl arbutin	-1.12	3
17	1.95	202.77		423.0928	313.0562, 169.0140	$C_{19}H_{20}O_{11}$	Galloyl arbutin	-1.12	3
18	1.95	172.19		325.0562	169.0141, 125.0242	$C_{14}H_{14}O_9$	Galloylshikimic acid	-0.89	3
19	2.03	192.99	191	483.0776	271.0459	$C_{20}H_{20}O_{14}$	Digalloyl glucose	-0.82	3
20	2.07	225.49	220	635.0887	407.0774, 289.0719	$C_{27}H_{24}O_{18}$	Trigalloyl glucose	-0.48	3
21	2.12	203.74		423.0927	313.0561, 169.0140, 125.0241	$C_{19}H_{20}O_{11}$	Galloyl arbutin	-1.43	3
22	2.18	192.89	191	483.0772	245.0453	$C_{20}H_{20}O_{14}$	Digalloyl glucose	-1.65	3
23	2.22	183.12	168.10	343.1031	135.0449	$C_{14}H_{18}O_7$	Picein	-1.29	2
24	2.25	233.13	220	635.0886	465.0671	$C_{27}H_{24}O_{18}$	Trigalloyl glucose	-0.69	3
25	2.28	300.59		951.0745	-	$C_{41}H_{28}O_{27}$	Geraniin isomer	-0.04	3

No.	RT (min)	Measured CCS (Å^2)	Reference CCS (Å^2)	*m/z* Observed	Fragments	Molecular formula	Candidate name	Mass error (ppm)	Confidence level
26	2.34	221.29	220	635.0886	483.0774, 271.0457, 169.0140	$C_{27}H_{24}O_{18}$	Trigalloyl glucose	-0.58	3
27	2.35	193.08	191	483.0775	331.0673, 271.0457, 211.0246	$C_{20}H_{20}O_{14}$	Digalloyl glucose	-1.11	3
28	2.35	156.61	157.25	289.0714	245.0817, 203.0711, 137.0242, 109.0290	$C_{15}H_{14}O_6$	(+)-Catechin	-1.39	1
29	2.39	183.77		423.0928	313.0559, 169.0140, 151,0035, 125,0241	$C_{19}H_{20}O_{11}$	Galloyl arbutin	-1.17	3
30	2.46	168.93	170.70	313.0925	273.0037, 247.0244, 151.0400, 108.0215	$C_{14}H_{18}O_8$	6-O-Acetylarbutin	-1.25	2
31	2.49	172.66		359.1344	182.0576, 151.0399	$C_{16}H_{24}O_9$	Unknown	-1.10	4
32	2.51	299.96		951.0745	933.0650, 463.0508, 300.9988, 273.0036	$C_{41}H_{28}O_{27}$	Geraniin isomer	0.00	3
33	2.55	180.53		423.0930	261.0401, 151.0035, 123.0086, 108.0215	$C_{19}H_{20}O_{11}$	Tetrahydroxy-glucopyranosyloxy-benzophenone	-0.60	3
34	2.64	249.56		729.1457	407.0773, 289.0717	$C_{37}H_{30}O_{16}$	Procyanidin gallate	-0.57	3
35	2.66	195.49		477.0667	271.0459	$C_{21}H_{18}O_{13}$	Quercetin glucuronide	-1.62	3
36	2.67	200.48	198.8	457.0773	169.0141	$C_{22}H_{18}O_{11}$	(-)-Epigallocatechin gallate	-0.80	1
37	2.68	270.27	264	787.0997	-	$C_{34}H_{28}O_{22}$	Tetragalloyl glucose	-0.26	4
38	2.69	224.53	220	635.0884	465.0669, 313.0562, 169.0141	$C_{27}H_{24}O_{18}$	Trigalloyl glucose	-0.90	3
39	2.76	221.94	222	633.0734	463.0513, 300.9989, 275.0196	$C_{27}H_{22}O_{18}$	Corilagin	0.11	2
40	2.81	282.73		881.1571	729.1438, 559.1251, 407.0774, 289.0716	$C_{44}H_{34}O_{20}$	Procyanidin di-gallate	0.07	3
41	2.82	204.29		447.1506	269.1026, 137.0242	$C_{19}H_{28}O_{12}$	Glucopyranosyl methyl arbutin	-0.54	3
42	2.82	195.59		477.0673	-	$C_{21}H_{18}O_{13}$	Quercetin glucuronide	-0.30	4
43	2.84	156.54	156.54	289.0716	245.0820, 109.0289	$C_{15}H_{14}O_6$	(-)-Epicatechin	-0.50	1
44	2.92	170.4		323.1346	203.0350, 166.0631	$C_{13}H_{24}O_9$	-Unknown	-0.51	4
45	2.93	215.24		575.1038	423.0930, 465.0670, 313.0564	$C_{26}H_{24}O_{15}$	Digalloyl arbutin	-0.70	3
46	2.93	304.37		965.0908	-	$C_{42}H_{30}O_{27}$	Unknown	0.66	4
47	3.01	266.22	264	787.1003	635.0892,617.0782, 465.0670, 447.0568, 313.0563, 169.0141	$C_{34}H_{28}O_{22}$	Tetragalloyl glucose	0.39	3
48	3.05	212.07		575.1041	423.0930, 261.0395, 151.0035	$C_{26}H_{24}O_{15}$	Digalloyl arbutin	-0.28	3
49	3.12	215.13		445.1716	327.1245, 179.0561, 161.0459	$C_{20}H_{30}O_{11}$	Unknown	0.08	3
50	3.19	198.42	196.87	441.0828	289.0719, 169.0143, 125.0244	$C_{22}H_{18}O_{10}$	(-)-Epicatechin 3-gallate	0.21	1
51	3.21	220.57		575.1045	423.0934, 169.0143	$C_{26}H_{24}O_{15}$	Digalloyl arbutin	0.51	3
52	3.28	247.3		727.1155	575.1048, 557.0946, 405.0819	$C_{33}H_{28}O_{19}$	Trigalloyl arbutin	0.48	3

No.	RT (min)	Measured CCS (Å^2)	Reference CCS (Å^2)	*m/z* Observed	Fragments	Molecular formula	Candidate name	Mass error (ppm)	Confidence level
53	3.29	195.91	200.50	437.1088	313.0564, 241.0351, 168.0062, 123.0085	$C_{20}H_{22}O_{11}$	6"-galloyl methyl arbutin	-0.31	3
54	3.30	137.63	141.00	197.0454	169.0141, 124.0163	$C_9H_{10}O_5$	Ethyl gallate	-0.54	2
55	3.30	209.43		461.1662	131.0347, 191.0563	$C_{20}H_{30}O_{12}$	Unknown	-0.58	4
56	3.35	266.49	264	787.0995	618.0821	$C_{34}H_{28}O_{22}$	Tetragalloyl glucose	-0.59	3
57	3.35	289.19	292.00	939.1108	787.0998, 769.0889, 617.0785, 599.0690, 465.0676	$C_{41}H_{32}O_{26}$	1,2,3,4,6-pentagalloyl glucose	-0.09	1
58	3.36	195.32	196.70	465.1035	313.0952, 303.0514, 285.0414, 151.0401	$C_{21}H_{22}O_{12}$	Taxifolin 3-glucoside	-0.64	2
59	3.37	205.92		493.0622	315.0150	$C_{21}H_{18}O_{14}$	Myricetin glucuronide	-0.36	3
60	3.38	193.22	190	433.0409	391.1023, 313.0952, 175.0398	$C_{19}H_{14}O_{12}$	Ellagic acid pentoside	-0.87	3
61	3.43	202.60		479.0833	316.0226, 287.0197, 271.0248,	$C_{21}H_{20}O_{13}$	Myricetin hexoside	0.39	3
62	3.51	302.27		955.1064	-	$C_{41}H_{32}O_{27}$	Unknown	0.65	4
63	3.55	222.00		615.0991	463.0879, 313.0565, 301.0346, 169.0142	$C_{28}H_{24}O_{16}$	galloyl-quercetin-hexoside	-0.14	3
64	3.56	196.65		477.1037	299.0200	$C_{22}H_{22}O_{12}$	isorhamnetin hexoside	-0.31	3
65	3.65	211.97		507.1147	315.0145	$C_{23}H_{24}O_{13}$	Syringetin hexoside	0.50	3
66	3.66	199.28		449.0726	316.0222/317.0283, 287.0197, 270.0171	$C_{20}H_{18}O_{12}$	Myricetin pentoside	0.07	3
67	3.69	192.21	190	433.0413	300.9986, 299.9914	$C_{19}H_{14}O_{12}$	Ellagic acid pentoside	0.21	3
68	3.76	210.06	208.14	447.0944	285.0401	$C_{21}H_{20}O_{11}$	Luteolin 7-glucoside	2.53	1
69	3.77	199.56	198.33	463.0884	300.0275, 271.0248, 255.0299, 243.0297	$C_{21}H_{20}O_{12}$	Quercetin 3-galactoside	0.32	1
70	3.80	199.70	199.75	463.0884	300.0276, 271.0249, 255.0298, 151.0037	$C_{21}H_{20}O_{12}$	Quercetin 3-glucoside	0.49	1
71	3.81	233.29	233.54	609.1462	463.0882, 301.0345, 271.0249, 255.0298	$C_{27}H_{30}O_{16}$	Rutin	0.10	1
72	3.83	222.49		599.1045	447.0920, 313.0564, 285.0393, 169.0144	$C_{28}H_{24}O_{15}$	Astragalin gallate	0.36	3
73	3.87	197.35		447.0567	314.0072, 299.9910, 160.016	$C_{20}H_{16}O_{12}$	Methyl ellagic acid pentoside	-0.49	3
74	3.93	151.08	149.78	300.999	283.9962, 257.0094, 229.0142, 185.0243	$C_{14}H_6O_8$	Ellagic acid	-0.01	1
75	4.06	196.78	196.50	433.0774	300.0275, 301.0344, 271.0245, 255.0295	$C_{20}H_{18}O_{11}$	Quercetin 3-arabinoside	-0.63	2
76	4.07	228.78		615.0988	463.0875, 433.0772, 301.0344	$C_{28}H_{24}O_{16}$	galloyl-quercetin-hexoside	-0.65	3
77	4.07	200.53	196.20	447.0928	301.0344, 271.0245, 255.0295, 151.0036	$C_{21}H_{20}O_{11}$	Quercetin 7-rhamnoside	-1.17	2
78	4.11	163.72	162.78	317.0306	287.0209, 151.0036, 109.0296	$C_{15}H_{10}O_8$	Myricetin	0.85	1
79	4.12	197.49		447.056	315.0142, 299.9908, 285.0394, 227.0347	$C_{20}H_{16}O_{12}$	Methyl ellagic acid pentoside	-2.03	3

No.	RT (min)	Measured CCS (Å^2)	Reference CCS (Å^2)	*m/z* Observed	Fragments	Molecular formula	Candidate name	Mass error (ppm)	Confidence level
80	4.13	180.92		349.0561	197.0453	$C_{16}H_{14}O_9$	Digallic acid ethyl ester	-1.07	3
81	4.14	230.58		593.1505	447.0925, 285.0393, 257.0430	$C_{27}H_{30}O_{15}$	Kaempferol-hexoside-rhamnoside	-1.25	3
82	4.16	200.10	199.59	447.0924	284.0323, 256.0344, 227.0347	$C_{21}H_{20}O_{11}$	Kaempferol 3-glucoside	-1.91	2
83	4.18	211.84		583.1089	463.0876, 300.0274	$C_{28}H_{24}O_{14}$	quercetin hydroxybenzoyl hexoside	-0.75	3
84	4.2	223.67		585.0888	433.0774	$C_{27}H_{22}O_{15}$	quercetin galloyl-pentoside	0.40	3
85	4.37	161	161.74	301.0347	255.0553	$C_{15}H_{10}O_7$	Morin	-2.18	1
86	4.38	195.59		417.0824	285.0399, 255.0299, 227.0349	$C_{20}H_{18}O_{10}$	Kaempferol 7-pentoside	-0.65	3
87	4.64	160.68	160.25	301.0354	271.0249, 245.0463, 151.0036	$C_{15}H_{10}O_7$	Quercetin	0.07	1
88	5.07	158.65	157.47	285.0405	267.0297, 227.0347	$C_{15}H_{10}O_6$	Kaempferol	0.23	1

The mass error and CCS variation of six of the reference compounds in the Test-Mix are shown in Figure III-6.1. It can be seen that the mass differences throughout this analysis were less than 3 ppm and CCS tolerances were within 2%, thus indicating the precision and stability of the Vion IMS-QTof.

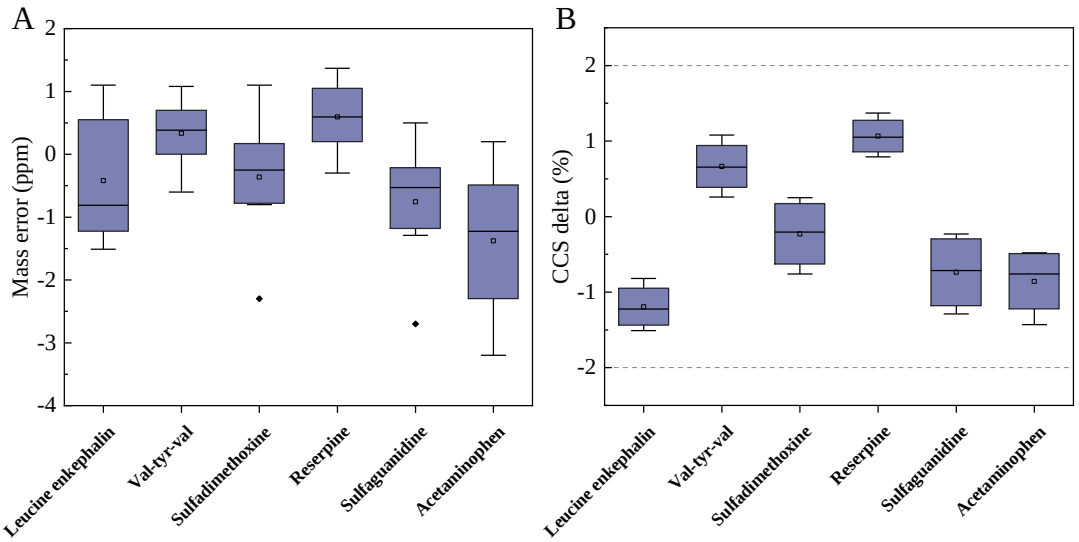


Figure III-6.1. Mass errors and CCS percentage deviations of six compounds in the Test-Mix solution measured throughout the analysis.

4.1. Arbutin and derivatives

Compound **6** was identified as arbutin by comparison with the reference standard. A fragment ion at *m/z* 109.0284 could be due to the loss of hexoside. The $[M-H]^-$ ion with *m/z* 423.0928 was detected for compounds **16**, **17**, **21** and **29**, together with fragment ions at *m/z* 313.0561 $[M-H-C_6H_6O_2]^-$, 169.0140 and 125.0241. The fragment ion with *m/z* 313.0561 corresponds to the loss of hydroquinone and the four compounds can be assigned as galloyl arbutin isomers. The high energy spectrum for compound 29, together with fragment assignments, is shown in Figure III-6.2A. Compound **23** was identified as picein by suspect screening, picein was targeted because it has been found in bearberry leaves in a previous study (Panusa et al., 2015). The precursor ion of picein was found to be $[M+HCOO]^-$ with a *m/z* value of 343.1031. The less abundant $[M-H]^-$ ion was also observed at *m/z* 297.0976 and a fragment ion at *m/z* 135.0449 $[M-H-C_6H_{10}O_5]^-$, corresponding to the loss of glucosyl group, was detected in the high energy spectrum. The $[M-H]^-$ ion was found for compound **30** with a *m/z* values of 313.0925,

and the compounds was identified as 6-O-acetylarbutin through suspect screening (N. and V., 2013). The predicted CCS values for this compound were found to be 170.70 $Å^2$, which perfectly matched with the measured CCS 168.93 $Å^2$.

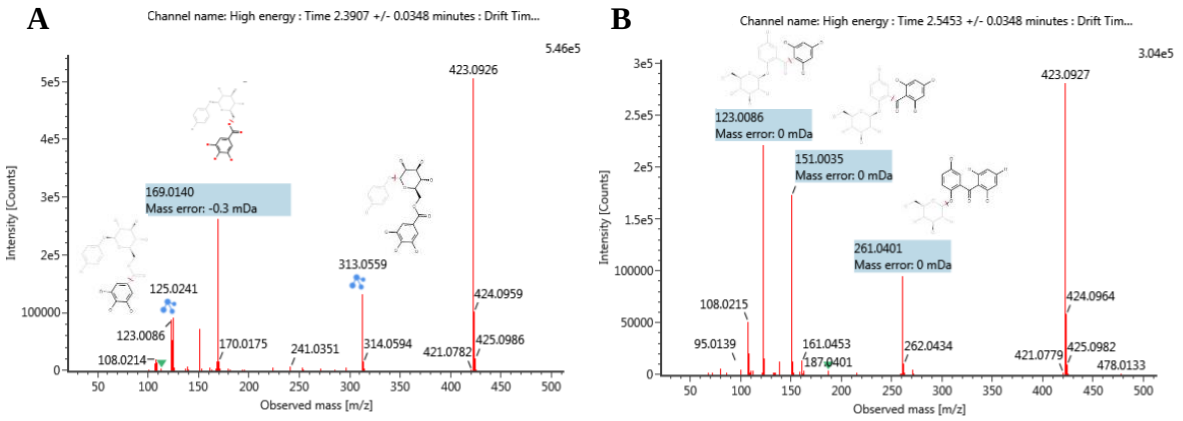


Figure III-6.2. High energy spectrum and fragment assignment for compound 29 (galloyl arbutin) and compound 33 (hypericophenonoside).

The $[M-H]^-$ ion was observed for compound **33** with a *m/z* value of 423.0930, which corresponds to the molecular formula of galloyl arbutin, however, the *m/z* values of the fragment ions detected for the compound were 261.0401 $[M-H-C_6H_{10}O_5]^-$, 151.0035, 123.0086 and 108.0215. In contrast to compounds 16, 17, 21 and 29, fragment ion *m/z* 169.0142 was not observed in the high energy spectrum (Figure III-6.2B). This indicated that a hexosyl moiety was located on one side of the structure and that the gallic acid group is not present. The compound was tentatively identified as tetrahydroxy-glucopyranosyloxy-benzophenone, among the different structures corresponding to the fragmentation information. Hypericophenonoside contains an arbutin unit in its structure, which was found in *Hypericum annulatum* previously (Kitanov and Nedialkov, 2001), and its possible fragments are shown in Figure III-6.2B.

The $[M-H]^-$ ion was detected for compound **41** at a *m/z* value of 447.1506, and this was tentatively assigned as $C_{19}H_{28}O_{12}$. Methyl arbutin has been identified in bearberry leaves in a previous study (N. and V., 2013). and the mass difference between methyl arbutin and this compound is 162.0528 Da, which corresponds to a hexosyl group. Therefore, compound 41 was tentatively identified as glucopyranosyl methyl

arbutin. The $[M-H]^-$ ion was observed for compound **45** at *m/z* of 575.1038. Three fragment ions were observed for this compound with *m/z* values 465.0670 $[M-H-C_6H_6O_2]^-$, 423.0930 $[M-H-C_7H_4O_4]^-$, 313.0564 $[M-H-C_6H_6O_2-C_7H_4O_4]^-$, corresponding to the loss of hydroquinone, the galloyl group and the two losses combined, respectively. Compound **48** and **51** have the same *m/z* value for the $[M-H]^-$ ion as compound **45**, and the three compounds were tentatively identified as digalloyl arbutin isomers.

The $[M-H]^-$ ion was detected for compound **52** at *m/z* value of 727.1155. Three fragment ions at *m/z* values 575.1048 $[M-H-C_7H_4O_4]^-$, 557.0946 $[M-H-C_7H_6O_5]^-$, 405.0819 $[M-H-C_7H_4O_4-C_7H_6O_5]^-$ were also observed and the compound was identified as trigalloyl arbutin. The $[M-H]^-$ ion was also detected for compound **53** at a *m/z* of 437.1088 and fragment ions at *m/z* values of 313.0564, 241.0351 and 168.0062 were also observed. Fragment 241.0351 could result from the loss of gallic acid and a methyl group. 6''-galloyl arbutin and methyl arbutin have previously been detected in bearberry,(N. and V., 2013) therefore compound **53** was inferred as 6''-galloyl methyl arbutin. The predicted CCS of this compound is 200.50 $Å^2$, which matched the measured CCS value of 195.91 $Å^2$ to within the 5% tolerance.

4.2. Hydrolysable tannins and derivatives

4.2.1. Gallic acid and gallotannins

The $[M-H]^-$ ion was detected for compound **4** with a *m/z* of 493.1194, together with fragment ions with *m/z* values of 323.0988 $[M-H-C_7H_60_5]^-$ and 161.0457 $[M-H-C_7H_60_5-C_6H_{10}0_5]^-$. The compound was tentatively assigned as Galloyl di-hexoside. Compound **7** was identified as gallic acid and this assignment was confirmed using a certified standard.

The $[M-H]^-$ ion, with a *m/z* of 331.0667, was detected for compounds **3, 9 10, 11** and exhibited a neutral loss of a hexose moiety to yield a fragment ion with a *m/z* of 169.0139 $[M-H-C_6H_{10}0_5]^-$. Other common fragment ions for these four species were detected with *m/z* of values of 125.0242, 211.0244 and 271.0456. The spectral data is consistent with that of 1-galloyl glucose and possible structures for the fragments have

been presented in a previous study (Sobeh et al., 2019). These four compounds were tentatively identified as monogalloyl glucose isomers. The variation of the measured CCS values for these four isomers was less than 2% and there was no available standard to be used for confirmation. The $^{TW}CCS_{N2}$ value monogalloyl glucose was detected as 168 Å^2 (Venter et al., 2019), showing a good match with measured CCS values herein (170.53-173.49 Å^2). The position of galloyl moiety on the hexose could not be inferred from the data since the different positional isomers had very similar arrival time distributions yielding almost identical CCS values.

The $[M-H]^-$ ion was detected for compounds **12** and **18** at a *m/z* of 325.0563, together with a fragment ion at *m/z* 169.0141 $[M-H-C_7H_8O_4]^-$, which corresponds to the neutral loss of a shikimate moiety. Based on these observations, the compounds were tentatively identified as galloylshikimic acid. The predicted CCS values of both 3- and 5-galloylshikimic acid is 169.30 Å^2 which is within 1.8% of the measured CCS value for these compounds of 172.33 Å^2. The precursor ion for compound **54** was found to have a *m/z* value of 197.0454. Two fragment ions were also determined for this compound with *m/z* values of 169.0141 and 124.0163, with the former fragment corresponding to the loss of an ethyl group. The compound was identified as ethyl gallate and the measured high energy spectrum was found to be consistent with that previously published (Gao et al., 2010). The $[M-H]^-$ ion for compound **80** was detected at *m/z* 349.0561 together with a fragment ion at *m/z* 197.0453 $[M-H-C_7H_4O_4]^-$, which corresponds to the loss of a galloyl group. The high energy spectrum for this compound was challenging to interpret due to its low ionization efficiency. The compound was tentatively assigned as digallic acid ethyl ester. The $[M-H]^-$ ion for compound **5**, **13**, **19, 22, 27** were observed with an *m/z* value of 483.0775. In addition, fragment ions were detected with *m/z* values of 331.0673, 271.0457 and 211.0246, with the first of these explained by the loss of a galloyl moiety. The compound 1,6-digalloyl glucose was previously isolated and identified in bearberry leaves (N. and V., 2013) and compound 27 showed a higher response than other ions. The identity of compound 27 possibly was 1,6-digalloyl glucose, however, due to the lack of reference standard, this identification was not confirmed. The $^{TW}CCS_{N2}$ of digalloyl glucose was detected as 191 Å^2 (Venter

et al., 2019), which matched well with the measured CCS values 192.89 - 194.11 $Å^2$.

The $[M-H]^-$ ion of compound **20, 24, 26, 38** was detected at *m/z* 635.0884, together with fragment ions at *m/z* 465.0669 $[M-H-C_7H_6O_5]^-$ and 313.0562 $[M-H-C_7H_6O_5-C_7H_4O_4]^-$, corresponding to the loss of gallic acid and a galloyl group, respectively. Fragments at *m/z* 169.0141 and 125.0242 also indicated the presence of gallic acid These four compounds were identified as trigalloyl glucose. Compound **38** showed a higher response, which can be 3,4,6-trigalloylglucose or 1,3,6-trigalloylglucose, because these two compounds were previously isolated in bearberry leaves (Kurkin et al., 2018; N. and V., 2013), but the reference standards were required for this identification.

The $[M-H]^-$ ion for compound **47** was detected at *m/z* 787.1003 together with fragment ions at *m/z* 635.0891 $[M-H-C_7H_4O_4]^-$, 617.0782 $[M-H-C_7H_6O_5]^-$, 465.0670 $[M-H-C_7H_6O_5-C_7H_4O_4]^-$ and 447.0568 $[M-H-2C_7H_6O_5]^-$. This compound was tentatively assigned as tetragalloyl glucose. The $[M-H]^-$ ions for compounds **37** and **56,** had *m/z* values of 787.0997, very similar to the *m/z* value for compound **47**; these two compounds were therefore assigned as tetragalloyl glucose isomers. The $^{TW}CCS_{N2}$ of tetragalloyl glucose was detected as 264 $Å^2$ (Venter et al., 2019), which matched well with the CCS values 266.49-270.27 $Å^2$ herein. The $[M-H]^-$ ion for compound **57** was detected with a *m/z* value 939.1108 and was confirmed as 1,2,3,4,6-pentagalloyl glucose using a reference standard.

4.2.2. Ellagic acid and ellagitannins

The $[M-H]^-$ ion for compound **32** was detected at *m/z* 951.0745 together with two major fragment ions with *m/z* values of 933.0650 and 300.9988, consistent with the fragments of geraniin (Sudjaroen et al., 2012). The presence of hexahydroxydiphenyl (HHDP) group was also confirmed by the observation of the characteristic fragment ion with *m/z* 300.9988. Therefore, this compound was identified as geraniin isomer. The $[M-H]^-$ ion of compound **39** was observed with a *m/z* value of 633.0734. This compound also exhibited the loss of 170 Da (gallic acid) and 162 Da (hexoside), producing two fragment ions at *m/z* 463.0513 and 300.9989, respectively, in the high energy spectrum.

The data matches well with that for corilagin presented in a previous publication (Zhu et al., 2015). Besides, corilagin was previously isolated in bearberry leaves (N. and V., 2013), and compound **39** was tentatively assigned as corilagin. The CCS value of corilagin was not found in literature, however, a $^{TW}CCS_{N2}$ of 222 $Å^2$ of galloyl-HHDP-glucose was detected in a previous study (Venter et al., 2019), which matched well with the measured value of 221.94 $Å^2$.

Compound **46** and **62** shows a $[M-H]^-$ ion with a *m/z* value of 965.0908 and 955.1064, respectively. The ionization efficiency of these two compounds was low, though, and no fragments of these two compounds were detected in the high energy spectrum. By searching their formulas in PubChem, many compounds containing unit of ellagic acid were found. As such, these two compounds were only tentatively assigned as ellagitannins and no more specific annotations were given.

The $[M-H]^-$ ion of compound **67** was detected with a *m/z* value of 433.0413. An associated fragment ion with *m/z* 300.9986 $[M-H-C_5H_8O_4]^-$ indicated the presence of ellagic acid, with the reduction in mass by 132.0428 Da corresponding to the loss of pentoside. The compound was identified as ellagic acid pentoside, with an experimental $^{TW}CCS_{N2}$ value (Venter et al., 2019) of 190 $Å^2$ within 2% of the measured value of CCS 192.21 $Å^2$. Compound **60** was also tentatively assigned as ellagic acid pentoside since the *m/z* of the $[M-H]^-$ ion for this compound was measured as 433.0409. Compound **74** was identified as ellagic acid by the reference standard. The $[M-H]^-$ ions of compounds **73** and **79** were both detected with a *m/z* value of 447.0560. Fragment ions with *m/z* values 315.0142 $[M-H-C_5H_8O_4]^-$ and 299.9908 $[M-H-C_5H_8O_4-CH_3]^-$ were detected in the high energy spectra with the mass reductions of 132.0417 and 15.0234 Da indicating the loss of pentoside and a methyl group, respectively. The two compounds were tentatively identified as methyl ellagic acid pentoside isomers.

4.3. Flavanols

The $[M-H]^-$ ions for compounds **28**, 36, 43 and 50 were identified as (+)-catechin, (-)-epigallocatechin gallate, (-)-epicatechin and (-)-epicatechin 3-gallate, respectively, by comparing with reference standards. The $[M-H]^-$ ions for compounds **34** and **40** were

detected with *m/z* values 729.1457 and 881.1571, respectively, with the mass difference of 152.0144 between these values corresponding to a galloyl group. The former compound had fragments with *m/z* values of 407.0773 and 289.0717, and was assigned as procyanidin gallate. The latter compound had fragment ions with *m/z* values of 729.1438, 711.1346, 559.1251, 541.1155, 407.0774, 289.0716 and was tentatively identified as procyanidin di-gallate.

4.4. Flavonols

The majority of flavonols exist naturally as glycosides. The aglycones, glycosyl groups and their relative position on the scaffold can affect the retention time of flavonoid glycosides. The general elution order of the aglycones was myricetin, quercetin followed by kaempferol; while the elution order of the glycosyl group was di-O-glycoside, O-galactoside, O-glucoside, O-rutinoside, O-neohesperidoside, with O-rhamnoside the last to elute. As for the influence of the sugar position, C-glycosides elute before O-glycosides and the elution order observed was 8-C-, 6-C-, 3',7-di-O-, 7-O-, 4'-O- and finally 3-O- (Abad-Garcia et al., 2009). The masses of the following glycosyl groups were also repeatedly observed: Hexoside ($C_6H_{10}O_5$, *m/z* 162.0528), pentoside ($C_5H_8O_4$, *m/z* 132.0423), rhamnoside ($C_6H_{10}O_4$, *m/z* 146.0579) and glucuronide ($C_6H_8O_6$, *m/z* 176.0473).

The $[M-H]^-$ ion for compound **59** was detected with a *m/z* value of 493.0622 and associated fragment ions, with *m/z* values of 315.0150, 151.0401 and 137.0245 were observed in the associated high-energy data. This compound was tentatively assigned as myricetin glucuronide. The $[M-H]^-$ ion was detected for compounds **35** and **42** with a *m/z* value of 447.0667, however, the high-energy spectrum was difficult to interpret due to the presence of coeluting compounds and the only fragment ion that could be reliably assigned was that with a *m/z* value of 271.0459 for compound 35. In this case, the structures of the two isomers were too similar to be separated using ion mobility. Since compound 59 had been identified as myricetin glucuronide, and taking into consideration the elution order discussed above, together with the mass spectral data, these two compounds were both tentatively assigned as isomers of quercetin

glucuronide, An experimental $^{TW}CCS_{N2}$ value (193.70 Å^2) of quercetin 3-glucuronide was reported previously,(Gonzales et al., 2016) which is similar to measured CCS (around 195.50 Å^2) of compound 35 and 42.

The $[M-H]^-$ ion for compound **58** was observed with *m/z* value 465.1035 together with fragment ions with *m/z* values 313.0952, 303.0514, 285.0414 and 151.0414 in the high energy data. These spectral data were in agreement with that for taxifolin 3-glucoside in MoNA. The identification was further confirmed by the comparison of the measured CCS value for this compound of 195.32 Å^2 with that of the predicted value of 196.70 Å^2. The $[M-H]^-$ ion detected for compound **61** had a *m/z* value of 479.0833 and three fragment ions, typical of myricetin were observed in the high energy data with *m/z* values of 316.0226, 287.0197 and 271.0248. The fragment with *m/z* 316.0226 is consistent with the mass of a radical myricetin anion, which can be produced by the loss of a hexosyl group through homolytic cleavage. This compound was identified as myricetin hexoside.

The $[M-H]^-$ ion for compound **63** was detected with a *m/z* value of 615.0991 together with fragment ions with *m/z* values of 463.0879 $[M-H-C_7H_4O_4]^-$, 313.0565 [M-H-301]$^-$, 301.0346 $[M-H-C_7H_4O_4-C_6H_{10}O_5]^-$ and 169.0142 [gallic acid-H]$^-$. The fragments and mass differences indicated the presence of quercetin aglycone, a hexosyl group and a galloyl group in the core structure. Based on these observations, the compound was assigned as galloyl-quercetin-hexoside. The observed $[M-H]^-$ ion for compound **76** also had a *m/z* value of 615.0988 but with a CCS values of 228.78 Å^2. It is tentatively suggested that this compound may be an isomer of galloyl-quercetin-hexoside. The $[M-H]^-$ ion of compound **64** was detected with a *m/z* value of 477.103 together with one fragment ion with a *m/z* value of 299.0200. This compound was tentatively assigned as isorhamnetin hexoside. The $[M-H]^-$ ion for compound **65** was observed with a *m/z* value of 507.1147 and a CCS measurement of 211.97 Å^2. A fragment ion with *m/z* 315.0145 was also observed in the high energy data. The compound was tentatively assigned as syringetin hexoside. The $[M-H]^-$ ion for compound **66** was detected with a *m/z* value of 449.0726 and a CCS value of CCS 199.28 Å^2, together with four fragment ions, typical of myricetin, at *m/z* values of

316.0222, 317.0283 [M-H-$C_5H_8O_4$]$^-$, 287.0197 and 270.0171. This compound was identified as myricetin pentoside. The [M-H]$^-$ ion for compound **68** was observed with a *m/z* value of 447.0944 together with a fragment ion at *m/z* 285.0401. This compound was identified as luteolin 7-glucoside by a reference standard.

Compound **69**, **70** and 71 were identified as quercetin 3-galactoside, quercetin 3-glucoside and rutin, respectively, by comparing with reference standards. The [M-H]$^-$ ion for compound **72** was detected with a *m/z* value of 599.1045 and a CCS value of 222.49 Å^2. Four fragment ions with *m/z* values 447.0920 [M-H-$C_7H_4O_4$]$^-$, 313.0564 [M-H-kaempferol aglycone], 285.0393 [M-H-$C_7H_4O_4$-$C_6H_{10}O_5$]$^-$ and 169.0144 were observed in the high energy data. The measurements consistent with astragalin gallate. The predicted CCS values of astragalin 2"-gallate and astragalin 6"-gallate are 220.60 and 220.20 Å^2, respectively, both within 1.1% of the measured value, therefore the position of gallate on structure cannot be deduced from the current data. The [M-H]$^-$ ion for compound **75** was observed with a *m/z* value of 433.0774 together with fragment ions with *m/z* values 300.0275, 271.0245, 255.0295 and 151.0036 in the high energy data. The fragment with *m/z* 300.0275 was associated with the loss of pentoside, while the lower mass fragments are typical of quercetin. Since avicularin (quercetin 3-arabinoside) has previously been detected in bearberry leaves (N. and V., 2013). the compound was assigned as quercetin 3-arabinoside which has a predicted CCS value of 196.50 Å^2, consistent with the measured here of 196.78 Å^2.

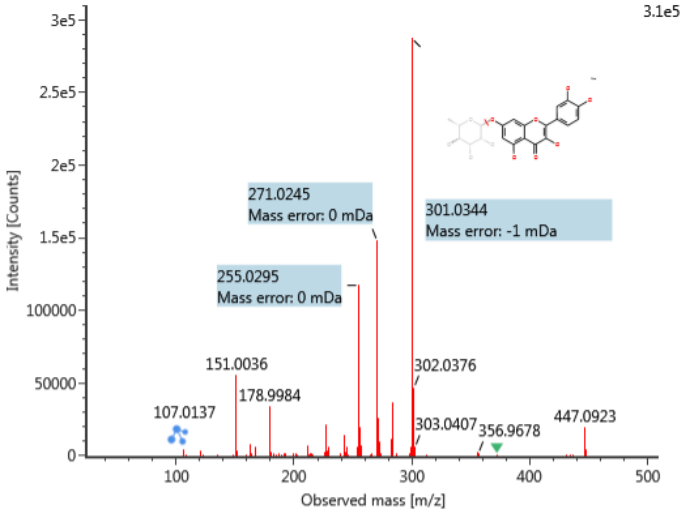


Figure III-6.3. High energy spectrum of compound 77, quercetin 7-rhamnoside, [M-H]- m/z 447.0923.

The $[M-H]^-$ ion for compound **77** was observed with a *m/z* value of 447.0928. The high energy spectrum for this compound is shown in Figure III-6.3. The most abundant fragment ion has a *m/z* value of 301.0344 $[M-H-C_6H_{10}O_4]^-$ which corresponds to the loss of rhamnoside. Additional fragments are observed with *m/z* values 271.0245, 255.0295 and 151.0036, which are typical of quercetin aglycone. As such the compound was initially identified as quercetin rhamnoside. A certified standard of quercetin 3-rhamnoside was injected to confirm the identification, however the most abundant fragment ion observed for the standard had a *m/z* of 300.0272, which corresponds to the radical quercetin aglycone. The glycosylation position can affect the relative abundance of radical aglycone to the aglycone and 3-O-glycoside can facilitate the formation of a radical aglycone (Ablajan et al., 2006). The high energy spectrum of compound 77 was subsequently compared to the MS/MS spectrum of quercetin 7-rhamnoside in MoNA with a high degree of correspondence between the two spectra obtained. Additionally, the deviation of 2.8% was observed between the measured CCS values of compound 77 (200.53 $Å^2$) and quercetin 3-rhamnoside (194.98 $Å^2$). Based on the fact that all the compounds identified at level 1 showed <2% CCS deviations with reference standards, compound 77 was not assigned as quercetin 3-rhamonoside. Generally, sugar moiety at 7-O position would lead to a more extended structure compared to the same sugar moiety at 3-O position of flavonols. For example, kaempferol 7-glucoside showed a higher experimental CCS value than kaempferol 3-glucoside (208.30 versus 199.59 $Å^2$) (Schroeder et al., 2019). Taking all of this information together, compound 77 was identified as quercetin 7-rhamnoside.

Compound **78**, **85**, **87** and **88** were identified as myricetin, morin, quercetin and kaempferol, respectively, by comparing the RT, accurate mass and CCS values with those obtained from certified standards. The $[M-H]^-$ ion for compound **81** was observed with a *m/z* value of 593.1505 together with fragment ions with *m/z* values 447.0925 $[M-H-C_6H_{10}O_4]^-$, 285.0393 $[M-H-C_6H_{10}O_4-C_6H_{10}O_5]^-$ and 257.0430 in the high energy data. The fragmentation information indicated the presence of kaempferol aglycone,

rhamnoside and hexoside in the structure. Based on these observations the compound could be kaempferol-hexoside-rhamnoside. The measured $^{TW}CCS_{N2}$ value of kaempferol 7-neohesperidoside of 227.65 Å^2 was found in a public CCS database (Corporation, 2020b). which showed a 1.3% deviation with the CCS value of compound 81 (230.58 Å^2). However, the measurement of a reference standard would be required to confirm this identification.

The [M-H]$^-$ ion for compound **82** was observed with a *m/z* value of 447.0924 together with fragment ions with *m/z* values 284.0323, 256.0344 and 227.0347 in the high energy data. The high energy spectrum matched well with the spectrum of kaempferol 3-glucoside in MoNA. In addition, the experimental $^{TIMS}CCS_{N2}$ values of 199.59 Å^2 and 208.30 Å^2 were found for kaempferol 3-glucoside and kaempferol 7-glucoside in the literature (Schroeder et al., 2019), the former was very close to the measured CCS value for compound **82** of 200.10 Å^2. Hence the compound was identified as kaempferol 3-glucoside with the relatively high confidence level of 2. The [M-H]$^-$ ion for compound **83** was detected with a *m/z* value of 583.1089 together with fragment ions at *m/z* values 463.0876 [M-H-120]$^-$ and 300.0274. The reduction in mass by 120.0213 Da to yield the larger fragment corresponds to the loss of a hydroxybenzoyl group. The compound was tentatively assigned as quercetin hydroxybenzoyl hexoside. The [M-H]$^-$ ion for compound **84** was observed with a *m/z* value of 585.088 together with one fragment ion with a *m/z* of 433.0774 [M-H-$C_7H_4O_4$]$^-$ which corresponds to the loss of galloyl group. The compound was tentatively assigned as quercetin galloyl-pentoside. The [M-H]$^-$ ion of compound **86** was detected with a *m/z* value of 417.0824. Fragment ions with *m/z* values of 284.0325 and 285.0399 corresponded to the loss of pentoside. Additionally, fragments with *m/z* 255.0299 and 227.0349 indicated the presence of kaempferol and therefore this compound was initially assigned as kaempferol pentoside. The response of 285.0399 (Y_0^-) in high energy data was higher than that for 284.0325 (radical kaempferol [Y_0-H]$^{-\bullet}$), indicating that the pentoside is likely to be at the 7-O- position of kaempferol. In addition, the predicted CCS values of kaempferol 3-arabinoside and kaempferol 7-arabinoside are 189.30 Å^2 and 194.00 Å^2, respectively. The latter was closer to the measured CCS value for this compound of

195.59 Å^2. Based on these observations the compound was tentatively identified as kaempferol 7-pentoside. The specific type of glycosyl group cannot be proposed based on current data.

4.5. Organic acids and saccharide derivatives

The $[M-H]^-$ ion for Compound **1** was detected with a *m/z* value of 191.0562 and was identified as quinic acid by the reference standard. The identities of compound 2, 8, 14, 15, 31, 44, 49, 55 were not confirmed due to their limited fragmentation information, however, by searching their formula in PubChem, many candidates belonging to saccharide derivatives were found. For example, the $[M-H]^-$ ion for compound **15** was detected with a *m/z* value of 315.0718. Two fragment ions were also observed with *m/z* values of 152.0113 and 108.0216 corresponding to the loss of hexoside and CO_2. Compound 44 was observed with a $[M-H]^-$ at m/z 323.1346, from which the elemental composition $C_{13}H_{24}O_9$ was determined, many saccharide derivatives with different types of sugar moiety were found in PubChem corresponding to this formula. The identification of these kinds of compounds may need standards or IMS with higher resolving power (Ben Faleh et al., 2019; Hofmann et al., 2015).

4.6. Quantification and antioxidant properties of phenolic compounds

The equation of the calibration line and the associate coefficient of determination (R^2), linear range, LOD and LOQ for fourteen compounds are shown in Table III-6.2. With the exception of quinic acid, gallic acid and quercetin 3-glucoside, the coefficients of determination were higher than 0.999 indicating good linearity of the calibration curves. The inclusion of the response of the $[2M-H]^-$ adducts further improved the linearity of the calibration data. For example, in the case of quercetin 3-glucoside, the $[2M-H]^-/[M-H]^-$ ratios at concentrations of 0.2, 0.5, 1, 2 and 5 mg/kg were 0.8%, 1.8%, 3.2%, 5.1% and 8.4%, respectively and the coefficient of determination increased from 0.9866 to 0.9911 upon considering the response of $[2M-H]^-$. Similar improvements in R^2 were also observed for (-)-epicatechin 3-gallate, (+)-catechin and quercetin.

Table III-6.2. Calibration data for fourteen standards.

Standards	Equation	R^2	Linear range	LOD/LOQ (μg/kg)
Quinic acid	y = 89.34x + 174247	0.9958	0.5-10[a]	1.0/3.2
Gallic acid	y = 15.77x + 5723.5	0.9977	0.5-10[a]	5.5/18.3
Ellagic acid	y = 136.38x + 21424	0.9990	0.2-7.5[a]	1.3/4.2
Arbutin	y = 0.7454x + 2763.2	0.9993	5-50[a]	8.7/28.9
(+)-Catechin	y = 332.92x + 8534.7	0.9996	0.05-2[a]	0.4/1.2
(-)-Epicatechin	y = 550.8x – 204.97	0.9994	1-20[b]	0.16/0.52
(-)-Epicatechin 3-gallate	y = 538.72x + 5177.8	1.0000	0.1-2[a]	0.3/1.0
(-)-Epigallocatechin 3-gallate	y = 236.59x – 14.244	0.9998	10-200[b]	0.7/2.4
Morin	y = 922.03x – 953.47	0.9991	1-30[b]	0.1/0.3
Kaempferol	y = 3758.4x – 1374	0.9991	1-20[b]	0.1/0.3
Quercetin	y = 1555.5x + 4674.5	0.9992	10-200[b]	0.2/0.7
Quercetin 3-glucoside	y = 750.94x + 22213	0.9911	0.2-5[a]	0.1/0.2
Rutin	y = 1013.8x + 1852	1.0000	10-300[b]	0.1/0.3
1,2,3,4,6-pentagalloglucose	y = 138.68x – 194199	0.9998	2-50[a]	10.8/35.9

(a) represents mg/kg, (b) represents μg/kg.

Table III-6.3. Concentrations of fourteen compounds in nine bearberry leaves.

Compounds	Bearberry 1	Bearberry 2	Bearberry 3	Bearberry 4	Bearberry 5	Bearberry 6	Bearberry 7	Bearberry 8	Bearberry 9
Quinic acid[2]	2.30 ± 0.31^{ab}	2.04 ± 0.34^{ab}	3.36 ± 0.40^{ab}	1.41 ± 0.16^{a}	5.35 ± 0.76^{b}	2.73 ± 0.19^{ab}	2.79 ± 0.83^{ab}	5.78 ± 0.77^{b}	9.38 ± 0.84^{c}
Gallic acid[2]	0.61 ± 0.09^{a}	0.92 ± 0.05^{ab}	1.51 ± 0.11^{abc}	1.67 ± 0.14^{bc}	1.78 ± 0.24^{bc}	2.14 ± 0.18^{c}	5.32 ± 0.70^{f}	3.28 ± 0.43^{d}	2.32 ± 0.12^{c}
Ellagic acid[2]	1.27 ± 0.16^{a}	1.29 ± 0.24^{a}	1.76 ± 0.28^{a}	1.29 ± 0.14^{a}	5.21 ± 0.94^{b}	2.35 ± 0.30^{a}	8.46 ± 0.59^{d}	6.85 ± 0.22^{c}	6.02 ± 0.13^{bc}
Arbutin[2]	53.92 ± 7.19^{a}	47.29 ± 6.26^{a}	67.57 ± 7.45^{a}	71.87 ± 5.89^{a}	101.62 ± 2.63^{b}	65.90 ± 3.17^{a}	97.90 ± 5.22^{b}	102.43 ± 9.71^{b}	107.04 ± 5.48^{b}
(+)-Catechin[1]	130.93 ± 10.99^{a}	299.85 ± 9.62^{b}	278.16 ± 27.45^{b}	460.20 ± 28.75^{c}	125.11 ± 12.91^{a}	586.01 ± 34.63^{d}	624.45 ± 76.48^{d}	315.98 ± 8.06^{b}	260.22 ± 14.43^{b}
(-)-Epicatechin[1]	5.47 ± 0.34^{a}	5.75 ± 0.29^{a}	8.86 ± 0.62^{b}	8.94 ± 1.13^{b}	6.69 ± 0.92^{a}	6.30 ± 0.45^{a}	12.24 ± 0.72^{c}	8.84 ± 0.74^{b}	9.09 ± 0.34^{b}
(-)-Epicatechin 3-gallate[1]	152.26 ± 14.99^{a}	357.54 ± 43.56^{de}	309.27 ± 23.21^{bcd}	317.62 ± 21.79^{cde}	259.62 ± 17.02^{bc}	239.58 ± 24.01^{b}	341.82 ± 31.71^{de}	288.36 ± 4.83^{bcd}	384.77 ± 20.81^{e}
(-)-Epigallocatechin 3-gallate[1]	13.99 ± 0.39^{a}	28.45 ± 1.71^{bc}	32.37 ± 0.95^{bc}	52.41 ± 3.26^{d}	26.58 ± 1.18^{b}	62.34 ± 7.90^{e}	24.86 ± 1.82^{b}	23.39 ± 1.98^{b}	36.03 ± 3.07^{c}
Morin[1]	1.62 ± 0.12^{a}	1.58 ± 0.06^{a}	2.14 ± 0.13^{b}	2.08 ± 0.07^{b}	2.49 ± 0.19^{bc}	2.65 ± 0.34^{c}	nd	nd	nd
Kaempferol[1]	1.01 ± 0.06^{a}	1.08 ± 0.08^{a}	1.82 ± 0.14^{bc}	1.47 ± 0.07^{ab}	1.68 ± 0.24^{ab}	3.54 ± 0.35^{d}	5.33 ± 0.48^{f}	2.45 ± 0.28^{c}	4.43 ± 0.25^{e}
Quercetin[1]	27.21 ± 1.89^{a}	30.66 ± 3.53^{a}	48.42 ± 4.52^{a}	36.70 ± 1.71^{a}	50.01 ± 5.30^{ab}	88.02 ± 6.39^{bc}	165.63 ± 19.34^{d}	105.86 ± 8.51^{c}	146.49 ± 5.09^{d}
Quercetin 3-glucoside[2]	2.72 ± 0.19^{a}	2.79 ± 0.24^{a}	3.23 ± 0.26^{a}	3.20 ± 0.13^{a}	5.13 ± 0.40^{b}	5.18 ± 0.28^{b}	5.52 ± 0.14^{b}	5.20 ± 0.08^{b}	5.69 ± 0.14^{b}
Rutin[1]	51.24 ± 2.83^{a}	67.59 ± 6.22^{b}	63.65 ± 3.37^{ab}	51.26 ± 0.76^{a}	64.22 ± 0.22^{ab}	142.57 ± 9.76^{c}	251.04 ± 5.01^{d}	152.28 ± 1.70^{c}	56.02 ± 2.99^{ab}
1,2,3,4,6-pentagalloglucose[2]	7.23 ± 0.45^{ab}	8.28 ± 0.66^{bc}	9.63 ± 0.69^{cd}	6.66 ± 0.42^{a}	10.89 ± 0.78^{de}	6.61 ± 0.71^{a}	12.04 ± 0.31^{ef}	12.89 ± 0.10^{f}	12.75 ± 0.30^{f}

(1) mg/kg dry leaves, (2) mg/g dry leaves, nd = not detectable. a, b, c, d, e, f: mean values with different letters are significantly different within a row.

The concentrations of the fourteen calibrated compounds in bearberry leaves are shown in Table III-6.3. The most abundant compound was arbutin, with concentration higher than 5% (w/w) on average, and even higher than 10% in the samples collected at Chelva, Pina and Albarracín (Spain). This observation is consistent with that reported in a previous study (Panusa et al., 2015). Arbutin is the most important antioxidant and antibacterial compound in bearberry leaves and it can suppress the biosynthesis of melanin in human skin, so is widely used in skin-lightening agents (Shang et al., 2020). Another compound showing a relatively high content in bearberry leaves is 1,2,3,4,6-pentagalloglucose with concentrations ranging from 6.61 to 12.89 mg/g dry weight. Furthermore, quercetin 3-galactoside/quercetin 3-glucoside, quinic acid and ellagic acid showed concentrations in the ranges 2.72-5.69, 1.41-9.38, 1.27-8.46 mg/g dry weight, respectively. Compounds such as quercetin 3-galactoside/quercetin 3-glucoside have been reported to play an important role in the antioxidant properties of bearberry leaves (Song et al., 2020). (+)-catechin and (-)-epicatechin 3-gallate showed concentrations of 125-624 and 152-385 mg/kg dry weight. Flavonols, such as morin and kaempferol, showed concentrations less than 5 mg/kg dry weight.

The antioxidant contribution of the detected compounds in bearberry leaves are related to their concentrations and antioxidant capacity. Therefore, the antioxidant activities of fourteen of the compounds were determined. Quinic acid did not show a scavenging effect on the DPPH radical even at a concentration of 10 g/kg. The IC_{50} values of the other thirteen compounds in the DPPH assay are shown in Figure III-6.4. It can be seen that gallic acid, ellagic acid, (-)-epicatechin 3-gallate, (-)-epigallocatechin 3-gallate, quercetin and 1,2,3,4,6-pentagalloglucose showed a relatively high antioxidant capacity, with IC_{50} values below 100 mg/kg. (+)-catechin, (-)-epicatechin, morin and kaempferol showed IC_{50} values in the range of 100-200 mg/kg, while the scavenging effect of arbutin on the DPPH radical was found to be relatively weak (IC_{50} 1694 mg/kg).

The antioxidant contribution of the compounds was defined as the ratio between the concentration in leaves and IC_{50}, both expressed as mg/kg. 1,2,3,4,6-pentagalloglucose was found to be the highest contributor to the total antioxidant

capacity of bearberry leaves, due to its high concentration and low IC_{50}. Although arbutin exhibited a weak scavenging effect on the DPPH radical, this compound presented a relatively high antioxidant contribution due to its high concentration in bearberry leaves. Other compounds, with antioxidant contributions were higher than 10, were gallic acid, ellagic acid and quercetin 3-glucoside. One important point to highlight is that DPPH scavenging assay is based on the electron donation of antioxidants to DPPH radicals, this assay may not fully mimic the radical scavenging capacity of antioxidant in read food system, due to the existence of other types of free radicals (Shahidi and Zhong, 2015).

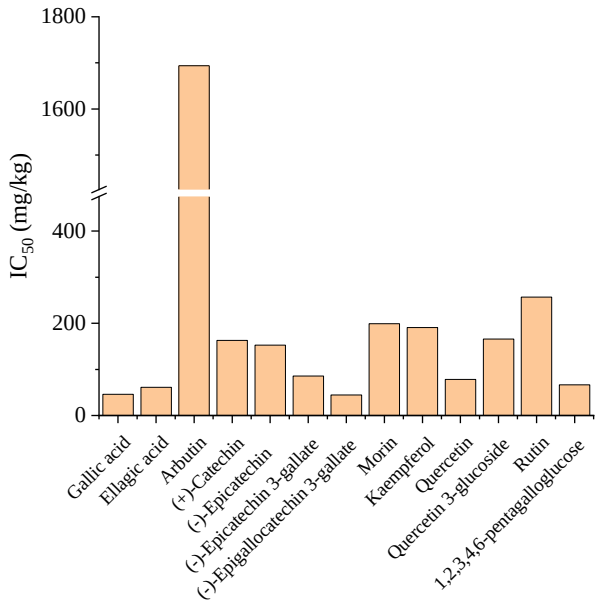


Figure III-6.4. IC50 values of thirteen standards in DPPH assay.

4.7. Advantages and limitations of IMS-HRMS in the identification of phenolics

IMS-QTof has proved to be a powerful tool for the analysis of small molecules (Belova et al., 2021; Celma et al., 2021; Celma et al., 2020; Hernandez-Mesa et al., 2018). The relatively high mass resolving power coupled to an ion mobility separation dimension increases peak capacity and can separate co-eluting compounds. This in demonstrated in Figure III-6.5, which should clear separation of the co-eluting compounds quercetin 3-glucoside and rutin. Data-independent acquisition (DIA) is a commonly mass spectrometry acquisition technique, in which all ions within a selected

m/z range are acquired and fragmented. Accurate assignment of product ions to the respective precursor ion can be challenging, since fragment ions could originate from multiple coeluting precursor ions. Arrival time alignment facilitated by IMS eliminates many of the interfering ions, resulting in cleaner high energy spectra, thus facilitating the spectral interpretation, without compromising the speed of acquisition. In addition to the arrival time alignment, CCS provided by IMS also increases the confidence of identification.

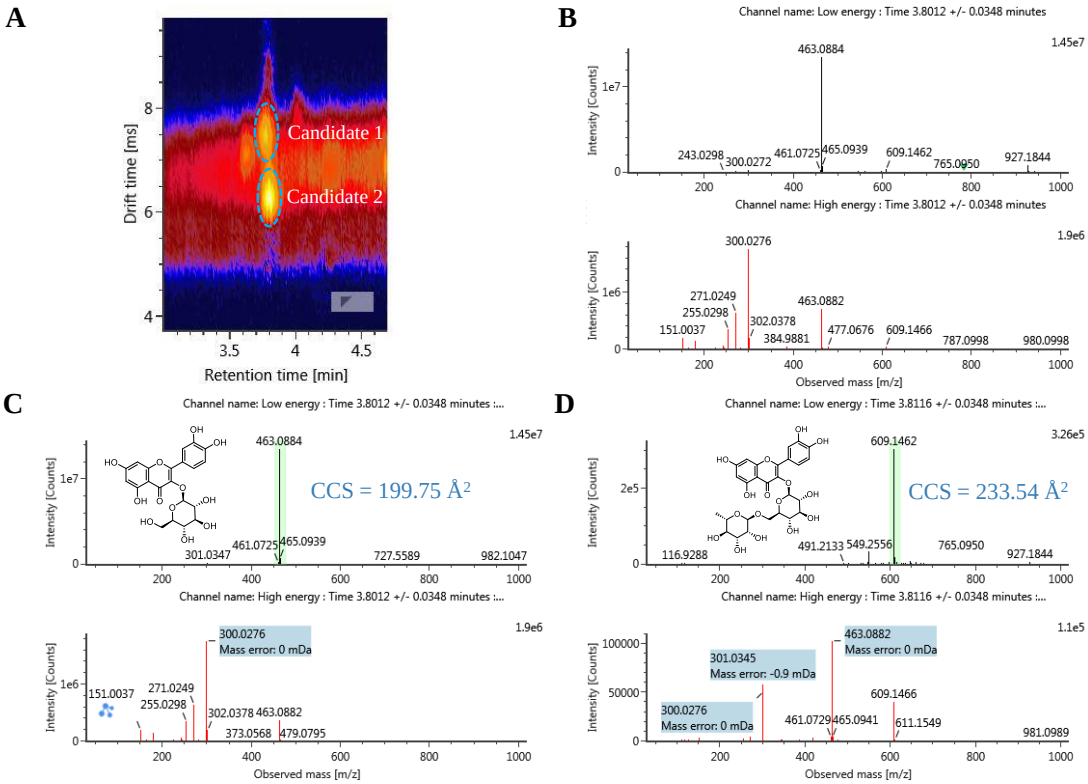


Figure III-6.5. Two components coeluting at RT 3.8 min (A), Low and high energy mass spectra at RT 3.81 min (B), clean mass spectra for quercetin 3-glucoside (C) and rutin (D) produced by arrival time alignment (resolved by IMS).

Despite these advantages, some limitations still exist. Co-eluting analytes cannot be separated if their arrival times are close to each other (similar mobility constant, *K*). This has been demonstrated here for quercetin 3-galactoside and quercetin 3-glucoside were not separated by IMS. In fact, in this study, the precursor and fragment ions were considered to be related if the difference in their arrival times was less than 0.23 ms. For example, quercetin 3-arabinoside ($[M-H]^-$ *m/z* 433.0774) and quercetin 7-

rhamnoside ([M-H]$^-$ *m/z* 447.0928) were detected with arrival times of 5.63 and 5.77 ms, respectively, and their retention times were almost identical at 4.06 and 4.07 minutes, respectively. The precursor ions for both compounds appear in the low energy spectrum even after the arrival time alignment, as shown in Figure III-6.6, and as such fragment assignment becomes challenging, especially if the two compounds share certain fragments. In order to separate these compounds by IMS, higher resolution in the ion mobility dimension is needed.

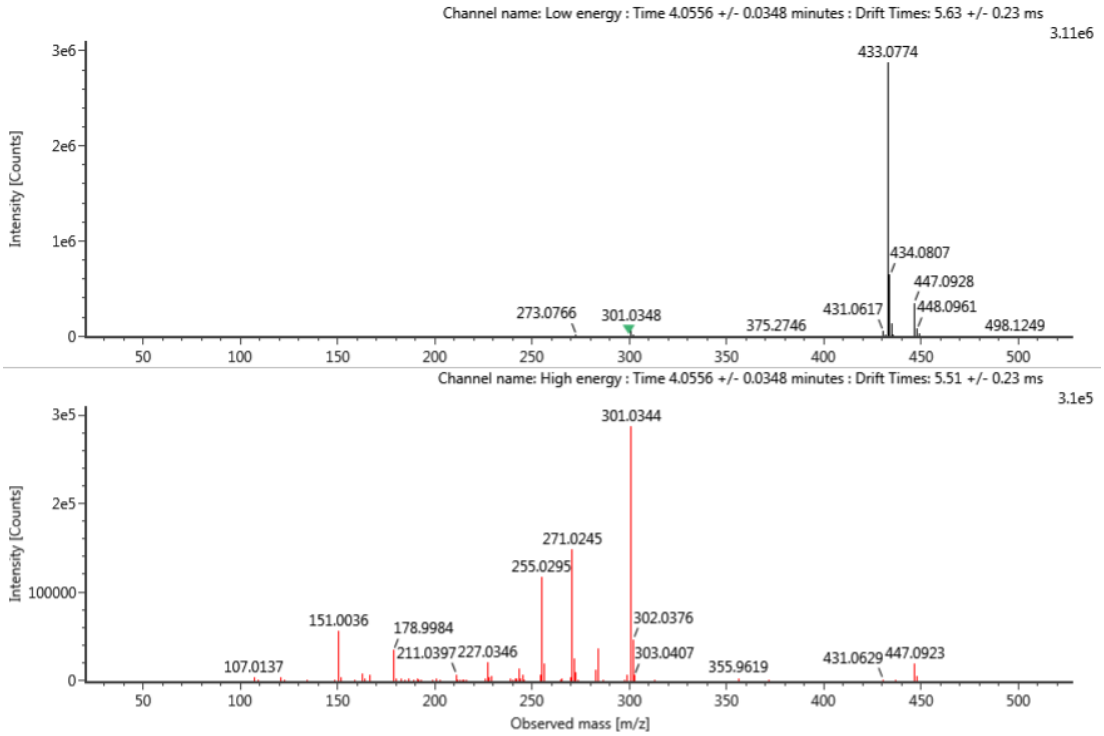


Figure III-6.6. Low energy and high energy spectrum of compound 75, quercetin 3-arabinoside, [M-H]$^-$ *m/z* 433.0774 and compound 77, quercetin 7-rhamnoside, [M-H]$^-$ *m/z* 447.0928.

An additional limitation for the confident use of CCS values, produced by IMS-QTof experiments for untargeted screening, is the lack of published empirical CCS values, it will be much easier and more reliable for the identification of compound 77 if there is an experimental CCS value available of quercetin 7-rhamnoside. As a matter of fact, many experimental CCS values of flavonols are unavailable due to the high price of commercial standards, such as quercetin 3-glucuronide and myricetin 3-glucuronide. In most cases, the distinction of isomers is based solely on predictive CCS values obtained from machine learning algorithms. Although current errors in the

predicted CCS values are within 5% for many compounds, this is still a margin of error that is too large to make a confident identification. As an example from this work, three tetragalloyl glucose isomers were detected, however CCS predictions were not sufficient to assign the position of the glycosyl moiety on the scaffold. Finally, although deviations of empirical CCS values from different TWIMS systems, and between different laboratories, have been shown to be less than 2% (Celma et al., 2020; Hernandez-Mesa et al., 2020; Righetti et al., 2020), stereoisomers can have CCS values which are within 2% of each other. For example (+)-catechin and (-)-epicatechin have CCS values of 156.61 and 156.54 $Å^2$, respectively and would require IMS resolving power capable of differentiating CCS values within 0.05% to separate them. Thus, the identification of challenging isomers still requires higher precision in the empirical of CCS values in addition to greater IMS resolution.

5. Conclusion

IMS-QTof was proved to be a powerful tool for the characterization of phenolic compounds, some structural isomer, such as quercetin 7-rhamnoside and quercetin 3-rhamnoside can be identified based on their different CCS values, the addition of ion mobility dimension makes the mass spectra clearer and more interpretable. A higher resolution of IMS and more empirical CCS values are needed to support the application of IMS in phenolic compounds characterization.

Section IV: Conclusions

1. General conclusions

According to the main objectives and research conducted during this thesis, the main conclusion of this thesis is that IMS has great potential applications in the identification of FCCs in suspect and untargeted screening analysis. Due to the wide chemical diversity of FCCs, the addition of another separation dimension: IMS, is very useful and necessary. CCS values are a new characteristic of the molecule and can be an excellent tool for identification purpose. Either experimental values or accurately predicted CCS incoporated in a wide library of compounds already detecte in both FCM and migration from them constitute a powerful advance for non-targeted analysis. In-silico prediction tools used in this work can prioritize the features obtained from IMS-QTOF platform and simplify the identification process.

Several important specific conclusions are also given below.

2. Specific conclusions

- Different types of additives present different correlation between CCS and *m/z* due to their different structural characteristics. Generally, linear-chain molecules could undergo more collisions with buffer gas and have higher CCS values, these molecules include slip agents and lubricants. The compounds containing halogens will have smaller CCS values because of lower atomic radius.

- The presence of protomers, post-ion mobility spectrometry dissociation of noncovalent clusters and inconsistent calibration of IMS cell can lead to high CCS deviations between different platform and laboratories. An improved CCS calibration approach should be performed to improve the accuracy of $^{TW}CCS_{N2}$ measurements.

- CCSondemand was found to be the most accurate CCS prediction tool for FCCs, followed by CCSbase and AllCCS. Approximately 90% of $[M+H]^+$ and $[M-H]^-$ ions show predicted errors lower than 5%.

- The presence of protomers, chemical diversity of training set can affect the accuracy

of CCS prediction. Collecting more CCS records from different chemical classes can further improve the prediction accuracy.

- The current reproducibility experimental CCS values, and deviations between $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ values can affect the quality of input data, thus affect the CCS prediction.

- The SVM-based model in conjunction with the CDK descriptors provided the better predictive performance than XGBoost algorithm and other descriptors (alvaDesc and RDKit). Approximately 93% of protonated molecules and 95% of sodiated molecules have prediction error less than 5%.

- A high structural similarity between the inputted chemicals and the training dataset can positively affect the CCS prediction accuracy, The prediction of CCS values of halogenated compounds is more accurate by the incorporation of more halogenated molecules in the training set.

- The merging of both $^{DT}CCS_{N2}$ and $^{TW}CCS_{N2}$ values in a model may bring some deviations, but meanwhile can lead to a higher chemical diversity of training dataset, thus improving the CCS prediction accuracy.

- Random forest can be used for retention time prediction, with prediction errors less than 1.22 min for testing set considering the 95^{th} confidence level.

- Approximately 75% and 29% of candidates was reduced by applying the predicted RT and CCS filters, when combining both predicted RT and CCS filter in the suspect screening, about 83% of candidates was reduced.

- The building of integrated FCM-related database is necessary for the identification of FCCs, some emerging substances associated with FCMs, such as PEG, PPG oligomers, that don't exist in CPPdb and FCCdb databases, these two databases need to be continually expanded and updated.

- The migration of PA oligomers from kitchenware to real food was below the

specified migration limit of 5 mg/kg. However, this limit was exceeded for food simulants, which therefore overestimated the oligomer migration.

- For all food simulants and real foods studied, the migration of the PA 6 monomer, caprolactam, was below the SML (15 mg/kg), the fat content is an important factor contributing to the migration of caprolactam.

- A total of 88 compounds were identified in bearberry leaves, including 14 arbutin derivatives, 33 hydrolyzable tannins, 6 flavanols, 26 flavonols, 9 saccharide derivatives, and glycosidic compounds. The most abundant compounds are arbutin (47-107 mg/kg), 1,2,3,4,6-pentagalloylglucose (6.6−12.9 mg/g), and quercetin 3-glucoside (2.7−5.7 mg/g). In addition, 1,2,3,4,6-pentagalloylglucose is considered the highest contributor to the total antioxidant capacity of bearberry leaves due to its high concentration and low IC_{50} value.

- Some structural isomers, such as quercetin 7-rhamnoside and quercetin 3-rhamnoside, kaempferol 3-glucoside and kaempferol 7-glucoside, can be identified based on their different CCS values.

- Arrival time alignment facilitated by IMS eliminates many of the interfering ions, resulting in cleaner high energy spectra, thus facilitating the spectral interpretation without compromising the speed of acquisition.

- The current CCS prediction tools have prediction errors around 5%, which cannot accurately describe the difference of CCS values for isomeric pairs.

Section V: Publications

Publications:

Song, X.C., Canellas, E., Dreolin, N., Goshawk, J., Nerin, C., 2022. A Collision Cross Section Database for Extractables and Leachables from Food Contact Materials. J. Agric. Food Chem. 70, 4457-4466. https://doi.org/10.1021/acs.jafc.2c00724

Song, X.C., Dreolin, N., Damiani, T., Canellas, E., Nerin, C., 2022. Prediction of Collision Cross Section Values: Application to Non-Intentionally Added Substance Identification in Food Contact Materials. J. Agric. Food Chem. 70, 1272-1281. https://doi.org/10.1021/acs.jafc.1c06989

Song, X.C., Dreolin, N., Canellas, E., Goshawk, J., Nerin, C., 2022. Prediction of Collision Cross Section Values for Extractables and Leachables from Plastic Products. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.2c02853

Song, X.C., Canellas, E., Goshawk, J., Dreolin, N., Nerin, C., 2022. A Workflow for the Identification of Non-Volatile Migrates from Food Contact Materials Based on Ion Mobility−High Resolution Mass Spectrometry and In-Silico Prediction Tools. J. Agric. Food Chem. Under review.

Canellas, E., Vera, P., **Song, X.C.**, Nerin, C., Goshawk, J., Dreolin, N., 2021. The use of ion mobility time-of-flight mass spectrometry to assess the migration of polyamide 6 and polyamide 66 oligomers from kitchenware utensils to food. Food Chem. 129260. https://doi.org/10.1016/j.foodchem.2021.129260

Song, X.C., Canellas, E., Dreolin, N., Nerin, C., Goshawk, J., 2021. Discovery and Characterization of Phenolic Compounds in Bearberry (Arctostaphylos uva-ursi) Leaves Using Liquid Chromatography-Ion Mobility-High-Resolution Mass Spectrometry. J. Agric. Food Chem. 69, 10856-10868. https://doi.org/10.1021/acs.jafc.1c02845

Song, X.C., Canellas, E., Asensio, E., Nerín, C., 2020. Predicting the antioxidant capacity and total phenolic content of bearberry leaves by data fusion of UV–Vis spectroscopy and UHPLC/Q-TOF-MS. Talanta 213, 120831.

https://doi.org/10.1016/j.talanta.2020.120831

Song, X.C., Canellas, E., Wrona, M., Becerril, R., Nerin, C., 2020. Comparison of two antioxidant packaging based on rosemary oleoresin and green tea extract coated on polyethylene terephthalate for extending the shelf life of minced pork meat. Food Packag. Shelf Life 26, 100588. https://doi.org/10.1016/j.fpsl.2020.100588

Song, X.C., Canellas, E., Nerin, C., 2021. Screening of volatile decay markers of minced pork by headspace-solid phase microextraction-gas chromatography-mass spectrometry and chemometrics. Food Chem. 342, 128341. https://doi.org/10.1016/j.foodchem.2020.128341

Chen, Z.-F., Lin, Q.-B., **Song, X.C.**, Chen, S., Zhong, H.-N., Nerin, C., 2020. Discrimination of virgin and recycled polyethylene based on volatile organic compounds using a headspace GC-MS coupled with chemometrics approach. Food Packag. Shelf Life 26, 100553. https://doi.org/10.1016/j.fpsl.2020.100553

Section VI: Bibliography

2019/BfR, N. (2019). Polyamide Kitchen Utensils: Keep contact with hot food as brief as possible.

Abad-Garcia, B., Berrueta, L.A., Garmon-Lobato, S., Gallo, B., Vicente, F., 2009. A general analytical strategy for the characterization of phenolic compounds in fruit juices by high-performance liquid chromatography with diode array detection coupled to electrospray ionization and triple quadrupole mass spectrometry. J. Chromatogr. A 1216, 5398-5415.

Abe, Y., Mutsuga, M., Ohno, H., Kawamura, Y., Akiyama, H., 2016. Isolation and Quantification of Polyamide Cyclic Oligomers in Kitchen Utensils and Their Migration into Various Food Simulants. Plos One 11 (7), e0159547.

Ablajan, K., Abliz, Z., Shang, X.Y., He, J.M., Zhang, R.P., Shi, J.G., 2006. Structural characterization of flavonol 3,7-di-O-glycosides and determination of the glycosylation position by using negative ion electrospray ionization tandem mass spectrometry. J Mass Spectrom. 41, 352-360.

Aćimović, M., Pezo, L., Tešević, V., Čabarkapa, I., Todosijević, M., 2020. QSRR Model for predicting retention indices of Satureja kitaibelii Wierzb. ex Heuff. essential oil composition. Ind. Crop. Prod. 154, 112752.

Alberto Lopes, J., Tsochatzis, E.D., Karasek, L., Hoekstra, E.J., Emons, H., 2021. Analysis of PBT and PET cyclic oligomers in extracts of coffee capsules and food simulants by a HPLC-UV/FLD method. Food Chem. 345, 128739.

Alin, J., Hakkarainen, M., 2011. Microwave heating causes rapid degradation of antioxidants in polypropylene packaging, leading to greatly increased specific migration to food simulants as shown by ESI-MS and GC-MS. J. Agric. Food Chem. 59, 5418-5427.

Ambrogi, V., Carfagna, C., Cerruti, P., Marturano, V., 2017. 4 - Additives in Polymers, in: Jasso-Gastinel, C.F., Kenny, J.M. (Eds.), Modification of Polymer Properties. William Andrew Publishing, pp. 87-108.

Arvanitoyannis, I.S., Bosnea, L., 2004. Migration of Substance from Food Packaging Materials to Foods. Crit. Rev. Food Sci. Nutr. 44, 63-76.

Aznar, M., Canellas, E., Nerin, C., 2009. Quantitative determination of 22 primary aromatic amines by cation-exchange solid-phase extraction and liquid chromatography-mass spectrometry. J. Chromatogr. A 1216, 5176-5181.

Aznar, M., Ubeda, S., Dreolin, N., Nerin, C., 2019. Determination of non-volatile components of a biodegradable food packaging material based on polyester and polylactic acid (PLA) and its migration to food simulants. J. Chromatogr. A 1583, 1-8.

Bauer, A., Jesus, F., Gomez Ramos, M.J., Lozano, A., Fernandez-Alba, A.R., 2019. Identification of unexpected chemical contaminants in baby food coming from plastic packaging migration by high resolution accurate mass spectrometry. Food Chem. 295, 274-288.

Bauer, A., Luetjohann, J., Hanschen, F.S., Schreiner, M., Kuballa, J., Jantzen, E., Rohn, S., 2018. Identification and characterization of pesticide metabolites in Brassica species by liquid chromatography travelling wave ion mobility quadrupole time-of-flight mass spectrometry (UPLC-TWIMS-QTOF-MS). Food Chem. 244, 292-303.

Beißmann, S., Stiftinger, M., Grabmayer, K., Wallner, G., Nitsche, D., Buchberger, W., 2013. Monitoring the degradation of stabilization systems in polypropylene during accelerated aging tests by liquid chromatography combined with atmospheric pressure chemical ionization mass spectrometry. Polym. Degrad. Stab. 98, 1655-1661.

Beldi, G., Senaldi, C., Robouch, P., Hoekstra, E., 2021. Testing conditions for kitchenware articles in contact with foodstuffs: Platics, Mentals, Silicone and Rubber. European Commission, Ispra JRC125894.

Belova, L., Caballero-Casero, N., van Nuijs, A.L.N., Covaci, A., 2021. 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. 93, 6428-6436.

Ben Faleh, A., Warnke, S., Rizzo, T.R., 2019. Combining Ultrahigh-Resolution Ion-Mobility Spectrometry with Cryogenic Infrared Spectroscopy for the Analysis of Glycan Mixtures. Anal. Chem. 91, 4876-4882.

Bijlsma, L., Bade, R., Celma, A., Mullin, L., Cleland, G., Stead, S., Hernandez, F., Sancho, J.V., 2017. Prediction of Collision Cross-Section Values for Small Molecules: Application to Pesticide Residue Analysis. Anal. Chem. 89, 6583-6589.

Bijlsma, L., Berntssen, M.H.G., Merel, S., 2019. A Refined Nontarget Workflow for the Investigation of Metabolites through the Prioritization by in Silico Prediction Tools. Anal. Chem. 91, 6321-6328.

Biryol, D., Nicolas, C.I., Wambaugh, J., Phillips, K., Isaacs, K., 2017. High-throughput dietary exposure predictions for chemical migrants from food contact substances for use in chemical prioritization. Environ. Int. 108, 185-194.

Blaženović, I., Kind, T., Sa, M.R., Ji, J., Vaniya, A., Wancewicz, B., Roberts, B.S., Torbašinović, H., Lee, T., Mehta, S.S., Showalter, M.R., Song, H., Kwok, J., Jahn, D., Kim, J., Fiehn, O., 2019. Structure Annotation of All Mass Spectra in Untargeted Metabolomics. Anal. Chem. 91, 2155-2162.

Blum, A., Behl, M., Birnbaum, L., Diamond, M.L., Phillips, A., Singla, V., Sipes, N.S., Stapleton, H.M., Venier, M., 2019. Organophosphate Ester Flame Retardants: Are They a Regrettable Substitution for Polybrominated Diphenyl Ethers? Environ. Sci. Technol. Lett. 6, 638-649.

Bolivar-Subirats, G., Cortina-Puig, M., Lacorte, S., 2020. Multiresidue method for the determination of high production volume plastic additives in river waters. Environ. Sci. Pollut. Res. Int. 27, 41314-41325.

Bolivar-Subirats, G., Rivetti, C., Cortina-Puig, M., Barata, C., Lacorte, S., 2021. Occurrence, toxicity and risk assessment of plastic additives in Besos river, Spain. Chemosphere 263, 128022.

Bonini, P., Kind, T., Tsugawa, H., Barupal, D.K., Fiehn, O., 2020. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. Anal. Chem. 92, 7515-7522.

Boschmans, J., Jacobs, S., Williams, J.P., Palmer, M., Richardson, K., Giles, K., Lapthorn, C., Herrebout, W.A., Lemiere, F., Sobott, F., 2016. Combining density functional theory (DFT) and collision cross-section (CCS) calculations to analyze the gas-phase behaviour of small molecules and their protonation site isomers. Analyst 141, 4044-4054.

Broeckling, C.D., Yao, L., Isaac, G., Gioioso, M., Ianchis, V., Vissers, J.P.C., 2021. Application of Predicted Collisional Cross Section to Metabolome Databases to Probabilistically Describe the Current and Future Ion Mobility Mass Spectrometry. J. Am. Soc. Mass Spectrom. 32, 661-669.

Browne, M.A., Galloway, T., Thompson, R., 2007. Microplastic--an emerging contaminant of potential concern? Integr. Environ. Assess. Manag. 3, 559-561.

Bush, M.F., Hall, Z., Giles, K., Hoyes, J., Robinson, C.V., Ruotolo, B.T., 2010. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. Anal. Chem. 82, 9557-9565.

Canellas, E., Vera, P., Nerín, C., 2014. Atmospheric pressure gas chromatography coupled to quadrupole-time of flight mass spectrometry as a tool for identification of volatile migrants from autoadhesive labels used for direct food contact. J Mass Spectrom. 49, 1181-1190.

Canellas, E., Vera, P., Nerin, C., 2015a. Risk assessment derived from migrants identified in several adhesives commonly used in food contact materials. Food Chem. Toxicol. 75, 79-87.

Canellas, E., Vera, P., Nerin, C., 2015b. UPLC-ESI-Q-TOF-MS(E) and GC-MS identification and quantification of non-intentionally added substances coming from biodegradable food packaging. Anal. Bioanal. Chem. 407, 6781-6790.

Canellas, E., Vera, P., Nerin, C., 2019. Ion mobility quadrupole time-of-flight mass spectrometry for the identification of non-intentionally added substances in UV varnishes applied on food contact materials. A safety by design study. Talanta 205, 120103.

Canellas, E., Vera, P., Nerin, C., Dreolin, N., Goshawk, J., 2020. Ion mobility quadrupole time-of-flight high resolution mass spectrometry coupled to ultra-high pressure liquid chromatography for identification of non-intentionally added substances migrating from food cans. J. Chromatogr. A 1616, 460778.

Canellas, E., Vera, P., Nerin, C., Dreolin, N., Goshawk, J., 2022. The detection and elucidation of oligomers migrating from biodegradable multilayer teacups using liquid chromatography coupled to ion mobility time-of-flight mass spectrometry and gas chromatography-mass spectrometry. Food Chem. 374, 131777.

Canellas, E., Vera, P., Nerin, C., Goshawk, J., Dreolin, N., 2021a. The application of ion mobility time of flight mass spectrometry to elucidate neo-formed compounds derived from polyurethane adhesives used in champagne cork stoppers. Talanta 234, 122632.

Canellas, E., Vera, P., Song, X.-C., Nerin, C., Goshawk, J., Dreolin, N., 2021b. The use of ion mobility time-of-flight mass spectrometry to assess the migration of polyamide 6 and polyamide 66 oligomers from kitchenware utensils to food. Food Chem. 350, 129260.

Carpenter, R., O'Grady, M.N., O'Callaghan, Y.C., O'Brien, N.M., Kerry, J.P., 2007. Evaluation of the antioxidant potential of grape seed and bearberry extracts in raw and cooked pork. Meat Sci. 76, 604-610.

Cecon, V.S., Da Silva, P.F., Curtzwiler, G.W., Vorst, K.L., 2021. The challenges in

recycling post-consumer polyolefins for food contact applications: A review. Resour. Conserv. Recycl. 167, 105422.

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

Celma, A., Sancho, J.V., Schymanski, E.L., Fabregat-Safont, D., Ibanez, M., Goshawk, J., Barknowitz, G., Hernandez, F., Bijlsma, L., 2020. Improving Target and Suspect Screening High-Resolution Mass Spectrometry Workflows in Environmental Analysis by Ion Mobility Separation. Environ. Sci. Technol. 54, 15120-15131.

Chalet, C., Hollebrands, B., Janssen, H.G., Augustijns, P., Duchateau, G., 2018. Identification of phase-II metabolites of flavonoids by liquid chromatography-ion-mobility spectrometry-mass spectrometry. Anal. Bioanal. Chem. 410, 471-482.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. 785-794. DOI: 10.1145/2939672.2939785.

Claassen, C., Ebel, E., Kuballa, J., Rohn, S., 2019. Impacts of Fungicide Treatment and Conventional Fertilization Management on the Potato Metabolome (Solanum tuberosum L.) Evaluated with UPLC-IMS-QToF. J. Agric. Food Chem. 67, 11542-11552.

Connolly, J., Munoz-Muriedas, J., Lapthorn, C., Higton, D., Vissers, J.P.C., Webb, A., Beaumont, C., Dear, G.J., 2021. Investigation into Small Molecule Isomeric Glucuronide Metabolite Differentiation Using In Silico and Experimental Collision Cross-Section Values. J. Am. Soc. Mass Spectrom. 32, 1976-1986.

Cowan-Ellsberry, C., Belanger, S., Dorn, P., Dyer, S., McAvoy, D., Sanderson, H., Versteeg, D., Ferrer, D., Stanton, K., 2014. Environmental Safety of the Use of Major Surfactant Classes in North America. Crit. Rev. Environ. Sci. Technol. 44, 1893-1993.

Curtzwiler, G.W., Silva, P., Hall, A., Ivey, A., Vorst, K., 2021. Significance of Perfluoroalkyl Substances (PFAS) in Food Packaging. Integr. Environ. Assess. Manag. 17, 7-12.

D'Atri, V., Causon, T., Hernandez-Alba, O., Mutabazi, A., Veuthey, J.L., Cianferani, S., Guillarme, D., 2018. Adding a new separation dimension to MS and LC-MS: What is the utility of ion mobility spectrometry? J. Sep. Sci. 41, 20-67.

Darbre, P.D., 2020. Chemical components of plastics as endocrine disruptors: Overview and commentary. Birth Defects Res. 112, 1300-1307.

Denberg, M., Mosbæk, H., Hassager, O., Arvin, E., 2009. Determination of the concentration profile and homogeneity of antioxidants and degradation products in a cross-linked polyethylene type A (PEXa) pipe. Polym. Test. 28, 378-385.

Devillers, J., Balaban, A.T., 1999. Topological Indices and Related Descriptors in QSAR and QSPR.

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., Wishart, D.S., 2016. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J. Cheminform. 8, 61.

Dodds, J.N., Alexander, N.L.M., Kirkwood, K.I., Foster, M.R., Hopkins, Z.R., Knappe, D.R.U., Baker, E.S., 2021. From Pesticides to Per- and Polyfluoroalkyl Substances: An Evaluation of Recent Targeted and Untargeted Mass Spectrometry Methods for Xenobiotics. Anal. Chem. 93, 641-656.

Dodds, J.N., Hopkins, Z.R., Knappe, D.R.U., Baker, E.S., 2020. Rapid Characterization of Per- and Polyfluoroalkyl Substances (PFAS) by Ion Mobility Spectrometry-Mass Spectrometry (IMS-MS). Anal. Chem. 92, 4427–4435

Dodds, J.N., May, J.C., McLean, J.A., 2017. Correlating Resolving Power, Resolution, and Collision Cross Section: Unifying Cross-Platform Assessment of Separation Efficiency in Ion Mobility Spectrometry. Anal. Chem. 89, 12176-12184.

Dong, J., Cao, D.S., Miao, H.Y., Liu, S., Deng, B.C., Yun, Y.H., Wang, N.N., Lu, A.P., Zeng, W.B., Chen, A.F., 2015. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. J. Cheminform. 7, 60.

Dorival-Garcia, N., Carillo, S., Ta, C., Roberts, D., Comstock, K., Lofthouse, S., Ciceri, E., D'Silva, K., Kierans, G., Kaisermayer, C., Lindeberg, A., Bones, J., 2018. Large-Scale Assessment of Extractables and Leachables in Single-Use Bags for Biomanufacturing. Anal. Chem. 90, 9006-9015.

Dreolin, N., Aznar, M., Moret, S., Nerin, C., 2019. Development and validation of a LC-MS/MS method for the analysis of bisphenol a in polyethylene terephthalate. Food Chem. 274, 246-253.

EC, 2004. European Commission Regulation (EC) No. 1935/2004 on materials and articles intended to come into contact with food.

EC, 2011. European Commission Regulation (EC) No 10/2011 on plastic materials and articles intened to come into contact with food.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802-813.

Elizalde, M., Aparicio, J., Rincón, M., 2020. Interpretation of the migration of benzophenone type photoinitiators into different food simulants and foodstuffs in terms of the physicochemical properties of the migrants. Food Packag. Shelf Life 23, 100444.

Fabregat-Safont, D., Ibáñez, M., Bijlsma, L., Hernández, F., Waichman, A.V., de Oliveira, R., Rico, A., 2021. Wide-scope screening of pharmaceuticals, illicit drugs and their metabolites in the Amazon River. Water Res. 200, 117251.

Fan, Y., Liu, F., He, W., Qin, Q., Hu, D., Wu, A., Jiang, W., Wang, C., 2022. Screening of multi-mycotoxins in fruits by ultra-performance liquid chromatography coupled to ion mobility quadrupole time-of-flight mass spectrometry. Food Chem. 368, 130858.

Felix, J.S., Isella, F., Bosetti, O., Nerin, C., 2012. Analytical tools for identification of non-intentionally added substances (NIAS) coming from polyurethane adhesives in multilayer packaging materials and their migration into food simulants. Anal. Bioanal. Chem. 403, 2869-2882.

Fenn, L.S., Kliman, M., Mahsut, A., Zhao, S.R., McLean, J.A., 2009. Characterizing ion mobility-mass spectrometry conformation space for the analysis of complex biological samples. Anal. Bioanal. Chem. 394, 235-244.

Fernandez-Lima, F., Kaplan, D.A., Suetering, J., Park, M.A., 2011. Gas-phase separation using a trapped ion mobility spectrometer. Int. J. Ion Mobil. Spectrom. 14(2-3), DOI:10.1007/s12127-011-0067-8.

Fries, E., Püttmann, W., 2002. Analysis of the antioxidant butylated hydroxytoluene (BHT) in water by means of solid phase extraction combined with GC/MS. Water Res. 36, 2319-2327.

Gabelica, V., De Pauw, E., Karas, M., 2004. Influence of the capillary temperature and the source pressure on the internal energy distribution of electrosprayed ions. Int. J. Mass Spectrom. 231, 189-195.

Gabelica, V., Marklund, E., 2018. Fundamentals of ion mobility spectrometry. Curr. Opin. Chem. Biol. 42, 51-59.

Gabelica, V., Shvartsburg, A.A., Afonso, C., Barran, P., Benesch, J.L.P., Bleiholder, C., Bowers, M.T., Bilbao, A., Bush, M.F., Campbell, J.L., Campuzano, I.D.G., Causon, T., Clowers, B.H., Creaser, C.S., De Pauw, E., Far, J., Fernandez-Lima, F., Fjeldsted, J.C., Giles, K., Groessl, M., Hogan, C.J., Jr., Hann, S., Kim, H.I., Kurulugama, R.T., May, J.C., McLean, J.A., Pagel, K., Richardson, K., Ridgeway, M.E., Rosu, F., Sobott, F., Thalassinos, K., Valentine, S.J., Wyttenbach, T., 2019. Recommendations for reporting ion mobility Mass Spectrometry measurements. Mass Spectrom. Rev. 38, 291-320.

Gallart-Ayala, H., Moyano, E., Galceran, M.T., 2011. Fast liquid chromatography-

tandem mass spectrometry for the analysis of bisphenol A-diglycidyl ether, bisphenol F-diglycidyl ether and their derivatives in canned food and beverages. J. Chromatogr. A 1218, 1603-1610.

Gao, S., Zhan, Q., Li, J., Yang, Q., Li, X., Chen, W., Sun, L., 2010. LC-MS/MS method for the simultaneous determination of ethyl gallate and its major metabolite in rat plasma. Biomed. Chromatogr. 24, 472-478.

García Ibarra, V., Rodríguez Bernaldo de Quirós, A., Sendón, R., 2016. Study of melamine and formaldehyde migration from melamine tableware. Eur. Food Res. Technol. 242, 1187-1199.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1-17.

Gelbke, H.-P., Buist, H., Eisert, R., Leibold, E., Sherman, J.H., 2019. Derivation of safe exposure levels for potential migration of formaldehyde into food. Food Chem. Toxicol. 132, 110598.

Getachew, P., Getachew, M., Joo, J., Choi, Y.S., Hwang, D.S., Hong, Y.-K., 2017. The slip agents oleamide and erucamide reduce biofouling by marine benthic organisms (diatoms, biofilms and abalones). Toxicol. Environ. Health Sci. 8, 341-348.

Geyer, R., Jambeck, J.R., Law, K.L., 2017. Production, use, and fate of all plastics ever made. Sci. Adv. 3, e1700782.

Ghose, A.K., Crippen, G.M., 1987. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. J. Chem. Inf. Comput. Sci. 27, 21-35.

Giles, K., Ujma, J., Wildgoose, J., Pringle, S., Richardson, K., Langridge, D., Green, M., 2019. A Cyclic Ion Mobility-Mass Spectrometry System. Anal. Chem. 91, 8564-8573.

Giles, K., Williams, J.P., Campuzano, I., 2011. Enhancements in travelling wave ion mobility resolution. Rapid Commun. Mass Spectrom. 25, 1559-1566.

Gong, X., Zhang, W., Zhang, S., Wang, Y., Zhang, X., Lu, Y., Sun, H., Wang, L., 2021. Organophosphite Antioxidants in Mulch Films Are Important Sources of Organophosphate Pollutants in Farmlands. Environ. Sci. Technol. 55, 7398-7406.

Gonzales, G.B., Smagghe, G., Coelus, S., Adriaenssens, D., De Winter, K., Desmet, T., Raes, K., Van Camp, J., 2016. Collision cross section prediction of deprotonated phenolics in a travelling-wave ion mobility spectrometer using molecular descriptors and chemometrics. Anal. Chim Acta 924, 68-76.

Gonzalez-Marino, I., Ares, L., Montes, R., Rodil, R., Cela, R., Lopez-Garcia, E., Postigo, C., Lopez de Alda, M., Pocurull, E., Marce, R.M., Bijlsma, L., Hernandez, F., Pico, Y., Andreu, V., Rico, A., Valcarcel, Y., Miro, M., Etxebarria, N., Quintana, J.B., 2021. Assessing population exposure to phthalate plasticizers in thirteen Spanish cities through the analysis of wastewater. J. Hazard. Mater. 401, 123272.

Gray, C.J., Thomas, B., Upton, R., Migas, L.G., Eyers, C.E., Barran, P.E., Flitsch, S.L., 2016. Applications of ion mobility mass spectrometry for high throughput, high resolution glycan analysis. Biochim. Biophys. Acta 1860, 1688-1709.

Grindler, N.M., Vanderlinden, L., Karthikraj, R., Kannan, K., Teal, S., Polotsky, A.J., Powell, T.L., Yang, I.V., Jansson, T., 2018. Exposure to Phthalate, an Endocrine Disrupting Chemical, Alters the First Trimester Placental Methylome and Transcriptome in Women. Sci. Rep. 8, 6086.

Grob, K., Biedermann, M., Scherbaum, E., Roth, M., Rieger, K., 2006. Food contamination with organic materials in perspective: packaging materials as the largest and least controlled source? A view focusing on the European situation. Crit. Rev. Food Sci. Nutr. 46, 529-535.

Groh, K.J., Backhaus, T., Carney-Almroth, B., Geueke, B., Inostroza, P.A., Lennquist, A., Leslie, H.A., Maffini, M., Slunge, D., Trasande, L., Warhurst, A.M., Muncke,

J., 2019. Overview of known plastic packaging-associated chemicals and their hazards. Sci. Total Environ. 651, 3253-3268.

Groh, K.J., Geueke, B., Martin, O., Maffini, M., Muncke, J., 2021. Overview of intentionally used food contact chemicals and their hazards. Environ. Int. 150, 106225.

Gunaalan, K., Fabbri, E., Capolupo, M., 2020. The hidden threat of plastic leachates: A critical review on their impacts on aquatic organisms. Water Res. 184, 116170.

Hahladakis, J.N., Velis, C.A., Weber, R., Iacovidou, E., Purnell, P., 2018. An overview of chemical additives present in plastics: Migration, release, fate and environmental impact during their use, disposal and recycling. J. Hazard. Mater. 344, 179-199.

Hall, L.H., Kier, L.B., 2007. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure‐Property Modeling. John Wiley & Sons, Ltd.,.

Hankin, J.A., Barkley, R.M., Zemski-Berry, K., Deng, Y., Murphy, R.C., 2016. Mass Spectrometric Collisional Activation and Product Ion Mobility of Human Serum Neutral Lipid Extracts. Anal. Chem. 88, 6274-6282.

He, Y.J., Qin, Y., Zhang, T.L., Zhu, Y.Y., Wang, Z.J., Zhou, Z.S., Xie, T.Z., Luo, X.D., 2021. Migration of (non-) intentionally added substances and microplastics from microwavable plastic food containers. J. Hazard. Mater. 417, 126074.

Heimrich, M., Nickl, H., Bönsch, M., Simat, T.J., 2015. Migration of cyclic monomer and oligomers from polyamide 6 and 66 food contact materials into food and food simulants: direct food contact. Packag. Technol. Sci. 28, 123-139.

Hernandez-Mesa, M., D'Atri, V., Barknowitz, G., Fanuel, M., Pezzatti, J., Dreolin, N., Ropartz, D., Monteau, F., Vigneau, E., Rudaz, S., Stead, S., Rogniaux, H., Guillarme, D., Dervilly, G., Le Bizec, B., 2020. Interlaboratory and Interplatform Study of Steroids Collision Cross Section by Traveling Wave Ion Mobility Spectrometry. Anal. Chem. 92, 5013-5022.

Hernandez-Mesa, M., Le Bizec, B., Monteau, F., Garcia-Campana, A.M., Dervilly-Pinel, G., 2018. Collision Cross Section (CCS) Database: An Additional Measure to Characterize Steroids. Anal. Chem. 90, 4616-4625.

Hernandez, L.M., Xu, E.G., Larsson, H.C.E., Tahara, R., Maisuria, V.B., Tufenkji, N., 2019. Plastic Teabags Release Billions of Microparticles and Nanoparticles into Tea. Environ. Sci. Technol. 53, 12300-12310.

Hines, K.M., May, J.C., McLean, J.A., Xu, L., 2016. Evaluation of Collision Cross Section Calibrants for Structural Analysis of Lipids by Traveling Wave Ion Mobility-Mass Spectrometry. Anal. Chem. 88, 7329-7336.

Hines, K.M., Ross, D.H., Davidson, K.L., Bush, M.F., Xu, L., 2017. Large-Scale Structural Characterization of Drug and Drug-Like Compounds by High-Throughput Ion Mobility-Mass Spectrometry. Anal. Chem. 89, 9023-9030.

Hinnenkamp, V., Klein, J., Meckelmann, S.W., Balsaa, P., Schmidt, T.C., Schmitz, O.J., 2018. Comparison of CCS Values Determined by Traveling Wave Ion Mobility Mass Spectrometry and Drift Tube Ion Mobility Mass Spectrometry. Anal. Chem. 90, 12042-12050.

Hofmann, J., Hahm, H.S., Seeberger, P.H., Pagel, K., 2015. Identification of carbohydrate anomers using ion mobility-mass spectrometry. Nature 526, 241-244.

Hopewell, J., Dvorak, R., Kosior, E., 2009. Plastics recycling: challenges and opportunities. Philosophical transactions of the Royal Society of London. Series B, Biol. Sci. 364, 2115-2126.

Hoppe, M., de Voogt, P., Franz, R., 2016. Identification and quantification of oligomers as potential migrants in plastics food contact materials with a focus in polycondensates – A review. Trends Food Sci. Technol. 50, 118-130.

Hoppe, M., de Voogt, P., Franz, R., 2018a. Oligomers in polyethylene naphthalate and polybutylene terephthalate–Identification and exploring migration. Food Packag. Shelf Life 17, 171-178.

Hoppe, M., de Voogt, P., Franz, R., 2018b. Oligomers in polyethylene naphthalate and polybutylene terephthalate – Identification and exploring migration. Food Packag. Shelf Life 17, 171-178.

Hu, Y., Du, Z., Sun, X., Ma, X., Song, J., Sui, H., Debrah, A.A., 2021. Non-targeted analysis and risk assessment of non-volatile compounds in polyamide food contact materials. Food Chem. 345, 128625.

Ilyas, M., Ahmad, W., Khan, H., Yousaf, S., Khan, K., Nazir, S., 2018. Plastic waste as a significant threat to environment - a systematic literature review. Rev. Environ. Health 33, 383-406.

Kappenstein, O., Ebner, I., Förster, C., Richter, S., Weyer, J., Pfaff, K., Luch, A., 2018. Validation and application of an LC-MS/MS method for the determination of cyclic oligomers originating from polyamide 6 and polyamide 66 in food simulant. Food Addit Contam Part A 35, 1410-1420.

Kirk, J., Mortishire-Smith, R., Wrona, M., 2017. Integrating Ion Mobility into Routine Metabolite Identification Studies Using the Vion IMS QTof Mass Spectrometer. Waters.

Kitanov, G.M., Nedialkov, P.T., 2001. Benzophenone O-glucoside, a biogenic precursor of 1,3,7-trioxygenated xanthones in Hypericum annulatum. Phytochemistry 57, 1237-1243.

Kozlowska, M., Goclon, J., Rodziewicz, P., 2016. Intramolecular Hydrogen Bonds in Low-Molecular-Weight Polyethylene Glycol. Chemphyschem 17, 1143-1153.

Krzywicki, K., 1979. Assessment of relative content of myoglobin, oxymyoglobin and metmyoglobin at the surface of beef. Meat Sci. 3, 1-10.

Kurkin, V.A., Ryazanova, T.K., Daeva, E.D., Kadentsev, V.I., 2018. Constituents of Arctostaphylos uva-ursi Leaves. Chem. Nat. Compd. 54, 278-280.

Lalli, P.M., Corilo, Y.E., Fasciotti, M., Riccio, M.F., de Sa, G.F., Daroda, R.J., Souza,

G.H., McCullagh, M., Bartberger, M.D., Eberlin, M.N., Campuzano, I.D., 2013. Baseline resolution of isomers by traveling wave ion mobility mass spectrometry: investigating the effects of polarizable drift gases and ionic charge distribution. J. Mass Spectrom. 48, 989-997.

Li, W.C., Tse, H.F., Fok, L., 2016. Plastic waste in the marine environment: A review of sources, occurrence and effects. Sci. Total Environ. 566-567, 333-349.

Lian, R., Zhang, F., Zhang, Y., Wu, Z., Ye, H., Ni, C., Lv, X., Guo, Y., 2018. Ion mobility derived collision cross section as an additional measure to support the rapid analysis of abused drugs and toxic compounds using electrospray ion mobility time-of-flight mass spectrometry. Anal. Methods 10, 749-756.

Liu, R., Mabury, S.A., 2018a. First Detection of Photoinitiators and Metabolites in Human Sera from United States Donors. Environ. Sci. Technol. 52, 10089-10096.

Liu, R., Mabury, S.A., 2018b. Unexpectedly High Concentrations of a Newly Identified Organophosphate Ester, Tris(2,4-di- tert-butylphenyl) Phosphate, in Indoor Dust from Canada. Environ. Sci. Technol. 52, 9677-9683.

Liu, R., Mabury, S.A., 2019a. Identification of Photoinitiators, Including Novel Phosphine Oxides, and Their Transformation Products in Food Packaging Materials and Indoor Dust in Canada. Environ. Sci. Technol. 53, 4109-4118.

Liu, R., Mabury, S.A., 2019b. Organophosphite Antioxidants in Indoor Dust Represent an Indirect Source of Organophosphate Esters. Environ. Sci. Technol. 53, 1805-1811.

Liu, R., Mabury, S.A., 2019c. Synthetic phenolic antioxidants and transformation products in dust from different indoor environments in Toronto, Canada. Sci. Total Environ. 672, 23-29.

Liu, R., Mabury, S.A., 2020. Synthetic Phenolic Antioxidants: A Review of Environmental Occurrence, Fate, Human Exposure, and Toxicity. Environ. Sci. Technol. 54, 11706-11719.

Liu, R., Mabury, S.A., 2021. Rat Metabolism Study Suggests 3-(3,5-Di-tert-butyl-4-hydroxyphenyl)propionic Acid as a Potential Urinary Biomarker of Human Exposure to Representative 3-(3,5-Di-tert-butyl-4-hydroxyphenyl)propionate Antioxidants. Environ. Sci. Technol. 55, 14051-14058.

Liu, R., Song, S., Lin, Y., Ruan, T., Jiang, G., 2015. Occurrence of synthetic phenolic antioxidants and major metabolites in municipal sewage sludge in China. Environ. Sci. Technol. 49, 2073-2080.

Liu, R.Z., Mabury, S.A., 2018c. Synthetic Phenolic Antioxidants and Transformation Products in Human Sera from United States Donors. Environ. Sci. Technol Lett. 5, 419-423.

Liu, X., Chen, D., Yu, Y., Zeng, X., Li, L., Xie, Q., Yang, M., Wu, Q., Dong, G., 2020. Novel Organophosphate Esters in Airborne Particulate Matters: Occurrences, Precursors, and Selected Transformation Products. Environ. Sci. Technol. 54, 13771-13777.

Liu, X., Peng, C., Shi, Y., Tan, H., Tang, S., Chen, D., 2019. Beyond Phthalate Diesters: Existence of Phthalate Monoesters in South China House Dust and Implications for Human Exposure. Environ. Sci. Technol. 53, 11675-11683.

Liu, X., Zeng, X., Dong, G., Venier, M., Xie, Q., Yang, M., Wu, Q., Zhao, F., Chen, D., 2021. Plastic Additives in Ambient Fine Particulate Matter in the Pearl River Delta, China: High-Throughput Characterization and Health Implications. Environ. Sci. Technol. 55, 4474-4482.

Luo, M.-D., Zhou, Z.-W., Zhu, Z.-J., 2020a. The Application of Ion Mobility-Mass Spectrometry in Untargeted Metabolomics: from Separation to Identification. J. Anal. Test.. 4, 163–174

Luo, Y.S., Aly, N.A., McCord, J., Strynar, M.J., Chiu, W.A., Dodds, J.N., Baker, E.S., Rusyn, I., 2020b. Rapid Characterization of Emerging Per- and Polyfluoroalkyl Substances in Aqueous Film-Forming Foams Using Ion Mobility Spectrometry-

Mass Spectrometry. Environ. Sci. Technol. 54, 15024-15034.

Martínez-López, B., Gontard, N., Peyron, S., 2018. Worst case prediction of additives migration from polystyrene for food safety purposes: a model update. Food Addit. Contam. Part A 35, 563-576.

Masike, K., de Villiers, A., de Beer, D., Joubert, E., Stander, M.A., 2022. Application of direct injection-ion mobility spectrometry-mass spectrometry (DI-IMS-MS) for the analysis of phenolics in honeybush and rooibos tea samples. J. Food Compos. Anal. 106, 104308.

Masike, K., de Villiers, A., Hoffman, E.W., Brand, D.J., Causon, T., Stander, M.A., 2020. Detailed Phenolic Characterization of Protea Pure and Hybrid Cultivars by Liquid Chromatography-Ion Mobility-High Resolution Mass Spectrometry (LC-IM-HR-MS). J. Agric. Food Chem. 68, 485-502.

May, J.C., Leaptrot, K.L., Rose, B.S., Moser, K.L.W., Deng, L., Maxon, L., DeBord, D., McLean, J.A., 2021. Resolving Power and Collision Cross Section Measurement Accuracy of a Prototype High-Resolution Ion Mobility Platform Incorporating Structures for Lossless Ion Manipulation. J. Am. Soc. Mass Spectrom. 32, 1126-1137.

May, J.C., McLean, J.A., 2015. Ion mobility-mass spectrometry: time-dispersive instrumentation. Anal. Chem. 87, 1422-1436.

Mazza, G., Fukumoto, L., Delaquis, P., Girard, B., Ewert, B., 1999. Anthocyanins, phenolics, and color of Cabernet Franc, Merlot, and Pinot Noir wines from British Columbia. J. Agric. Food Chem. 47, 4009-4017.

McCullagh, M., Douce, D., Van Hoeck, E., Goscinny, S., 2018. Exploring the Complexity of Steviol Glycosides Analysis Using Ion Mobility Mass Spectrometry. Anal. Chem. 90, 4585-4595.

McCullagh, M., Giles, K., Richardson, K., Stead, S., Palmer, M., 2019a. Investigations into the performance of travelling wave enabled conventional and cyclic ion

mobility systems to characterise protomers of fluoroquinolone antibiotic residues. Rapid Commun. Mass Spectrom. 33 Suppl 2, 11-21.

McCullagh, M., Pereira, C.A.M., Yariwake, J.H., 2019b. Use of ion mobility mass spectrometry to enhance cumulative analytical specificity and separation to profile 6-C/8-C-glycosylflavone critical isomer pairs and known-unknowns in medicinal plants. Phytochem. Anal. 30, 424-436.

Mevik, B.H., Wehrens, R., 2007. The pls package: Principal component and partial least squares regression in R. J. Stat. Softw. 18, 1-23.

Michelmann, K., Silveira, J.A., Ridgeway, M.E., Park, M.A., 2015. Fundamentals of trapped ion mobility spectrometry. J. Am. Soc. Mass Spectrom. 26, 14-24.

Mohd Amin, A.M., Mohd Sauid, S., Ku Hamid, K.H., 2015. Polymer-Starch Blend Biodegradable Plastics: An Overview. Adv. Mater. Res. 1113, 93-98.

Mohd Azman, N.A., Gallego, M.G., Segovia, F., Abdullah, S., Shaarani, S.M., Almajano Pablos, M.P., 2016. Study of the Properties of Bearberry Leaf Extract as a Natural Antioxidant in Model Foods. Antioxidants 5, 11.

Mollerup, C.B., Mardal, M., Dalsgaard, P.W., Linnet, K., Barron, L.P., 2018. 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 1542, 82-88.

Moore, C.J., 2008. Synthetic polymers in the marine environment: a rapidly increasing, long-term threat. Environ. Res. 108, 131-139.

Mosier, P.D., Counterman, A.E., Jurs, P.C., Clemmer, D.E., 2002. Prediction of Peptide Ion Collision Cross Sections from Topological Molecular Structure and Amino Acid Parameters. Anal. Chem. 74, 1360-1370.

Mullin, L., Jobst, K., DiLorenzo, R.A., Plumb, R., Reiner, E.J., Yeung, L.W.Y., Jogsten, I.E., 2020. Liquid chromatography-ion mobility-high resolution mass spectrometry

for analysis of pollutants in indoor dust: Identification and predictive capabilities. Anal. Chim. Acta 1125, 29-40.

Muncke, J., 2016. Chemical Migration from Food Packaging to Food. Reference Module in Food Science. Elsivier: 2016

Muncke, J., Andersson, A.M., Backhaus, T., Boucher, J.M., Carney Almroth, B., Castillo Castillo, A., Chevrier, J., Demeneix, B.A., Emmanuel, J.A., Fini, J.B., Gee, D., Geueke, B., Groh, K., Heindel, J.J., Houlihan, J., Kassotis, C.D., Kwiatkowski, C.F., Lefferts, L.Y., Maffini, M.V., Martin, O.V., Myers, J.P., Nadal, A., Nerin, C., Pelch, K.E., Fernandez, S.R., Sargis, R.M., Soto, A.M., Trasande, L., Vandenberg, L.N., Wagner, M., Wu, C., Zoeller, R.T., Scheringer, M., 2020. Impacts of food contact chemicals on human health: a consensus statement. Environ. Health 19, 25.

Muncke, J., Backhaus, T., Geueke, B., Maffini, M.V., Martin, O.V., Myers, J.P., Soto, A.M., Trasande, L., Trier, X., Scheringer, M., 2017. Scientific Challenges in the Risk Assessment of Food Contact Materials. Environ. Health Perspect. 125, 095001.

Olennikov D.N., Chekhirova G.V., 2013. 6″-Galloylpicein and other phenolic compounds from Arctostaphylos uva-ursi. Chem. Nat. Compd. 49, 1-7.

Naczk, M., Pegg, R.B., Amarowicz, R., 2011. Protein-precipitating capacity of bearberry-leaf (Arctostaphylos uva-ursi L. Sprengel) polyphenolics. Food Chem. 124, 1507-1513.

Nerin, C., Alfaro, P., Aznar, M., Domeno, C., 2013. The challenge of identifying non-intentionally added substances from food packaging materials: a review. Anal. Chim. Acta 775, 14-24.

Nerin, C., Bourdoux, S., Faust, B., Gude, T., Lesueur, C., Simat, T., Stoermer, A., Van Hoek, E., Oldring, P., 2022. Guidance in selecting analytical techniques for identification and quantification of non-intentionally added substances (NIAS) in food contact materials (FCMS). Food Addit. Contam. Part A 39, 620-643.

Nerin, C., Su, Q.Z., Vera, P., Mendoza, N., Ausejo, R., 2020. Influence of nonylphenol from multilayer plastic films on artificial insemination of sows. Anal. Bioanal. Chem. 412, 6519-6528.

Nichols, C.M., Dodds, J.N., Rose, B.S., Picache, J.A., Morris, C.B., Codreanu, S.G., May, J.C., Sherrod, S.D., McLean, J.A., 2018. Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry. Anal. Chem. 90, 14484-14492.

Nye, L.C., Williams, J.P., Munjoma, N.C., Letertre, M.P.M., Coen, M., Bouwmeester, R., Martens, L., Swann, J.R., Nicholson, J.K., Plumb, R.S., McCullagh, M., Gethings, L.A., Lai, S., Langridge, J.I., Vissers, J.P.C., Wilson, I.D., 2019. A comparison of collision cross section values obtained via travelling wave ion mobility-mass spectrometry and ultra high performance liquid chromatography-ion mobility-mass spectrometry: Application to the characterisation of metabolites in rat urine. J. Chromatogr. A 1602, 386-396.

Omer, E., Cariou, R., Remaud, G., Guitton, Y., Germon, H., Hill, P., Dervilly-Pinel, G., Le Bizec, B., 2018. Elucidation of non-intentionally added substances migrating from polyester-polyurethane lacquers using automated LC-HRMS data processing. Anal. Bioanal. Chem. 410, 5391-5403.

Osorio, J., Aznar, M., Nerin, C., Birse, N., Elliott, C., Chevallier, O., 2020. Ambient mass spectrometry as a tool for a rapid and simultaneous determination of migrants coming from a bamboo-based biopolymer packaging. J. Hazard. Mater. 398, 122891.

Osorio, J., Dreolin, N., Aznar, M., Nerin, C., Hancock, P., 2019. Determination of volatile non intentionally added substances coming from a starch-based biopolymer intended for food contact by different gas chromatography-mass spectrometry approaches. J. Chromatogr. A 1599, 215-222.

Padron, J., Carrasco, R., Pellon, R., 2002. Molecular descriptor based on a molar refractivity partition using Randic-type graph-theoretical invariant. J. Pharm.

Pharmaceut. Sci. 5, 258-266.

Paglia, G., Angel, P., Williams, J.P., Richardson, K., Olivos, H.J., Thompson, J.W., Menikarachchi, L., Lai, S., Walsh, C., Moseley, A., Plumb, R.S., Grant, D.F., Palsson, B.O., Langridge, J., Geromanos, S., Astarita, G., 2015. Ion Mobility-Derived Collision Cross Section As an Additional Measure for Lipid Fingerprinting and Identification. Anal. Chem. 87, 1137-1144.

Paglia, G., Astarita, G., 2017. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. Nat. Protoc. 12, 797-813.

Paglia, G., Smith, A.J., Astarita, G., 2021. Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics. Mass Spectrom Rev. 1:44.

Panusa, A., Petrucci, R., Marrosu, G., Multari, G., Gallo, F.R., 2015. UHPLC-PDA-ESI-TOF/MS metabolic profiling of Arctostaphylos pungens and Arctostaphylos uva-ursi. A comparative study of phenolic compounds from leaf methanolic extracts. Phytochemistry 115, 79-88.

Plante, P.L., Francovic-Fontaine, E., May, J.C., McLean, J.A., Baker, E.S., Laviolette, F., Marchand, M., Corbeil, J., 2019. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. Anal. Chem. 91, 5191-5199.

PlasticsEurope, 2021. Plastics - the Facts 2021: An analysis of European plastics production, demand and waste data.

Poland, J.C., Leaptrot, K.L., Sherrod, S.D., Flynn, C.R., McLean, J.A., 2020. Collision Cross Section Conformational Analyses of Bile Acids via Ion Mobility-Mass Spectrometry. J. Am. Soc. Mass Spectrom. 31, 1625-1631.

Prata, J.C., da Costa, J.P., Lopes, I., Duarte, A.C., Rocha-Santos, T., 2020. Environmental exposure to microplastics: An overview on possible human health effects. Sci. Total Environ. 702, 134455.

Pringle, S.D., Giles, K., Wildgoose, J.L., Williams, J.P., Slade, S.E., Thalassinos, K., Bateman, R.H., Bowers, M.T., Scrivens, J.H., 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. Int. J. Mass Spectrom. 261, 1-12.

Puype, F., Samsonek, J., Knoop, J., Egelkraut-Holtus, M., Ortlieb, M., 2015. Evidence of waste electrical and electronic equipment (WEEE) relevant substances in polymeric food-contact articles sold on the European market. Food Addit. Contam. Part A 32, 410-426.

Qadeer, A., Kirsten, K.L., Ajmal, Z., Jiang, X., Zhao, X., 2022. Alternative Plasticizers As Emerging Global Environmental and Health Threat: Another Regrettable Substitution? Environ. Sci. Technol. 56, 1482-1488.

Qi, W., Wang, Y., Cao, Y., Cao, Y., Guan, Q., Sun, T., Zhang, L., Guo, Y., 2020. Simultaneous Analysis of Fatty Alcohols, Fatty Aldehydes, and Sterols in Thyroid Tissues by Electrospray Ionization-Ion Mobility-Mass Spectrometry Based on Charge Derivatization. Anal. Chem. 92, 8644-8648.

Rahman, M., Brazel, C., 2004. The plasticizer market: an assessment of traditional plasticizers and research trends to meet new challenges. Prog. Polym. Sci. 29, 1223-1248.

Rajalahti, T., Arneberg, R., Kroksveen, A.C., Berle, M., Myhr, K.M., Kvalheim, O.M., 2009. Discriminating variable test and selectivity ratio plot: quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles. Anal. Chem. 81, 2581-2590.

Regueiro, J., Negreira, N., Berntssen, M.H., 2016. Ion-Mobility-Derived Collision Cross Section as an Additional Identification Point for Multiresidue Screening of Pesticides in Fish Feed. Anal. Chem. 88, 11169-11177.

Regueiro, J., Negreira, N., Hannisdal, R., Berntssen, M.H.G., 2017. Targeted approach

for qualitative screening of pesticides in salmon feed by liquid chromatography coupled to traveling-wave ion mobility/quadrupole time-of-flight mass spectrometry. Food Control 78, 116-125.

Reingruber, E., Buchberger, W., 2010. Analysis of polyolefin stabilizers and their degradation products. J. Sep. Sci. 33, 3463-3475.

Revercomb, H.E., Mason, E.A., 1975. Theory of plasma chromatography/gaseous electrophoresis. Review. Anal. Chem. 47, 970-983.

Richardson, K., Langridge, D., Dixit, S.M., Ruotolo, B.T., 2021. An Improved Calibration Approach for Traveling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. Anal. Chem. 93, 3542-3550.

Richardson, K., Langridge, D., Giles, K., 2018. Fundamentals of travelling wave ion mobility revisited: I. Smoothly moving waves. Int. J. Mass Spectrom. 428, 71-80.

Ridgeway, M.E., Lubeck, M., Jordens, J., Mann, M., Park, M.A., 2018. Trapped ion mobility spectrometry: A short review. Int. J. Mass Spectrom. 425, 22-35.

Righetti, L., Bergmann, A., Galaverna, G., Rolfsson, O., Paglia, G., Dall'Asta, C., 2018. Ion mobility-derived collision cross section database: Application to mycotoxin analysis. Anal. Chim. Acta 1014, 50-57.

Righetti, L., Dreolin, N., Celma, A., McCullagh, M., Barknowitz, G., Sancho, J.V., Dall'Asta, C., 2020. Travelling Wave Ion Mobility-Derived Collision Cross Section for Mycotoxins: Investigating Interlaboratory and Interplatform Reproducibility. J. Agric. Food Chem. 68, 10937-10943.

Ross, D.H., Cho, J.H., Xu, L., 2020. Breaking Down Structural Diversity for Comprehensive Prediction of Ion-Neutral Collision Cross Sections. Anal. Chem. 92, 4548-4557.

Ross, D.H., Xu, L., 2021. Determination of drugs and drug metabolites by ion mobility-mass spectrometry: A review. Anal. Chim Acta 1154, 338270.

Rouvray, D.H., King, R.B., 2002. Topology in chemistry: Discrete mathematics of molecules. Elsevier.

Ruotolo, B.T., Benesch, J.L.P., Sandercock, A.M., Hyung, S.-J., Robinson, C.V., 2008. Ion mobility–mass spectrometry analysis of large protein complexes. Nat. Protoc. 3, 1139-1152.

Rusko, J., Perkons, I., Rasinger, J.D., Bartkevics, V., 2020. Non-target and suspected-target screening for potentially hazardous chemicals in food contact materials: investigation of paper straws. Food Addit. Contam. Part A 37, 649-664.

Saleem, A., Harris, C.S., Asim, M., Cuerrier, A., Martineau, L., Haddad, P.S., Arnason, J.T., 2010. A RP-HPLC-DAD-APCI/MSD method for the characterisation of medicinal Ericaceae used by the Eeyou Istchee Cree First Nations. Phytochem. Anal. 21, 328-339.

Samsonek, J., Puype, F., 2013. Occurrence of brominated flame retardants in black thermo cups and selected kitchen utensils purchased on the European market. Food Addit. Contam. Part A 30, 1976-1986.

Scarsella, J.B., Zhang, N., Hartman, T.G., 2019. Identification and Migration Studies of Photolytic Decomposition Products of UV-Photoinitiators in Food Packaging. Molecules 24, 3592.

Schmidt, N., Fauvelle, V., Ody, A., Castro-Jimenez, J., Jouanno, J., Changeux, T., Thibaut, T., Sempere, R., 2019. The Amazon River: A Major Source of Organic Plastic Additives to the Tropical North Atlantic? Environ. Sci. Technol. 53, 7513-7521.

Schroeder, M., Meyer, S.W., Heyman, H.M., Barsch, A., Sumner, L.W., 2019. Generation of a Collision Cross Section Library for Multi-Dimensional Plant Metabolomics Using UHPLC-Trapped Ion Mobility-MS/MS. Metabolites 10, 13.

Schweighuber, A., Fischer, J., Buchberger, W., 2021. Differentiation of Polyamide 6, 6.6, and 12 Contaminations in Polyolefin-Recyclates Using HPLC Coupled to

Drift-Tube Ion-Mobility Quadrupole Time-of-Flight Mass Spectrometry. Polymers 13, 2032.

Shahidi, F., Zhong, Y., 2015. Measurement of antioxidant activity. J. Funct. Foods 18, 757-781.

Shang, Y., Wei, W., Zhang, P., Ye, B.C., 2020. Engineering Yarrowia lipolytica for Enhanced Production of Arbutin. J. Agric. Food Chem. 68, 1364-1372.

Sharma, V., Goswami, R., Madan, A.K., 1997. Eccentric Connectivity Index: A Novel Highly Discriminating Topological Descriptor for Structure−Property and Structure−Activity Studies. J. Chem. Inf. Comput. Sci. 37, 273-282.

Shaw, S., Harris, J., Berger, M., Subedi, B., 2014. Brominated Flame Retardants and Their Replacements in Food Packaging and Household Products: Uses, Human Exposure, and Health Effects, pp. 61-93.

Shi, J., Xu, C., Xiang, L., Chen, J., Cai, Z., 2020. Tris(2,4-di-tert-butylphenyl)phosphate: An Unexpected Abundant Toxic Pollutant Found in PM2.5. Environ. Sci. Technol. 54, 10570-10576.

Shvartsburg, A.A., Smith, R.D., 2008. Fundamentals of Traveling Wave Ion Mobility Spectrometry. Anal. Chem. 80, 9689-9699.

Silveira, J.A., Ridgeway, M.E., Park, M.A., 2014. High resolution trapped ion mobility spectrometery of peptides. Anal. Chem. 86, 5624-5627.

Sobeh, M., Rezq, S., Sabry, O.M., Abdelfattah, M.A.O., El Raey, M.A., El-Kashak, W.A., El-Shazly, A.M., Mahmoud, M.F., Wink, M., 2019. Albizia anthelmintica: HPLC-MS/MS profiling and in vivo anti-inflammatory, pain killing and antipyretic activities of its leaf extract. Biomed. Pharmacother. 115, 108882.

Song, H.J., Chang, Y., Lyu, J.S., Yon, M.Y., Lee, H.-s., Park, S.-J., Choi, J.C., Kim, M., Han, J., 2018. Migration study of caprolactam from polyamide 6 sheets into food

simulants. Food Sci. Biotechnol. 27, 1685-1689.

Song, X.-C., Canellas, E., Asensio, E., Nerín, C., 2020. Predicting the antioxidant capacity and total phenolic content of bearberry leaves by data fusion of UV–Vis spectroscopy and UHPLC/Q-TOF-MS. Talanta 213, 120831.

Song, X.-C., Canellas, E., Dreolin, N., Goshawk, J., Nerin, C., 2022a. Prediction of Collision Cross Section Values for Extractables and Leachables from Plastic Products. Environ. Sci. Technol. Accept.

Song, X.-C., Wrona, M., Nerin, C., Lin, Q.-B., Zhong, H.-N., 2019. Volatile non-intentionally added substances (NIAS) identified in recycled expanded polystyrene containers and their migration into food simulants. Food Packag. Shelf Life 20, 100318.

Song, X.C., Canellas, E., Dreolin, N., Goshawk, J., Nerin, C., 2022b. A Collision Cross Section Database for Extractables and Leachables from Food Contact Materials. J. Agric. Food Chem. 70, 4457-4466.

Song, X.C., Canellas, E., Dreolin, N., Nerin, C., Goshawk, J., 2021. Discovery and Characterization of Phenolic Compounds in Bearberry (Arctostaphylos uva-ursi) Leaves Using Liquid Chromatography-Ion Mobility-High-Resolution Mass Spectrometry. J. Agric. Food Chem. 69, 10856-10868.

Song, X.C., Dreolin, N., Damiani, T., Canellas, E., Nerin, C., 2022c. Prediction of Collision Cross Section Values: Application to Non-Intentionally Added Substance Identification in Food Contact Materials. J. Agric. Food Chem. 70, 1272-1281.

Soper-Hopper, M.T., Petrov, A.S., Howard, J.N., Yu, S.S., Forsythe, J.G., Grover, M.A., Fernandez, F.M., 2017. Collision cross section predictions using 2-dimensional molecular descriptors. Chem. Commun. 53, 7624-7627.

Soper-Hopper, M.T., Vandegrift, J., Baker, E.S., Fernandez, F.M., 2020. Metabolite collision cross section prediction without energy-minimized structures. Analyst

145, 5414-5418.

Stärker, C., Welle, F., 2019. Migration of bisphenol A from can coatings into beverages at the end of shelf life compared to regulated test conditions. Beverages 5, 3.

Stander, M.A., Van Wyk, B.E., Taylor, M.J.C., Long, H.S., 2017. Analysis of Phenolic Compounds in Rooibos Tea (Aspalathus linearis) with a Comparison of Flavonoid-Based Compounds in Natural Populations of Plants from Different Regions. J. Agric. Food Chem. 65, 10270-10281.

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., 2017. An Interlaboratory Evaluation of Drift Tube Ion Mobility-Mass Spectrometry Collision Cross Section Measurements. Anal. Chem. 89, 9048-9055.

Su, Q.-Z., Vera, P., Nerín, C., Lin, Q.-B., Zhong, H.-N., 2021a. Safety concerns of recycling postconsumer polyolefins for food contact uses: Regarding (semi-)volatile migrants untargetedly screened. Resour. Conserv. Recycl. 167, 105365.

Su, Q.-Z., Vera, P., Salafranca, J., Nerín, C., 2021b. Decontamination efficiencies of post-consumer high-density polyethylene milk bottles and prioritization of high concern volatile migrants. Resour. Conserv. Recycl. 171, 105640.

Sudjaroen, Y., Hull, W.E., Erben, G., Wurtele, G., Changbumrung, S., Ulrich, C.M., Owen, R.W., 2012. Isolation and characterization of ellagitannins as the major polyphenolic components of Longan (Dimocarpus longan Lour) seeds. Phytochemistry 77, 226-237.

Sushko, I., Novotarskyi, S., Korner, R., Pandey, A.K., Rupp, M., Teetz, W., Brandmaier, S., Abdelaziz, A., Prokopenko, V.V., Tanchuk, V.Y., Todeschini, R., Varnek, A., Marcou, G., Ertl, P., Potemkin, V., Grishina, M., Gasteiger, J., Schwab, C., Baskin, II, Palyulin, V.A., Radchenko, E.V., Welsh, W.J., Kholodovych, V., Chekmarev, D., Cherkasov, A., Aires-de-Sousa, J., Zhang, Q.Y., Bender, A., Nigsch, F., Patiny,

L., Williams, A., Tkachenko, V., Tetko, I.V., 2011. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J. Comput.-Aided Mol. Des. 25, 533-554.

Szöcs, E., Stirling, T., Scott, E.R., Scharmüller, A., Schäfer, R.B., 2020. webchem: An R Package to Retrieve Chemical Information from the Web. J. Stat. Softw. 93, 1-17.

Tejada-Casado, C., Hernandez-Mesa, M., Monteau, F., Lara, F.J., Olmo-Iruela, M.D., Garcia-Campana, A.M., Le Bizec, B., Dervilly-Pinel, G., 2018. Collision cross section (CCS) as a complementary parameter to characterize human and veterinary drugs. Anal. Chim. Acta 1043, 52-63.

Thomson, J.J., Rutherford, E., 1896. On the passage of electricity through gases exposed to Röntgen rays. Philos. Mag. 42, 392-407.

Tisler, S., Christensen, J.H., 2022. Non-target screening for the identification of migrating compounds from reusable plastic bottles into drinking water. J. Hazard. Mater. 429, 128331.

Tsochatzis, E.D., Alberto Lopes, J., Kappenstein, O., Tietz, T., Hoekstra, E.J., 2020. Quantification of PET cyclic and linear oligomers in teabags by a validated LC-MS method - In silico toxicity assessment and consumer's exposure. Food Chem. 317, 126427.

Ubeda, S., Aznar, M., Nerin, C., 2018. Determination of oligomers in virgin and recycled polyethylene terephthalate (PET) samples by UPLC-MS-QTOF. Anal. Bioanal. Chem. 410, 2377-2384.

Ubeda, S., Aznar, M., Rosenmai, A.K., Vinggaard, A.M., Nerin, C., 2020. Migration studies and toxicity evaluation of cyclic polyesters oligomers from food packaging adhesives. Food Chem. 311, 125918.

Venter, P., Causon, T., Pasch, H., de Villiers, A., 2019. Comprehensive analysis of chestnut tannins by reversed phase and hydrophilic interaction chromatography

coupled to ion mobility and high resolution mass spectrometry. Anal. Chim Acta 1088, 150-167.

Vera, P., Canellas, E., Barknowitz, G., Goshawk, J., Nerin, C., 2019. Ion-Mobility Quadrupole Time-of-Flight Mass Spectrometry: A Novel Technique Applied to Migration of Nonintentionally Added Substances from Polyethylene Films Intended for Use as Food Packaging. Anal. Chem. 91, 12741-12751.

Vera, P., Canellas, E., Nerin, C., 2018. Identification of non volatile migrant compounds and NIAS in polypropylene films used as food packaging characterized by UPLC-MS/QTOF. Talanta 188, 750-762.

Warnke, S., Seo, J., Boschmans, J., Sobott, F., Scrivens, J.H., Bleiholder, C., Bowers, M.T., Gewinner, S., Schollkopf, W., Pagel, K., von Helden, G., 2015. Protomers of benzocaine: solvent and permittivity dependence. . Am. Chem. Soc 137, 4236-4242.

Waters Corporation, 2020a. Metabolic Profiling CCS Library https://marketplace.waters.com/apps/177290/metabolic-profiling-ccslibrary#!overview.

Waters Corporation, 2020b. Natural Products Profiling CCS Library https://marketplace.waters.com/apps/255816/naturalproductsprofiling-ccs-library#!overview.

Waters Corporation, 2021. FDA approved drugs profling CCS library, https://marketplace.waters.com/apps/338581/fda-approved-drugs#!overview.

Webb, I.K., Garimella, S.V.B., Tolmachev, A.V., Chen, T.-C., Zhang, X., Norheim, R.V., Prost, S.A., LaMarche, B., Anderson, G.A., Ibrahim, Y.M., Smith, R.D., 2014. Experimental Evaluation and Optimization of Structures for Lossless Ion Manipulations for Ion Mobility Spectrometry with Time-of-Flight Mass Spectrometry. Anal. Chem. 86, 9169-9176.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R.,

Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686.

Wrona, M., Blasco, S., Becerril, R., Nerin, C., Sales, E., Asensio, E., 2019. Antioxidant and antimicrobial markers by UPLC((R))-ESI-Q-TOF-MS(E) of a new multilayer active packaging based on Arctostaphylos uva-ursi. Talanta 196, 498-509.

Wrona, M., Nerin, C., 2020. Analytical Approaches for Analysis of Safety of Modern Food Packaging: A Review. Molecules 25, 752.

Yang, Y., Hu, C., Zhong, H., Chen, X., Chen, R., Yam, K.L., 2016. Effects of Ultraviolet (UV) on Degradation of Irgafos 168 and Migration of Its Degradation Products from Polypropylene Films. J. Agric. Food Chem. 64, 7866–7873.

Zhang, A., Wan, L., Wu, C., Fang, Y., Han, G., Li, H., Zhang, Z., Wang, H., 2013. Simultaneous determination of 14 phenolic compounds in grape canes by HPLC-DAD-UV using wavelength switching detection. Molecules 18, 14241-14257.

Zhang, X., Quinn, K., Cruickshank-Quinn, C., Reisdorph, R., Reisdorph, N., 2018. The application of ion mobility mass spectrometry to metabolomics. Curr. Opin. Chem. Biol. 42, 60-66.

Zhao, Y.H., Abraham, M.H., Zissimos, A.M., 2003. Fast Calculation of van der Waals Volume as a Sum of Atomic and Bond Contributions and Its Application to Drug Compounds. J. Org. Chem. 68, 7368-7373.

Zheng, X., Aly, N.A., Zhou, Y., Dupuis, K.T., Bilbao, A., Paurus, V.L., Orton, D.J., Wilson, R., Payne, S.H., Smith, R.D., Baker, E.S., 2017a. A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. Chem. Sci. 8, 7724-7736.

Zheng, X., Renslow, R.S., Makola, M.M., Webb, I.K., Deng, L., Thomas, D.G., Govind, N., Ibrahim, Y.M., Kabanda, M.M., Dubery, I.A., Heyman, H.M., Smith, R.D.,

Madala, N.E., Baker, E.S., 2017b. Structural Elucidation of cis/trans Dicaffeoylquinic Acid Photoisomerization Using Ion Mobility Spectrometry-Mass Spectrometry. J. Phys. Chem. Lett. 8, 1381-1388.

Zhou, Z., Luo, M., Chen, X., Yin, Y., Xiong, X., Wang, R., Zhu, Z.J., 2020. Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. Nat. Commun. 11, 4334.

Zhou, Z., Shen, X., Tu, J., Zhu, Z.J., 2016. Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry. Anal. Chem. 88, 11084-11091.

Zhou, Z., Tu, J., Xiong, X., Shen, X., Zhu, Z.J., 2017. LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision To Support Ion Mobility-Mass Spectrometry-Based Lipidomics. Anal. Chem. 89, 9559-9566.

Zhu, M., Dong, X., Guo, M., 2015. Phenolic Profiling of Duchesnea indica Combining Macroporous Resin Chromatography (MRC) with HPLC-ESI-MS/MS and ESI-IT-MS. Molecules 20, 22463-22475.

Zimmermann, L., Bartosova, Z., Braun, K., Oehlmann, J., Völker, C., Wagner, M., 2021. Plastic Products Leach Chemicals That Induce In Vitro Toxicity under Realistic Use Conditions. Environ. Sci. Technol. 55, 11814-11823.