Food Monitoring by Untargeted Screening with Semi Quantification — Feasibility Study on Analysis of Cereals and Honey

Wang, Tingting

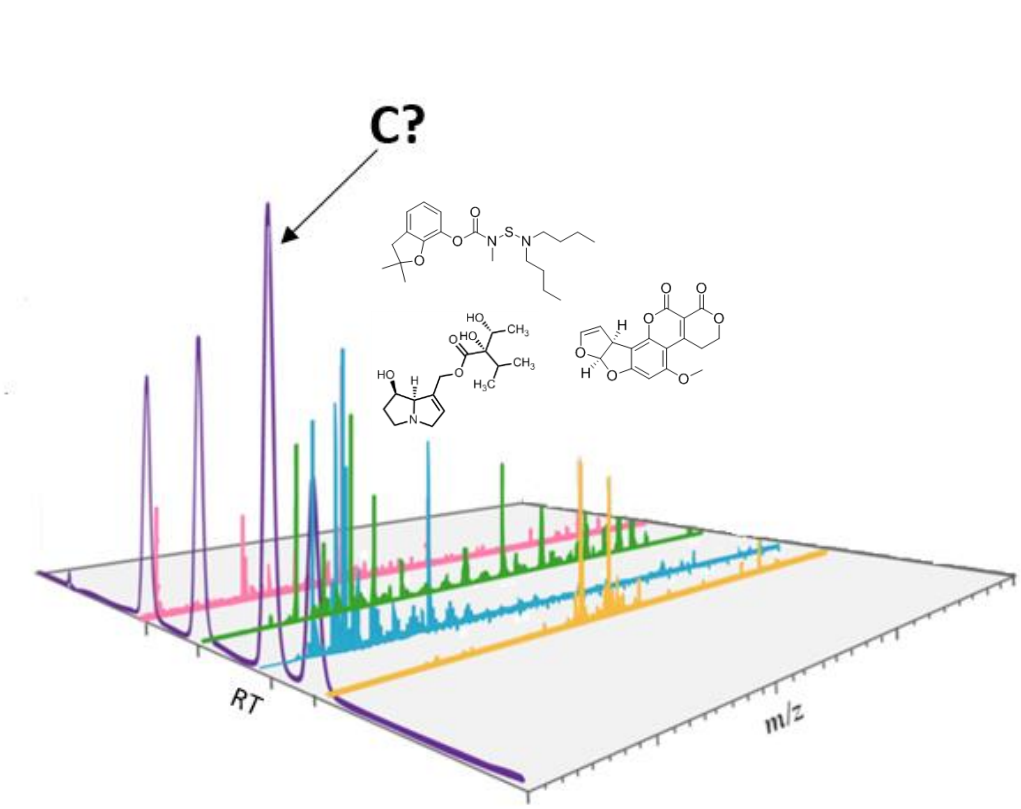
Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

National Food Institute
Technical University of Denmark

Food Monitoring by Untargeted Screening with Semi-Quantification

— Feasibility Study on Analysis of Cereals and Honey



Tingting Wang

PhD Thesis

2020

This page is intentionally left blank to preserve thesis layout of the electronic version.

Please **do not print this page**, as this will produce a wrong layout for the printed version.

Danmarks Tekniske Universitet

Food Monitoring by Untargeted Screening with Semi-Quantification

— Feasibility Study on Analysis of Cereals and Honey

PhD thesis

Tingting Wang

Bibliographic information

DTU Food
National Food Institute

1st Edition

http://www.food.dtu.dk

May, 2020

Title	Food Monitoring by Untargeted Screening with Semi-quantification – Feasibility Study on Analysis of Cereals and Honey
Author	Tingting Wang
Supervision	*Lene Duedahl-Olesen, Associate Professor, Technical University of Denmark* *Jørn Smedsgaard, Professor (former). Technical University of Denmark* *Henrik Lauritz Frandsen, Senior researcher (former). Technical University of Denmark*
Evaluation Committee	*Kit Granby, Associate Professor, Technical University of Denmark* *Nikolaos S. Thomaidis, Professor, National and Kapodistrian University of Athens* *Christian Janfelt, Associate Professor, University of Copenhagen*
Funding	The Ph.D. project was part of the Fødevareforlig 3 – Asterix project, funded by the Danish Ministry of Environment and Food.
Copyright	Tingting Wang National Food Institute, Technical University of Denmark
Cover	Tingting Wang
ISBN	
Availability	This thesis is available at: National Food Institute Technical University of Denmark Kemitorvet Building 202 2800 Kgs. Lyngby Denmark Web: www.food.dtu.dk E-mail: tiwa@food.dtu.dk Tel: + 45 35 88 70 00

Preface

This thesis presents the primary results of three years research from a PhD project entitled “Food Monitoring by High Resolution Mass Spectrometry”. The research was performed at the Research Group for Analytical Food Chemistry in the National Food Institute at the Technical University of Denmark as a PhD program from 1st June of 2016 until 20th May of 2020. This research was funded by Fødevareforlig III – Asterix project, partly funded by the ministry of environment and food.

The project was carried out under the supervision of Professor Dr. Jørn Smedsgaard and co-supervision by Senior Researcher Dr. Henrik Frandsen from June 2016 until June 2019. From July 2019 until May 2020, The supervision responsibility was transferred to Associate Professor Dr. Lene Duedahl-Olesen.

The external research stay was supervised by Associate Prof. Dr. Anneli Kruve and carried out in the Institute of Chemistry, Faculty of Science and Technology, University of Tartu (from August to November of 2018).

Acknowledgements

Now, I would like to use this opportunity to express all my gratitude to everyone who has supported and helped me during my PhD journey. I am also really appreciate that I got this chance to have my new chapter of my life here in this wonderful group from four years ago. There are a lot of people to thank for helping me compete the research program of this thesis.

I would like to first thank all my supervisors. Former supervisors, Jørn Smedsgaard is thanked for taking me into this group and providing me with great supports and guidance for my whole PhD, thanks for the in-depth, great discussion and advices on this project. Henrik L. Frandsen is thanked for the patient supervision and kindly help in or out of the lab, and also thanks for all the great scientific discussion. I would also like to thank my current supervisor, Lene Duedahl-Olesen, thanks for taking me as your student during my last half year of this PhD study. Thanks for your invaluable assistance in setting up everything and all the kindly support during the last part of my PhD journey.

Anneli Kruve is particularly thanked for her initial inspiration and great supervision during my external stay in the University of Tartu. Thanks for the nice cooperation we have had especially the semi-quantification model topic. Thanks for all the caring during my stay there.

I also thank all my dear colleagues and friends who accompany with me along with my PhD journey, my great DTU PhD fellows, Amelie, Demi, Petra, Philipp, Eelco, Michaela, Agnieszka, and Ida, my sweet colleagues and lunch partners, Helen, Elena. Amelie, I am so grateful to have you as such a sweet and sincere colleague and friend, thanks for all the greetings and smiles everyday. Demi, thanks for your occasionally home visit during my maternity leave and all the nice conversation. Petra, thanks for your nice cakes and also all your visit together with Amelie and Demi, which let me feel that I was not alone during that year with baby at home. Philipp, thanks for the cooperation during PhD and all your nice cookies. Helen, thanks for your kindly help for my paper proof and the lunch accompany outside when all my PhD colleagues went to external stay, especially the tough period when I suffered a lot sickness during my pregnancy. Special thanks for Jaanus from Estonia, who is a talented data scientist and really helped me a lot during my stay there, thanks for the cooperation. Mikael is thanked for the kindly help with validation report writing. There are

also many other sweet colleagues I would like to thank, Faranak, Liljana, Lisbet, Merete, Maud, Hoang, Alin, and all the colleagues I forget to mention here, thanks you all for the warm that I got from the big group at DTU.

There is one person I would like to thank the most, my husband Wanwan, thanks for all your kindly support, especially the most difficult one year after we had our little Mier. It is amazing that we still survive and stay together after the tough life with so much of expected and unexpected events happened. It has never been an easy road, the joy, trust, happiness, sadness, autocriticism, and experience we earned and leaned from this whole period that deserves us to keep and cherish in our whole life. It will always remind myself to be a better “me”.

Last but not least, thanks for the support of my whole family, my parents and my parents in-law. Thanks for the coming of my little Mier, you bring a lot of joy for your mum’s life, although it is not an easy way to take care of such a picky “you”. Thanks a lot to my mother in law, who helped a lot during my pregnancy and also the caring after Mier came. Thanks for all your kindly help.

Summary (English)

Trust of food is of great importance to our society. In order to ensure the food safety for now and for the future, the analytical strategy should be developed in parallel with the growing health concern and increased complexity in chemicals occurring in food. Many chemical compounds can be found in food, and a large number of them are unwanted. It is hard to include too many unwanted compounds into one specific targeted method. Therefore there has been growing interests to identify and quantify not only targeted and known compounds but also untargeted or “unknown” compounds in food that are missed in those targeted analysis. There is clearly a need for a robust and reliable untargeted analytical methods with wide scope that can cover more compounds, even standards are not available.

In this project, cereal and honey were selected as cases for the untargeted analysis feasibility study. The perspectives of this PhD project is to deal with the massive and complicate data come out of high resolution mass spectrometry (HRMS) and develop a generic, wide scope, and efficient analytical strategy for untargeted analysis, in particularly to get quantitative data from these when standards are not available. In the end, providing more data for food monitoring and risk assessment with less efforts. Mainly four parts were carried out in my project, namely, analytical method optimization, data processing method optimization, identification, and semi-quantification method development and application.

In analytical method optimization part, to achieve better sensitivity, ion source parameters and the eluent compositions were studied and optimized. Ion source optimization was first operated in flow injection mode on APCI and ESI. As no obvious advantage in APCI was found compared with ESI from the results, ESI was selected for the further study. The effects of four key source parameters in ESI on common food contaminants and residues were also studied with chromatography separation (Manuscript 1). A 40% increase in positive mode and 20% increase in the negative mode of response factor compared with central points was achieved by ion source optimization. In eluent composition optimization, pH 3.0 and pH 3.3 with ammonium formate buffer achieved the best performance for all test compounds.

In data processing method development part (Manuscript 1), the commonly used metabolomics approaches were optimized and compared regarding peak detection capacity and accuracy. These approaches were also compared with the performance of suspected

screening. Developed and highly automated workflow based on XCMS package was setup and achieved highest mass accuracy, highest detection rate (96%) and made a clear distinction between the control and spiked groups by multivariate analysis in a true untargeted way (metabolomics-like approach), even for the concentration of 5 µg/kg. Such approach could be a good complement to routine targeted analysis in a view of rapidly detecting potential contaminated food products without prior information and sacrificing too much accuracy.

In identification parts, a generic database were constructed including common chemical residues and contaminants according to EU regulations, standards available from our collection, identified or tentatively identified mycotoxins and metabolites from literature. In addition, the predicted log*P*/log*D* values were also registered into the database to minimize the number of false candidates. The tentative identification of compounds was also demonstrated by correlating fragmentation spectra from QToF-MS to *in silico* generated spectra from database to find the compounds with the best match. The constructed database was applied in the screening in both manuscript 1 and 2.

Semi-quantification method development and application took the largest proportion of my study. Quantitative prediction model (QPM) was applied for pyrrolizidine alkaloids (PAs) in honey, and pesticides and mycotoxins in cereals without using standards. In honey study, a targeted method was proposed that can be extended to a general strategy to achieve estimation of quantitative data when standards are unavailable (manuscript 2). The QPM by using multiple linear regression (MLR) was successfully quantified eight PAs without standards the first time and only 50.8% prediction error was achieved on QToF-MS. Another QPM using random forest was applied for cereal matrices and quantified 134 pesticides and 5 mycotoxins without standards (manuscript 3 and 4). This model was fist time validated on different instruments from different labs in this study, a high correlation was also achieved (R^2=0.86). Less than 4 times prediction errors were achieved for all test compounds, which is quite promising to be used in future food monitoring.

The untargeted strategy proposed in this study shows the semi-quantitative untargeted analysis is valuable tools for helping ensure food safe in the future, while the quality of identification and semi-quantification can always be further developed until it is more close to truth, but we moved one step further on this way.

Resumé (Dansk)

For forbrugerne er det vigtig at have tillid til fødevarerne. For at sikre denne tillid skal fødevaresikkerheden opretholdes og den analytiske strategi for kontrollen bør udvikles parallelt med den øgede bekymring for fødevaresundheden og den øgede kompleksitet af kemiske stoffer i fødevarerne. Mange af de kemiske stoffer der findes i fødevarerne er uønsket. Analytisk er det udfordrende at inkludere alt for mange kemiske stoffer i en og samme fødevareanalyse. Der er derfor en voksende interesse for at identificere og kvantificere ikke kun target, og dermed kendte stoffer, men også untarget, og dermed ukendte stoffer, som ikke detekteres med target analyse af fødevarerne. Der er behov for robuste og pålidelige untarget analyse metoder, som bredt kan dække flere stoffer, også når standard materiale er utilgængelig.

I dette projekt er cerealier og honning udvalgt som praktiske eksempler for anvendelsen af untarget analyse. Formålet med dette PhD projekt er at behandle den enorme mængde kompliceret data som højt opløsende masse spektrometri (HRMS) resulterer i og udvikle en generisk, effektiv og bredt dækkende analytisk strategi for untarget analyse med særlig fokus på at få kvantitative data hvor standardmateriale ikke er tilgængelige. Overordnet er målet at skaffe flere data til overvågning og risikovurdering af fødevarer ved samme arbejdsindsats som hidtil. Projektet fokuserede på fire dele; optimering af instrumentmetode og af databehandling, udvikling og anvendelse af identifikation og semi-kvantitative metoder.

Instrumentoptimering inkluderede ionkilde parametre og eluent sammensætningen for at opnå en bedre analytisk følsomhed. Optimeringen af ionkilden blev først udført med kontinuerlig direkte injektion i MS-delen med APCI og ESI. Da der ikke var nogen klar fordel ved anvendelsen af APCI blev den mest almindelig, nemlig ESI, valgt til de efterfølgende eksperimenter. Effekten af fire nøgle parametre i ESI på den kromatografiske separation af almindelige fødevare forureninger og restkoncentrationer blev undersøgt (Manuskript 1). En sammenligning med fastsatte udgangsparametre viste at positiv ESI resulterede i en 40% forøgelse af respons faktoren (peak areal/koncentration), mens negativ ESI kun viste 20% forøgelse. Den optimale eluent sammensætning var en eluent med pH 3.0 og pH 3.3 med ammonium formiat buffer, uanset stofgruppe.

Under optimering og udvikling af databehandlingen (Manuskript 1) blev almindelige metabolomics analyser optimeret og sammenlignet med hensyn til kromatografisk detektions kapacitet og nøjagtighed. Resultaterne blev ligeledes sammenlignet med screening for forventede stoffer, såkaldt suspect screening. Udviklet automatiseret databehandling baseret på XCMS software (ofte anvendt til metabolomics) gav størst masse nøjagtighed, detektion (96%) og kunne tydelig adskille kontrol grupper fra grupper med tilsætninger på 5 µg/kg ved multivariate analysis. Metoden kan blive et godt supplement til rutine target analyser p.g.a den hurtige, nøjagtige detektion af potentielle forurenede fødevarer uden forudgående information om forureningstypen.

For identification af ukendte stoffer blev en generel database, baseret på EU lovgivning og tilgængelige standarder i laboratoriet samt identificeret og delvis identificeret mykotoksiner med metabolitter fra litteraturen, udarbejdet for restkoncentrationer og forureninger. For at minimere antallet af falske positive kandidater og øge kvaliteten af identifikationen blev estimater af log*P*/log*D* værdier inkluderet i databasen. Den foreløbige identifikation af stofferne blev ligeledes påvist ved at korrelerer fragment spektre fra QToF-MS til *in silico* genereret spektre fra databasen for at finde de bedst match. Databasen blev anvendt til screening i både manuskript 1 og 2.

Udvikling og anvendelse af semi-kvantificering udgjorde hovedparten af mit projekt. De udviklede modeller blev anvendt til kvantificering af både pyrrolizidin alkaloider (PA) i honning samt pesticider og mykotoksiner i cerealier; begge uden brug af kommercielle standarder. Den udviklede kvantificeringsmetoden fra honning studiet, kan anvendes som en general strategi til estimering af kvantitative data når kommercielle standarder ikke er tilgængelige (Manuskript 2). En succesfuld kvantitativ estimeringsmodel (QPM) med multiple lineær regression (MLR) til kvantificering af otte PA'er i honning uden standarder resulterede i en estimeret fejl på 50.8% procent med QToF-MS. En anden kvantificeringsmodel baseret på random forest blev anvendt på cerealier til kvantificering af 134 pesticider og 5 mykotoksiner uden standarder (Manuskript 3 og 4). Den første validering af modellen med forskellige laboratorier og udstyr viste en god korrelation mellem de to laboratorier (R^2=0.86). Et fejlestimat under en faktor 4 var resultatet for samtlige inkluderet stoffer. Begge lovende resultater for anvendelsen i den fremtidige fødevarekontrol.

Den foreslåede untarget strategi i projektet viste at semi-kvantitativ untarget analyse er et værdifuldt værktøj til at sikre den fremtidige fødevaresikkerhed. Kvaliteten og fejlmargen af identifikation og semi-kvantificeringen vil som altid kunne udvikles yderligere. Dette projekt har ført os et skridt nærmere målet.

Abbreviations

3-MCPD	3-chloropropane-1,2-diol
MeCN	Acetonitrile
AFs	Aflatoxins
AOAC	Association of Official Agricultural Chemists
APCI	Atmospheric pressure chemical ionization
ANN	Artificial neural networks
bbCID	Broadband collision-induced dissociation
BfR	German Federal Institute for Risk Assessment
CA	Cluster analysis
CE	Capillary electrophoresis
CID	Collision-induced dissociation
CONTAM	European Food Safety Authority Panel on Contaminants in the Food Chain
DAS	Diacetoxyscirpenol
DDA	Data-dependent-acquisition
DFA	Discriminant function analysis
DIA	Data-independent acquisition
DMPF	N-2,4-Dimethylphenyl-N-Methyl-formamidine
DON	Deoxynivalenol
EI	Electron impact
EFSA	European Food Safety Authority
ESI	Electrospray ionization
EU	European Union
FAO	Food and Agriculture Organization
FA	Formic acid
FDA	The US Food and Drug Administration
FMs	Fumonisins
GC	Gas chromatography
HAC	Acetic acid
HRMS	High-resolution mass spectrometry
IAS	Intentionally added substances
LC	Liquid chromatography
LDA	Linear discriminant analysis
LOD	Limit of detection
LOQ	Limit of quantification
MeOH	Methanol
MLR	Multiple Linear Regression
MLs	Maximal levels
MS/MS	Multi-stage mass spectrometry

MS	Mass spectrometry
MRM	Multiple reaction monitoring
NH4OH	Ammonium hydroxide
NIST	National Institute of Standards and Technology
NIAS	Non-intentionally added substances
NIV	Nivalenol
NMR	Magnetic resonance spectroscopy
OTA	Ochratoxins
OPLS-DA	Orthogonal partial least squares discriminant analysis
PAs	Pyrrolizidine alkaloids
PAH	Polycyclic aromatic hydrocarbons
PAT	Patulin
PCA	Principal component analysis
PC	principal components
PCB	polychlorinated biphenyl
PLS-DA	Partial least squares-discriminant analysis
Q	Quadrupole
QIT	Quadrupole ion trap
QqQ	Triple-quadruple
QToF	quadrupole-time of flight
QPM	Quantitative prediction model
QuEChERS	Quick, Easy, Cheap Effective Rugged and Safe
SL	supervised learning
SRM	Selective reaction monitoring
SMILES	Simplified molecular-input line-entry system
SPE	Solid-phase extraction
ToF	Time of flight
UL	Unsupervised learning
WAPS	Weighted average positive sigma
WHO	The World Health Organization
ZEN	Zearalenone
η0	the viscosity at atmospheric pressure

List of manuscripts

Published manuscripts

Manuscript 1: **T. Wang**, H. L. Frandsen, N. R. Christiansson, S.E. Rosendal, M. Pedersen, J. Smedsgaard*, Pyrrolizidine alkaloids in honey: Quantification with and without standards, ***Food Control***, 2019, 98, 227-237.

Manuscript 2: **T. Wang**, J. Liigand, H. L. Frandsen, J. Smedsgaard, A. Kruve*, Standard substances free quantification makes LC/ESI/MS non-targeted screening of pesticides in cereals comparable between labs. ***Food Chemistry**, 2020.* 318, 126460

Manuscript 3: J. Liigand, **T. Wang**, J. J. Kellogg, J. Smedsgaard, N. B. Cech, A. Kruve*, Quantifying the unquantifiable: Quantification for non-targeted LC/MS screening without standards. ***Scientific Reports**,* 2020*.* 10, 5808

Submitted manuscript

Manuscript 4: **T. Wang***, H. L. Frandsen, L. Duedahl-Olesen, Non-targeted food unexpected compounds analysis by LC/HRMS: feasibility study on rice. ***Food Chemistry** (under review)*

Dissemination

Conference abstract I:

Date: June 6, 2019

Conference: 67th ASMS Conference on Mass Spectrometry and Allied Topics

Organiser: ASMS

Location: Atlanta, US

Type: Poster

Title: Standard substance free quantification of LC/ESI/MS on the example of pesticides in cereal

Authors: Jaanus Liigand; Tingting Wang; Piia Liigand; Mari Ojakivi; Anneli Kruve

Conference abstract II:

Date: June 1st – June 12th , 2020

Conference: ASMS 2020 Reboot/ 68th ASMS Conference on Mass Spectrometry and Allied Topics

Organiser: ASMS

Location: Online form

Type: Poster

Title: Food monitoring feasibility study on cereal: non-targeted food contaminants detection method development and semi-quantification model application using LC/HRMS

Authors: Tingting Wang, Jaanus Liigand, Anneli Kruve, Lene Duedahl-Olesen

Table of Contents

1. Introduction

1.1. Food control and food monitoring

In an age of increasing global food trade and rapidly changing food technologies, controlling the food hazards along the food chain has taken on a critical role. Food control is a mandatory regulatory activity of enforcement by national or local authorities to provide consumer protection, fulfil food safety and quality requirements, and to ensure that food is honestly and accurately labelled as the legislations (FAO, 2020a). A food control system is the integration of mandatory regulatory approaches with preventive and educational strategies along the whole food chain (production, handling, storage, processing and distribution) as shown in Figure 1-1 (WHO & FAO, 2019). The protection of consumer's health is of outmost importance. Therefore, it is crucial to have an efficient food control system in place that ensures the trade of food complying with high safety and quality standards both internationally and nationally.

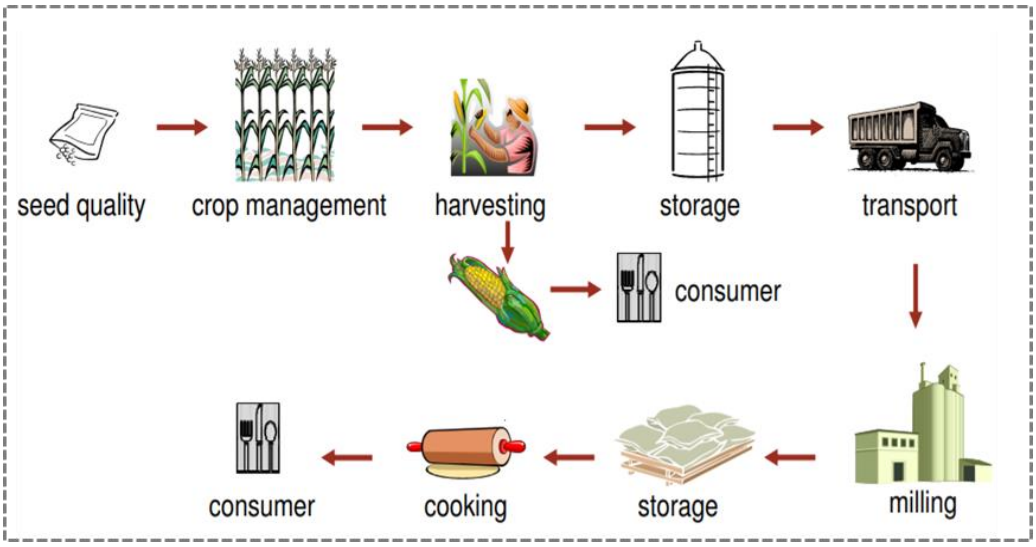


Figure 1-1. Example of value chain for maize: various stages along a food chain at which efforts are required to ensure food safety (coordinated approach by businesses operators, authorities, customers etc. can ensure the safety) (FAO, 2020b)

In order to operate effectively, a food control system requires applicable legal and policy instruments, well-qualified human resources, sound institutional frameworks, and equipment and infrastructure as its foundation. The framework for the development of a national food control system is shown in Figure 1-2, continuous improvement from policy setting, system

designing and monitoring parts is needed for the successful implementation of the system. In addition, in order to manage the emerging challenges and maintain the quality and safety of food, it is important for a food control system to incorporate risk analysis principles, fulfils the international standards and keep pace with the new scientific developments to continuously improve the efficiency and effectiveness through food monitoring (FAO, 2020a).

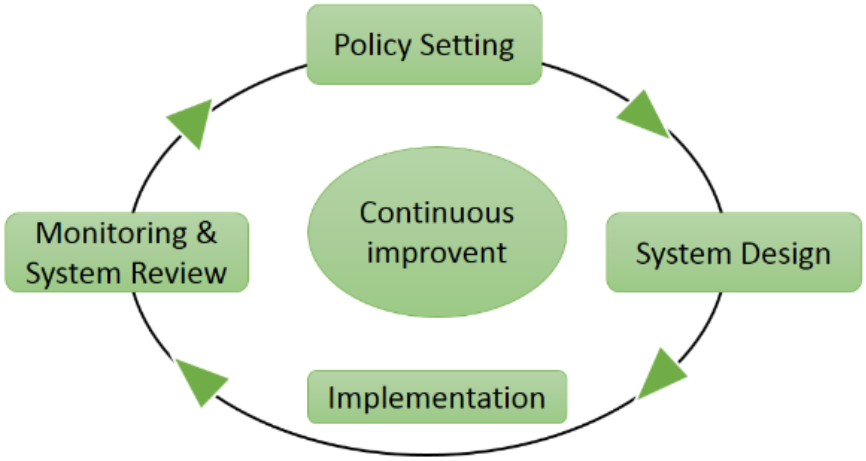


Figure 1-2. Framework of the national food control development, the design and implementation of a national food control system should include the consistent application of a systematic framework for the identification, evaluation and control of food safety risks associated with food hazards (FAO & WHO, 2010)

Food monitoring provides the information on occurrence and the level of chemical contaminants or residues hazards in predefined legislations that statistically based sampling, processing and analysis of the food samples to (WHO & FAO, 2009). The purpose of food monitoring is to ensure that food safety hazards are under control, all procedures are being correctly implemented by regularly verifying and tracking of the implementation of surveillance activities, overall performance of the surveillance and response system (FAO, 2020a; WHO & FAO, 2009). We would not be able to support the food safety status of the food production and supply without food monitoring.

All participants (food business operators, consumers, the national government, academics and scientific institution) in a food control system are supposed to have specific roles and responsibilities. Academics and scientific institutions as well as organizations have an important role in contributing to a food control system. In the fact that they are a source of

expertise to support scientific foundation of this system. As the scientific research part, the responsibility is to make use of the information generated by the official controls, contribute to the legislation drafting, continuously improve food monitoring strategy, keep pace of the ever-increasing global food market and the increasing food safety concerns. The new scientific developments in food monitoring analytical methods are certainly areas that need to be further developed in the next iterations of the food control system.

1.2. Why untargeted screening is so attractive?

1.2.1. The chemical challenge of food

Chemical food safety issues have always been great concerned to the society, and to researchers who are working in agriculture and food scientific community. In the past several decades, chemical food safety has emerged as a significant global problem and international trade implications. From chemical safety point view, many chemical compounds can be found in food samples, and a large number of them are unwanted. Unwanted chemicals in food may originate from various sources and possess various different chemical structures, which can be divided into residues and contaminants. Residues originate from intentionally added or used substances (IAS), which might come along food production (e.g. pesticides and veterinary drugs), storage or transportation. Contaminants are usually non-intentionally added substances (NIAS) of natural origin that have not been intentionally added to or formed by food itself (e.g., alkaloids), formed during production, processing (e.g. polycyclic aromatic hydrocarbons-PAH), transportation or storage (e.g., mycotoxins), or from the environment. A third group are compounds originating from malpractice or deliberate fraud (e.g. melamine) (Jackson, 2009). Unwanted compounds occurring in food can also be classified into known and unknown compounds: The known-knowns are the compounds that are known to investigator and likely present in the sample. Information of known-knowns can be searched in database and confirmed with standards. In contrast, the unknown compounds are those unknown to the investigator, they can be separated in two categories: known unknowns (unknown to investigator but actually known in the chemical literature) and unknown unknowns (unknown to an investigator and the structures are not yet identified or reported in literature) (Little et al., 2012). Maximum levels (MLs) for contaminants and residues such as, pesticides, mycotoxins (aflatoxins, ochratoxin A, fusarium-toxins, patulin,

citrinin), dioxins, polychlorinated biphenyl (PCBs), PAH, 3-chloropropane-1,2-diol (3-MPCD), and melamine are set in the European Union Legislation (European Commission, 2020).

A large number of unwanted chemicals in food can constitute into a food safety risk. Some chemicals have intrinsic hazards and some are toxic at certain concentrations. Many unwanted compounds in food are formed by the industry, e.g. pesticides, IAS from packaging materials, within which most of them may find their way into the final food supply. Similarly, byproducts like dioxins which are very toxic and persistent organic pollutants may also find their way into food as they are common existing in the environment. Pesticides may cause acute and chronic health effects in exposed humans, such as simple irritation of the skin and eyes, affecting the nervous system and inducing cancer ("United States Environmental Protection Agency," 2019). Fungi growing on food either in fields or in storage form mycotoxins, these food spoilage organisms are associated with a huge loss of food and numerous health disorders in human and animals. Mycotoxins may cause acute or chronic disease episodes, with carcinogenic, mutagenic, teratogenic, estrogenic, nephrotoxic, hepatotoxic, neurotoxic or immunosuppressive effects (Pereira, Fernandes, & Cunha, 2014). In addition, unwanted chemical compounds in food also undergo their transformation and/or degradation, and many of them are not even known yet and not included in the control and screening method, and they can also be hazardous. Therefore, how to monitor this large number of unwanted compounds in food and ensure the food safety and quality is always a challenge for both, authorities and researchers.

1.2.2. The important role of untargeted screening

In order to monitor the numerous residues and contaminants in food, analytical methodologies must be able to identify and accurately quantify usually very low concentration compounds in food samples. Mass spectrometry (MS) particularly high-resolution mass spectrometry (HRMS) is extremely useful for measuring food hazardous compounds within low abundance by providing valuable information. The current MS based approaches for food monitoring are mainly based on risk-based analytical programs using specific analytical methodologies (targeted analysis) that detect only what the methods are developed for.

“Targeted analysis” is the conventional analysis based on establishing a method with reference standards prior to analysis (Garcı, Hernando, & Molina-dı, 2007) (Figure 1-3), which is sensitive, standardized and very specific. Targeted analysis has been the main methodology to safeguard food against a number of chemical hazards. This analytical strategy usually needs a number of specific targeted methods to monitor all relevant regulated chemicals, which is expensive, time consuming and also limited in scope due to difficulty of developing a wide-scope method (Tengstrand, Rosén, Hellenäs, & Åberg, 2013). One example was when melamine which was used as an adulterant in milk to increase the apparent protein content. As melamine has previously not been monitored by the routine targeted analysis (Skinner, Thomas, & Osterloh, 2010), this scandal was responsible for multiple illnesses and deaths among infants and pets, which underline the limits of current targeted analytical method and need for new analytical approaches.

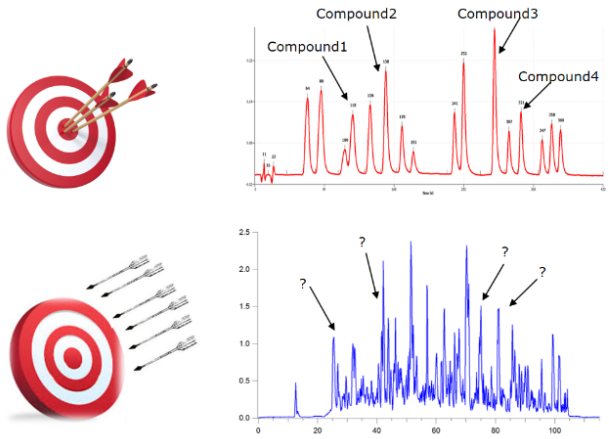


Figure 1-3. Targeted and untargeted analysis. Targeted analysis is based on establishing a method with reference standards prior to analysis (Garcı, Hernando and Molina-dı, 2007). Untargeted analysis starts with full-scan data, no priori information or reference standards is available

To keep pace with the requirement for chemical food monitoring, there is clearly a need for robust and reliable untargeted analysis methods with a wide scope that can cover more compounds and are available to many stakeholders. As the unwanted compounds that can be found in food are numerous, it is hard to include a large number of unwanted compounds into one specific targeted method. Therefore, there has been growing interests to identify and quantify not only targeted compounds but also untargeted or “unknown” compounds in food samples that are missed in targeted analysis. All in all, makes untargeted screening so attractive to the society and to the research community.

Table 1-1 Comparison between targeted analysis, suspected screening and untargeted screening

Analytical techniques	Targeted analysis	Suspected screening	Untargeted screening
Terms in literatures	Targeted analysis	Targeted screening/untargeted (non-targeted) screening/ /post-targeted analysis/ descreening	Untargeted (non-targeted) analysis/ true untargeted (non-targeted) analysis/true untargeted (non-targeted) screening
Category used in this project	Targeted	Untargeted	Untargeted
Questions	Are compounds a, b, c present in the samples?	Which compounds of a defined list are present in the sample?	Which compound are present in the samples
Compounds type	Known-knowns	Known-unknowns	Known-unknowns & unknown-unknowns
Standard needed or not	Yes	No	No
Pre-selected list needed or not	No	Yes	No
Possibility of identification	Full identification	Tentative identification	Tentative identification
Possibility for quantification	Yes	Semi-quantification	Semi-quantification

Untargeted screening starts with full-scan data. No priori information or reference standards are available in untargeted analysis. So called “suspected screening” partially correspond to untargeted screening, it is based on a preselected list of possible compounds that summarized from literatures, and the samples are screened for those potential chemical candidates. As pre-selected list is still needed, suspected screening sometimes is also named as “post-targeted screening” (Díaz, Ibáñez, Sancho, & Hernández, 2012). In many studies, “suspected screening” is classified into untargeted analysis (Hird, Lau, Schuhmacher, & Krska, 2014; Veenaas & Haglund, 2017). Descreening of known compounds in suspected screening is important to find all those are known and start looking for what remains unknown afterwards. True untargeted screening starts with the MS data and aims to reveal what is in the sample, or what is the difference between samples, and it is not limited to the preselected list/database.

The difference between these screening techniques are summarized in Table 1-1. Its further explanation on workflow and difficulties can be found in section 2.3.2. Both suspected screening and true untargeted screening correspond to “untargeted”, and both of them are

not limited to the standards and database, which are highly needed as a complement to targeted analysis (Croley, White, Callahan, & Musser, 2012).

1.3. Pitfalls and gains using untargeted approach

Untargeted analysis provides the great possibility of analysing a wide range of compounds that is not limited to the availability of reference standards in a specified method. One could extend monitoring to any new suspected compounds, metabolites or other transformation products. It also allows retrospective analysis after months or even years.

However, on the one hand, due to the diversity of unwanted compounds in food such as pesticides and mycotoxins, it is impossible to detect all compounds by a single analytical method. By this fact, how to develop a generic method that includes as many compounds as possible is the first big challenge. Therefore, both generic sample extraction methods and generic analytical method are highly demanded. Getting compound out of the sample is a huge and complex challenge that takes a lot of efforts, while it is not the focus in this project. The focus of this project is to address the biggest problem and challenge in developing generic untargeted analytical method after the sample preparation.

On the other hand, it has to be concerned that investigators who tried to solve their analytical problems in MS are sometimes disappointed by the lack of speed and the limited user-friendly data processing software or workflows. Data treatment of untargeted analysis is usually most time consuming part. There could be hundreds or even thousands of peaks present in a single chromatogram, and there is no common definition in terms data treatment procedures leading to difficultly in interpretation of data. To many users, as well as authorities, chemometrics is a “BlackBox” in where very limited understanding of what is happening inside. Meanwhile, the development and implementation of untargeted screening workflows are particularly challenging, especially for food matrices due to the sample complexity, diversity and a large analytes concentration range (McGrath et al., 2018). A more cost effective, automatic, accurate and all-inclusive data processing approach is needed. Therefore, how to extract and integrate the useful information from the massive data that comes from HRMS is also a big challenge.

Furthermore, quantification of compounds in untargeted analysis is another big challenge. It is usually quite simple to do quantification in targeted analysis with the aid of reference

standards, just requires simple calculation of the concentrations based on the calibration curves using peak area of reference and internal standards. However, in the reality of food untargeted analysis, many standards are not commercially available, newly discovered, or there are too many compounds need to be quantified or the standards are not included in the run. In some cases, a certain number of compounds can be identified and then ended up with a few structure candidates. Lacking chemical standards makes the quantification in untargeted of compounds very difficult or even impossible. For example, it is common that pesticide residue analysis includes 500 or more analytes in the routine scope and the same number of reference standards need to be purchased and included in such methods (X. Wang, Wang, & Cai, 2013). It becomes either impossible or too expensive to purchase all these compounds as numerous tentatively identified compounds are often found in a single untargeted analysis. How we can know the concentration without using numerous and costly standards s is a big problem that needs to be solved in untargeted analysis. Therefore, a robust and reliable semi-quantification method applied for the unwanted compounds in untargeted food analysis is of high value and importance.

In this project, we selected cereal and honey as examples for the feasibility study of untargeted analysis. Cereal is one of the largest food commodity and consumed worldwide (González-Curbelo et al., 2012), which usually suffers the problems of containing various unwanted compounds such as pesticides, mycotoxins and environmental contamination. Honey is widely consumed as pure or used as food ingredients, while it is often neglected that it is one of the food products that contain the the naturally occurring plant toxins belong to pyrrolizidine alkaloids (PAs). PAs are suspected to be hepatotoxic and genotoxic carcinogens in human (Mulder et al., 2018). Both cereal and honey are facing the problems with lacking efficient analytical methods and reference standards.

In general, the objective of this project is to deal with the massive and complicate data that come out of HRMS and develop a generic and efficient analytical strategy with a wide scope for untargeted analysis in the case of honey and cereal. In particular to get quantitative data from these when reference standards are not available (either not commercially available or not included in the method). In brief, this research project envisioned to address three key challenges in the chemical food monitoring:

1. The increased complexity and rate of changes in chemicals occurring in food call for a generic analytical method with wider scope in analytical procedures;
2. The analytical methods and data processing method used for chemical food monitoring are in general complex, expensive and relative slow;
3. It is very difficult to get quantitative data from complex HRMS data without using reference standards and to carry out quantification and identification simultaneously in untargeted analysis.

1.4. Hypotheses

For the future food monitoring, it is possible to develop a generic analytical strategy with wide scope that combines targeted analysis and untargeted analysis. A more generic, efficient, and economical monitoring strategy compared with current methods in cereal and honey can be developed. This strategy that based on screening analyses, multivariate data analysis and response prediction model is able to qualify and quantify a wider range of unwanted compounds, even if the authentic standards are not unavailable. Ultimately, this strategy can be applied in the food monitoring and provide more data for the future food risk assessment.

1. The response factor of compounds in mass spectrometry can be optimized by carefully choosing the MS source parameters and LC parameters for simultaneous analysis and quantification.
2. By exploiting the rich data from HRMS in an untargeted screening strategy, an analytical method including efficient and highly automatic data processing method with less efforts can be provided and used to obtain fast information readout about unwanted compounds in food matrices by LC/HRMS.
3. A prediction model that allows the estimation of concentrations of compounds in HRMS samples without using standards can be provided and applied in the quantification of unwanted compounds in honey and cereal.
4. Untargeted food safety assessment of cereals can be supported by applying multivariate analysis to datasets that were obtained by LC/HRMS analyses.

2. Background

2.1. Honey and cereal case studies

In this project, we selected cereal and honey as examples of case studies to the feasibility study of untargeted analysis. Both cereal and honey are widely consumed food products and possess huge possibilities of containing unwanted compounds, which is illustrated in the following sections. Meanwhile, both honey and cereal are facing the challenge of lacking efficient untargeted screening methods as well as lacking authentic standards. It can be found that pesticides, mycotoxins, plant toxins and their degraded products, and metabolites are the main issues in cereal and honey that need to be monitored, and the number is very large. It would be highly valuable to develop generic and efficient analytical strategies to deal with the food monitoring problems involved in those two food species.

2.1.1. Honey

Honey is widely consumed pure or used as an ingredient in fruit products, breakfast cereals, sweets, bakery, cosmetics and for medicinal purposes. It enjoys the reputation of being a natural and healthy product. However, it is often neglected that honey is one of the food products that contain the naturally occurring plant toxins belonging to the pyrrolizidine alkaloids (PAs).

PAs are a group of at least 600 different secondary plant metabolites that are produced by more than 6000 plant species worldwide (Bolechová, Čáslavský, Pospíchalová, & Kosubová, 2015). About 95 % of the known PAs are found in the botanical families of *Asteraceae*, *Boraginaceae*, *Fabaceae Orchidaceae* and *Apocynaceae* (EFSA, 2011). PAs are of natural origins, and are produced by plants as a protection against e.g. herbivores. Bees collect pollen or floral nectar from a wide variety of these plants and therefore diverse profiles of PAs are observed in honey samples (Avula et al., 2015; EFSA, 2011; Mulder et al., 2018). 1, 2-Unsaturated PAs have been reported as toxic both to human and animals. They are suspected to be hepatotoxic and genotoxic carcinogens in human in the factor that these PAs are activated to electrophilic compounds through the liver metabolic enzymes (Mulder et al., 2018). Single or continuous intake of 1, 2-unsaturated PAs causes

hemorrhagic necrosis and hepatic venous occlusion, and increases the incidence of liver haemangiosarcoma for chronic risk assessment (EFSA, 2017). Acute poisoning of PAs in human is associated with the liver damage, and a subacute or chronic onset poisoning may lead to liver cirrhosis and pulmonary arterial hypertension (EFSA, 2011). The German Federal Institute for Risk Assessment (BfR) recommends an exposure limit of different foods to 0.007 µg of PAs/kg body weight per day, which corresponds to 0.49 µg of PAs for an average of a 70 kg European person (BfR, 2011).

In 2012, the EFSA Panel on Contaminants in the Food Chain (CONTAM) adopted the Scientific Opinion on PAs in feed and food (EFSA, 2011). The Committee agreed with the recommendation from the CONTAM Panel in EFSA that LC-MS/MS is the method of choice. The LOQ to be achieved for PAs compounds is 1 µg/kg for honey, and only for the available analytical standards of retrorsine, seneciphylline, senecionine, echimidine and isomers, lycopsamine and isomers, heliotrine, lasiocarpine, monocrotaline, trichodesmine, (integerrimine) (EFSA, 2016). Based on the data from EFSA, 64 PAs were reported in food products (EFSA, 2011), while lacking authentic standards hampers an unequivocal quantification. This is particularly a problem in the PAs analysis of honey, where a complex and variable profile of PAs can be expected. Up to now, very limited PA standards are commercially available at very high cost (e.g., in average 200~500 Euro per 10 mg and no more than 44 different PAs in one method have been reported (Kaltner, Stiglbauer, Rychlik, Gareis, & Gottschalk, 2019). Although we have identified a study which achieved quantification with isotopic labelled PAs (Chung & Lam, 2018), isotopic labelled PAs are still not commercially available. This is also unlikely to change in the near future since there is no facile chemical access, *in vitro* biosynthesis or isolation (EFSA, 2011). Therefore, measuring the concentrations of a wide variety of PAs within a single experimental detection continues to remain very challenging.

The CONTAM Panel identified four groups of PAs (the tertiary amine as well as the corresponding *N*-oxide forms) of particular importance as regards the monitoring of their presence in food and feed (EFSA, 2011). According to the taxonomy of PAs and the available literatures regarding PAs occurrence data, where the PAs found in honey are also mainly belong to these four groups (EFSA, 2016), as shown in Figure 2-1, senecionine, lycopsamine, heliotrine and monocrotaline types. A detailed PAs list including the commonly found PAs reported from literature (85 PAs) were summarized in Table S2 in Appendix I.

Senecionine-type PAs: they represent the largest and most diverse class (e.g., senecionine, retrorsine, seneciphylline, acetylerucifoline, erucifoline, integerrimine, jacobine, jacoline, jaconine, jacozine, retrorsine, senecivernine as well as their N-oxide), and consists of macrocyclic diesters mostly. The main necine base is retronecine, sometimes replaced by otonecine or a 1,2-saturated analogues (they occur primarily in the Asteraceae tribe Senecioneae (*Asteraceae* family), but are also found in *Crotalaria* spp. (*Fabaceae* family).

Lycopsamine-type PAs: (e.g., intermedine, intermedine N-oxide, acetylechimidine and isomers, lycopsamine and isomers, echimidine and isomers, echivulgarine, vulgarine). These are characterized by open chain mono- or diesters PAs containing at least one necic acid unit of 2-isopropyl-2,3-dihydroxybutyrate, which occur predominantly in the *Boraginaceae* family, and also found in the *Eupatorieae* spp. (*Asteraceae* family).

Heliotrine-type PAs: (e.g., lasiocarpine, lasiocarpine N-oxide, fulvine, monocrotaline, retusamine, trichodesmine, europine, heliotrine). They occur in *Heliotropium* spp. (*Boraginaceae* family).

Monocrotaline-type PAs: (e.g., fulvine, monocrotaline, retusamine, trichodesmine). They are characterized by an eleven-membered macrocyclic ring instead of a twelve-membered ring found in macrocyclic senecionine-type PAs. This type of PAs is produced by *Crotalaria* species (*Fabaceae* family)

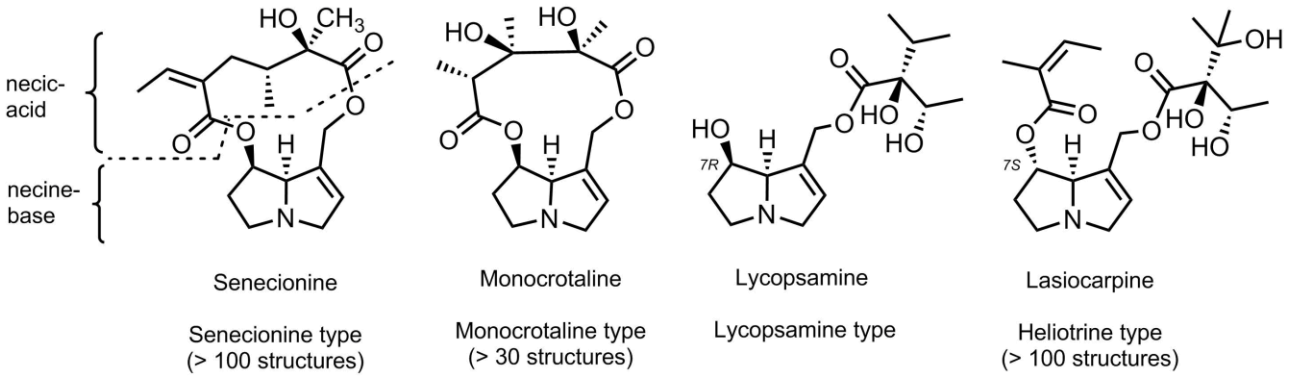


Figure 2-1. Four groups of PAs mainly found in honey according to the taxonomic of PAs and the available literatures (EFSA, 2016)

2.1.2. Cereals

Cereal crops were the first cultivated plant species by human and for over ten thousand years have been the staple food and feed source. Currently cereal is one of the largest food commodity and consumed worldwide, particularly wheat, rice and maize are of prime importance (González-Curbelo, Herrera-Herrera, Ravelo-Pérez, & Hernández-Borges, 2012). The main cereals grown in EU are wheat, barley, corn, rye, oat, sorghum, and triticale, from the EU-33 grain production forecast, the total production of cereal crops is 297 and 322 million tonnes for 2018 and 2019 (Coceral, 2019). To ensure the safety and quality of such a large food commodity, regulatory agencies and authorities are pushing the cereal producers, manufacturers, and researchers to pay serious attention to the production processes and to develop comprehensive quality policies, management systems and efficient analytical methods to improve the cereal safety. Unwanted compounds in cereal can be various, including chemical residues, biological contaminants, and physical contaminants. The biological contaminants include microorganisms, natural occurring toxins (i.e. mycotoxins from fungi, toxins from cyanobacteria, histamine, vegetal alkaloids, etc.), environmental (i.e. dioxins, PCBs, etc.) and contaminants (e.g., PAH) from environment and processing (e.g., improper drying), and chemical residues contain agrochemicals as pesticides, plant growth regulators from crop protection, all of which get more concerned in cereal food safety (Y. Tang, Lu, Zhao, & Wang, 2009). Therefore, there is an intense attention to the occurrence of unwanted compounds in cereal to maintain in a high food safety standard (Alldrick, 2017; Pereira et al., 2014). In the following section, pesticides and mycotoxins that have been found as the dominant unwanted compounds in cereals are illustrated.

2.1.2.1 Pesticides

WHO defined a pesticide as a chemical compound that is used to kill pests, including insects, rodents, fungi and unwanted plants (WHO, 2017). Pesticides are widely applied in numerous agricultural, commercial, residential, and industrial applications. Since pesticides are directly applied on cereals, infants, children, and adults can all be exposed to residues of pesticides by the ingestion of those pesticide-contaminated foods. In general, pesticides can be divided into herbicides, fungicides, insecticides, acaricides, nematicides, molluscicides, rodenticides, growth regulators, repellents, rodenticides and biocides based on the targeted

usage (*EU - Pesticides database*). Based on the chemical structure of pesticides, pesticides can be classified into organochlorines, organophosphates, carbamates, pyrethroids, phenyl amides, phenoxy alkonates, trazines, benzoic acids, phtalimides, dipyrids, and some unclassified ones (Jayaraj, Megha, & Sreedev, 2016).

All matters related to legal limits for pesticide residues in food and feed are covered by the regulation (EC) No 396/2005. Plant protection products in cereals are mainly regulated by the framework regulation (EC) No 1107/2009. This regulation also contains provisions on official controls of pesticides residues in food of plant and also animal origin that may arise from their use in plant protection.

From the pesticides regulations of the main cereals consumed in EU, there are over 1000 pesticides regulated to be used in crop production according to the mentioned regulation that are reported in an EU database (*EU - Pesticides database*), however according to GAP – good agriculture praxis – only a limited (but still large) number of pesticides are used in the individual crops. It is not feasible nor possible to develop a method to monitor all approved pesticides in one analysis. It is common that modern pesticide residue analysis can include a few hundreds or more analytes in a routine method (X. Wang et al., 2013). Even if the same number of standard substances are available, they would need to be purchased and, it is not really feasible to include all standards in each of the sequence as this will increase analysis time significantly. Unfortunately, all standard compounds are rarely available in each laboratory and it is even more of a predicament when new active substances are put on the market, residue definitions are changed to include metabolites of existing compounds or when compounds that are newly discovered or unstable. In addition, most pesticides have regulatory guidelines as a set of MRLs accepted in food to help protecting the community against contamination and potential negative health effects. For the pesticides regulated for cereals, clear MRLs were established for common cereals such as barley, maize, oat, rice, rye, and wheat in the range from 0.001 to 20 mg/kg (*EU - Pesticides database*). These low MRLs require the development of more powerful, sensitive analytical methods to meet these requirements in foodstuffs.

2.1.2.2 Mycotoxins

Infection of stored cereals by fungi may lead to production of toxic secondary metabolites, which usually referred to mycotoxins. Mycotoxins are a major concern for food safety due to

their impacts on human health. According to FAO, it is estimated that approximately 25% of the world's cereal crops are contaminated with mycotoxins (Luttrell, 2002). Mycotoxins in food samples are mostly produced by various fungi species belonging to the genera *Fusarium*, *Aspergillus*, and *Penicillium*. Several hundreds of mycotoxins have already been identified, but only a few of them are addressed by food legislation. The most important ones regarding to their occurrence and toxicity are: aflatoxins (AFs), fumonisins (FMs), trichothecenes (eg., Type A: T-2, DAS and HT-2; Type B: 3-acetyl-deoxynivalenol, 15-acetyl-deoxynivalenol, deoxynivalenol (DON), nivalenol (NIV), ochratoxins (OTs), patulin (PAT) and zearalenone (ZEN), and also their metabolites (Beccari, Caproni, Tini, Uhlig, & Covarelli, 2016). The main fungi and mycotoxins are shown in Figure 2-2 (Pereira et al., 2014).

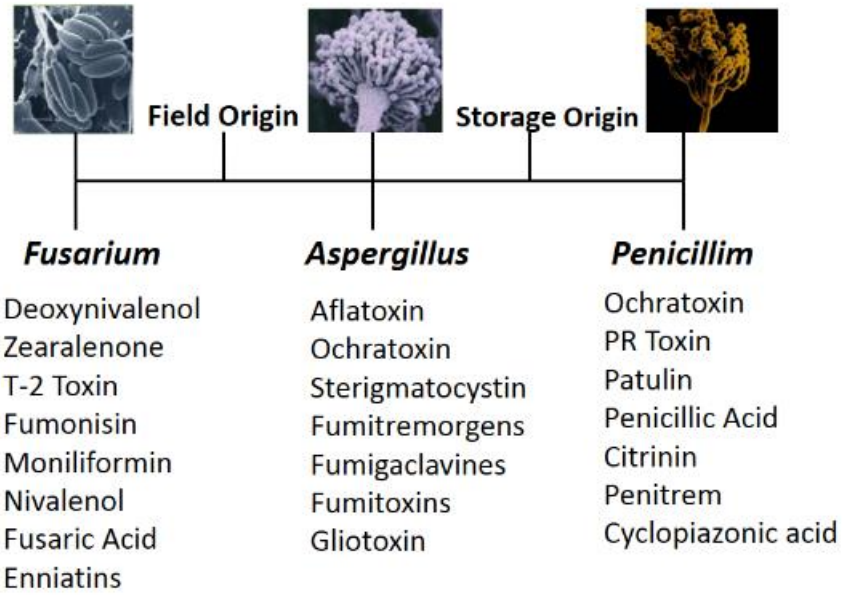


Figure 2-2. Main fungi and mycotoxins. Mycotoxins are produced by various fungi species belonging, mainly, to the Aspergillus, Penicillium and Fusarium genera. The most important ones regarding their occurrence and toxicity are AFs, FMs, TRC, OTs, PAT and ZEN. (Pereira, Fernandes and Cunha, 2014)

In 2001, the Commission's Scientific Committee for Food firstly established MLs for AFs, OTA and PAT in food stuffs (European Commission, 2001). It has been updated several times and substituted in 2006 by the EU regulation 1881/2006, which was further updated in e.g. 2007, 2010, 2012, and 2014. The regulations of mycotoxins are listed in Table 2-1. The Commission Regulation (EC) No. 401/2006 explained the methods of sampling and

analysis for the official control of the levels of mycotoxins in foodstuffs. The mycotoxins currently listed in the Regulation (EC) No 1881/2006 include AFs, OTA, PAT, DON, ZEN, FMs, T-2 and HT-2 toxin.

Apart from these very limited registered mycotoxins, there are also a large number of mycotoxins and second metabolites that are not listed in the regulation. For example, enniatins that originates from species in the genera *Fusarium* fungi are in a very high level in many cereal samples, and it has been shown that after exposure to enniatins might lead to acutely toxic or cardiac symptoms (Cigić & Prosen, 2009). However, they are currently still not included in EU regulation. An analytical method with a wide scope is needed to take these unregulated but toxic compounds into account.

Table 2-1. EU legislation concerning mycotoxins in food and feed

Legislative Reference	Matrix	MLs (Y/N)	Compounds
COMMISSION REGULATION (EC) No 1881/2006 on setting maximum levels for certain contaminants in foodstuffs	food	Y	AFs, OTA, DON, ZEN, FUM, and T-2 and HT-2 toxin, PAT
COMMISSION REGULATION (EC) No 466/2001 of 8 March 2001 setting maximum levels for certain contaminants in foodstuffs	food	Y	AFs, OTA, PAT
Worldwide regulations for mycotoxins in food and feed in 2003: Summary of study	Food and feed	Y	12 mycotoxins(van Egmond & Jonker, 2005)
Commission Recommendation (2006/583/EC) of 17 August 2006 on the prevention and reduction of Fusarium Toxins in cereals and cereals products	Cereals (preventive measures)	N	Fusarium toxins
Commission Regulation (EC) No 401/2006	Food	N	AFs (Sum of B1, B2, G1 and G2) Aflatoxin M1 OTA, PAT, DON, ZEN, Fumonisin B1 and B2 T-2 and HT-2 toxin

Note: MLs is the Maximal levels. Y/N means the Maximal levels of specific mycotoxins were defined or not.

2.2. Chromatography, mass spectrometer, and the theory behind

MS is widely used in food monitoring (Antignac et al., 2011) especially for measuring hazardous compounds with low abundance due to its higher performance in the term of sensitivity and selectivity. A modern mass spectrometer is a device that typically includes an ion source producing ions in the gas phase, a mass analyzer that can separate all the ions according to their mass to charge ratio or m/z value, a detector that can count the ions for every m/z ratio and a computer that can convert the data from analyzer (m/z) to detector (intensity) in mass spectrum. In addition, the mass spectrometer requires a device to introduce the sample, either directly or through a separation device such as gas chromatography (GC), liquid chromatography (LC), or capillary electrophoresis (CE) (Hoffmann, 2005). Furthermore, the mass spectrometer may contain several analyzers, which allow MS/MS or MSn experiments to be performed (Hird et al., 2014). In this part, the basic principles of separation device (GC, LC and chromatographic separation) and mass spectrometer including ion source and mass analyzers are illustrated.

2.2.1. Separation device

2.2.1.1 GC and LC

GC with electron impact (EI) ionization and LC using electrospray ionisation (ESI) are techniques that most often applied in food monitoring currently. Most analytical methods for quantitative analysis of food contaminants or residues like mycotoxins and pesticides are based on LC or GC coupled to MS (Annunziata et al., 2017; Cunha & Fernandes, 2010; Martos, Thompson, & Diaz, 2010). Both of them can be connected with different single or tandem mass spectrometers. The choice of the most appropriate separation device to handle the majority of analytes is one of the most important decisions on investments in food monitoring in analytical laboratories.

There are many advantages of using GC. First, the main advantage is its high chromatographic separation power, therefore the method optimisation is more simple as the separation is primarily based on the boiling points of solute molecules. Second, GC is usually coupled with EI, it causes extensive fragmentation which may contain useful information of compound structure, and GC-MS generates more reproducible molecular

fragmentation patters. Moreover, unlike LC-MS, the identification in GC-MS is usually more easier, reliable and faster than LC-MS by comparing a query mass spectrum with reference mass spectra in a libraries (e.g., NIST) via spectrum matching.

However, it is difficult to exclusively use GC-MS method for the screening analysis in food monitoring. Limitations in GC–MS/MS arise from the absence of a universal soft ionization mode, which could be used for the efficient production of molecular ions (Alder, Greulich, Kempe, & Vieth, 2006). The total ion current in EI ionization is often spread on many fragments, resulting in a low intensity of parent ions of MS/MS experiments. Also, GC is usually used for separating vaporizable or volatile compounds, and it deals poorly with polar thermally unstable, or little volatile compounds. Therefore it is limited in the analysis of food contaminants and residues which are not volatile without derivatization (Masiá, Suarez-Varela, Llopis-Gonzalez, & Picó, 2016).

When compounds are not amenable to GC, the application of LC is the best alternative. Developments in chromatography enable more rapid, highly efficient LC separations. Recent developments in both LC and MS have resulted in very powerful instrumentation for sensitive and selective determination of other more polar or ionic unwanted compounds at trace levels in food, such as pesticides, mycotoxins, veterinary medicines and PAs (L. Cramer, Schiebel, Ernst, & Beuerle, 2013; Delatour, Racault, Bessaire, & Desmarchelier, 2018; Frenich, Romero-González, Aguilera-Luiz, & Research, 2014; Sulyok, Krska, & Schuhmacher, 2010). Many LC-MS based methods are used for determination of specific compounds with targeted analysis methods. There is also an increase of using LC-MS screening methods, as LC-MS is capable of covering a wide range of food contaminants and residues by using untargeted screening (León, Pastor, & Yusà, 2016). These progresses illustrate that LC-MS is becoming an important screening approach in food monitoring. However, LC-MS based methods are limited by the lack of generic spectra database and efficient data processing approaches which needs to be addressed.

2.2.1.2. Chromatographic separation

Food samples are usually the mixture of numerous compounds and highly complex, thus the corresponding analytical method must be selective towards the analytes. The co-eluted compounds from samples can cause suppression and also interference in the MS detector which also called matrix effects, and this a particularly a problem in low resolution LC-MS.

Chromatography is an important step to separate the numerous compounds and increase the selectivity of the method. An analytical method is regarded selective when its results are not affected by other components from sample to any significant extent. Development of chromatographic method is one of the important way to improve the method selectivity. It is common to evaluate the selectivity regarding the chromatographic separation. There are a few ways to characterize how well the chromatographic peaks are separated. One of the most common way to quantitatively express the chromatographic selectivity is using the peak resolution (R_S), which takes peak widths at half height ($w_{1/2}$) into account, it is more accurate than the well-known separation factor ($\alpha = \frac{t_{RB}-t_M}{t_{RA}-t_M}$) as R_S reflect the more actual situation in chromatograms (Figure 2-3). Some guidelines of validation require the R_S value exceeds certain threshold, e.g., FDA requires $R_S > 2$ (FDA, 1994) and AOAC requires $R_S > 1.5$ as the adequate peak separation (AOAC, 2002).

$$R_S = \frac{t_{RB} - t_{RA}}{0.85(w_{1/2B} - w_{1/2A})}$$

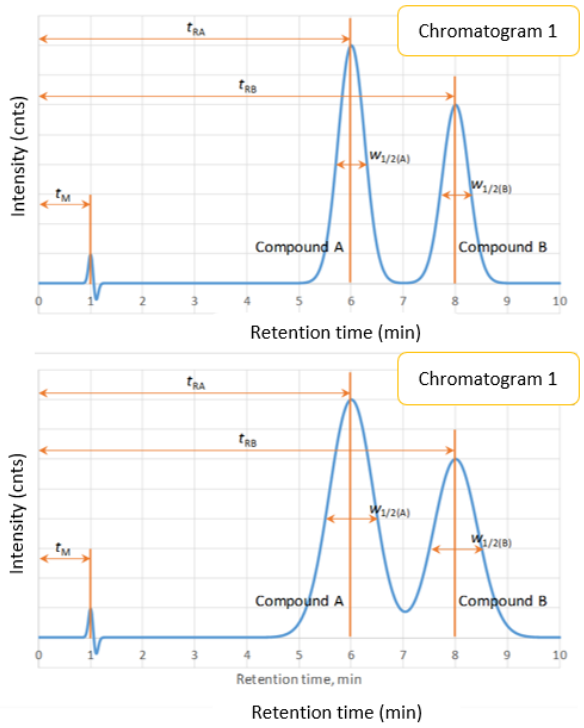


Figure 2-3. Chromatograms with similar separation factor but different resolution (Herodes, 2020).

The chromatographic requirement becomes less strict as the development of high resolution mass spectrometers with high selectivity, high accuracy and high resolving power. However, it is still needed sometime to optimize the chromatographic separation for the

highly complex samples when the co-eluted compounds have the same mass or too much matrix effects from compounds in specific retention window observed. This can be achieved by prolonging the retention window or modifying the eluent composition in mobile phase (Hird et al., 2014). Although the chromatographic separation was not specifically optimized in this project, the chromatographic separation is still need to be concerned sometimes.

2.2.2. Theory behind Ionisation

An ion source is the device that creates atomic and molecular ions. Analyzed samples are ionized before analysis in the mass spectrometer (Hoffmann & Stroobant, 2004). There are various ionization techniques used for MS. The direct ion sources are divided into two types: liquid-phase ion sources and solid-state ion sources. In solid-state ion sources, the analyte is in an involatile deposit; ions are extracted by an electric field and focused towards the analyzer. Matrix-assisted laser desorption, secondary ion mass spectrometry, plasma desorption and field desorption sources belong to this type (Freeman, 2013). In liquid-phase ion sources, ions are produced at atmospheric pressure from solution that are introduced by nebulization and focused into the mass spectrometer through some vacuum pumping stages (Hoffmann & Stroobant, 2004). Electrospray ionisation (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photoionization (APPI) correspond to this type. ESI, APCI and APPI sources are by far the most versatile and well known ionisation techniques. APCI and ESI are specifically illustrated in this part, which were used in this project.

2.2.2.1 ESI ion source

ESI is the most common ion sources used for all types of samples and analytes, which allows very high sensitivity analysis to be reached and is easy to couple to a LC. ESI is robust, simple to operate, generate ions from a wide range of compounds that are highly polar, least volatile and thermally unstable (Banerjee & Mazumdar, 2012). ESI is capable of dealing with both small and large molecules like proteins in biological chemistry. ESI is also a soft ionisation technique that can keep the original structures of chemical compounds to a large degree (Awad, Khamis, & El-Aneed, 2015).

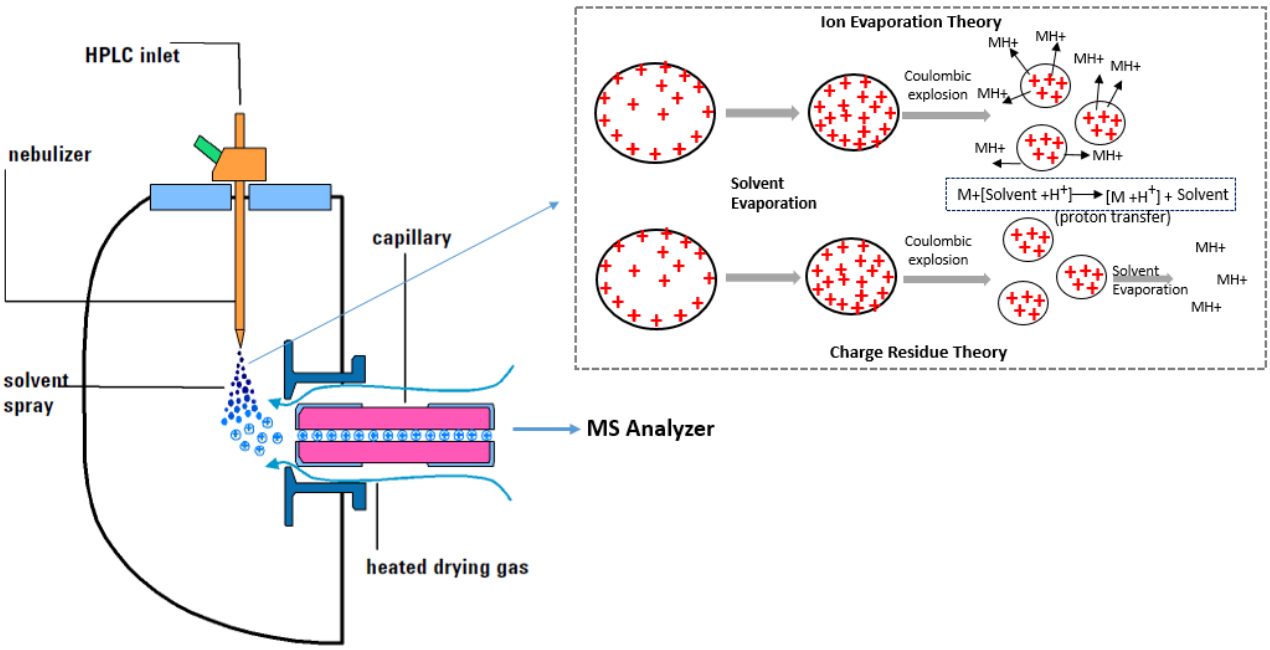


Figure 2-4. Schematic representation of the ESI ionization process, modified from Agilent Technologies (2014). The analyte undergoes three major procedures, 1)production of charged droplets. 2) repeat solvent evaporation from charged droplet; 3) gas phase ion formation. Solvent evaporation happens until it reaches the "Rayleigh limit", "Coulomb explosion" occurs (Banerjee and Mazumdar, 2012). The formed gas phase ions are attracted to and pass through a capillary sampling orifice into mass analyzer. Two mechanisms for formation of gas phase ions in ESI: charge residual model and ion evaporation model (Awad et al., 2015)*. Gas-phase proton transfer from protonated solvent make an important contribution to the formation of gaseous ions in the ESI process (Yang et al., 2013)*

In ESI, the analyte is introduced in solution either from the syringe pump or as eluent flow from a LC column at a flow rate normally 1-1000 µL min^{-1} into a spray needle/capillary at a high voltage (Awad et al., 2015). The ionization process in electrospray can be summarized as Figure 2-4, the LC eluent is sprayed (nebulized) into a spray chamber at atmospheric pressure in the presence of heated drying gas and a strong electrostatic field (Agilent Technologies, 2014). The electrostatic field occurs between the nebulizer and the capillary that at high voltage (typically in the range of 2 to 6 kV). The analyte that transferred from the solution to the gas phase via ESI undergoes three major procedures, namely, production of charged droplets (nebulization); repeat solvent evaporation from charged droplet and droplet fission (desovation); gas phase ion formation by a mechanism (Banerjee & Mazumdar, 2012). In nebulization, analytes solution enters the spray chamber though nebulizer, and nebulizing gas enters the spray chamber concentrically through a tube that

surrounds the needle. Analytes solution are broken into droplets by the strong shear forces generated by the nebulizing gas and the high voltage in the spray chamber. As a result, the analytes are simultaneously charged and dispersed into charged droplet. Charged droplet evaporation follows droplet generation, and solvent continuously evaporates from the droplet surface and all charges retained. This process continues until the droplet size reduces to a critical value at which the total charges on the droplet surface reach Rayleigh limit, coulomb explosions occurs, the droplet explodes and produce smaller charged droplets (K. Tang, Page, Kelly, & Marginean, 2016).

Two principal mechanisms have been proposed for the formation of gas phase ions in ESI, one is the charge residual model, one is ion evaporation (Awad et al., 2015; El-Aneed, Cohen, & Banoub, 2009) (Figure 2-4). In the charge residual model, the analytes will not desorb from the charged droplet. Instead it will be freed by complete evaporation of the solvent, and this is more likely to happen for large molecules (Iribarne & Thomson, 1976). When the last few solvent molecules evaporate, the charges are delocalized on the analyte, and giving rise to the most stable ions in the gas phase (Crotti, Isak, & Traldi, 2017). In ion evaporation mechanism, a direct emission of ions from the droplet is considered. It includes solvent evaporation, coulomb fissions (or Coulomb explosion) of the charged droplets, forming smaller droplets. The ions in the gas-phase are directly desorbed/released from the surface of the small droplets when the repulsion between charges at the droplet surface overcomes the cohesive force of the surface tension (Iribarne & Thomson, 1976).

In ESI process, different compounds usually yield different ionization efficiencies. The ionization efficiencies can be defined as the different responsiveness of compounds during the gas-phase ions generation (Kruve, 2020). It was hypothesized that the gas-phase proton transfer from protonated solvent make an important contribution to the formation of gaseous ions in the ESI process (Yang et al., 2013), it happens when the proton affinities between two gaseous molecules are large enough that a proton can be abstracted. It also has been proved that the solvent proton affinity influence the proton transfer and the analyte response (Amad, Cech, Jackson, & Enke, 2000). ESI can also be used in the case of molecules without any ionisable site through the formation of sodium, potassium, ammonium, chloride, acetate or other adducts (Hoffmann & Stroobant, 2004).

ESI current is limited by the electrochemical process that occurs at the probe tip and is sensitive to concentration. In ESI, the variables that can influence the droplet formation

include e.g., chemical-physical characteristics of the solvent (e.g., viscosity, surface tension, pK_a), concentration and nature of analytes, nature and size of the capillary, distance to the counter-electrode, presence or not of other analytes (matrix effects), applied potential and so on (Cole, 2000; P. Liigand et al., 2017). The most important consideration in the ionisation process is the ion source settings, characteristics of the solvent and the physicochemical properties of the analytes that can be ionized. Therefore, on the one hand, it provides the possibility to optimize the ion source settings and eluent compositions to improve the ionization efficiency or the sensitivity of the analytes. Studies have shown that source parameters such as drying gas temperature, nebulizer gas, sheath gas temperature, sheath gas flow, capillary voltage, pH of mobile phase can influence the ionization efficiency or response factor on some specific compounds (Kruve, 2016; J. Liigand, Laaniste, & Kruve, 2017; Pieke, Granby, Trier, & Smedsgaard, 2017). It means that the response factor of compounds in MS can be optimized by carefully choosing the MS source parameters and mobile phase to maximize the sensitivity of the entire analysis. This will be illustrated and studied in section 3.1.1. In addition, it also provides the possibility to investigate the correlation between chemical responses and physicochemical properties of analytes and thereafter used as the basis for predicting ionization efficiency by potential semi-quantification model development, which will be further discussed in section 2.3.5.

2.2.2.2 APCI ion source

APCI can be considered as an analogous ionisation method which is used at atmospheric conditions compared to chemical ionization (CI), and it is a natural evolution of CI (commonly used in GC–MS, the ionisation at reduced pressure) (Hoffmann & Stroobant, 2004). APCI is currently the second most used ionization mode (after ESI) in MS connected with a LC (Rebane et al., 2016), which has its primary applications in the areas of ionisation of polar and relative non-polar compounds with moderate molecular weight (up to 1500 Da) (Khatal, Gaur, Naphade, Kandikere, & Mookhtiar, 2016). In view of the high temperature and need for vaporization, the analytes have to be thermally stable and volatile. which is a compensation of ESI (suitable for high polar and thermally unstable compounds) (Khatal et al., 2016). As ionization in APCI occurs in the gas phase not the liquid phase, it is not necessary for the solvent to be polar and able to carry a charge. The analyte in ESI is ionized in the liquid phase (charged droplets) and then introduced to gas phase, thus ion formation is more susceptible to the matrices that may compete for the charge or the space on the

droplet surface or change droplet properties. However, in APCI, the analyte is introduced to the gas phase in neutral form to be ionized via chemical reactions with the reactant ions, it therefore suffers less matrix effects compared to ESI (Awad et al., 2015). In addition, APCI works better at higher flow rates than ESI.

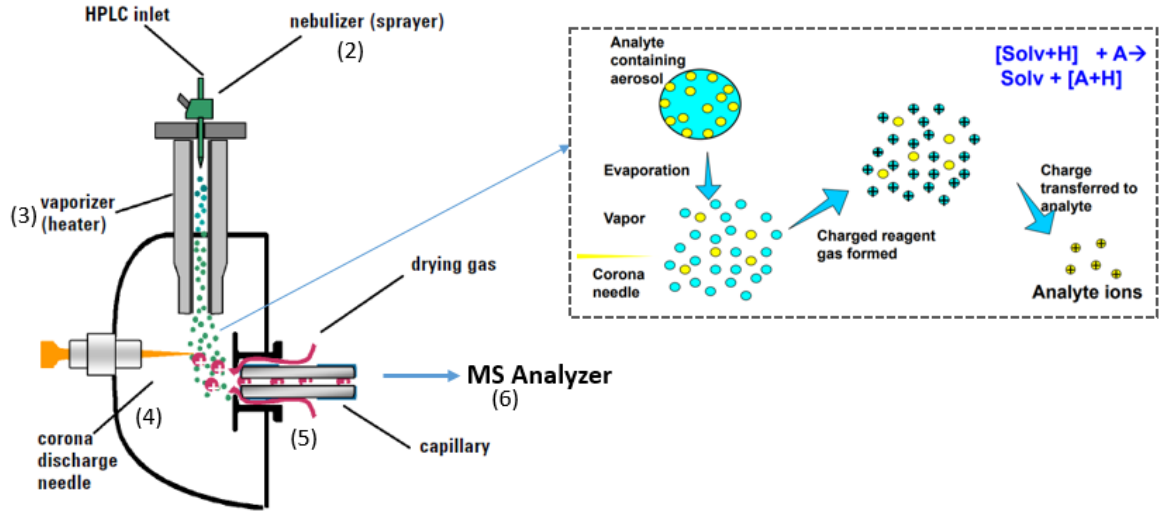


Figure 2-5. Schematic representation of the APCI ionization process in positive mode, modified from Agilent Technologies (2014). The analyte solution from LC part is introduced into a nebulizer, and dried by gas flow and desolvated in vaporizer. Ionization is done by ionisation of solvent molecules by a corona discharge electrode at atmospheric pressure. The produced ions then enter the mass analyzer through an inlet, or heated capillary

The ionization mechanism (Figure 2-5) of APCI is typically the same as present in CI. The solvent molecules with high abundance are statistically privileged to interact with the electrons originating from the corona discharge; the formed ions react with other solvent molecules, leading to protonated or deprotonated species, which are the reactants for the analyte ionization (Banerjee & Mazumdar, 2012). The general source set-up is shown in Figure 2-5, APCI is commonly used in conjunction with LC or other flow separation techniques. In APCI, the analyte solution from a direct inlet probe or LC part (see 1 in Figure 2-5) at a flow rate typically of 200-2000 µL/min is introduced into a pneumatic nebulizer (2) (Awad et al., 2015). Analyte droplets are dried by gas flow and desolvated in a vaporizer (3) before interacting with the corona discharge (4). The temperature of the vaporization chamber is controlled, which makes the vaporization conditions independent of the flow and from the nature of the mobile phase. Ionization is done by the ionization of solvent molecules

with a corona discharge electrode at atmospheric pressure. The ions produced at atmospheric pressure enter the mass analyzer (6) through an inlet, or through the heated capillary (5), and are then focused towards the MS analyzer (Hoffmann & Stroobant, 2004).

2.2.3. Choice of Mass analyzers

The most important change in the past decades has been in the increase in choice of mass analyzers for LC-MS and it highly influence the method to monitor chemical compounds in food (Hird. *et al.*, 2014). Various mass analyzers are used in combination with LC and GC for analyte screening, and structural elucidation. Many of them have been applied to the food monitoring, such as quadrupole (Q), quadrupole ion trap (QIT), time of flight (ToF), triple/tandem quadrupole (QqQ), quadrupole-time of flight (QToF), orbitrap, and fourier transform ion cyclotrone resonance (FT-ICR). Various combinations of mass analyzers can be assembled for providing useful structural information by MS/MS. In this part, the commonly used tandem mass spectrometry and HRMS have been demonstrated.

2.2.3.1 Tandem mass spectrometry

Tandem mass spectrometry also known as MS/MS or MS^2. The basic principle of MS/MS is the selection of precursor ion, fragmentation of this ion. The hybrid mass spectrometer combines analyzers of different types. Two fundamentally different approaches to MS/MS are used: tandem in space and tandem in time. Tandem-in-space means the MS spectrometer have separate independent mass analyzers in physically different locations. Such type of tandem mass spectrometers include, e.g. QqQ, QToF and Orbitrap hybrid mass spectrometer. Tandem-in-time means different stages of MS are conducted within the same physical storage device but at different times, ion-trapping mass spectrometers, which include e.g., 3-D QIT, linear ion traps (LIT), FT- ICR are correspond to this type (Hoffmann & Stroobant, 2004). The specificity can be increased significantly when tandem MS (e.g. QqQ, QIT) in combination with SRM (MRM) is applied compared with acquiring full scan data. Tandem MS such as QqQ and QIT offer good sensitivity and selectivity. LC-QqQ–MS is mostly focusing on targeted compounds using MRM mode with low resolution, and MRM is blind to the compounds which is not included in the method, therefore LC-QqQ –MS has its limitation in untargeted analysis (Wang et al., 2013). Compared to QqQ-MS, Q-ToF-MS has the major advantage in structure identification and confirmation with wide scope, as it

allows accurate mass measurement of both molecular and fragment ions, it therefore provides increased efficiency (Hird et al., 2014).

2.2.3.2 High resolution mass spectrometry

Among the various MS techniques, high-resolution mass spectrometry (HRMS) is a troubleshooting tool for screening and confirmation with high accuracy and sensitivity. It has been defined that HRMS is the mass spectrometer with resolving power greater than 10000 (Hoffmann & Stroobant, 2004). HRMS is capable of providing both full spectral information and high mass-resolving power, and provide the capability for accurate mass measurement to aid identification with the additional advantage of retrospective analysis. The drawbacks such as limited number of compounds per analysis, no retrospective data, reliance on the availability of standards, inability to screen for unknowns in the use of SRM mode on QqQ are overcome by HRMS.

HRMS that used for contaminants and residue analysis in food, has gained wider acceptance and a significant increase in reported studies in recent years (Jamin, Bonvallot, & Tremblay-franco, 2014; Kaufmann, 2012; Pérez-Ortega et al., 2016). As the benefits provided by HRMS are considerable, e.g., the full-scan spectra collection with more comprehensive information, screening samples with little preconception or without preconception on what might be present, the possibility for retrospective data analysis, and unknown compounds structural elucidations (Renaud, Sabourin, Topp, & Sumarah, 2017). HRMS is usually necessary for untargeted analysis, as it allows for the screening and the efficient identification with high accuracy, high sensitivity and quantification purpose afterwards. HRMS also allows for the combination between targeted, suspected and true untargeted screening.

Various MS analyzers, e,g. ToF, QToF, IT-ToF, Orbitrap, and Q-orbitrap correspond to HRMS. Orbitrap achieves the highest mass resolving power among these analyzers mentioned, while it comes at the expense of speed. ToF and QToF have the potential for faster acquisition rate than Orbitrap, and achieve a compromise in sensitivity. The usage of a Q-ToF-MS allows the certain identification, which is based on retention time, accurate mass of the quasi molecular ion selected by the quadrupole mass filter, and the complete collision induced mass spectrum obtained by the ToF analyzer (García-Reyes, Ferrer, Thurman, Molina-Díaz, & Fernández-Alba, 2005). LC-ToF-MS is amenable for analyzing

and identifying chemical contaminations or residues in food, not only targeted but also untargeted analytes.

Although HRMS has been an ideal tool for screening and structure confirmation in untargeted analysis (X. Wang et al., 2013), the analysis usually stops after spectra interpretation, more HRMS approaches should be extensively evaluated for quantitative analysis.

2.3. Determination of unwanted compounds in food

The three main questions need to be answered in food motoring in general, namely, “how to get all relevant compounds out of the sample?”, “what compounds are in the samples?”, and “how high is the concentration of the compounds?” Each question can be very challenging to all the investigators. In this part, the research progress of determination of unwanted compounds in food monitoring are summarized and illustrated, which includes the sample preparation, data acquisition and data treatment, identification, and quantification/semi-quantification.

2.3.1. Sample preparation

Sample preparation, as the first key step prior to the analysis of unwanted compounds, is usually the bottleneck in untargeted food analysis. The determination of these unwanted compounds with low concentration in complex food matrices usually requires extensive sample extraction and preparation prior to instrumental analysis. An optimal sample preparation method can reduce potential sources of errors in each step, improve sensitivity and enable reliable identification and quantification (Ridgway, Lalljie, & Smith, 2007). The progress of sample preparation is a long story and will be briefly discussed here.

The typical procedure in sample preparation include sampling/homogenisation, extraction, clean-up and pre-concentration which is a pre-requisite in the final analysis (Ridgway et al., 2007). The selection of sample preparation methods are usually based on the physiochemical properties of selected analytes, such as solubility, polarity, p*K*a, volatility (Dzuman, Zachariasova, Veprikova, Godula, & Hajslova, 2015; Ridgway et al., 2007).

Many sample extraction methods mainly focused on a single type of compounds, e.g., pressurized liquid extraction, solid-liquid extraction, solid-phase extraction and solid-phase micro extraction techniques are the common procedures have been used (Ridgway et al., 2007). However, there are a large number of compounds that need to be considered in untargeted analysis and the properties of compounds in untargeted analysis are very diverse. For this purpose, there is a trend towards developing generic extraction and clean-up procedures, which allow the extraction of a wide range of compounds in food samples. Among these different generic extraction approaches (Dzuman et al., 2015; Masiá et al., 2016; Mol et al., 2008), the QuEChERS method (Quick, Easy, Cheap Effective Rugged and Safe) is a reference for determining e.g., pesticides residues, mycotoxins, veterinary drugs in food, which was introduced originally from 2003 for pesticide residues analysis in fruits and vegetables (Mol et al., 2008). The main advantages of QuEChERS method are low solvent consumption, good speed, sample throughput and the possibility to obtain high recoveries for a wide spectrum of compounds.

Nevertheless, there is still a need to minimize and simplify the sample preparation procedure and improve the extraction efficiency and to ensure the high quality of further chemical analysis. For example, increasing automation of the current techniques to allow online extraction, increasing recovery, reproducibility and robustness as well as decreasing the sample preparation time.

2.3.2. Procedure of targeted and untargeted screening

Depending either on the way the analysis is approached or on its final purpose, determination of unwanted compounds in food using MS is divided into targeted analysis, suspected screening and untargeted screening (Garcı et al., 2007). The concepts of these approaches have been explained in previous section 1.2.2. In our study, we tend to classify suspected screening into part of untargeted screening. These analytical method deals with the unwanted compounds in food of both knowns and unknowns. Untargeted analysis mainly focus on unknowns, namely, known unknowns and unknown unknowns. Known unknowns, can be identified from literature data and may be registered into databases that generally contain MS or MS/MS (MS^2) spectra. The identification of known unknowns is the routine work of screening analysis in food safety. The pure reference standards can be used

to confirm the structures in the sample. Unknown unknowns are the most difficult compounds to identify and the identification demanding comprehensive analytic work with structural elucidation.

Figure 2-6 shows the workflow of targeted analysis, suspected and untargeted screening assembled from the literatures (Bletsou, Jeon, Hollender, Archontaki, & Thomaidis, 2015; Mollerup, Dalsgaard, Mardal, & Linnet, 2017). Known and unknown compounds can be analyzed by targeted analysis, suspected screening, and true untargeted screening depending on the availability of standards and the preselected compound list. The known compounds can be included and monitored in a defined targeted analysis method, both identification and quantification can be easily done in targeted analysis by the comparison with reference standards.

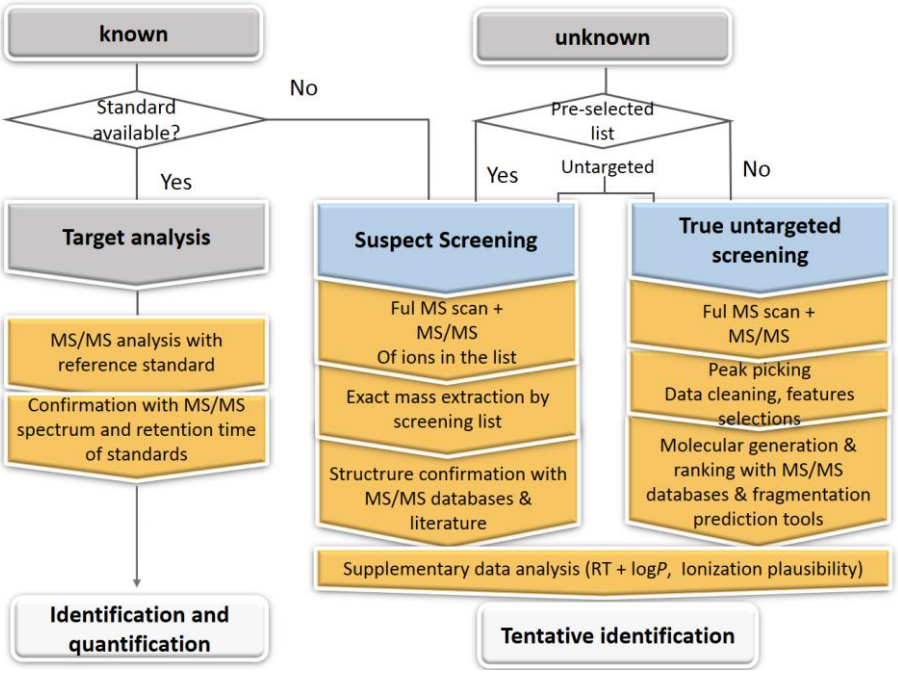


Figure 2-6. Flow chart of targeted analysis, suspected screening and untargeted screening procedure of compounds, modified from reference (Bletsou et al., 2015)*. For a comprehensive sample analysis by targeted analysis, suspected screening, and untargeted screening depending on the availability of standards and the preselected list.*

Targeted analysis carried out by multiple reaction monitoring (MRM) [also named as selective reaction monitoring (SRM) mode by some supplier] in MS is the most widely applied targeted method that can be found in many studies (Aguilera-Luiz, Vidal, Romero-González, & Frenich, 2008; Martos et al., 2010). MRM has been used for the identification

and quantification of various unwanted compounds such as pesticides, PAs, mycotoxins, PAHs (Al-Alam, Fajloun, Chbani, & Millet, 2017; Annunziata et al., 2017; Griffin, Danaher, Elliott, Glenn Kennedy, & Furey, 2013; Z. Li et al., 2017). However, MRM is blind to all the other compounds that are not included in the method, and the data does not allows for retrospective analysis if investigators are aiming to find more information from the raw data months or years back. Therefore, targeted analysis usually fails to discover new issues in the complex and rapid changing food market. When the targeted analysis is carried out by a HRMS full scan mode, all the ions present in the sample can be determined at the same time. Then it comes to suspected screening.

Suspected screening allows the compounds that included in a database or from literatures to be screened with screening list that based on the mass accuracy, isotopic pattern, retention time and fragment pattern. Reference standards are not necessary in suspected screening. It has been applied in the study of Klitgaard et al., (2014), in which suspected screening method was used to dereplicate and find novel fungal secondary metabolites. One could extend targeted analysis to certain metabolites, transformation products or impurities from commercial formulations by suspected screening for which reference standards might be not available (Hird et al., 2014). The compounds are identified using established screening list or databases. Screening list or database is needed for suspected screening, the more comprehensive the database are; the more efficient is the 'suspected screening' procedure.

In true untargeted screening, there is no previous knowledge like preselected list and standards available and it is usually carried out after targeted analysis and suspected screening. True untargeted analysis is not limited to target lists and the molecular information can be unknown (Knolhoff & Croley, 2016). For the unknown unknowns, the chemical formulas and structures can be obtained or predicted by a database match and *in silico* prediction tools for predicting fragments and their abundances from chemical structures information (Smith, Want, Maille, Abagyan, & Siuzdak, 2006; Wolf, Schmidt, Müller-Hannemann, & Neumann, 2010). It is a critical step in untargeted analysis to exclude the irrelevant peaks and pick out the relevant suspected peaks by comparing the control and case samples to figure out what is in the sample. There are some successful examples of using untargeted approaches to identify the food contaminants (Delaporte, Cladière, Bouveresse, & Camel, 2019; Kunzelmann, Winter, Åberg, Hellenäs, & Rosén, 2018). The

identified compounds from untargeted analysis without the confirmation of authentic standards are tentative identified compounds. Untargeted and suspected screening offer the possibility to be combined with semi-quantification methods when standards are not available. However, it is less sensitive compared to targeted analysis, and it is still a big challenge in both identification and quantification part, the data processing parts.

2.3.3. Data acquisition and data treatment approaches in untargeted analysis

2.3.3.1 Data acquisition in untargeted analysis

In soft ESI, a few fragments can be obtained by inducing fragmentation in a collision cell (or the souce) that can help to elucidate the structure of compounds. As routine targeted MS/MS analysis (e.g, with MRM mode in Figure 2-7) with limited scope and compounds-specific knowledge is insufficient to ensure the food safety. Therefore, untargeted analysis is highly needed to cover as many compounds as possible. However, the sensitivity and specificity of untargeted analysis is obviously lower than that of targeted analysis. Thus, it becomes more important to select an efficient data acquisition mode to ensure the high MS/MS data quality that is used for structure elucidation.

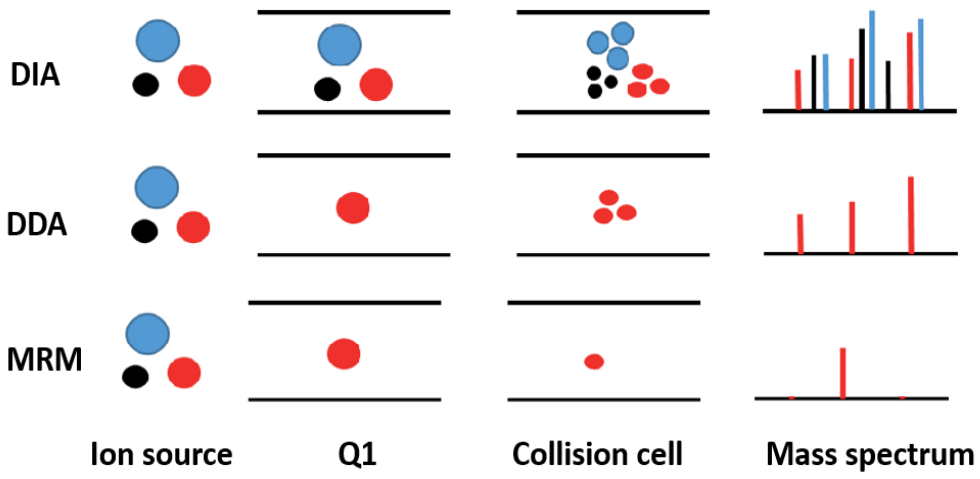


Figure 2-7. Data-acquisition approaches for HRMS: DDA uses the wide quadrupole (Q1) passing mode for precursor ions, precursors selected by the quadrupole are fragmented in collision cell. DIA, all precursor ions within a certain window of m/z values are fragmented. MRM, selected precursor ions are isolated, fragmented, and selected fragment ions are detected. Modified from reference (Zhu, Chen, & Subramanian, 2014)

Untargeted MS/MS acquisition modes include data-dependent-acquisition (DDA), and data-independent acquisition (DIA) (Renaud et al., 2017). Both approaches aim to characterize analytes comprehensively by MS/MS, and get as much information (mass of compounds and fragments) as possible for identification by inducing fragmentation in a collision cell. They all generate data that can be used for qualitative, retrospective analysis of suspected analytes after the samples have been analyzed, increasing the return from the initial investment of collecting and processing the samples. .

DDA means that the data-acquisition software directs the acquisition of MS/MS data based on real-time evaluation of MS-level data. By using DDA, fragmentation is performed only on precursors that are known to exist in a recent MS-level scan and that meet certain user-guided criteria (Broeckling, Hoyes, Richardson, Brown, & Prenni, 2018; Koopmans, Ho, Smit, & Li, 2018). DDA is generally used to increase annotation confidence in untargeted analysis. As shown in Figure 2.6, in standard DDA mode, usually MS/MS scans of the most intense ions in a full MS survey scan are sequentially fragmented. This method is biased toward the most abundant signals and may not record product ion spectra of compounds that present at low signal levels. Particularly for trace analysis, signals of ions in low level might be neglected (Renaud et al., 2017). The number of precursor ions to be fragmented is limited by the cycle time per scan (Zhu et al., 2014).

DIA has developed in parallel with DDA and emerged as a promising alternative to DDA, The principle of DIA is to acquire fragment ions for all precursor ions within a certain window of m/z values, sequentially covering the entire range of relevant m/z values (Tsou, Tsai, Teo, Chen, & Nesvizhskii, 2016). Figure 2-7 illustrates the difference between DDA and DIA. In DIA the quadrupole mass filter sequentially isolates spectral windows across a mass range for MS/MS throughout the LC run. DIA collects data of product ion of all precursors. However, this approach suffers from a lack of selectivity, as no precursor selection is used (Renaud et al., 2017). Figure 2-8 illustrates an example of the schematic of DIA data acquisition on QToF. After DIA data acquisition, libraries database match with spectral libraries achieved interpretation of obtained complex fragmentation spectra. Then the tentatively identified compounds are further move to quantification. The identification and quantification parts will be discussed in section 2.3.4 and 2.3.5.

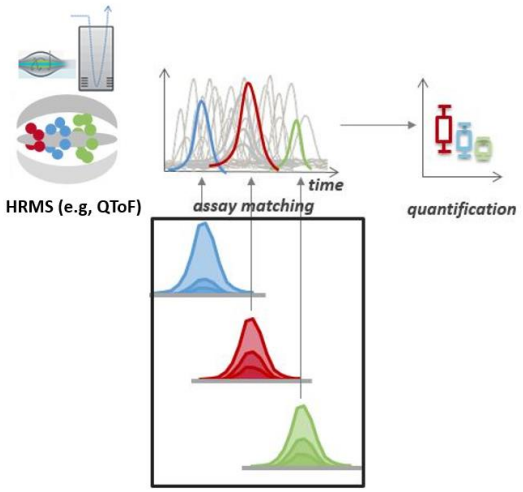


Figure 2-8 Schematic of the DIA in untargeted analysis, interpretation of the obtained complex fragmentation spectra by database match, and quantification on the example of QToF, modified from reference (Biognosys, 2016)

DIA strategy is exemplified by broadband CID (bbCID) data acquisition from Bruker TargetScreener (Figure 2-9), which is achieved by continuously cycling between ToF-MS and bbCID-MS/MS resulting in the seamless collection of a comprehensive catalogue of precursor and fragment ion data in single analysis. It ensures all ionisable compounds eluting have molecular ion data and fragmentation spectra acquired all of the time. Therefore no compounds will be missed. The ability of QToF to collect “all of the ions, all of the time” in both MS and MS/MS full scan modes results in a dataset that is complete and unrestricted. Once collected, detailed targeted and untargeted investigations as desired can be performed (T. Wang et al., 2019). Nevertheless, the performance and feasibility of these approaches on untargeted food screening still need to be clarified to obtain the MS/MS data with a high quality.

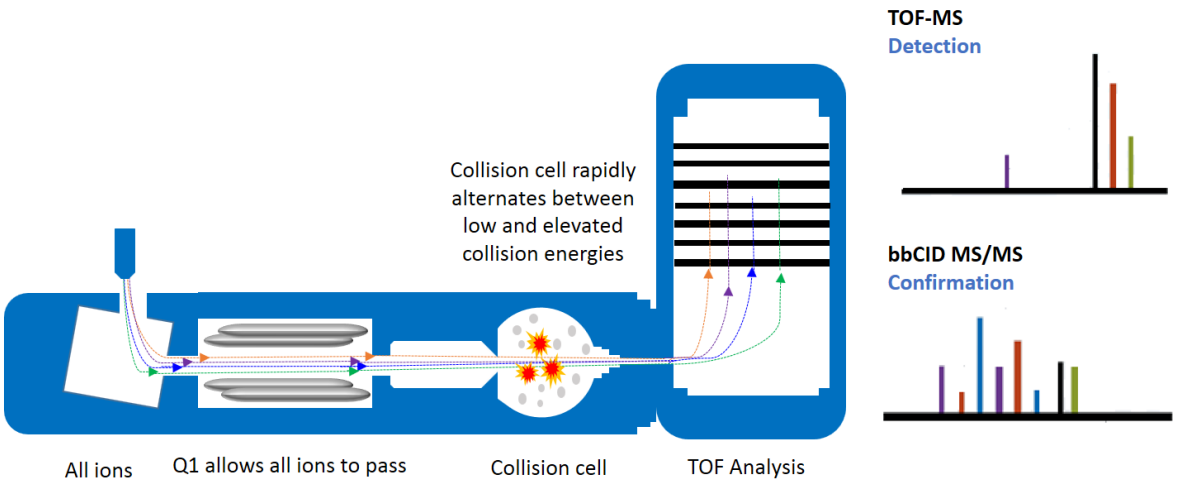


Figure 2-9. Schematic of Bruker bbCID data acquisition: The QToF mass spectrometer rapidly alternates between two full scan accurate mass channels, one at low collision energy and the other at elevated collision energy.

2.3.3.2 Untargeted Statistical Methodology

Unsupervised learning (UL) and supervised learning (SL) statistical approaches are main methods used for data analysis. UL is the analysis that conducted on only X-data (variables) with the goal to generate clusters from the input data (Hastie, Tibshirani, & Friedman, 2009). This process is often referred to dimension reduction or simplification. After clustering, ordination plots or dendrograms are used to visualise the clusters. SL is the analysis conducted on both X-data and Y-data (objects) (Reinholds, Bartkevics, Silvis, van Ruth, & Esslinger, 2015). This process uses some mathematical transformation to associate the X-data with Y-data, to determine the optimal number of components (e.g., R^2 and Q^2 coefficients are used to evaluate the performance of the (O)PLS-DA model). SL is often achieved by reducing the errors between the models output prediction and the actual known target trait which needs to be predicted. Therefore it is essential that the Y-data must be clearly known during the model construction. The prediction model can be either used for the class prediction or specific value prediction (Daglia, Antiochia, Sobolev, & Mannina, 2014).

Two of the main methods used in UL statistical approaches are principal component analysis (PCA) and cluster analysis (CA). PCA is commonly used as an explorative technique in multivariate analysis (R.G. Brereton, 2003). It allows the transformation of the multidimensional space of experimental variables into a space with reduced dimensions with

linear combinations of variables, which are named as principal components (PC). PC explained the variation of the linear combinations of variables, e.g., the first PC explains the most variation, and the second one the second most. Using two or three principal components as variables, data can be visualized in a new system of coordinates and basic features of data such as grouping, the presence of outliers and clustering can be revealed. Hierarchical clustering analysis (HCA) joins together objects into larger clusters by using different rules for clusters and some measurements of the distance between single objects (R.G. Brereton, 2003). Their results are reported as a dendrogram and classifications in a certain number of groups.

In SL, the aim is to train the machine using data that are "labelled." SL approach allows to collect data or produce data output from the previous experience with the correct answer, and also helps to further predict outcomes (Y-data) based on the given variables (X-data). It is more desirable to use the method that produce X-data than the one that uses Y-data, as the method to obtain Y-data are more expensive and difficult to generate. Various regression models are belong to SL. The regression models are of quantitative and classifications of qualitative nature. Stepwise-MLR is a multiple linear regression equipped with variable selection scheme. The procedure starts by first selecting the variable that has the highest correlation with Y (Raji et al., 2009). A variable is considered for addition to or subtraction from the variables set based on testing the significance using T-test. Forward selection and backward elimination are repeated until no improvement of the model is achieved by adding new variables. The criterion used in stepwise-MLR procedure is based on data fitting, and the To avoid overfitting, cross-validation (CV) is applied, and the predictive abilities of two models, one built using all variables selected with stepwise-MLR and the other with all variables except the one (leave one out) that yields the lowest partial correlation coefficient are evaluated (Caetano et al., 2005). Y data could be predicted after the linear regression model has been built.

PLS regression can also be chosen to create a prediction model (R.G. Brereton, 2003), in which the features from PCA and MLR are generalized and combined. PLS studies the correlation between two data matrices, in which one matrix contains data obtained from one analytical methodology (X-data), and the second one contains extra information on samples obtained using reference methodology (Y-data). In this way, the relationship between the analytical data (X-data) and the other properties (Y-data) of samples can be used to predict

the properties directly using the analytical methodology. PLS-DA (Barker & Rayens, 2003) and orthogonal projections to latent structures (OPLS) (Trygg & Wold, 2002), take advantages of statistical approaches of both PLS and LDA avoiding their weaknesses. The utilization of class memberships in PLS-DA allows the algorithm to better expose separations between classes in score space. There is another supervised model called random forests. Random forests are an ensemble leaning method for classification or regression by constructing a multitude of decision trees at training time and outputting the classes (the mode of classes for classification or mean prediction for regression of individual trees) (Ho, 1995). The commonly used chemometrics strategies can be used for MS data analysis are illustrated in Table 2-2. PCA, Stepwise-MLR, PLS-DA, and random forests are the chemometrics strategies applied in this project. For example, PCA and PLS-DA were used for the purpose of unwanted compounds detection and rice samples discrimination in manuscript 1; Stepwise MLR model was used for the semi-quantification model development of PAs in honey in manuscript 2; random forest model is the basic idea of the semi-quantification model of manuscript 3 and 4.

Table 2-2 Chemometric strategies and methods can be used for the analysis of MS data, supervised or unsupervised leaning method, simplified description and the model type are presented

Modelling method	Model type	Unsupervised or supervised	Comments	Ref
Principal components analysis (PCA)	Exploratory	Unsupervised	Explores the data to look for any obvious clusters and for the detection of outliers. Reveal the most important factors of variability characterizing the data set. Relevant chemical information can be extracted from the loadings matrix.	(Fernández Pierna et al., 2016)
Hierarchical cluster	Exploratory/ Summative	Unsupervised	Generates a dendrogram which is a tree-like structure, the leaves are the samples and the branches shows their relationships	(Hastie et al., 2009)
Stepwise Multiple Linear Regression	Quantitative	Supervised	Multivariate linear regression method. A semi-automated process of building a model by adding or removing variables based solely on the t-statistics of their estimated coefficients.	(Raji et al., 2009)
Partial least squares regression (PLS)	Quantitative	Supervised	A canonical covariance method, finds X-components with maximal Y-covariance. finds a linear regression model by projecting the predicted variables and the observable variables to a new space	(Martens & Næs, 1989)
Discriminant analysis (DA)	Classification	Supervised	Series of algorithms that perform classification resulting in scores plots where the goal is to separate clusters according to group classifications.	(Richard G. Brereton, 2009)
Artificial neural networks (ANNs)	Quantitative or classification	Supervised	Multilayer perceptrons are powerful approaches that can map non-linear functions from X-data to Y-data.	(Cartwright, 2015)
Random forests	Classification or regression	Supervised	Combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest.	(Breiman, 2001)

2.3.3.3 Data treatment of the untargeted MS data

Data treatment of untargeted chromatogram-spectrum is usually the most time consuming part of identification and also challenging, as there could be hundreds or even thousands of peaks present in a single LC/HRMS chromatogram. Each feature can be described with retention time, m/z, and intensity as shown in Figure 2-10. In the case of looking for contaminants or residues in food samples, case and blank matrices samples are needed, the first step in MS data treatment is to extract the features from the massive MS chromatogram based on reliable algorithms, and then assign them to species according to their, isotope peaks and adducts of the protonated compounds. Thereafter the peaks originate from blank matrices, instrument, solvent, or sample preparation should be removed from the peak list. After grouping the peaks and irrelevant peaks filtration, a few or hundreds of suspected peaks may still present, and these peaks are used for further analysis with the purpose of structure identification. How to keep as many relevant peaks as possible, and exclude as many irrelevant peaks as possible are the main challenges to interpret the MS data in untargeted analysis.

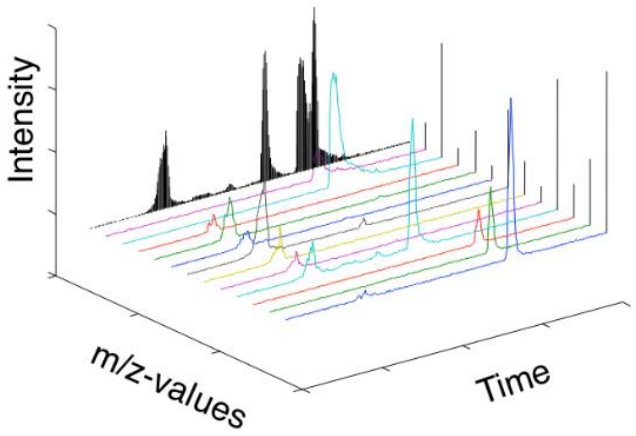


Figure 2-10. Description of data points by retention time, m/z value and intensity
https://github.com/bioinformaticsdotca/Metabolomics_2017/blob/master/mod2_lab.md

In untargeted analysis, metabolomics-like approaches earn more attention in recent years to deal with the MS data from food monitoring research with unrestrictive descriptive methodologies (Delaporte, Cladière, Bouveresse, et al., 2019; Mie et al., 2014). In the field of chemical risk, the conceptual issue linked to metabolomics is that the generated fingerprints are expected to reflect the presence of chemical contaminants or residues (also including their possible degradation and biotransformation products) by chemical profiles

(Broeckling et al., 2018). Although the application of metabolomics to food chemical analysis is still in its initial phase, and no consensus has been reached on defining many important parameters in data processing workflow such as mass assignment and data cleaning (Antignac et al., 2011). It still appears attractive for food analysis.

Currently, two complementary approaches are commonly used for metabolomics study: metabolic profiling and metabolic fingerprinting. Metabolic profiling is a hypothesis-driven approach, and it highly depends on prior knowledge of the biological system, while metabolic fingerprinting is to compare patterns or ‘‘fingerprints’’ of all metabolites accessible to the analysis that change in response to disease, toxin exposure, environmental or genetic alterations. The general workflow of metabolomics study is shown in Figure 2-11 (Antignac et al., 2011) . It starts with sample preparation by different methods, then according to mass fingerprinting carried out by different MS techniques, the data obtained can be further pre-processed in different platforms and software such as XCMS, MetaboAnalyst, and MZMine. After data pre-processing, multivariate analysis methods such as PCA, PLS-DA, and OPLS-DA can be applied for identification of clusters, detection of outliers, and understanding of which variables or features carry the most class separating information. Then features of interests can be selected and used for further identification and marker selection (K. L. Nielsen, Telving, Andreasen, Hasselstrøm, & Johannsen, 2016; Watrous et al., 2017). The interested features might correspond to potential residues or contaminants in food monitoring.

Data-processing steps including retention time alignment, features grouping, and data cleaning methodologies are very important for the final interpretation especially in metabolomics or metabolomics-like studies. Therefore, particular attention should be paid on it. The datasets are often very complex, not only because they contain a large number of chromatographic peaks that are separated according to retention time, but also because of fluctuations in the detected mass-to-charge ratio of ions between each scan, and the high influence of chemical and instrumental noise to the data acquired. Any problems during the data processing (e.g., default in peak alignment, peak integrations, or mass assignment) can therefore have major influences for the further statistical analysis, which overall would be erroneous (Eliasson et al., 2012; Jaeger, Hoffmann, Schmitt, & Lisec, 2016). To avoid such outcomes, the software parameters must therefore be undertaken carefully with regards to data-acquisition conditions. For example, depending on the chromatographic

separation and the resolution of the mass spectrometer used, the parameters for data processing need to be optimized simultaneously. One drawback of such a data-processing step is that the dedicated software may sometimes appear to be ‘‘black box’’ and it is tricky to identify which parameters are likely to have a stronger influence on the obtained results.

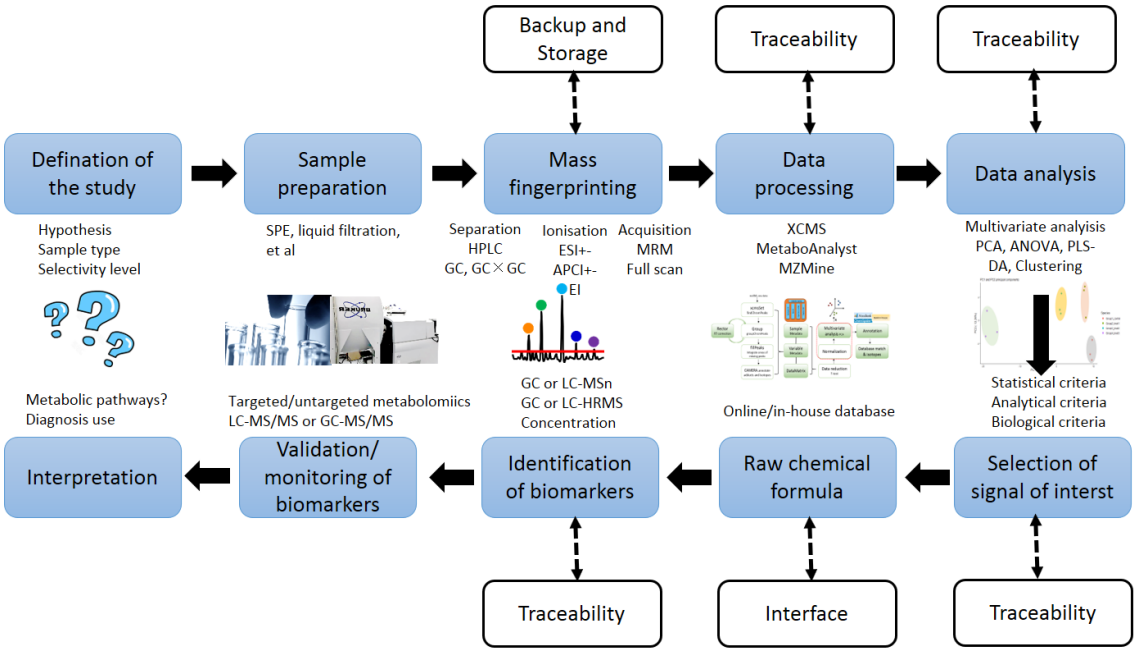


Figure 2-11. Analytical procedure of metabolomics study. 1, definition of the study including hypothesis, sample type and selectivity level. 2, sample preparation via various approaches. 3. Mass fingerprinting by various separation and MS techniques. 4, data processing via various platforms or software to extract the relevant features. 5, data analysis based on multivariate analysis, 6. Selection of interested signal after data analysis based on different criteria. 7, interpret the chemical formula preliminarily. 8, confirmation of biomarker 9, method validation. 10, interpretation of final results

Several tools have been developed to convert the initial three-dimensional raw data (m/z, retention time, ion intensity) into a two-dimensional data table reporting time-aligned and mass-aligned abundances of chromatographic peaks. These software or platforms may come from instrument vendors or other available sources for metabolomics analysis. The advantage and disadvantage of each of them are shown in the Table 2-3. e.g., XCMS is an open-source bioinformatics software that designed for statistical analysis of MS data, created by Scripps Research (Smith et al., 2006; Zheng, Clausen, Dalsgaard, Mortensen, & Bertram, 2013). XCMS provides a solution for the untargeted metabolomics workflow, while it is command line based and demands the knowledge of R. Meanwhile, there are limited options for customizing graphics in MetaboAnalyst, TracMass2 and MetaboliteDetect, and

some are not comparable with advanced analysis or have problems with handling a large dataset. Due to the advantages and drawback of each software and platforms, it would be a promising and efficient way to combine some different approaches and make it more feasible and efficient to do an advance analysis. XCMS and MetaboliteDetect were selected relevant software to this project, and this part is illustrated in Manuscript 1.

Table 2-3. Tools and platforms used for metabolomics study.

Platform	Description	advantage	disadvantage	reference
XCMS	-R-based platform for raw LCMS data processing and visualization	-Adjustable parameters -Streamlined workflow -open source--Comparable with multivariate analysis and other advanced analysis	-knowledge of R language -Command line based	(Smith et al., 2006; Zheng et al., 2013)
MetaboAnalyst	Online statistical analysis	-wide variety of statistical tests available - Open source - Comparable with Multivariate analysis and Other advanced analysis	-Relies on pre-processed data -Limited options for customizing graphics -Problem with handling large dataset	(*MetaboAnalyst*, 2020)
TracMass2	A Modular Suite of Tools for processing -Full Scan MS Data	- Quick visual -Feedback at every Processing step -User friendly -Open source	- Runs under MATLAB -Low quality graphics - Not comparable with advanced analysis	(Tengstrand, Lindberg, & Åberg, 2014)
MZmine	Raw data processing and visualization	-Java based -User friendly -Project batching	-Limited options for - Customizing graphics -numerous options can be overwhelming	(Katajamaa & Orešič, 2005)
Metabolite Detect	Data processing for raw LC-MS data	-User friendly -Fast process	- Instrument vendor based -Not comparable with advanced analysis Problem with handling large dataset	(Bletsou et al., 2015)

2.3.4. Chemical identification

2.3.4.1 Identification principles

The typical untargeted analysis usually includes sample collection, sample preparation, sample analysis, and data treatment steps. Each step is prone to errors as some compounds

may be overlooked. To obtain the maximum identification reliability, each of these steps needs to be investigated and considered in detailed. For example, minimizing the storage time of sample between collection, preparation and analysis, using solid-phase extraction (SPE) with mixed polarity or combine extracts obtained from different SPE cartridge, optimizing the LC separation, and combining with log*P* values (Kruve, 2019). The observed LC/HRMS spectrum should be correlated with a structure, and the process of the identification here is usually complex and tedious. Key parameters can be obtained from HRMS helping to identify a known unknown compounds usually need pure standards or additional analysis including the following points: mass accuracy of the precursor and product ions; relative isotopic abundance; fine isotopic distribution; ratios between product ions; collision cross-section (ion mobility) (Rochat, 2016).

Schymanksi et al., (2014) proposed a five-level scheme for interpreting the confidence achieved during identification (Figure 2-12). The exact mass of interest molecules serves as the lowest level of confidence (Level 5) and it does not contain much other unequivocal information even a formula. The reliable exact mass of the compounds can be obtained from HRMS. In this step, high precision of the instrument is important to narrow down the number of false hits. A precision of 1 to 2 ppm is usually sufficient, and up to 5 ppm is still acceptable. However, even with the precision of less than 2 ppm, a number of possible formula candidates are still commonly observed, e.g., a formula query using the formula $C_{10}H_{16}N_4O_2$ rendered 7080 hits in ChemSpider. Then information regarding adducts, isotopes could be helpful and worth presenting before carrying out MS/MS (Level 4). For example, in the study of Andersen et al., (2016), an unbiased technique which can identify Cl, Br, S, and other A + 2 elements containing metabolites in LC-HRMS data was developed based the special isotope patterns of these atoms.

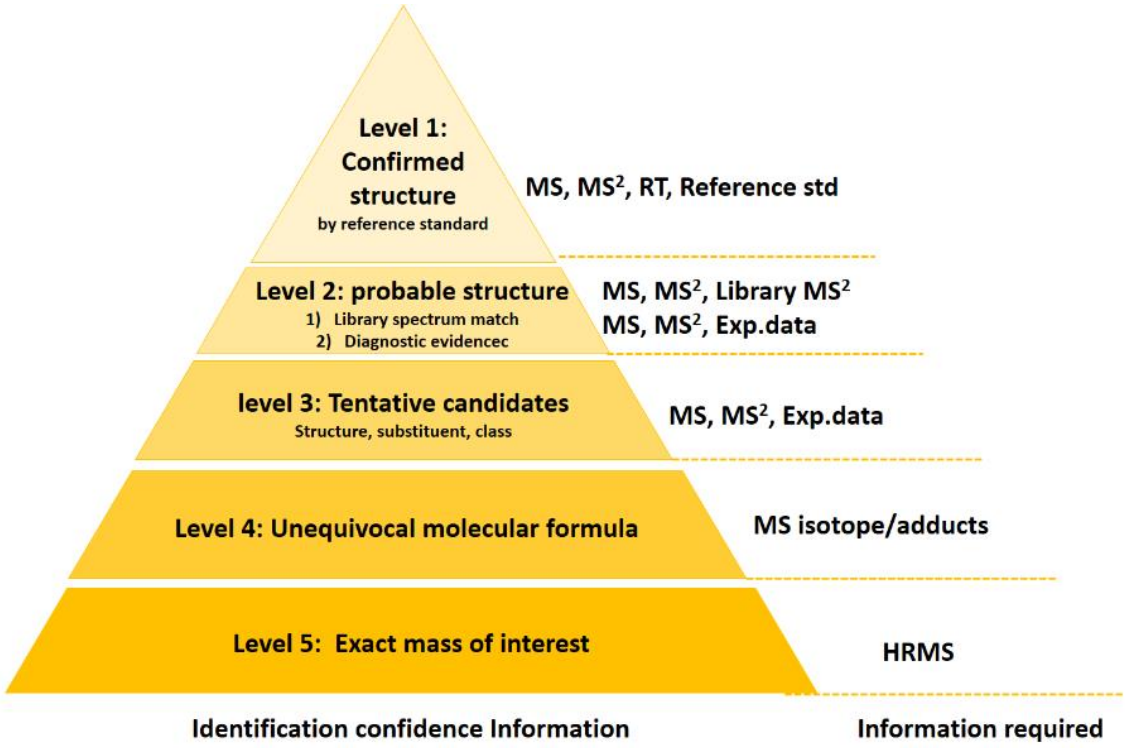


Figure 2-12: The identification levels for non-target analysis proposed by Schymanski et al (2014). Higher confidence levels (e.g. level 2) require the information obtained from the lower confidence levels (e.g. level 5, 4, 3)

In order to reveal structure of a compound, more information from CID and MS/MS experiments would be needed. This step was critical as it allows for tentative compound identification, despite no chemical standards being available. MS/MS is usually necessary to get the primary structure information of a compound. Based on the MS and MS/MS experimental data, some functional groups present in the structures might be revealed, although they are not always observed in MS/MS spectra. This information is helpful to get some structure candidates or classes of the compounds (Level 3). In this step, if only one possible structure match with MS/MS is obtained, then confidence Level 2-2 diagnostic evidence, might be achieved (Figure 2-12). Further investigation of structures can be obtained by matching the experimental data with MS and MS/MS library or database, fewer candidate structures can be obtained. Database is important here, and it can be either in-house built or obtained from various online sources (level 2-1). The final confirmation of the structure (Level 1) normally needs the aid of reference standards, as relating the MS spectrum to a specific structure can be very tricky due to the complexity of the MS spectra. Database hits are then verified by means of reference substances and MS/MS experiments.

For the compounds that can not be identified by this way, the structure identification is commonly carried out via more sophisticated MS scan techniques, e.g., MSn experiments (Liu & Hop, 2005), ion mobility spectrometry (Holčapek, Jirásko, & Lísa, 2012). In addition, the analytical techniques such as nuclear magnetic resonance spectroscopy (NMR) also helps to confirm the structure. However, the detection limit of NMR is far higher than the detection limit in MS analysis, and it is sometimes impossible to isolate and separate the pure suspected compounds with low concentration from the food matrices.

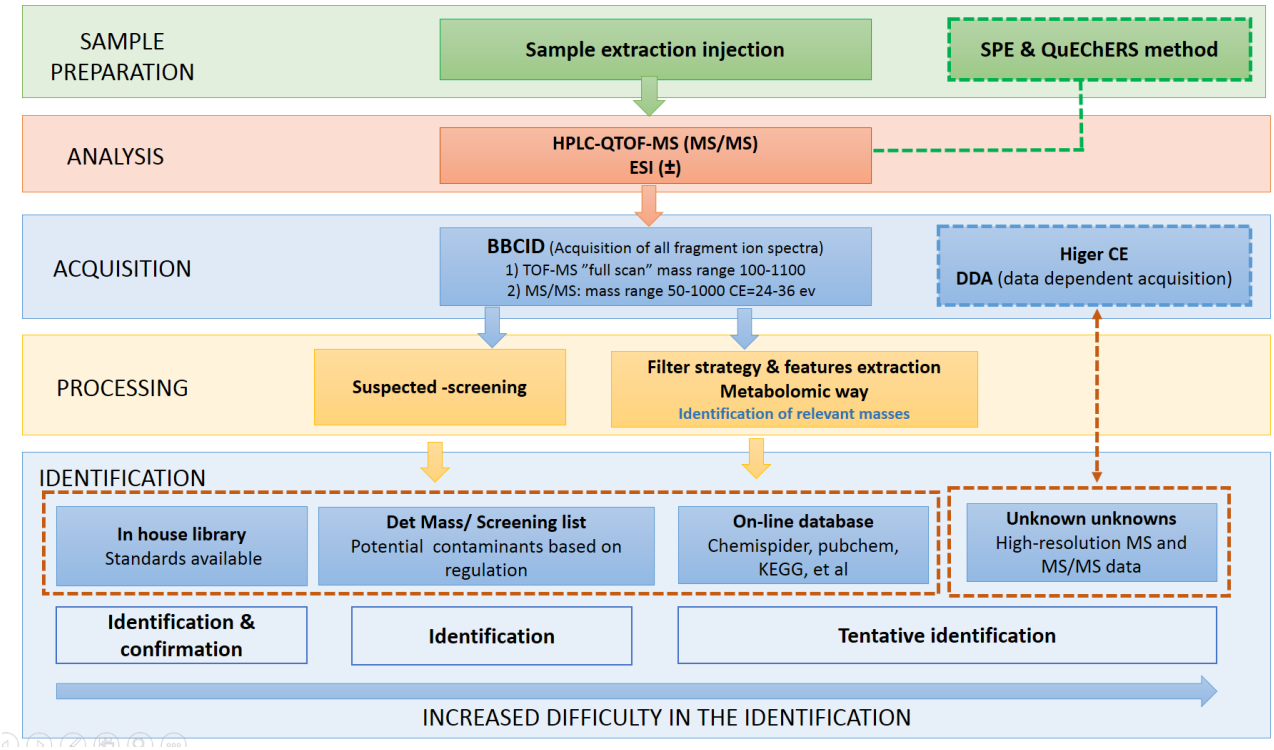


Figure 2-13. Workflow of suspected and untargeted screening including sampling, analysis, acquisition, processing and final identification based on LC-QToF-MS. The difficulty of identification increases from left to right depending on the database used and the availability of the standards

The whole workflow of suspected and untargeted screening including sampling, analysis, acquisition, processing and final identification based on LC-QToF-MS is shown in Figure 2-13. Data can be processed by either suspected screening or untargeted analysis inspired from metabolomics to exact the relevant features. The difficulty of identification is depending on whether the structures confirmation is carried out with reference standards or in-house database/online database, and whether the structure is known. Due to the lack of MS/MS data in mass spectral databases, the ability to identify unknown compounds through databases match is limited. To fill the gap between spectra and database, *in silico* MS/MS

spectra approach for predicting fragment and their abundance from chemical structure is an alternative strategy (Smith et al., 2006; Wolf et al., 2010).

2.3.4.2 Importance of database

Database are essential in untargeted screening especially suspected screening, as they allow the collection of information that can be leveraged for various types of analysis. The more comprehensive the databases are, the more successful dereplication can be obtained. In this study, a large amount of work based on the databases that contain the information of names, structures, formulas. log*P* values in some specific database to reduce false matches. The database primarily worked with in this project are listed in following Table 2-4. The in-house database is based on the active substances regulated by EU regulation and a survey of literatures. A selection of “relevant” substances was carried out to exclude the compounds are not relevant for MS. The detailed information about database construction is illustrated in section 3.3.1. Chemspider and Pubchem are both very comprehensive database including millions of entries, and they are challenging with too many matches sometimes for even one specific mass. However, it is still helpful sometimes to identify compounds with additional *in silico* MS/MS connection when the suspected compound is not available in the local database. The self-created (in-house) database with specific focus would be more useful than Chemspider or Pubchem, as it keeps the most relevant information for the suspected samples.

Table 2-4. Database could be used in untargeted food analysis in this project. The compounds focus area, entries of registered compounds and the access permission are included. The difficulty of identification is higher in online database such as Chemspider and Pubchem than local in-house database

Database	Area of focus	Entries	Availability	Ref
Chemspider	Compounds in commercial, environmental, forensic, and natural product samples	>26 million entries (elemental composition, molecular weight, or monoisotopic mass)	Free access	(Pence & Williams, 2010)
Pubchem	world's largest collection of freely accessible chemical information	93.9 million entries compounds	Free access	(Bolton, Wang, Thiessen, & Bryant, 2008)
METLIN	lipids, steroids, plant & bacteria metabolites, peptides, exogenous drugs/metabolites	over 500,000 molecular standards with MS/MS data at multiple energies	Free access	(Colin A Smith et al., 2005)
KEGG Compounds	Small molecules, biopolymers, and other chemicals relevant to biological systems	~ 18000 compounds entries	Free access	(Wolf et al., 2010)
In-house database	Pesticides, mycotoxins and metabolites, PAs	~ 7000 (e.g. structures, accurate mass, log*D*, log*P* values)	Free access	("EU - Pesticides database," n.d.; K. F. Nielsen & Smedsgaard, 2003; T. Wang et al., 2019)

2.3.5. Quantification and semi-quantification in food monitoring

2.3.5.1 Quantification in food monitoring

LC-QqQ-MS earns most applications focusing on the targeted analysis quantification (X. Wang et al., 2013), while it has limitations for untargeted analysis as explained in section 2.2.3. Thereafter, LC-Q-ToF-MS has been frequently reported in recent years in food

monitoring for untargeted screening, identification and quantification purposes (Frenich et al., 2014; Jin et al., 2020). It is usually very easy to carry out absolute quantification with reference standards in targeted analysis. The routine quantification step is to obtain pure standard of analyte, generate a calibration curve and get the concentration directly by calibration curve calculation with peak area. Improvement of quantification capability or performance, such as LOD, LOQ, linearity, reproducibility, specificity, are the main focus of quantification method development with standards. However, in reality of untargeted analysis, many standards are not commercially available, or there are simply too many compounds that need to be quantified or the standards are just not included in the run. It is not feasible to get all standards we need. How can we know the concentration without using standards is remaining a big challenge.

2.3.5.2 Semi-quantification method from literatures

Recently, quantification without using standards attracts a lot of attention (Gong et al., 2017; Pieke et al., 2017). The challenge in quantifying these compounds in LC−HRMS comes from the vastly different responsiveness of the compounds (Oss, Kruve, Herodes, & Leito, 2010). Even when two compounds are at the same concentration, they may produce absolutely different signals (Figure 2-14). The signals of some compounds may be already saturated while the signal of other compounds may be under the LOD. In addition, the signals produced from one instrument are not directly comparable to the signals measured on another as the factor that response scales are highly depend on different ionization source geometries and instrument vendors. In most cases, the sample matrices (Chalcraft, Lee, Mills, & Britz-McKibbin, 2009; P. Liigand, Liigand, Cuyckens, Vreeken, & Kruve, 2018) also influence the ionization efficiency. Therefore, many factors need to be taken into account in semi-quantification.

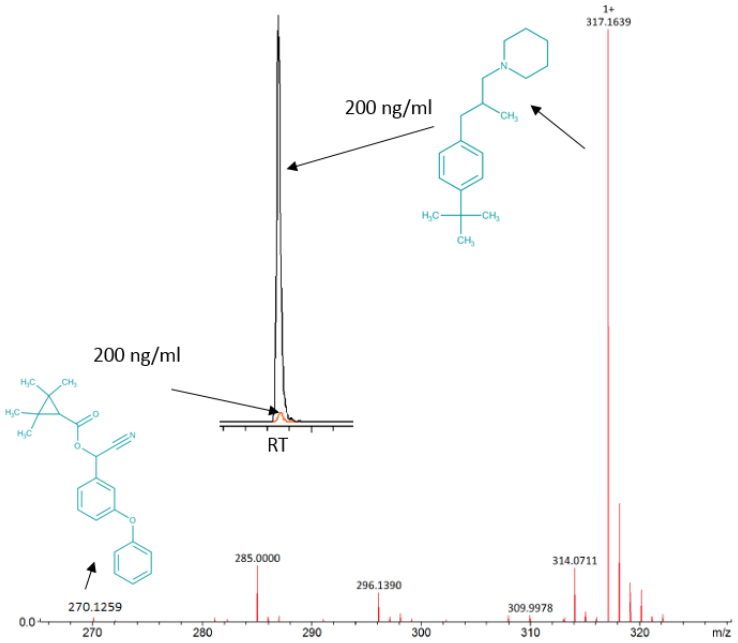


Figure 2-14. MS signals of fenpropathrin and fenpropidin that elute from the same retention time on LC-QToF

ESI response is considered to be a function of chemical structure along with molecular properties, as well as the constituents and pH of the mobile phase, flow rates, ion source parameters, protonation/deprotonation, resonance, inductive effects in compounds and other compounds from samples, which affect chemical reactivity for ionization (Hatsis, Waters, & Argikar, 2017). Measuring and quantifying ionization efficiency have been carried out in various studies to understand the ionization process. Many studies attempted to correlate the chemical responses with the physicochemical properties (Alymatiri, Kouskoura, & Markopoulou, 2015; Kruve & Kaupmees, 2017; Kruve, Kaupmees, Liigand, & Leito, 2014; P. Liigand et al., 2018; Oss et al., 2010). Due to the complexity of the ESI process, it is difficult to correlate responsiveness of small molecules to ESI-MS with any single parameter. Therefore, some recent attempts to predict ESI response have considered multiple physiochemical parameters simultaneously to combine ionization efficiency with multi-descriptors. An overview of studies reporting investigations between chemical response and physicochemical properties are summarised in Table 2-5. Many researchers found the correlation between chemical response and their physiochemical properties, e.g. p*K*a, molecular volume, WAPS, log*P*, α, proton affinity, gas-phase basicity, polarity, surface tension, and surface areas were ever found to influence ionization efficiency or chemical response, in despite sometimes their conclusions or results were contradicted in some way

(Table 2-5). For instance, log*P*, surface tension, surface area or p*K*a were found to be correlated with ionization efficiency in some studies (Golubović, Birkemeyer, Protić, Otašević, & M., 2016; Hermans, Ongay, Markov, & Bischoff, 2017; Oss et al., 2010). However, some studies also proved that no significant influence was found to correlate with ionization efficiency (Chalcraft et al., 2009; Gioumouxouzis, Kouskoura, & Markopoulou, 2015; Huffma, Poltash, & Hughey, 2012). In addition, many software tools used to calculate descriptors are either only commercially available or takes much efforts therefore it is still time consuming in the calculation part, such as COSMOtherm, or Turbomole.

Models based on the correlation between response and physiochemical properties analytes of analytes, such as MLR, ANN, and PLS models that mentioned in section 2.3.3.2, have somehow shown the potential of making semi-quantification without using standards come true. The developed models were mostly based on specific types of compounds, such as peptides (Raji et al., 2009), acid (Kruve & Kaupmees, 2017), steroids (Alymatiri et al., 2015), drugs molecules (Cramer, Johnson and Kamel, 2017), phenols (Henriksen and Juhler, 2005), and sartans (Golubović et al., 2016).

There are also some approaches used for the semi-quantification without using standards which are not based on the prediction models as described in Table 2-6. For example, chemical similarities (Kalogiouri, Aalizadeh, & Thomaidis, 2017; J. Yang et al., 2016), relative peak areas (Zendong et al., 2017), adjusted mass of the compounds (Nguyen et al., 2013), retention time differences and markers selection (Pieke et al., 2017), isotope labelling (Chung & Lam, 2018; Gong et al., 2017). These approaches in Table 2-6 also have the potential to do a semi-quantification, although most of them lack efficient validation to verify the quantification quality. It is also a challenge for quantification that many isotopic labelled compounds in general (e.g., PAs) are still not commercially available. This fact is unlikely to change in the near future since there is no facile chemical access, *in vitro* biosynthesis and isolation.

Therefore, in order to solve the quantification problems for a large number of unwanted compounds in food monitoring, the solution is either endlessly including more and more reference standards in the analytical methods, or developing a reliable and robust response prediction model for semi-quantification without using standards. Apparently, the second solution seems to be a better one with lower cost and higher feasibility.

Table 2-5. Studies about the correlation between ionization efficiency and physicochemical properties in literatures.

Compounds	Ion-source	Conc range	Dependence founds	Other parameters	Software and methods	Quan-information	matrices	Ref
62 compounds of different chemical nature	ESI	10^{-8} to 10^{-3} mol/L	p*K*a, MV	log*P*, PSA dipole moment (δ)	COSMOtherm Spartan MLR	logIE	Solvent	(Oss et al., 2010)
62 acidic compounds validated for 17 acids in two solvents	ESI	10^{-5} M to 10^{-7} M	WAPS, α, Hb_acc	p*K*a, log*P*, Klamt parameters (Hb,_don3)	COSMOtherm Turbomole MLR	log*IE*	Solvent	(Kruve & Kaupmees, 2017)
29 phenols, 18 benzoic acids, and 3 bifunctional compounds	ESI	2.2×10^{-5} to 2.4×10^{-7} M	α and WAPS	MV, p*K*a	COSMOtherm Turbomole MLR	log*IE*	Solvent	(Kruve et al., 2014)
40 compounds with wide range	APCI	14 to 1.6 µM.	Ionization degree, WANS, p*K*a, log*P*, MV, polarizability	sig2, sig3, Hb_acc3, Hb_don3	COSMOtherm Turbomole PLS, PCA	log*IE*	Solvent	(Rebane et al., 2016)
10 compounds mainly acidic drugs	ESI	1-200 µM	WAPS, α,	p*K*a	COSMOtherm Turbomole MLR	log*IE*	Solvent, biological matrices	(P. Liigand et al., 2018)
30 steroids (pK_{basic} < -2.65, pK_{acidic} > 10.6	ESI	1 µg/mL	-OH GB, solubility, HBA, log*P*, and log*D*, HBD, density, ST	31 descriptors (functional groups, topological geometrical descriptors)	PLS	Signal intensity	Solvent	(Alymatiri et al., 2015)
46 cationic/zwitterionic metabolites	ESI	2-100 µM	MV, log *P*, absolute ion mobility, effective charge	log*D* p*K*a	Ultra software, Hubbard-Onsager, ACD ALOGPS PCA, PLS	IE	Solvent, red blood cell	(Chalcraft et al., 2009)
77 drugs	ESI	10 and 100 µM	PA, SA	PSA, aqueous solvation free energies, gas-phase and aqueous, ΔG	MLR, PC Model program, M06-L/MIDI, SMD aqueous solvation model	log*RF*	Solvent	(C. J. Cramer, Johnson, & Kamel, 2017)

12 GXG peptides	ESI	0.5–100 µM	GPB; TSA, SA, solvation energy, NPSA; log*D*, , PSA.	PA	MLR, decision tree, support vector regression	Signal intensity	Solvent	(Raji et al., 2009)
7 sartans	ESI	5–100 µM	log*P*, P*K*a, SA, number of proton acceptors, Polarity, chargeability, H-acc	CSEV	MLR ANNs STATISTICANN, ChemAxon Ltd, Chem3D® Pro, Marvin Sketch	Signal intensity	Solvent	(Golubović et al., 2016)
99 compounds	ESI	0.5 µg/mL	PSA, log*P*, log*D* values, HBA, HBD, WA, density of solid, ST, GFE, P*K*a	log*S*	Simca-P ACD/labs PLS	Signal intensity	Solvent	(Mandra, Kouskoura, & Markopoulou, 2015)
25 organic acids	ESI	0.05–10 µg/mL	HBC, HOMO energy, HBD, ratio of organic phase, PSA	more than 200 descriptors	MLR Napierian logarithm Discovery Studio. Sybyl-X	Response factor	Solvents and pharmaceutical adjuvants	(Wu et al., 2013)
127 amino acids, derivatives	ESI	50 µM	MV,log*P*, p*K*a, ST	-	QSPR, Dragon 5.5, Chemicalize, van der Waals radii, Marvin- sketch, Chemspider, ACD	IE	Solvent	(Hermans et al., 2017)
1248 compounds	ESI	5 mM	Oxygen atoms, negatively charged atoms, descriptors linked to aromaticity	63 topological, electronic, count, and structural descriptors, 10 molecular descriptors	Program Corina, PLS model, knowledge-based model	Signal Intensities	Solvent	(Mauser, Roche, Stahl, & Müller, 2005)
15 oligonucleotide	ESI	2.5 to 15 mM	BP, PA, partition coefficient, WS, Henry's law constants, hydrophobic thymine content	Vapor pressure, gas phase basicity, density, MW	SAS JMP-SAS, HENRYWIN model, PLS	Normalized signal intensity	Solvent	(Basiri, Murph, & Bartlett, 2017)
6 peptides	ESI	10^{-3} M	NPSA, GFE	Affinity for the surface phase,	Equilibrium partitioning model, Linear regression	Log Response	Solvent	(Cech & Enke, 2000)
58 polar metabolites (amino acide, amines. Peptides, etc)	CE-ESI	2-100 µM	MV, log*D*, absolute mobility, effective charge	p*K*a,	PCA, MLR, Chem3D Ultra software, ALOGPS 2.1, ACD/Laboratort	Relative response factor	Solvent	(Chalcraft et al., 2009)

19 drug like compounds	ESI	10^{-4} M	p*K*b	log*P*, p*K*a, GB	NIST Web Book, ChemSketch	Log Response	Solvent	(Ehrmann, Henriksen, & Cech, 2008)
25 neutral acylsucrose metabolites	ESI	0.08–50 µM	NSP, the length of the longest acyl chain, number of hydroxyl groups	-	ChemBio3D Ultra software, Linear regression	Response factor	Solvent	(Ghosh & Jones, 2015)
110 drugs	ESI	10 to 200 ng/mL	p*K*a, ionization percentage, PSA, HBA, COOH, WS , ST	log*P*, log*D*, Enthalpy of vaporization, MV, Refractivity, Polarizability, H bond acceptor, H bond donor, solvent accessible area	PLS Simca-P , ACD/Labs Percepta Platform	Signal Intensities	Solvent	(Gioumouxouzis et al., 2015)
71 drugs/metabolites	ESI	1 to 1000 nM	log*D*	105 descriptors	PCA, univariate analysis, VolSurf+	Response factor	Solvent	(Hatsis et al., 2017)
35 phenols	ESI	10^{-6} M 10^{-7} M	log*P*	PA, pKa,	Chem- sketch	Signal Intensities	Solvent	(Henriksen & Juhler, 2005)
49 acidic molecules	ESI	1 µM	Nonpolar character and MV	pKa, PSA, NSA, MV, solvation energy, gas phase PA, etc	equilibrium-partitioning model	Normalized response	Solvent	(Huffma et al., 2012)
56 small N-containing compounds	ESI	1 Mm - 40 µM	Solution basicity, polarity, vaporability, log*D*,	-	ChemSpider, Spartan software, DFT, R	Signal Intensities	Solvent	(Kiontke, Oliveira-Birkmeier, Opitz, & Birkemeyer, 2016)

Ionization efficiency (IE).Hydrogen bond acceptor ability (Hb_acc), ionization degrees (α), WAPS (weighted average positive sigma), weighted average negative sigma (WANS), number of halogen atoms in the molecule n(Hal), Gas basicity (GB), Hydroxy groups (-OH), Carbonyl groups (CO), Hydrogen bond acceptor (HBA), Hydrogen bond donor (HBD, nonpolar surface area (NSA), polar surface area (PSA), total surface area (TSA), Connolly solvent-excluded volume (CSEV), the ability of analytes to acts as hydrogen bond donors or acceptors (AHbd, AHba), hydrogen bond acidity (HBC), quantitative structure−property relationship (QSPR) model, surface tension (ST), nonpolar surface areas (NPSA). Boiling point (BP), water solubility (WS), proton affinity (PA), Gibbs free energy (GFE), gas-phase basicity (GPB)

Table 2-6 studies used for the semi-quantification without using standards that are not based on prediction models

Compounds	Ion-source	Conc range	Software and methods	Matrices	ref
8 glycans	MALDI	0.5 nM to 6.5 µM	Reference to Exogenous Standards	Biological matrices	(Mehta et al., 2016)
2 Drug Metabolites	RAD	10.0 to 5000 ng/mL	Radio isotopologues ^{13}C	Biological matrices	(Gong et al., 2017)PA
28 PAs	ESI	0.005 µg/L to 1 µg/L	Isotopically labeled PAs	Food	(Chung & Lam, 2018)
34 lipids	ESI	$0\sim2^{15}$ ng/mL	RF-based approach with internal standards, lipidAnalyzer	Plasma, mice tissue, cell pellet	(Tu, Yin, Xu, Wang, & Zhu, 2018)
10 lipophilic marine algal toxins	ESI	162 nM to 176 nM	Differences in relative molar response	Solvent environmental samples	(Zendong, Sibat, Herrenknecht, Hess, & McCarron, 2017)
7 polar lipids	ESI	3.3 mg/mL	Linear regression, internal standard	Biological tissues	(Cífková et al., 2012)
32 pesticides	DESI	20 to 100 µg/L	Normalization to one ISTD - Linear regression	Solvent and fruit sample	(Gerbig et al., 2016)
14 polyphenol compounds	ESI	0.5 up to 1.5 µg L^{-1}	LDA, ACO-RF , Select markers based chemical similarity	Olive oil	(Kalogiouri et al., 2017)
29 flavonoids	ESI	2-100 µg/mL	SIMCA-P+, PLS-DA, OPLS-DA chemical profiling strategy, similarity in common standards	Ginkgo extracts	(J. Yang et al., 2016)
18 analytes	ESI	0.25 to 1.0 µmol/L	Marvin software, JMP, normalization, choose markers by closed retention time.	Solvent	(Pieke et al., 2017)
6 metabolites	ESI	5.0–200.0 ng/ml	Calibration curve of one similar compounds	Solvent and Plasma	(Du, Liu, Shen, Zhu, & Xing, 2012)
4 nature compounds	ESI	10^{-5} M 10^{-3} M	Using the *adjusted mass* of the compound, MLR	Environmental samples	(Nguyen, Nizkorodov, Laskinc, & Laskin, 2013)
132 structurally diverse drug and their 233 metabolites	ESI	2 mM	Relative peak areas	Solvent	(Blanz et al., 2017)
9 urinary metabolites	ESI	20 - 5000 nM	Extending MS analysis times during slow FAIMS scanning, extrapolation of suppression regime sensitivity	Urine	(Beach & Gabryelski, 2013)

3. Scientific Research

This chapter briefly summarizes the outcome of the research performed in this project. The main results from the scientific work are shown as supplemented in three published manuscripts and one submitted manuscript.

3.1. Analytical strategy development

The whole project was carried out based on following flowchart as described in Figure 3-1, which includes analysis, data processing, identification and (semi) quantification. Sampling and sample exaction is also a very important elements of a successful untargeted analysis, while it was not the focus in this project. This project focused on the analytical method development (optimization of ion source setting and eluent composition), data processing strategy development (comparison of different data processing approaches and develop more efficient strategy), identification development, and also semi-quantification model development and its application. Honey and cereal were selected as the case studies for the untargeted analysis and semi-quantification model application.

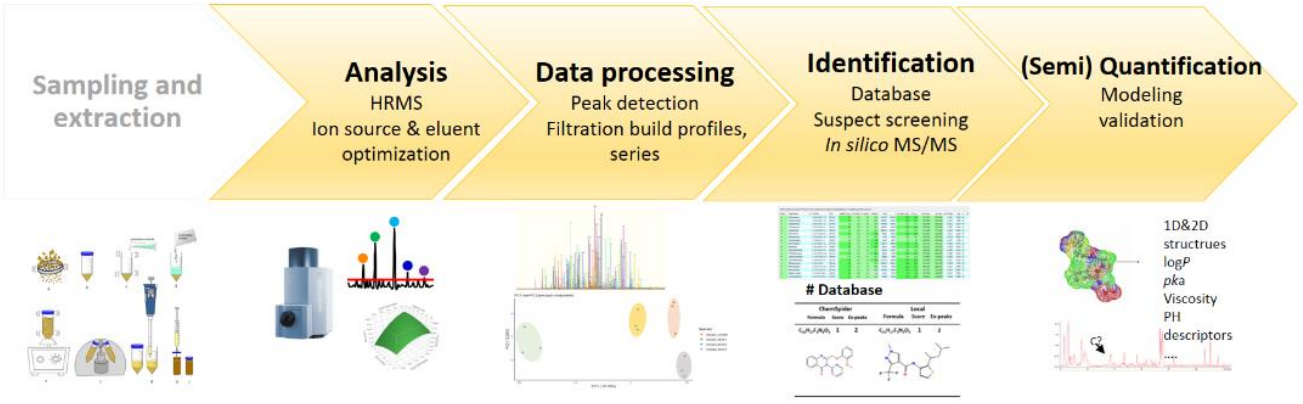


Figure 3-1. Workflow of this project. The whole procedure includes analysis, data processing, identification and (semi)-quantification

3.1.1. Study on Ion source and eluent composition

Any study to improve the sensitivity of a LC/MS method will require the ionization process of the ion source to be clearly understood to maximally optimize its efficiency. The ionization efficiency depends on compounds properties, solvent properties and electrospray conditions (Kruve, 2016; J. Liigand, Kruve, Leito, Girod, & Antoine, 2014). Therefore, three major parameters have to be taken into account, namely, the instrument setting, the analyte characteristics, and the eluent composition. To improve the detection and sensitivity of a wide range of compounds, instrument setting particularly ion source, eluent composition of the mobile phase were studied in this project.

A standardized ESI source design would be prerequisite for improving the transferability of ESI-MS analytical methods from one instrument to another, and for the development of ESI responsiveness prediction (Bruins, 1991). On the other hand, the selection of the mobile phase particularly the eluent compositions is also important to obtain good chromatographic separation, ionization and sensitivity (Picó, Blasco, & Font, 2004), Mobile phase additives in LC/MS are usually used to improve peak shape appearance, ionization efficiency of analytes and MS coverage. Since the eluent composition can be handled with the most flexibility, the water phase in which the analyte is introduced into the ion source becomes of utmost importance for the optimization of the sensitivity of a LC/MS method.

ESI and APCI ion sources are universally available and have been used for developing analytical methods of determining e.g. pesticides, mycotoxins and other contaminates in food samples. It is a common procedure to try both ion sources on analytes of interest to see which one works better. Meanwhile, due to the high flexibility of eluent composition, it is needed to evaluate which eluent composition works the best to maximize the MS coverage. To achieve better sensitivity, the influences and interaction between different ionization source parameters in APCI and ESI, and the influences of different eluent compositions with different pH on response factors were all studied. The details are illustrated in the following sections.

3.1.1.1 Experimental setup

a. Ion source optimization

Comparison between APCI and ESI ion source was carried out on a set of 200 compounds in Bruker-micrOToFq. The corresponding compounds and information used in the study (e.g., name, formula, log*P*, molecular volume) are shown in Table S1 in Appendix IV. The log*P* values of all the tested compounds for ion source optimization is in a range of -0.07 to 6.76, molecular volume is in a range of 125.3 to 772.5. The standard mixture (1 µg/ml mixture dissolved in methanol) was introduced into the ion source via direct flow injection using a syringe pump at a flow rate of 0.18 mL/h and a mobile phase (composed of 50% 0.1% formic acid in water and 50% MeCN) at a flow rate of 0.3 mL/min. This setup resulted in a standard concentration of 10 ng/mL going into the ion source. The dry gas temperature and vaporizer temperature factors were set to hard-to-change parameters, thus the generated settings with same dry gas temperature or vaporizer temperature were assigned to the same block to reduce the waiting time of temperature increasing or decreasing. ESI in line with the trial performed using the APCI source, a Custom designer split plot experiment was generated using JMP® Version 13 software from SAS. The design was generated from a random seed with 10000 starts and D-optimal criterion (de Aguiar, Bourguignon, Khots, Massart, & Phan-Than-Luu, 1995). Each combination of ion source parameters was repeated three times and an average of the obtained results were used for the data analysis. Six factors in APCI and four factors in ESI were entered as continuous 2-level parameters. The parameters that need to be optimized are shown in Table 3-1 and Table 3-2 for APCI (6 factors out of 7 in APCI source settings) and ESI (4 factors out of 5 in ESI source settings). The full design matrix included 40 different setups in APCI and 20 different setups in ESI of ion source are presented in Table 3-3 and Table 3-4. All the response factor ($\frac{peak\ area}{\text{Concentration (ng/mL)}}$) were scaled to normalize the range of independent variables of data. Three targeted parameters, namely, Min25% (sum of the n response factors of 25% lowest-response compounds), NC (number of detected compounds), Aver (average response factor of all detected compounds) were selected to be the indicators for evaluating the optimization performance.

Table 3-1. D-optimal Custom design on APCI

	Parameter	Abbreviation	Lev-1	Lev-0	Lev+1
1	Capilary voltage (V)	Cap	2500	3500	4500
2	Corona (nA)	Corona	3000	4000	5000
3	Nebulizer (Bar)	Neb	2	2.5	3
4	Dry gas flow rate (L min-1)	DryF	3	4	5
5	Dry gas temperature (℃)	DryT	200	225	250
6	Vaporizer temperature (℃)	VapoT	300	375	450

Table 3-2. D-optimal Custom design on ESI

	Parameter	Abbreviation	Lev-1	Lev-0	Lev+1
1	Capillary Voltage (V)	Cap	2500	3500	4500
2	Nebulizer Pressure (bar)	Neb	2	2.5	3
3	Dry Gas Flow Rate (L min-1)	DryF	8	10	12
4	Dry Gas temperature (˚C)	DryT	150	185	220

Table 3-3. Custom designer split plot experiment generated from a random seed with 10000 starts and D-optimal criterion (APCI) using JMP® Version 13

	Whole Plots	Cap	Corona	Neb	DryF	DryT	VapoT	min25%	NP	Aver
1	1-1	4500	5000	2.5	5	225	300	0	37	2168
2	1-2	3500	4000	3	3	225	300	34800	58	14106
3	1-3	2500	5000	2	3	225	300	44002	58	19217
4	1-4	2500	3000	2.5	5	225	300	0	36	2305
5	1-5	3500	4000	2.5	4	225	300	2573	47	6469
6	2-1	2500	5000	3	5	250	450	0	37	2668
7	2-2	4500	4000	2.5	5	250	450	0	18	683
8	2-3	4500	3000	3	3	250	450	0	44	7459
9	2-4	2500	3000	2	4	250	450	0	36	3534
10	2-5	3500	5000	2	3	250	450	6874	51	9486
11	3-1	3500	4000	2.5	4	250	375	0	40	4899
12	3-2	3500	5000	2	5	250	375	0	33	1483
13	3-3	2500	4000	2.5	3	250	375	29200	58	13970
14	3-4	4500	5000	3	3	250	375	17308	56	9996
15	3-5	3500	3000	3	5	250	375	0	38	2713
16	4-1	4500	5000	3	5	200	450	0	38	3295
17	4-2	3500	3000	2.5	5	200	450	0	31	1805
18	4-3	2500	3000	3	3	200	450	15674	56	11820
19	4-4	2500	5000	2	4	200	450	0	43	5834
20	4-5	4500	4000	2	3	200	450	12880	54	11731
21	5-1	4500	4000	3	5	250	300	0	35	2248
22	5-2	3500	3000	2	5	250	300	0	27	1055
23	5-3	4500	4000	2	3	250	300	27555	57	14866
24	5-4	2500	3000	3	3	250	300	40641	58	17389
25	5-5	2500	5000	2.5	4	250	300	3788	48	6962
26	6-1	4500	3000	2.5	3	225	375	19618	56	11863
27	6-2	3500	4000	2.5	4	225	375	1056	46	6512
28	6-3	3500	4000	2.5	4	225	375	0	45	5623
29	6-4	2500	5000	3	3	225	375	29214	59	14196
30	6-5	4500	5000	2	4	225	375	0	43	5884
31	7-1	4500	4000	2	5	200	300	0	34	1818
32	7-2	2500	5000	3	5	200	300	0	42	4488
33	7-3	4500	3000	3	4	200	300	9676	53	10247
34	7-4	2500	3000	2	3	200	300	45141	58	19646
35	7-5	4500	5000	2.5	3	200	300	51383	59	22178
36	8-1	3500	3000	2.5	3	225	376.5	27387	58	14437
37	8-2	4500	3000	2	4	225	376.5	0	41	4973
38	8-3	3500	4000	3	4	225	376.5	13326	55	9932
39	8-4	2500	4000	2	5	225	376.5	0	33	1733

40	8-5	3500	4000	2.5	4	225	376.5	0	45	5648

Note: NP, total number of detected compounds peaks. Average, average response of all detected compounds. Capi, capillary voltage. Neb, Nebulizer. DryF, dry gas flow rate. DryT, dry gas temperature. Whole Plots number is the serial number of each source setting combination in different temperature block. Min25% is sum of the response factors of 25% lowest-response compounds.

Table 3-4. Custom designer split plot experiment generated from a random seed with 10000 starts and D-optimal criterion (ESI) using JMP® Version 13.

	Capi	Neb	DryF	DryT	NP	Min25%	Aver
1-1	2500	3	12	150	59	0	12056
1-2	4500	2	12	150	75	18920	18332
1-3	2500	2	8	150	81	31403	23523
1-4	4500	3	8	150	86	35410	20917
2-1	2500	2	12	150	60	0	11708
2-2	2500	3	8	150	82	25505	23614
2-3	4500	3	12	150	84	35293	22061
2-4	4500	2	8	150	79	27546	18674
3-1	2500	3	10	220	60	0	11606
3-2	4500	3	8	220	70	8185	11785
3-3	4500	2.5	12	220	66	2187	9566
3-4	3500	2	8	220	70	7737	12475
4-1	4500	2	9.66	220	68	5390	10588
4-2	3500	3	12	220	63	0	9445
4-3	2500	2.5	8	220	68	4629	15088
4-4	2500	2	12	220	46	0	5977
5-1	2500	3	12	185	53	0	9493
5-2	3500	2.5	10	185	75	15824	18863
5-3	2500	2	8	185	78	18840	21906
5-4	4500	2	12	185	73	13748	16102

Note: NP, total number of detected compounds peaks among 200. Average, average response of all detected compounds. Capi, capillary voltage. Neb, Nebulizer. DryF, dry gas flow rate. DryT, dry gas temperature. Whole Plots number is the serial number of each source setting combination in different temperature block. Min25% is the sum response factor of the 25% lowest signals

b. Eluent composition optimization

As acetonitrile and methanol are the most commonly used organic phase in reversed phase chromatography, the influence of acetonitrile and methanol on tested compounds were compared first. In order to compare the performance of different organic phases, three different phases with triplicates (A: MeCN-0.1%FA, B: MeCN -2.5mM NH4F, C: MeOH-0.1%FA) were

set up for comparing their effects on the signal of test compounds. All the data were standardized to make a final comparison in the heat map.

Standard score (Empereur-Mot, Zagury, & Montes, 2016), $Z = \frac{\mathrm{X}-\mu}{\mathrm{SD}}$,

where μ is the mean value of the response factor ($\frac{peak\ area}{\mathrm{Concentration\ (ng/mL)}}$) of each compound in 3 different phases, X is the actual value of the response factor of each compound, and SD is the corresponding standard variation.

The eluent setup is shown in Table 3-5. To investigate the influence of pH of water phase on the response factor of test compounds, the response factor in ESI positive mode using each of the 11 mobile phase compositions were determined in triplicates, i.e. 33 runs. The pH is in the range from pH 2.7 to 5.7. The teste standard solution was a mixture of 200 pesticides and 15 mycotoxins (Table S1 in Appendix IV) dissolved in 50% MeOH (v/v). The concentration of each pesticide was 200 ng/mL. The log*P* values of all the tested compounds for ion source optimization is in a range of -0.17 to 6.76, molecular volume is in a range of 125.3 to 695.8. The pH of the mobile phase was adjusted by adding acetic acid (HAC) or formic acid (FA) and/or ammonium hydroxide (NH4OH). The ion source setting were: capillary voltage 3.5 kV, nebulizer pressure 2 bar, dry gas temperature 180 °C, and dry gas flow rate 10 L/min. The acquisition range was m/z 50 to 1000 at a rate of 2 Hz. The LC gradient has been used is same as described in manuscript 1.

Table 3-5. Eluents composition setup, pH was measured by pH meter

	A	B	pH
A	0.02% FA	MeCN	3.1
B	0.1% FA	MeCN	2.7
C	0.02FA-2.5N	MeCN	4.1
D	2.5 mM NH4OH, 0.1% FA	MeCN	3.0
E	5 mM NH4OH, 0.1% FA	MeCN	3.3
F	10 mM NH4OH, 0.1% FA	MeCN	3.9
G	0.02% HAC	MeCN	3.3
H	0.1% HAC	MeCN	3.7
I	2.5 mM NH4OH, 0.1% HAC	MeCN	4.1
J	5 mM NH4OH, 0.1% HAC	MeCN	4.6
K	10 mM NH4OH, 0.1% HAC	MeCN	5.6

3.1.1.2 Results

a. APCI optimization

As the difficulty of the compounds separation in flow injection, only a maximum 60 compounds out of 200 were detected in APCI mode. From the heat map as shown in Figure 3-2, settings of 7-5, 7-4, 1-3, 5-4 corresponding to Table 3-3 obtained the highest response for most of the test compounds (the more red the colour, the more intense the response), and the average response of them are 22178, 19646, 19217, and 17389 counts, Ion source settings 2-2 and 5-2 show the lowest response for most compounds, the average responses for all compounds are 1055 and 683 counts. There were 33 and 42 compounds among these 60 compounds were not detected at all in 2-2 and 5-2, which means that the response factor of compounds in APCI are quite sensitive to the ion source setting, where many compounds suffer a very high risk of being non-detected when an unsuitable setting for them are used. The log*P* value of the compounds have been detected in APCI is in the range from 0.62 to 6.76, although the response of compounds of low log*P* value (e.g., 0.62 for dazomet and 1.06 for acetamiprid) was very low and detected only in a few settings. The compounds with higher log*P* value (log*P*>*4*) generally have relative higher signals.

Min25%, NC, Aver were then selected to be the indicators for evaluating the optimization performance. It can be found that DryF (*P*<0.001) has the most significant influence on the Min25%, NP, and Aver, as all the settings with high flow rate showed lower response than others, by using maximizing desirability function in the model, the optimal ion source setting is shown in (Figure 3-3), while all the source settings with low DryF show relatively higher response factor in comparison to the source settings with high DryF. DryT and VapoT also have large influence on the Min25%, NP, and Aver, and the settings of high DryT and VapoT tend to have negative effects on these indicators, and vice versa. Parameter Estimates (Parameter Estimates reports the t test and it compares the estimate to the value of the null hypothesis) of the influence of source setting on Min%25 is shown in Table 3-6, the DryF (*P* <0.0001), DryT (*P* < 0.02), VapoT (*P* <0.0001) have the biggest influence on Min25%, which is consistent with Figure 3-3. Interactions between DryT and DryF (*P*<0.05), DryF and VapoT (*P*<0.001) have also been found..

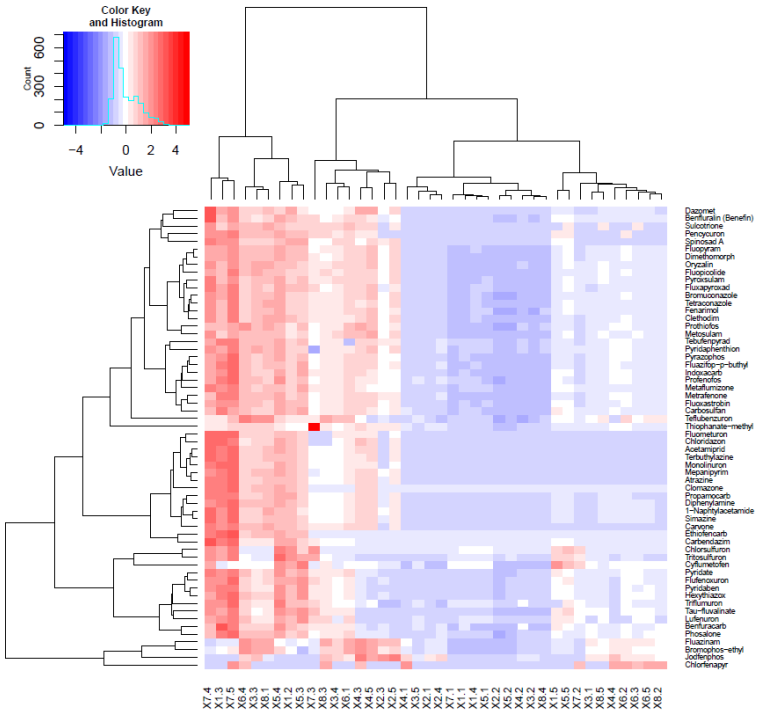


Figure 3-2. Heat map generated using the statistical software "R" from the scaled response factor when introduced via direct flow injection into the APCI source. Red areas indicate a relative increase in the instrumental response and blue areas indicate a relative decrease in the instrumental response

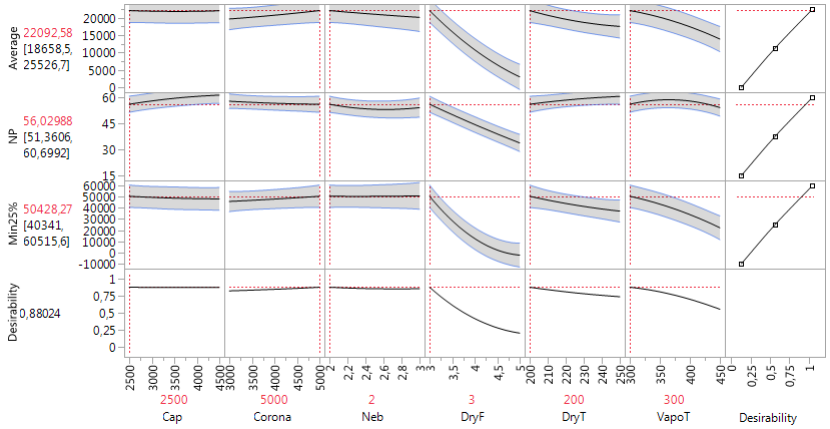


Figure 3-3. Prediction profile of ion source parameters in APCI after maximizing the desirability by using JMP® Version 13 software from SAS

Table 3-6. Parameter estimates by T-test for the influence of the optimized factors on Min25% in APCI

Term	t Ratio	Prob>|t|
Cap(2500,4500)	-2,33	0,0383*
Corona(3000,5000)	0,26	0,8023
Neb(2,3)	0,17	0,8670
DryF(3,5)	-13,86	<,0001*
DryT(200,250)	-2,90	0,0132*
VapoT(300,450)	-7,03	<,0001*
Cap*Cap	0,27	0,7913
Cap*Corona	0,92	0,3743
Corona*Corona	0,11	0,9115
Cap*Neb	0,32	0,7546
Corona*Neb	0,02	0,9826
Neb*Neb	0,13	0,8971
Cap*DryF	1,88	0,0851
Corona*DryF	-0,92	0,3749
Neb*DryF	0,43	0,6759
DryF*DryF	6,29	<,0001*
Cap*DryT	-1,29	0,2223
Corona*DryT	-1,50	0,1596
Neb*DryT	-0,11	0,9159
DryF*DryT	2,56	0,0252*
DryT*DryT	0,61	0,5502
Cap*VapoT	-0,75	0,4677
Corona*VapoT	-0,39	0,7001
Neb*VapoT	-0,58	0,5724
DryF*VapoT	6,72	<,0001*
DryT*VapoT	0,99	0,3414
VapoT*VapoT	-1,41	0,1842

t-ratio is the estimate divided by the standard error. t-ratios greater than 1.96 (in absolute value) suggest that the coefficient is statistically significantly different from 0 at the 95% confidence level. Prob > |T| labels the P values or the observed significance levels for the t statistics

b. ESI optimization

In total 86 compounds were detected in ESI flow injection mode. It can be found that 2-2, 1-3, 2-3 5-3 (Table 3-4) showed the highest average response compared with other settings in ESI ion source. Regarding the number of detected compounds and the overall performance as shown in Figure 3-4, 1-4, 2-3, 2-2 perform the best, which means the high average response value does not mean high response for every compound. For example in setting 1-4, although the average response factor was not the highest, it still detected with the most compounds (86

compounds). Making compromise among compounds sometimes is more important in untargeted screening, as the aim of untargeted screening is to detect and find as many compounds as possible. Settings like 5-1 (9493), 4-2 (9445), 4-4 (5977) show the lowest response.

According to the parameter estimation by JMP, there was no significant influence of source parameters on the Aver or the NP was found. The significant influence of source parameters on Min25% was found, as shown in Table 3-7, Cap (P<0.02) tend to have biggest influence on Min25%, and the interaction between Cap and DryF (P<0.001), Cap and DryT (P<0.05) will also influence Min25%. However, this is no clear tendency was found from the prediction profile, so it is hard to predict which setting is optimal for the entire analysis in ESI by the model prediction. This is probably due to the factor by using flow injection mode in ESI, all the tested compounds were ionized in liquid phase and introduced to the gas phase together. The ion formation is more susceptible to the matrix with a lot of other compounds that may compete the charge and or the space on the droplet surface, and it is not like APCI which suffers less matrix effect (Awad et al., 2015). Therefore, the results obtained from flow injection mode in ESI might not be precise enough to evaluate and predict the effects of source parameters. In order to further investigate the effects of ESI source parameters and get optimal settings for the further untargeted analysis, chromatography separation is required. This part was illustrated in manuscript 1, in which 34 compounds were used for optimization (Table 1 in manuscript 1) and selected to be as diverse as possible in terms of log$P$ (-0.07~5.69), retention times (3.6~21.0 min), elemental compositions (Cl, F, P, S), and molecular volumes (127.82 ~ 772.53). In positive mode, Cap and DryF influence Min25% the most than other factors with P value of 0.02 and 0.005, respectively. Cap and DryF also showed second-order interaction with a P value of 0.02, which is consistent with the flow injection mode in Table 3-4, indicating that Min25% is optimal when Cap is at a high value or DryF is at relative low values without concern of interaction between these two factors. By using “maximize desirability” function in the model, 40% increase was achieved compared with the predicted value of central points. In negative mode, 20% increase was achieved compared with the predicted value of central points. The log$P$ value of all the compounds have been detected in ESI is in the range from 0.48 to 6.76,

in contrast to APCI, the response of compounds of high log*P* value was very low and only detected only in a few settings.

In flow injection, even ESI suffered high matrix effect, there were still 86 compounds detected in ESI. 53 compounds of them could be detected both in ESI and APCI ion source, and 34 compounds can only be detected in ESI but not found in APCI, and only 7 compounds that can be detected in APCI (weak response) but not found in ESI. It is interesting to observe that the log*P* values of 85% compounds (six out of seven) that can only be detected in APCI are higher than 3.6, and the log*P* values of 57% compounds are higher than 5.1, the lowest value is 1.2. However, for the 34 compounds that can only be detected in ESI, the log*P* values of 50% of these compounds are lower than 3, the lowest log*P* is 0.62, the log*P* values of 91% of these compounds are lower than 5. According to the generic data base has been created in section 3.2.1, the log*P* values of 89% compounds (pesticides, maycotoxns, PAs) in the database is lower than 5. Based on the current results from the tested compounds with a wide range, no obvious advantage of detection was found in APCI by comparing to ESI. Due to this fact, ESI was therefore used for the further method development.

Table 3-7. Parameter estimates by T-test for the influence of the optimized factors on Min25% in ESI

Term	t Ratio	Prob>|t|
Intercept	1,81	0,1302
Cap(2500,4500)	3,75	0,0133*
Neb(15,45)	1,78	0,1349
Dry F(8,12)	-0,69	0,5210
DryT(150,220)	-0,41	0,6991
Cap*Cap	-3,46	0,0180*
Cap*Neb	3,58	0,0159*
Neb*Neb	1,77	0,1367
Cap*Dry F	13,90	<,0001*
Neb*Dry F	-0,42	0,6941
Dry F*Dry F	-2,70	0,0426*
Cap*DryT	-3,15	0,0252*
Neb*DryT	-0,10	0,9256
Dry F*DryT	1,04	0,3467
DryT*DryT	-4,04	0,0099*

t-ratio is the estimate divided by the standard error. t-ratios greater than 1.96 (in absolute value) suggest that the coefficient is statistically significantly different from 0 at the 95% confidence level. Prob > |T| labels the P values or the observed significance levels for the t statistics

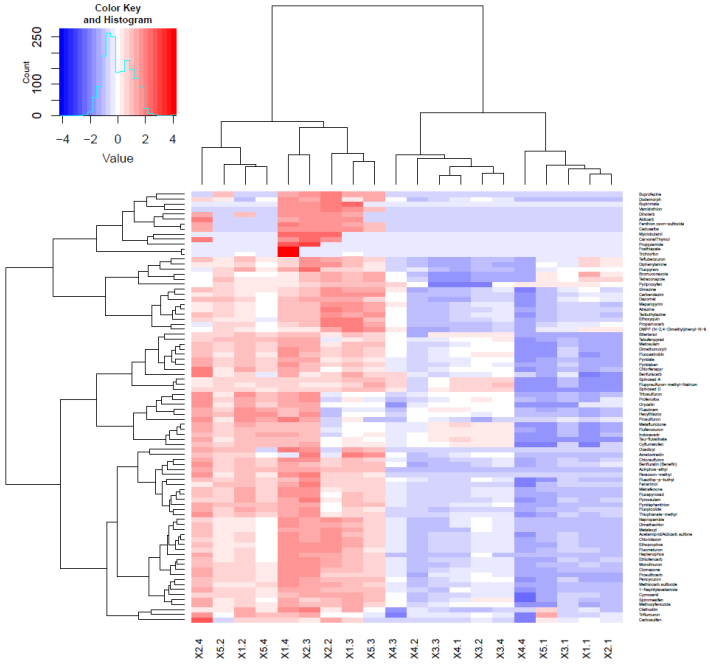


Figure 3-4. Heat map generated using the statistical software "R" from the scaled responses factor when introduced via direct flow injection into the ESI source (86 compounds detected). Red areas indicate a relative increase in the instrumental response and blue areas indicate a relative decrease in the instrumental response.

c. The influence of organic solvents and modifiers in mobile phase

First, we compared the performance of acetonitrile and methanol on the separation and sensitivity of teste compounds and kept the water phase composition constant. From the results in Figure 3-5, when keeping the water phase constant, the overall signals of compounds in A (MeCN -0.1%FA) are higher than C (MeOH-0.1%FA), which means our teste compounds in acetonitrile tends to result in a better sensitivity and higher signals than that of methanol. Therefore, acetonitrile was chose for further analysis and optimization in our study.

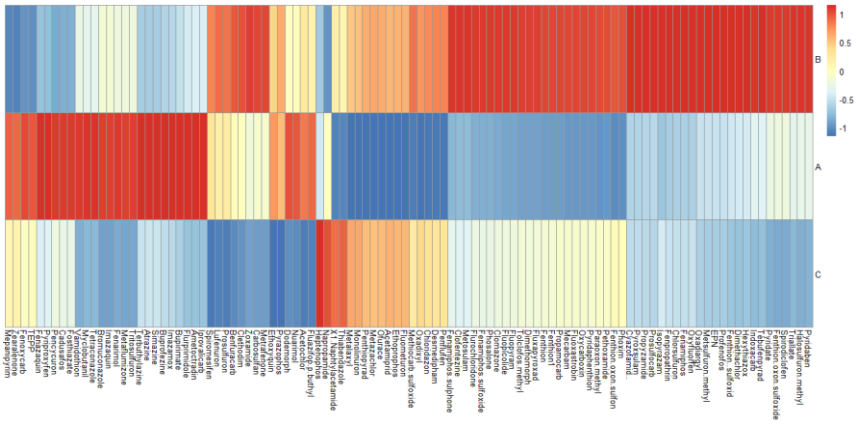


Fig 3-5. Heatmap of the influence of mobile phase on response factor of compounds (A: MeCN -0.1%FA, B: MeCN: MeCN -2.5mM NH4F, C: MeOH-0.1%FA). Red areas indicate a relative increase in the instrumental response and blue areas indicate a relative decrease in the instrumental response. The response were scaled and relative to the maximum

Changes in the retention times (RT) with a gradient elution using the 11 different mobile phase compositions have been determined. The RT of the majority of the compounds changed relatively in a small degree while using the different mobile phases. Most compounds were stable when the pH of the mobile phase changed. However, RT for about 28 of tested compounds were still affected by pH as shown in Figure 3-6. Compounds such as propamocarb, fluazinam and DMPF were even non-detectable at some pH particularly at higher pH more than 4.2. Therefore, it needs to be careful to select the modifiers for the mobile phases. Compounds in the mobile phase pH between 2.7 to 3.7 have relatively stable retention times.

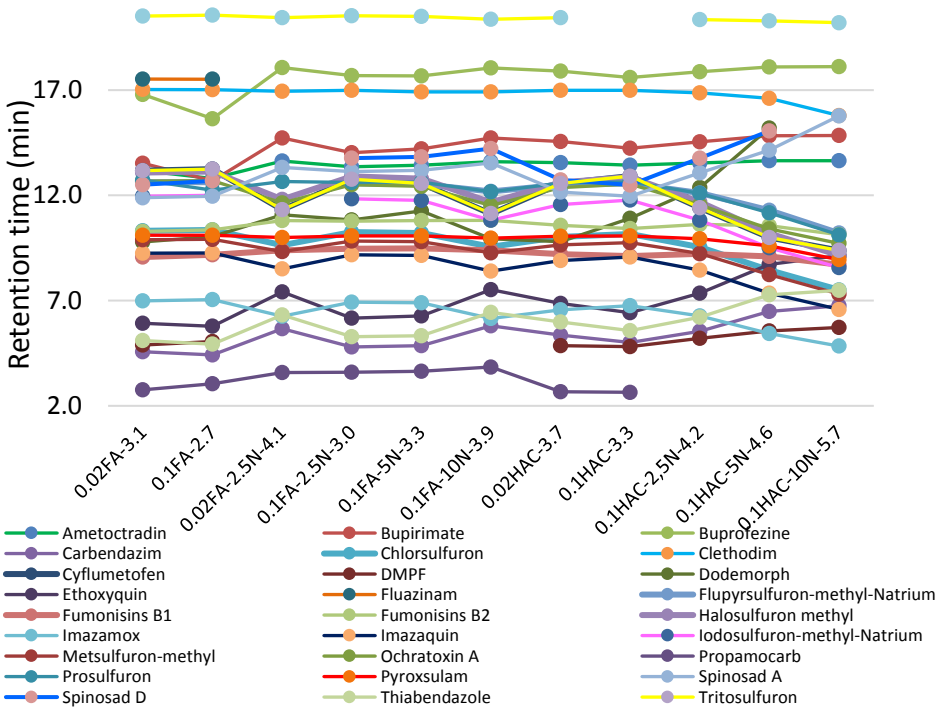


Figure 3-6. Changes in the RT when performing the gradient elution using the 11 different mobile phase compositions

A heat-map, illustrating the effects of the 11 different eluent compositions (A-K in Table 3-5) on the chemical response, is included in Figure 3-7. Red or orange areas indicate a relative increase for the response and green areas indicate a relative decrease for the response. Thus composition D (0.1FA-2.5 mM NH4OH), E (0.1FA-5 mM NH4OH), G (0.02% HAC) and H (0.1% HAC) with pH of 3.0, 3.3, 3.3 and 3.7, respectively, show the overall higher responses for most teste compounds than others. C (0.02FA-2.5 mM NH4OH, PH 4.1), J (0.01HAC-5 mM NH4OH, PH 4.6), and K (0.1HAC-10 mM NH4OH, PH 5.7) show the weakest response in general for most of the compounds, which means that most pesticides and mycotoxins ionize better at a lower PH. Some compounds out of tested compounds analyzed in 0.02% HAC and 0.1% HAC ionized very well and give higher signals compared with the compounds using formic acid in eluent, but the rest compounds also suppressed a lot. In conclusion, mobile phase composition D and E resulted in the highest responses and achieved a better compromise for most of the tested compounds, D was therefore selected for further studies on data processing method and semi-quantification method development.

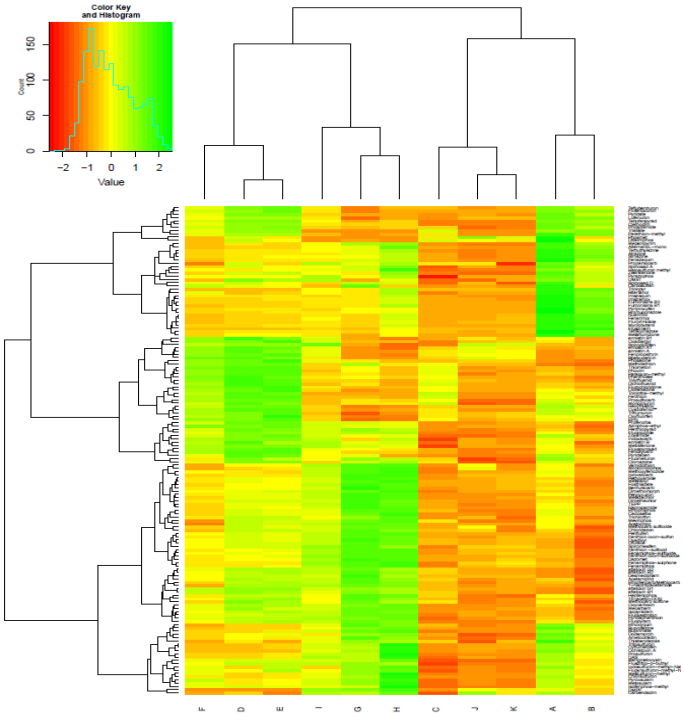


Figure 3-7. Effect of the 11 different eluents compositions. Green areas indicate a relative increase in the instrumental response and red areas indicate a relative decrease in the instrumental response. Yellow area is response in between.

3.1.2. Data acquisition and treatment method development

3.1.2.1 Data acquisition of untargeted analysis

As explained in section 2.3.3, the data acquisition of untargeted analysis can be classified into DDA and DIA. In our study on Bruker micrOTOFq, the acquisition mode belonging to DIA is bbCID, and the acquisition mode belonging to DDA is auto-MS/MS mode. To compare their MS/MS data quality, the rice sample spiked with the selected compounds as used in manuscript 1 (100 ng/mL for each) was run both in DDA and DIA mode on the same instrument (Bruker micrOTOFq). In auto-MS/MS mode, five most abundant precursor ions in cycle for fragmentation were selected and an MS/MS summation time adjusted to the precursor intensity, and the collision energy was set to 24-36 eV, the threshold of MS intensity as set to 100. In

bbCID mode, only the MS/MS mode was different from auto-MS/MS as bbCID was achieved by continuously cycling between ToF-MS and bbCID-MS/MS, and all the other settings were set to be consistent with auto-MS/MS, e.g., the same collision energy, same MS threshold. The range of the MS scan was from m/z 50 to 1100 for both acquisition mode.

In general, the MS/MS spectra of compounds analyzed in DDA have much lower signal than using DIA mode. As shown in Figure 3-8, it is an example of carbosulfan that fragmented under 24-36 eV in DIA (bbCID) and DDA (auto-MS/MS) acquisition mode. The intensity of parent ions are almost same for carbosulfan in DIA and DDA mode. However, the intensity of product ions (calculated m/z, 118.0690, 128.1439, 160.1160) of carbosulfan in DIA mode (experimental m/z, 118.0685, 128.1425, 160.1150) was about 15 folds higer than the intensity that obtained in DDA mode (experimental m/z, 118.0739, 128.1346, 160.1135). From the results, DIA mode achieved much higher sensitivity or higher accuracy than DDA, although DDA usually gets less noise from matrices than DIA mode in MS/MS spectra. In the case in our study for food untargeted analysis, the contaminants or residues are usually in a trace level which could be lower than the test concentration for carbosulfan (100 ng/ml). Then it means by using DDA mode (auto-MS/MS), some fragments from compounds might not be detected in MS/MS spectra or in a very low level with poor quality, even though DDA is more selective. Therefore, in this project, DIA (bbCID) with higher sensitivity was selected to be used in further study, which could be found in Manuscript 2, where bbCID was used to get the MS/MS spectra of PAs in honey. Nevertheless, DIA is still useful tool for the compounds interpretation from complex food sample in specific cases, the method just needs to be further developed here to increase the sensitivity.

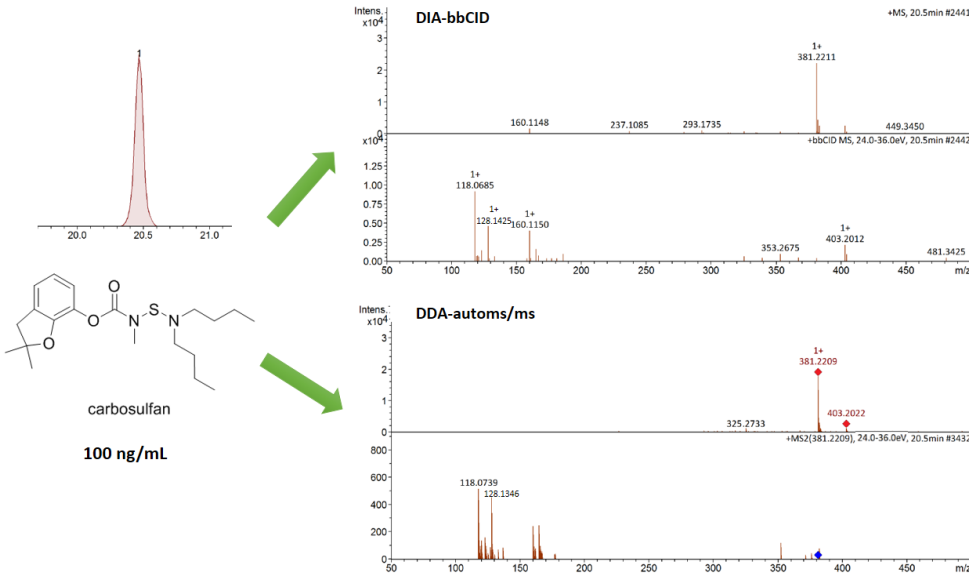


Figure 3-8. MS and MS/MS spectra of carbosulfan in DDA (auto-MS/MS) and DIA (bbCID) mode with the collision energy of 24-36eV

3.1.2.2 Data treatment method development

The quality of suspected screening highly depends on the database that has been used. Therefore there is usually no specific data processing requirement in targeted analysis approach or suspected screening and only raw data is needed. However, in untargeted screening, the data treatment of true untargeted analysis without a screening list or any prior information is the most arduous part. It is critical to extract the useful information from the massive HRMS data. In this project, the different data processing strategies were compared and developed. The eventual aim is to to get as much useful information as possible and as few irrelevant information as possible from the complex HRMS data, and also further develop into a more robust, accurate data processing procedure to provide high quality data. From a chemical point of view, it can also be rephrased as getting as much true positives as possible, and as fewer false negatives as possible. In most cases, the false negatives are the serious issue of highest concern and need more efforts because of the underestimated true risk.

Similar to searching biomarkers for diagnosis, treatment, and prevention in Metabolomics (Cunsolo, Muccilli, Saletti, & Foti, 2014), it is possible to identify and monitor specific variables by metabolomics-like approaches in food monitoring, to confirm and quantify the potential

chemical risks. In this study, the commonly used Metabolites analysis software and platforms in MS, such as Metabolites Detect and the XCMS based approaches were optimized and compared regarding to peak detection capacity and accuracy on case sample and control samples, which both of them also compared to the routine suspected screening. Case samples were spiked control samples with 29 model compounds (pesticides, mycotoxins, and drugs) to mimic the unknown contaminants or residues at 3 different concentration levels (from 5 µg/kg to 100 µg/kg). The detailed design can be found in Manuscript 1.

Metabolite Detect 2.0 from Bruker Daltonics is a common metabolite detection software for the identification and confirmation of expected and known metabolites as well as the annotation of unknowns. Many parameters in the software influence the detection results, such as smoothing width, S/N, maximum peaks number. These parameters were optimized to maximize the number of true positive ions. Although it is well known that smoothing width is an important factor in detection and chromatographic resolution, it highly depends on the shape and width of the signal. The digitization interval and also the different algorithms have been used in different software (O'Haver, 2019). It therefore needs to get an optimal smoothing value for a specific software before its application to an analysis. The details about the optimal setting of each data processing software are shown in Appendix I of manuscript 1.

XCMS is an R language command based approach, which can be difficult for users without statistics background. Although there is also an online XCMS version developed by Scripps Research (https://xcmsonline.scripps.edu), there are still limited options for customizing graphics. Therefore, this project tried to combine another XCMS based platform with R. This data processing procedure in Workflow4Metabolomics and R was set up to be as automated as possible. The general workflow was shown in the Figure 2 in Manuscript 1. The raw data file in .d format produced from Bruker micrOTOFq were first converted to mzXML format by ProteoWizard (Toolkit, 2012). Then the data files were uploaded to Workflow4Metabolomics, which is a virtual research environment built upon the Galaxy web-based platform (Giacomoni et al., 2014; Smith et al., 2006), while it is more user-friendly than the XCMS online.

As the number of output ions generated by Workflow4Metabolomics is high, data filtration step is important not only to minimize relevant chemical information loss but also minimize false

positive ions. A three-steps data cleaning strategy developed in this study, including an automated T-test pre-filtration step, a second variation filtration, and a third further cleaning. After data cleaning, a very clear distinction between the four groups without any prior information can be found in PCA (Figure 3-9), even for the spiked sample group with the lowest concentration (5 µg/kg for most). The explained variance of PC1 increases from 15.5% to 49.6% after the data cleaning. PLS-DA model was also carried out on the same data (before and after data cleaning) to further evaluate the efficiency of data cleaning. Typically, a Q2 score >0.4 and an R2 > 0.5 indicates a robust model, whereas scores between 0.7 and 1.0 indicates a highly robust model (K. L. Nielsen et al., 2016). In Figure 3-10, it can be found that after data cleaning, R2X increased from 0.46 to 0.95, and Q2Y increased from 0.38 to 0.58. A much better model performance was achieved by efficient data filtration. By using multivariate analysis, the variables which contribute most to the group distinction of PCA and PLS-DA could be the potential suspected contaminants or residues what we are looking for. The contribution of each variable can be found in PCs in PCA, and variable importance parameters (VIPs) in PLS-DA.

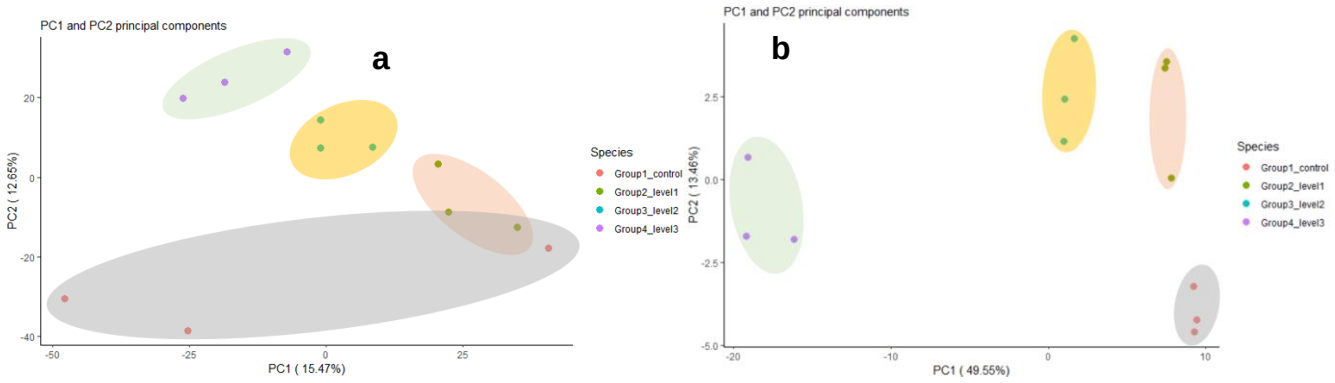


Figure 3-9. Principal component analysis plot of the features with an ANOVA p-value > 0.05 upon comparison between control (rice blank) and spiked rice samples with different concentrations. (Group 1_control: blank rice matrices; Group 2_level1: spliked with 5 µg/kg; *Group 3_level2: spliked with 5* µg/kg; *Group 4_level3: spliked with 100* µg/kg*). a: Before filtration. b: After filtration*

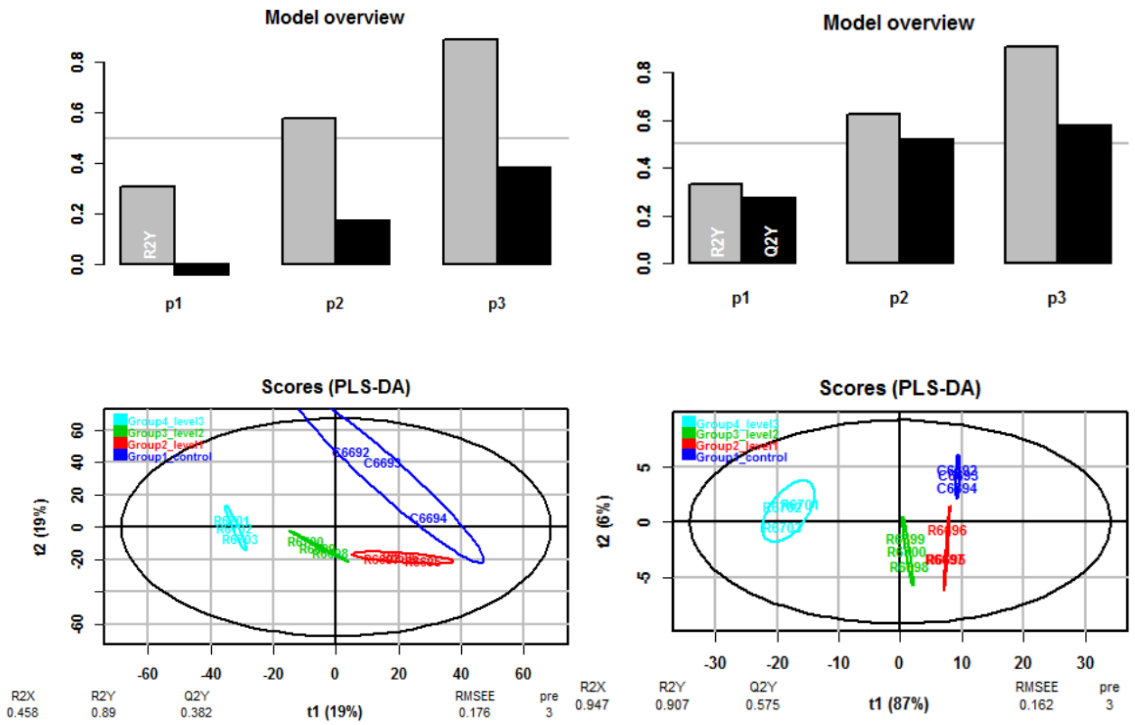


Figure 3-10. PLS-DA of control and spiked rice samples spiked with different levels (Group 1_control: blank rice matrices; Group 2_level1: spliked with 5 µg/kg; *Group 3_level2: spliked with 5* µg/kg; *Group 4_level3: spliked with 100* µg/kg*)*

Regarding the detection results, the number of true positives and false negative and also the mass accuracy were calculated. Comparison of different data processing methods showed that, suspected screening has the highest mass accuracy, highest detection rate (100% for all 3 levels), while the commonly used Metabolites Detect has the lowest mass accuracy (Table 2 in Manuscript 1), lowest detection rate (66% for level 1) and gives the largest number of false negative. By using the developed metabolomics-like approaches based on XCMS, the detection rate of true positive is more than 96% for all all test concentration levels, and the accuracy is very close to suspected screening, but no preselected list is needed. It means that, even in the worst case that most compounds are in 5 µg/kg, these contaminants or residues can still be successfully detected by the developed true untargeted method with prior information.

3.1.3. Discussion

Optimization of LC/MS conditions for small molecules is particularly important in untargeted food analysis in which the aim is to maximize the number of compounds detected in complex food matrices. Increased ionization efficiency will lead to increased response, lower LOD and LOQ, and increase the production of higher quality MS/MS spectra for improving compounds identification. A number of studies have reported methods with summaries of the advantages and weaknesses of APCI and ESI (Chen et al., 2014; Thurman, Ferrer, & Barceló, 2001). Nevertheless, still no well defined procedure exists to select the best ion source for food monitoring. Although the ionization source settings were frequently optimized during a method development to achieve a better sensitivity (Khatal et al., 2016; Kruve, 2016; Maragou, Thomaidis, & Koupparis, 2011), the influences and interaction between different ionization source parameters on ionization have not been sufficiently studied and the compounds used for these studies are very limited.

Our study further investigated the interaction between different ionization source parameters and the influence of these setting on the chemical response on a larger compounds set with a wide coverage. Although compounds suffer more matrix effects than APCI, ESI still has a better coverage for all compounds with diverse properties. It is more suitable for untargeted analysis of unwanted compounds in food samples which demands a wide scope. The parameters such as Cap, DryT, Dry F were found to influence the chemical response, and it is consistent with some previous studies somehow. For example, in the study of Kruve (2016), drying gas temperature, nebulizer gas were found to significantly influence the ionization efficiency. It has also been reported that the ionization efficiency of selected pesticides in ESI is influenced by capillary temperature and auxiliary gas flow rate in the study of Maragou, et al., (2011). This study also proposed a method to rapidly evaluate which ion source (APCI or ESI) to choose based on the theoretical log*P* values, which is highly correlated with the polarity of the compounds (Kiontke et al., 2016). It is also consisted with the speciality of ESI and APCI as explained in section 2.2.2 (ESI for polar compounds, APCI for less polar and non-polar compounds) (Khatal et al., 2016). From our results, ESI is very suitable for studying a wide range compounds with diverse properties. The choice of ion source could just simply based on

log*P* values rather than looking at the structure one by one to get a compromise method for most compounds in a untargeted analysis.

For the mobile phases, we selected acetonitrile as the organic phase after by comparing the sensitivity with a big set of compounds. There are also other consideration to select acetonitrile from the studies from the literatures. Acetonitrile and methanol are fully miscible with water and have their certain advantages and disadvantages (Pattison, 2019). Acetonitrile/water mixes have lower viscosity (e.g. 80% acetonitrile in water, η0 =0.53 cP) than methanol/water mixes (e.g. 80% methanol in water, η0 =1.05 cP). Acetonitrile/water generate lower back pressures on the LC column, puts less strain on the LC system, and therefore provides additional possibilities to increase the flow rate and further reduce run times (Thompson, Kaiser, & Jorgenson, 2006). Acetonitrile has also higher elution strength than methanol in reversed-phase chromatography, results in a shorter analyte retention for equal proportions of organic phase (Pattison, 2019). Methanol is a polar-protic solvent, whereas acetonitrile is a polar-aprotic solvent and possesses a stronger dipole moment. Sensitivity is not the only factor to consider in a LC/MS method development. Some factors such as, solvent cost, environmental impact, run time, and chromatographic resolution are also important factors to be considered. Methanol costs significantly less and is more environmental-friendly than acetonitrile, while larger volumes may be consumed due to its lower eluotropic strength and longer run times as analytes eluted much later in a water-methanol gradient (Huffma et al., 2012).

For the aqueous phase, it has been found that ionization efficiency depends on solvent properties somehow, mobile phase additives have a strong influence on the sensitivity and spectral quality when using LC-MS (Leitner, Emmert, Boerner, & Lindner, 2007; J. Liigand et al., 2014). Creydt & Fischer have studied the influence of 17 mobile phase additives (e.g., 0.02-0.1% formic acid, 0.02-0.1% acetic acid, ammonium formate and ammonium acetate from pH 3.5 to 6.5) on maximizing untargeted plant metabolite coverage and improving the electrospray ionization process. Metabolite coverage could be significantly increased by using 0.1% acetic acid in positive mode and 0.1 % formic acid in negative mode (Creydt & Fischer, 2017). Yang et al., (2013) demonstrated that mobile phase additives with pH-adjusting may significantly influence ionization efficiency and showed that ammonia increased the signal of protonated

analytes and suppresses the adducts formation out of the studied additives (formic acid, ammonia, and combinations). It has also been found that in the study of Liigand et al., (2017) different pH has stronger influence on ionization efficiency than different organic solvent proportion in a mobile phase (J. Liigand et al., 2017). In brief, the solvent composition plays a very important role in ionization efficiency and is a powerful method to improve the sensitivity of ESI-MS method. It is therefore worthwhile to investigate the influence of different mobile phase additives on the compounds that we are interested before further study being conducted. In addition, the influence of eluent composition on adducts formation will be also interesting to study with. For example some compounds only form $[M+NH4]^+$ or $[M+Na]^+$ adducts, and the adducts formation might influence the MS/MS spectra quality as reported (Pieke, Smedsgaard, & Granby, 2018), while this point was not investigated in this study.

The data processing development has been explained in detailed in Manuscript 1. Our study provided a rapid metabolomics-like approach to assess the food quality and find out suspected contaminants or residues. Compared with the common suspected screening method (Hird et al., 2014; Veenaas & Haglund, 2017), it is a big advantage that no prior information is needed, meanwhile almost reaches equally high detection rate (96%) as that suspected screening was achieved. Some unknown chemical risks might be discovered even these compounds are not included in any database or screening lists. Less steps were required for data cleaning part and higher detection rate was achieved, this metabolomics-like approach achieved the discrimination of contaminated food samples with control samples on an even lower level (5 µg/kg) than the lowest reported value (10 µg/kg) (Delaporte, Cladière, Bouveresse, et al., 2019) (Delaporte et al., 2019). This level is lower than the MLs for most pesticides (e.g.10 µg/kg for most pesticides in regulation) and mycotoxins (e.g. 2-4000 µg/kg, the MLs of aflatoxin B1 in cereals is 2 µg/kg, all MLs of other regulated mycotoxins are higher than 5 µg/kg) regulated for unprocessed cereals in EU regulation (Cheli, Battaglia, Gallo, & Dell'Orto, 2014). However, it still needs to be concerned that a sufficient blank matrices should be carefully selected, and the relevant parameters in processing need to be assessed and validated carefully as well based on the specific method before analysis to avoid important information being missed.

3.2. Exploration of untargeted compounds

3.2.1. Identification of untargeted compounds

The main bottleneck in the interpretation of HRMS based data is the identification of unknown features due to the significant structural diversity of potential compound (Edmands et al., 2017). For untargeted compounds it is often laborious and sometimes requires reference standards or at least a recorded spectrum. In the 5-levels identification rule proposed by Schymanski (2014), the importance of database/library has been demonstrated. Both online database library and in-house database are important for untargeted analysis. The more comprehensive the databases is, the more successful dereplication can be obtained.

In my study, first, a generic database were constructed, to make it as generic as possible, chemical residues and contaminants (pesticides, mycotoxins, dioxins, PCB, PAH, acrylamide, PAs) according to various EU regulations which mentioned in section 2.1, standards available from our institute collection, identified compounds relevant to mycotoxins and their metabolites in literature were all included in the database.

For pesticides, there are in total 1366 active substances pesticides regulated in the EU database (492 of them have been approved while 828 of them have not been approved, with 26 are pending and 20 are banned). In order to select the most relevant ones, the substances including dithiocarbamates, biological agents (bacteria, fungi, viruses, etc.), inorganic compounds, which can not be analyzed by GC-MS or LC-MS were excluded from the list. It results in 693 pesticides left in the end. For mycotoxins, there are 13 different mycotoxins are regulated in EU regulation. 'Dioxins' encompass a group of 75 polychlorinated dibeno-*p*-dioxins and 135 polychlorinated dibenzofuran congeners. PCBs are a group of 209 different congeners. For the accessible standards in department, there are 836 pesticides and 23 mycotoxins included. For PAs, there are 93 PAs from literature were registered into the database, and 17 of them with standards available in our lab. There are also 6011 identified compounds relevant to mycotoxins and their metabolites from literature were registered into database. The names, chemical formulas, m/z that used for screening are included. In total, the whole database contains around 7000 compounds. Figure 3-11 has an overview of the m/z distribution of the

compounds from the database has been constructed. It can be found that the m/z with the highest distribution is in the range from m/z 200 to 500, which covers about 70% of all the compounds. There are only 0.7% compounds are below m/z 100 and 1.1% compounds are higher than m/z 1100. Therefore, when the untargeted screening was carried out on HRMS, the mass range was set to m/z 100 to 1100, which covers more than 98% compounds in the database.

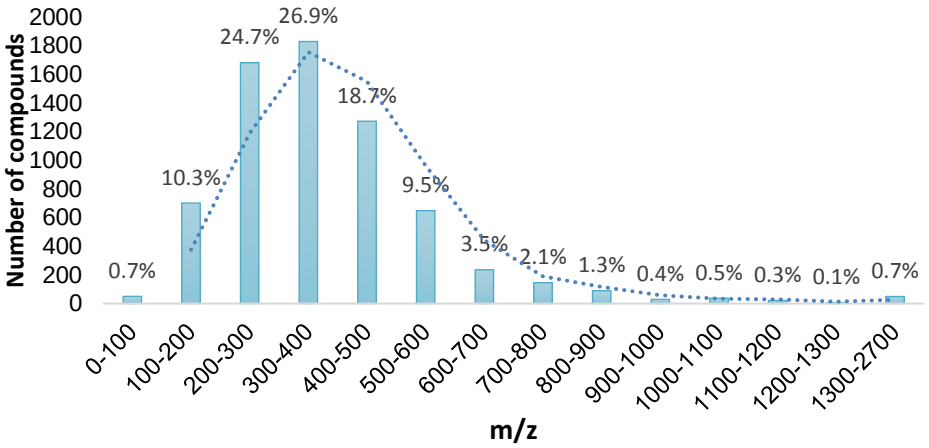


Figure 3-11. The molecular weight distribution of the compounds in the constructed database.

In addition to the structures, predicted log*P*/log*D* values were also registered into the database to minimize the number of false candidates, log*P*/log*D* were calculated with JChem for Office (https://chemaxon.com/products/jchem-for-office) by using SMILES code achieved from PubChem. As shown in Figure 3-12, the plot is based on a set of 152 compounds. A high correlation (R^2= 0.84) was observed between measured and predicted retention time based on log*P* values. A few compounds (4 or 5) with wide retention time spread among these compounds can be used for retention time calibration during each run. By using log*D* prediction, some irrelevant compounds candidates with huge log*P* bias from database match can be excluded, which is promising and easy way to improve the identification quality. Figure 3-13 is an example of the overview of the database has been constructed. For all the registered compounds in the database with calculated log*P* values.

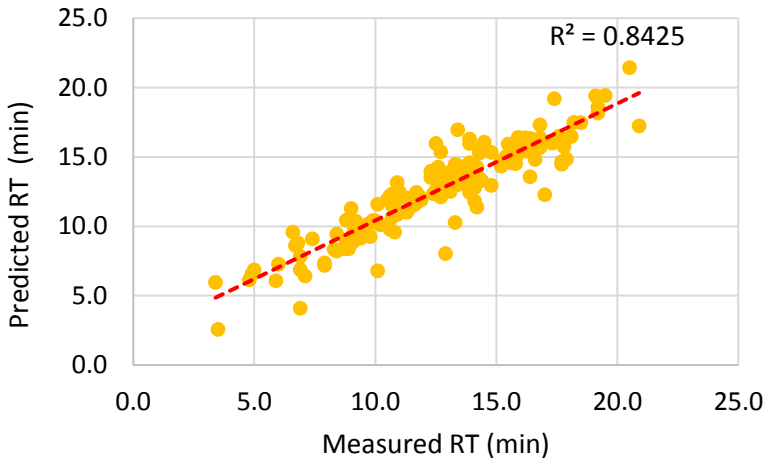


Figure 3-12. Predicted retention time vs measured retention time by logD calculation

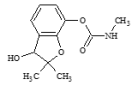
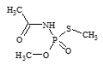
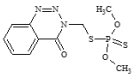
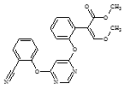
Structure	Smiles	*ID	Custom	LogP	LogD (pH = 3,00)	LogD (pH = 3,50)	Molecular Weight
	Oc1c(cccc1)c1ccccc1	1	2-Phenylphenol	3.08	3.08	3.08	170.21
	CNC(=O)Oc1cccc2c1OC(C)(C)C2O	2	3-Hydroxycarbofuran	0.7	0.7	0.7	237.25
	COP(=O)(NC(C)=O)SC	3	Acephate	-0.53	-0.53	-0.53	183.17
	COP(=S)(OC)SCN2N=Nc1ccccc1C2=O	4	Azinphos-methyl	3.01	3.01	3.01	317.33
	CO\C=C(/C(=O)OC)c1ccccc1Oc2cc(Oc1ccccc1C#N)ncn2	5	Azoxystrobin	3.54	3.54	3.54	403.39
	Cc2c(COC(=O)[C@@]1([H])[C@]([H])(\C=C(/[Cl])C(F)(F)F)C1(C)C)cccc2c1ccccc1	6	Bifenthrin	6.93	6.93	6.93	422.87

Figure 3-13. Example of in-house database has been constructed including the compound name, chemical structure, SMILES, logP, logD, molecular weight.

We applied the suspected screening in the food contaminants/residues identification in rice samples in Manuscript 1, and the detection of PAs compounds in various honey samples in Manuscript 2. The workflow of suspected screening based on in-house database is shown as Figure 3-14. It provides an untargeted approach to chemical identification without using

authentic standards. The broadband CID (bbCID) data acquisition from Bruker TargetScreener was applied for the fragmentation of PAs in honey samples. This acquisition method ensures that all ionisable compounds eluting have molecular ion data and fragmentation spectra acquired all of the time, thus no compounds would be missed.

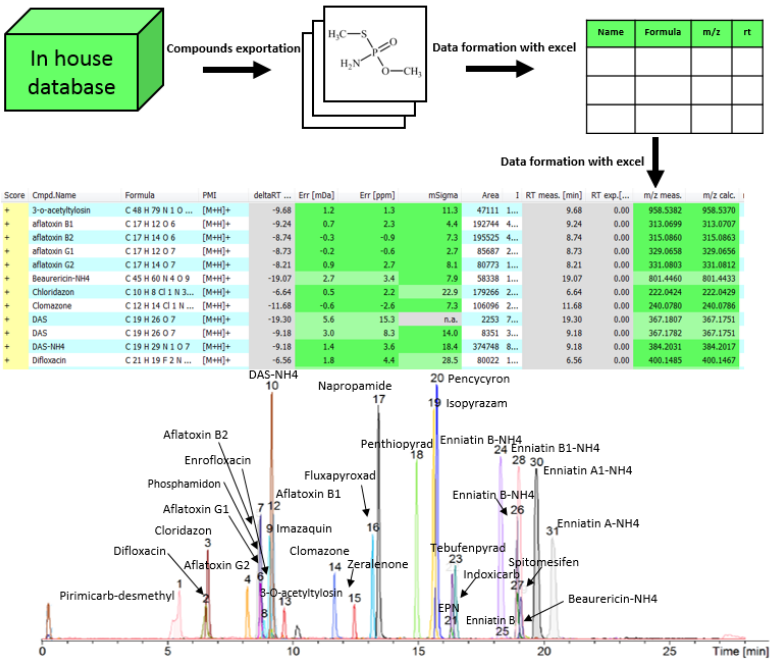


Figure 3-14. Example of workflow for screening of contaminants and residues in rice with suspected screening. Analysis via screening yields both a graphical interpretation of the results and a table of the data

To achieve an identification level higher than level 3, MS/MS fragments pattern is usually necessary to do the further confirmation, even MS/MS spectra do not directly provide the structure information. Particularly in untargeted screening, the features that need to be identified can be numerous, thus there is a need for an efficient and automatic identification method that enable identification of unknowns without using chemical standards. However, there are no commercially available standards or no experimental spectra exist in any database for a large number of compound. Herein, *in silico* MS/MS spectra approach (Wolf et al., 2010) exists from a few different sources either from instrument vendors or from some available online database connected with different platforms (Giacomoni et al., 2014; Smith et al., 2006). For

example, MetFrag is an open-source software for the annotation of high precision tandem mass spectra of metabolites using the bond disconnection approach. Fragmentation data with the parent and product ions from Bruker raw data file can be sent directly to this platform. Figure 3-15 is an example of *in silico* MS/MS identification one compound from rice. Such type of software can be able to automatically correlate the fragments to *in silico* bond breaking data retrieved from e.g., KEGG, PubChem or ChemSpider databases (Wolf et al., 2010).

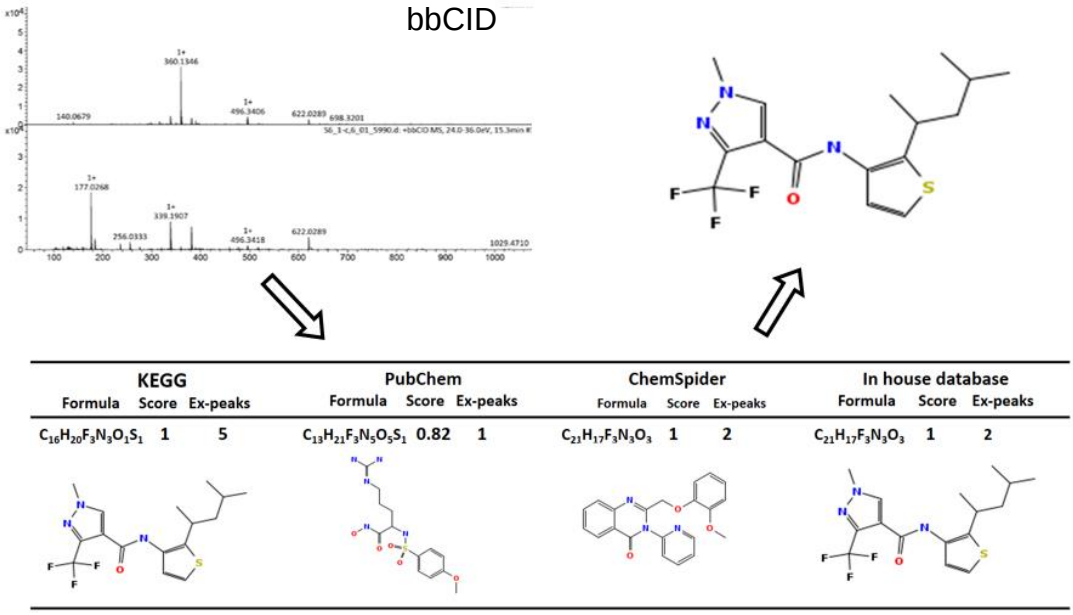


Figure 3-15. Example of the in silico MS/MS identification of penthiopyrad from rice sample by database match with bbCID MS/MS acquisition mode. KEGG, PubChem, ChemSpider, and self-created in-house database were used. The match obtained from KEGG explained the most peaks, which is consistent with the result obtained from in-house database. Ex-peaks means the number of explained peaks from the database match.

3.2.2. Bridge between targeted and untargeted analysis

To build a bridge between targeted and untargeted analysis, it is important to take the advantages of each approach and make them compensate for each other. Targeted here means the data acquired in a full scan mode by HRMS and used for targeted analysis or suspected screening. Therefore, there is a possibility to carry out untargeted analysis as well

on the same dataset after the targeted analysis. This allows targeted and untargeted run being carried out in parallel. As standards are available in targeted analysis, there is almost no limit to do both identification and quantification at the same time, while it is a challenge to do both in untargeted screening due to the lack of standards. Herein, a comprehensive screening database, and a practical and reliable quantification model without using standards are able to address the challenge to the greatest extent. In manuscript 2, a validated targeted PAs analytical method in honey was demonstrated. The validated method was based on 12 PAs with authentic standards that applied for the recovery, repeatability, matrix effects, LOD, and LOQ determination. Then, a stepwise-MLR prediction model was built based on the validated targeted analysis method to figure out the correlation between the ionization efficiency and physicochemical properties of these 12 PAs.

For the same sample profiles, untargeted screening was carried out to find out the other potential PAs which we did not have available standards by using self-created in-house database. Quantitative prediction model (QPM) was used to quantify the potential PAs without costly standards. By carrying out both targeted quantitative and untargeted semi-quantitative study, a bridge was successfully built between a targeted and untargeted analysis in this study. We used this combined method to analyze 32 honey samples collected from Danish supermarkets successfully. In total, 14 different PAs were found in these tested honey, six of these PAs were included in our validated method and can be quantified with authentic standards, while eight of them without having standards were not included in our validated method and were semi-quantified with QPM. The results from the honey samples had been screened are shown in Figure 3-16.

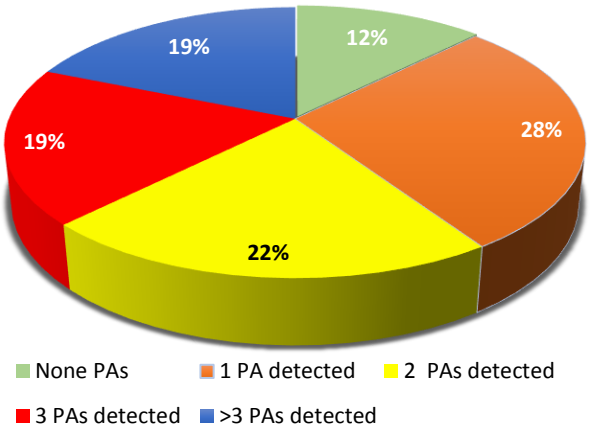


Figure 3-16. PAs in 34 collected honey samples from Danish supermarkets. The number of PAs includes both tentative identified and identified PAs with authentic standards. Fourteen different PAs were found in all tested honey, six of these detected PAs were included in the validated method and can be quantified with authentic standards, while eight of them having no standards were semi-quantified with QPM

3.2.3. Semi-quantification model development and application

In some cases, one spectra end up with a few structure candidates, how to move forward from these identification results to quantification without using standards is another big challenge that we are facing (Figure 3-17). As discussed in the previous section 2.3.5, the challenge of quantification in LC−HRMS comes from the vastly different responsiveness of the compounds. Even at the same concentration, two compounds may produce very different signals and in different scales. Even in some cases with compounds of similar structures, they may still have very different response. What factors need to be considered in semi-quantification method development and how we deal with them in our studies are demonstrated in the following sections. The semi-quantification model development and application have been carried out on the case of honey and cereal samples.

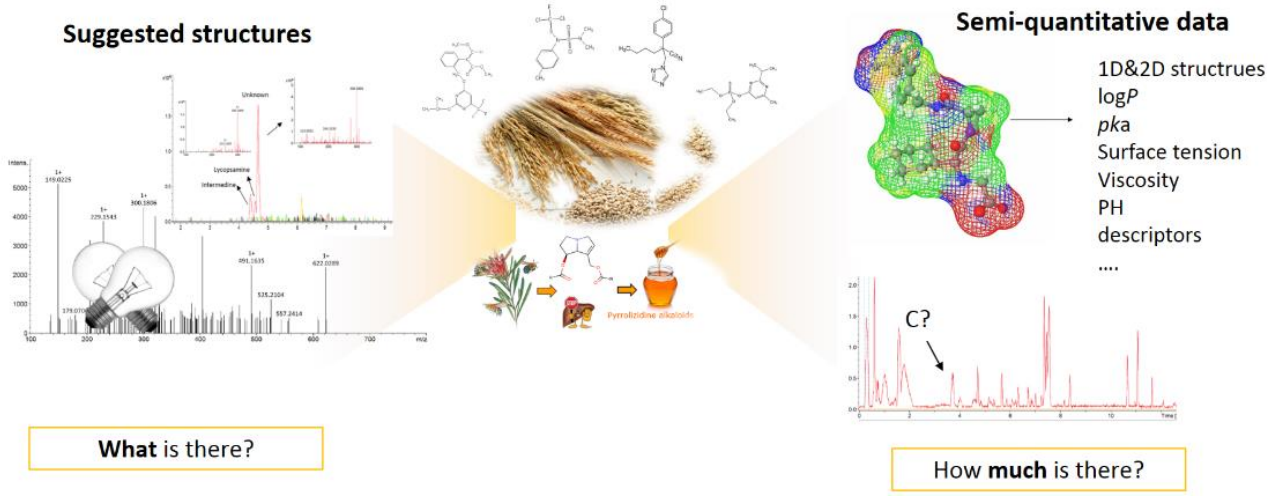


Figure 3-17. Semi-quantification of non-targeted compounds. After answering the question about what is in the sample with few suggested structure candidates, how to move forward to the question "how much is there" with only structure information and the peak area obtained from the chromatograms.

The first factor that will influence the ionization efficiency is physicochemical properties of the compounds themselves. Semi-quantification models based on the correlation between physicochemical properties and ionization efficiency of compounds have already been successfully proved in some studies (P. Liigand et al., 2018; Oss et al., 2010), although the application scopes have been confined to the specific type of compounds in those studies. Most approaches have hardly been used for real quantification with chromatography separation in matrices. In the study of Liigand et al.,(2018) ionization efficiencies of compounds was predicted in case of the analysis of biological matrices with a flow injection analysis. Most developed prediction models in literatures are challenging to calculate the relevant descriptors used in the model (C. J. Cramer et al., 2017; Kruve et al., 2014).

To our knowledge, there are no quantitative prediction model developed specific for PAs in honey samples up to now. Although all PAs are based on the similar structural skeleton, it does not mean their response is close by that. By using the approach simply reported every alkaloid as equivalent of one alkaloid, the Max25% values (the sum of 25% biggest absolute prediction error) were in the range of 232.2%–603.1%, 202.5%–502.2%, 170.3%–621.2% with the concentration of 5, 10, 25 ng/mL, respectively (T. Wang et al., 2019). In manuscript 2, we

proposed a semi-quantification strategy without available standards through a MLR quantitative prediction model as the first time for PAs in honey samples. When we used our QPM to quantify PAs, the Max25% values were 196.5%, 142.2%, 162.3% respectively, it is more accurate and controllable than simply reported every alkaloid as equivalent of one alkaloid. To evaluate this prediction model, its feasibility and reliability have been validated by leave-one-out (LOO) cross validation and four additional PAs that were not included in the original dataset, the maximum concentration prediction error was 50.8%. Finally, we estimated the contents of PAs in honey samples combining the available reference standards and the QPM. The log*P*, p*K*a, p*K*b, and exact mass were found to be significant for the prediction and was included in the model.

In manuscript 3, another more generic prediction model proposed by Liigand et al.,(2020) was first time applied to the concentration prediction of pesticides and mycotoxins in various cereal matrices. This approach used random forest regression to predict the ionization efficiency of the compounds in ESI/HRMS, in which 450 descriptors in ESI positive mode and 145 descriptors in ESI negative mode were shown to be significant. These descriptors were all calculated with open source platform called PaDEL (Yap, 2011). From the first study on a set of 35 compounds (30 pesticides and 5 mycotoxins) (manuscript 3), the average quantification error was 5.9 times on QqQ in all six cereal matrices. When the tested compounds extended to larger compounds set onto 134 pesticides, the average quantification error was 3.8 times in all six cereal matrices on QqQ and 2.4 times on QToF (manuscript 4).

The second factor is the solvent effects which is often neglected in relevant studies. It includes the solvent type of mobile phase, the proportion of organic phase, also the pH and the content of aqueous phase (Kruve, 2016; Leito et al., 2008). For example, the eluent including more organic phase tended to have higher ionization efficiencies (J. Liigand et al., 2014). When using gradient elution by LC, as the decrease of the proportion of organic phase, the ionization efficiency of compounds usually increase. Therefore, the solvent factor needs to be taken into account as well in some cases when developing the semi-quantification prediction model. In this project, when prediction model applied to PAs in honey matrices, as the proportion of organic phase of all eluted PAs was closed (5% acetonitrile to 38% acetonitrile), changes in eluents were not taken into account here, and a high prediction accuracy was still achieved (1.3

times mismatch by cross validation). When the prediction model was applied to the semi-quantification of pesticides and mycotoxins in cereals, the composition of eluents could not be ignored anymore since the compounds were of high diversity that distributed from 2% organic phase to 100%. Five empirical eluent descriptors (viscosity, surface tension, polarity index, pH, NH_4 content) calculated based on retention time and mobile phase composition were taken into consideration in the prediction model to correct the solvent effects, which was explained in both manuscript 3 and Manuscript 4.

The third and final important key factor to be considered was the instrument effects. The signals obtained from one instrument are not directly comparable to the signals measured on the other instrument as the intensity scales strongly depend on the instrumentation. Different instruments have different ionization source geometries and parameters. The same compounds may yield absolute different scales of response on another instrument. For instance, QqQ, QToF and QIT are commonly used MS analyzer, while the profile of one sample obtained from these MS analyzers can be very different, it depends on e.g. the sensitivity, selectivity, dynamic range of these MS analyzer. As a result, the absolute ionization efficiency values are not always constant from one instrument to another. Even for the exactly same instrument during different days, the response of same compound can also vary. For the same instrument, quality control or relative ionization efficiency can be used to correct daily variations, which was applied in manuscript 2. When building the model, the relative ionization efficiencies of all PAs relative to one anchor compound (Seneciphylline) was used. In the study of Liigand et al., (2015), a good correlation of relative ionization efficiency values between different instruments was observed, which means it is very promising to make the ionization efficiency prediction comparable between instruments. In manuscript 4, 134 pesticides in 6 cereal matrices were semi-quantified on both QqQ and QToF using the same prediction model that described in manuscript 3. The prediction model was first time validated for pesticides in cereal on different instrument in different labs. A set of 30 transformation compounds were used to correct the instrument differences and to get the results comparable. We have shown that results obtained from suspect screening can be made quantitatively comparable between different labs using different instruments as well as different analytical methods. The consistency between the results obtained independently in

two labs was very high (R^2 = 0.85) in spite of the potential enormous variations involved in the instrumentation and methods.

3.2.4. Discussion

In some cases in untargeted analysis, some compounds can be identified at level 1 or 2 or even higher and end up with one or few structure candidates. “what compounds are in the samples?”, and “how high is the concentration of the compounds?” are questions need to be addressed. These questions can be answered only if the whole data processing procedure becomes more robust, automated, compatible with the semi-quantitative analysis, identification quality is improved and after a robust semi-quantification prediction model has been developed. Although the semi-quantitative untargeted analysis (Figure 3-18) is still a bit idealized, as there are many features from one sample in untargeted analysis that can only be identified at confidence level 5, and the MS/MS interpretation is usually tedious, and the semi-quantification methods also takes much efforts to develop and validate. To address these questions, our study provided an efficient and automated untargeted data processing approach (manuscript 1), potential semi-quantification model (manuscript 2), also the relevant prediction model validation and application (manuscript 3 and 4), and it took one step further on the way to semi-quantitative untargeted analysis.

MS/MS interpretation is often restricted by limited reference MS/MS data in spectral databases. Hereby *in silico* MS/MS spectra approach for predicting fragment and their abundance from chemical structure served as an good alternative for the structure confirmation, which was illustrated in section 3.3.1 (Smith et al., 2006; Wolf et al., 2010). The advantage of semi-quantitative untargeted analysis is that firstly it is always possible to trace back the raw data file and quantify the newly interested compounds without using reference standards. The interconnectivity between semi-quantification and untargeted analysis is based on the physiochemical properties of tentatively identified structures and the peak area of the identified/tentatively identified compounds, as shown in Figure 3-18. A response model based on the correlation between chemical structures and response (peak area) is capable of connecting untargeted analysis and semi-quantification. The model was proposed by our

studies and applied on the quantification of PAs in honey samples, and the model proposed by Liigand et al., (2020) was also successfully applied to the concentration prediction of pesticides and mycotoxins in various cereal matrices between instruments. It is promising to apply these models for the untargeted food monitoring to deal with the problems of lacking reference standards and efficient data processing method.

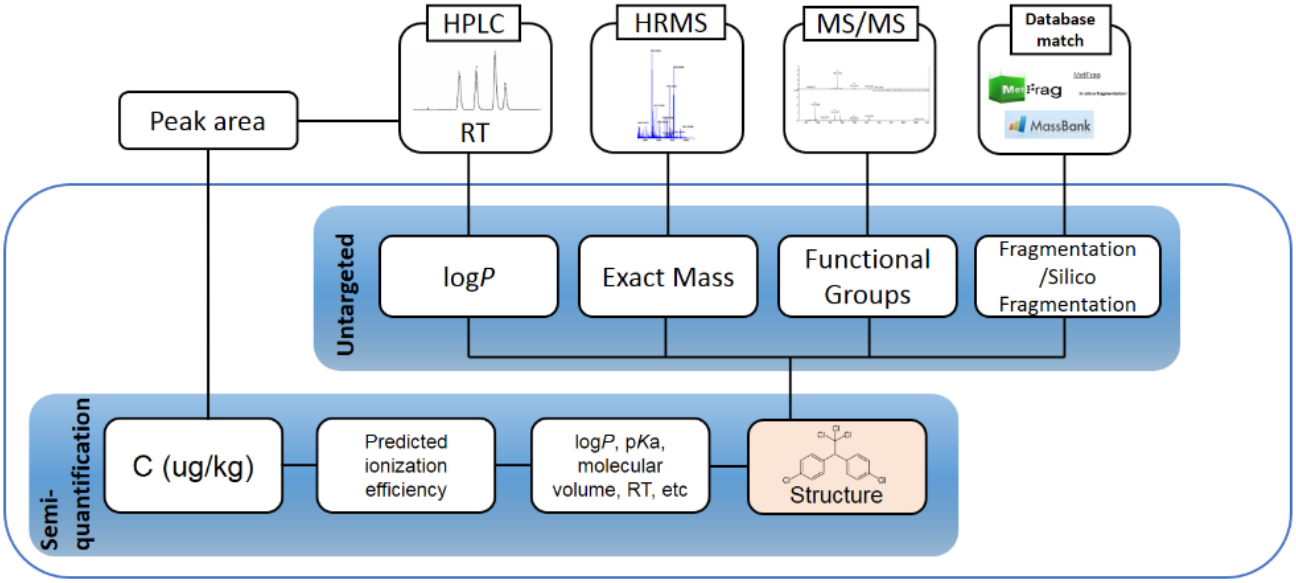


Figure 3-18. The combination of the workflow of untargeted analysis and semi-quantitative analysis. modified from reference (Kruve, 2019)*. Untargeted analysis is carried out based on LC-HRMS and MS/MS, the untargeted analysis ends up with one or few structure candidates and the physiochemical properties of the proposed structures can be calculated.*

4. Conclusion

The perspectives of this PhD project is to deal with the massive and complicate data come out from HRMS analyses and to develop a generic, wide scope, and efficient analytical strategy for untargeted analysis, in particularly to get quantitative data from these when standards are not available. In the end, providing more data for food monitoring and risk assessment with less efforts. Mainly four parts were carried out in this project, namely, analytical method optimization, data processing method optimization, identification and semi-quantification method development.

In analytical method optimization part, to achieve better sensitivity, the influences and interaction between different ionization source parameters, and the influences of eluent compositions on response factor have been studied. Ion source optimization was first operated in flow injection mode (APCI and ESI). No obvious advantage in APCI was found compared to ESI for the test compounds, and ESI has significantly wider coverage than APCI. However, more than half tested compounds that only detectable in APCI have a high log*P* value (log*P* > 5). On the contrary. most compounds can only be detectable in ESI have a lower log*P* value (log*P* < 5), which means that the choice of ion source could be possible just simply based on theoretical log*P* values rather than looking at the structure one by one to get a compromise method for most compounds in untargeted analysis. ESI was selected for the further study with a LC separation. The effects of key ESI ion source parameters on an example of 34 common diverse food contaminates and residues have been studied. A 40% increase in positive and 20% increase in the negative mode of response factor compared with central points was achieved by ion source optimization. In eluent composition optimization, better separation by acetonitrile, pH 3.0 and pH 3.3 with ammonium formate buffer in eluents, has achieved the best performance for most tested compounds.

In data processing method optimization and development, the commonly used metabolomics analytical methods such as Metabolites Detect and the XCMS based approach have been optimized and compared regarding to peak detection capacity and accuracy based on case and control samples. Both of them have also been compared with the suspected screening.

Case samples were spiked control samples with 29 spiked model compounds to mimic the unknown contamination at different concentration levels. Comparison of different data processing methods showed that, suspected screening has the highest mass accuracy, highest detection rate (100% for all 3 levels), while Metabolite Detect widely used in metabolomics has the lowest mass accuracy, lowest detection rate (66% for 5 µg/kg) and shows the largest number of false negatives. Developed and highly automated workflow based on XCMS package has been setup and achieved highest mass accuracy, highest detection rate (96%) and made a clear distinction between the control and spiked groups by PCA and PLS/DA in a untargeted way, even for the spiked samples with concentration of 5 µg/kg. This study can finally provide recommendations for Quan/Qual untargeted analysis by LC/HRMS. Such approach as XCMS based method could be a good complement to the routine targeted analysis in a view of rapidly detecting potential contaminated food products without the prior information and without sacrificing too much accuracy or increasing the number of false negatives.

In identification parts, a generic database were constructed including chemical residues and contaminants according to the EU regulations, standards available from our collection, identified or tentatively identified mycotoxins from literature. In addition, the predicted log*P*/log*D* values have been registered into the database to minimize the number of false candidates, which showed promising to improve the identification quality. The tentative identification of compounds was also demonstrated by correlating fragmentation spectra from QToF-MS to *in silico* generated spectra from database to figure out the compounds with the best match.

Semi-quantification method development and application have formed the largest proportion of this PhD project. We used the semi-quantification model for the quantification of both PAs in honey, and also pesticides and mycotoxins in cereals without using standards. In PAs study, we proposed a method that can be extended to a general strategy to achieve an estimation of quantitative data where standards are unavailable. This approach has been demonstrated by the simultaneous quantification of 12 PAs and also the further analysis of honey containing other high content PAs without standards. From validation. In honey samples, the QPM by using MLR was successfully quantified 12 PAs without standards for the first time and the maximum concentration prediction error on QToF was maximally 50.8%. The QPM using

random forest proposed by Liigand et al., (2020) was successfully applied and validated in cereal matrices for the first time and quantified 134 pesticides and 5 mycotoxins without standards. We even validated the model on different lab and instrument, a high correlation between labs was also achieved (R^2=0.86). On QqQ, 81% compounds achieved less than 5 times prediction errors, and in an average of 3.8 times for all compounds in all cereal matrices. On QToF, 94% compounds obtained less than 5 times prediction error, in an average of 2.4 times. The quantitation of pesticides in different cereals matrices without standard substances is feasible with the accuracy of less than 4 times on average for chromatographic analyses. which is quite promising to be used in food monitoring.

The untargeted strategy proposed in this study shows the semi-quantitative untargeted screening analysis is valuable tools for helping ensure food safe in the future.

5. Future perspective

Trust of food is of great importance to our society, and in order to ensure the food safety for now and for the future, the analytical strategy should be developed in parallel with the growing health concerns and increased complexities in chemicals occurring in daily food. A generic untargeted analytical strategy should include generic sample preparation method, generic data processing method, and also generic analytical method for both identification and quantification. Each part of them is critical for a successful untargeted analysis. This project has focused on the data processing and analytical method, which is probably the biggest problem in untargeted analysis, even though the focus can often seen on the sample extraction part. However, improving the extraction discovery and coverage is still important and it is the first step to a generic untargeted strategy of great importance that needs to be further studied.

Although untargeted analysis methods are becoming more and more popular, the algorithms used in untargeted compounds identification are also being developed fast in recent years, while the quantitative data is still very limited. The challenge can only be addressed when the data treatment process becomes more automated, robust, and compatible with semi-quantitation feature. The outcome of this thesis provides an efficient workflow to detect and quantify the unwanted compounds in the cereal and honey samples in an untargeted manner. In food monitoring research, it could be a good complement to the routine targeted analysis in a view of rapidly screening and quantification of potential contaminated food products without sacrificing too much accuracy. The next step is to increase the proportion of quantitative data in untargeted analysis, while there is still a lack of guidelines and legislation governing both for the development and validation of untargeted methodologies. Then it means that a semi-quantification model, which is robust, reproducible and accurate enough to involve more and more compounds, will be needed and we has moved forward to this final goal one step further. It might be implemented someday in the future, the untargeted qualitative and semi-quantitative data can also be included in guidelines and legislation with the development of validated methods,

On the one hand, our workflow allows the combination between untargeted analysis and semi-quantification, and also allows the combination between targeted and untargeted analysis. The models that proposed or applied in this thesis also have the potential to be further extended to the semi-quantification of other food products and other food contaminants or residues. It should not only be limited to cereals and honey, and not only limited to pesticides and mycotoxins. On the other hand, lacking of reference standards is not only the problem in food monitoring, but also the problem of many other research fields of metabolomics, environmental research, and so on. The application of semi-quantification model would be valuable for researches as well.

The quality of identification and semi-quantification method can always be further developed until it is more and more closed to “truth”. Specific and in depth studies on this research field is still highly needed, which requires the development of instrument, software, machine leaning method, as well as analytical methods. While it has been a good start in this study to demonstrate the potential and possibility to go deeply into this topic and widely apply them in the future to some content. In the end, taking all this information together would allow calculation of contaminant dynamics, toxicity and human exposure, then the chemical risk assessment possesses that related human health can be assessed (Figure 5-1). In addition, data transparency, data storage, sharing and accessibility regarding to identification and semi-quantification data will also be important in the future, both vendors and analysts wishing to communicate the data quality will benefit from such. Analytical scientists who work on untargeted food analysis and semi-quantitation are recommended to work cooperatively with toxicology experts, data scientists, software engineers and also food authority hand in hand, to ensure the food trust worldwide.

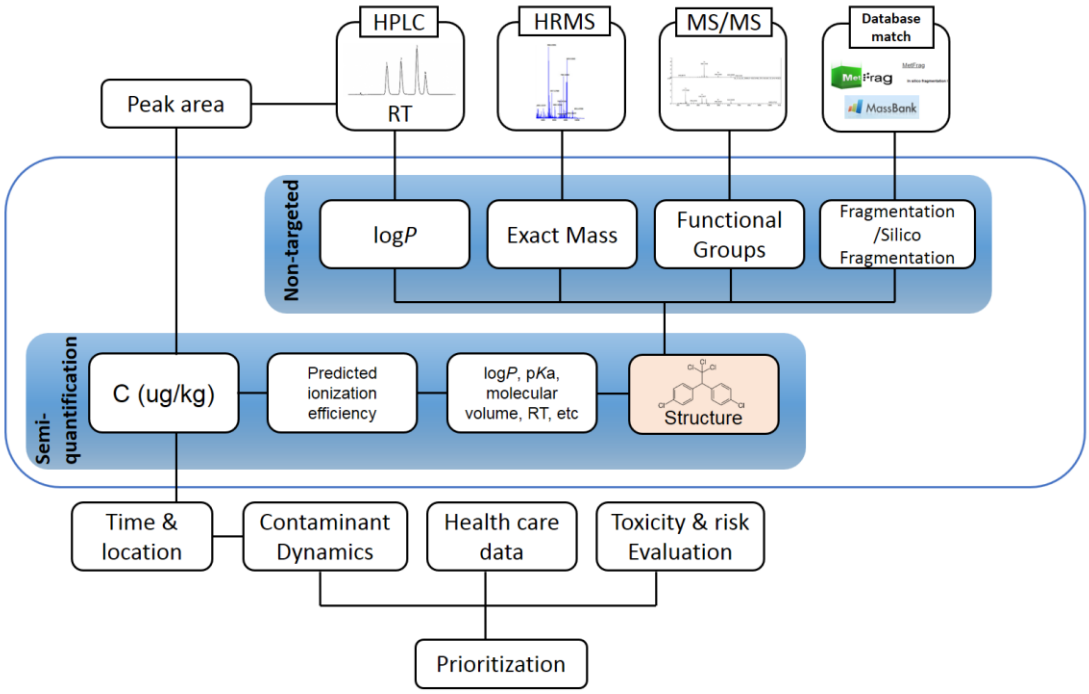


Figure 5-1. Combination between untargeted and semi-quantitative analysis contributes to the risk assessment.

Manuscripts

Manuscript 1: Development of detection and data treatment method

(Submitted to Food Chemistry)

Non-targeted food unexpected compounds analysis by LC/HRMS: feasibility study on rice

Tingting Wang, * Lene Duedahl-Olesen, Henrik Lauritz Frandsen

National Food Institute, Research Group for Analytical Food Chemistry, Technical University of Denmark, Kemitorvet Building 202, Kgs. Lyngby, DK-2800, Denmark

Highlights

- Optimization of ESI ion source and data treatment methods in LC-HRMS
- Powerful and rapid non-targeted data treatment strategies inspired from Metabolomics.
- PCA and PLS-DA for samples discrimination.
- Applicability to a wide range of food contaminants and residues.
- Detection of contaminants down to 5 $\mu g.kg^{-1}$ in a complex food matrix (rice).

Abstract

A widely applicable analytical LC/HRMS method based on ion source optimization, data treatment optimization on rice matrix was developed. The effects of key parameters in ion source, and also their interaction on an example of 34 common diverse food unwanted compounds were studied on HPLC-QTOF. A 40% increase in positive and 20% increase in the negative mode of response factor compared with center points can be achieved by ion source optimization. Data processing strategies inspired from metabolomics and suspected screening were compared and developed based on case and control rice samples. Developed and highly automated workflow based on XCMS achieved highest mass accuracy, highest detection rate (96%) and made a clear distinction between the control and contaminated groups by PCA and PLS-DA in a non-targeted way, even for the contaminated concentration of 5 µg/kg. This study can finally provide recommendations for Quan/Qual non-targeted analysis by LC/HRMS.

KEYWORDS: Food monitoring; HRMS; non-targeted analysis; method optimization; data treatment; metabolomics; food unexpected compounds; XCMS

1.Introduction

Non-targeted screening methods with HPLC-ESI-QToF-MS have been extensively applied to environmental science, metabolomics studies, diverse scientific issues in pharmacometabolomics and food monitoring (Fraser et al., 2014; Blum et al., 2017; Fu et al., 2018; Liu et al., 2019). From a chemical point of view, unwanted compounds in food products are mostly small molecules (<1200 Da) of many various substance classes including pesticides, mycotoxins, plant toxins, veterinary drug, and other contaminants originate from food processing (European Commission, 2007). These substances can be present in many different concentrations. Due to their diversity, it is impossible to detect all these compounds by a single analytical method. Despite of the frequent application of LC/MS, the ionization process of individual compounds are still not very clear, and fundamental studies are still needed. Meanwhile, research is continually performed to identify emerging risks in food safety (Brug et al., 2014), but analytical methods and data treatment method need to be developed in tandem to be able to analyze samples for the presence of unexpected hazardous compounds in a non-targeted way.

Some important insights have been gained by several research groups, to enhance ionization and MS coverage, the influence of mobile phase additives (Kruve, 2016; Creydt and Fischer, 2017), source type, source parameters (Chen et al., 2014) and data acquisition method on chemical analysis were studied (Wang et al., 2017). Only limited information about the effects and interaction of source parameters on the ionization efficiency is available.

The data treatment process is usually the most time consuming part of the investigation and also challenging, and it plays a critical role in the a successful final annotation. Routine target and non-target screening are the most common approaches to be used for food monitoring. However, both of them are targeted in the sense in many studies due to they start with compound information followed by analyzing the MS data (Hernández, F., Sancho, J. V., Ibáñez, M., & Grimalt, 2008; Tengstrand et al., 2013; Wang et al., 2019). Non-target here normally means that data are acquired in a non-targeted way, but the pre-selected compounds list are still necessary to detect compounds. The unexpected compounds remain non-detected until it is too late to prevent them from being widely spread, such as the milk powder scandals

with melamine adulteration in China in 2008 (Tyan, Yang and Jong, 2009), and fipronil in Dutch eggs spread in Europe in 2017 (Stafford et al., 2018).

Throughout a number of studies (Coble and Fraga, 2014; Vergeynst, Langenhove and Demeestere, 2015; Bader et al., 2016), it is evident that the peak finding step is important for the annotation, and it is even more challenged when dealing with low abundant signals. Various approaches have been developed in order to distinguish true positive from interference (Smith et al., 2006; Yu et al., 2008; Tengstrand et al., 2013) For example, multivariate statistics has been used to emphasize the variances among different samples (Müller et al., 2011; Delaporte et al., 2019), while it requires that the variance is attributed to true differences between the samples and availability of numerous replicates, and the calculation in between are generally complicate. There are also different software or platforms from instrument vendors or some other sources available for metabolomics analysis and environmental research, such as MZmine (Katajamaa and Orešič, 2005), XCMS (Zheng et al., 2013), Metabolite Detect (Bletsou et al., 2015), TracMass2 (Kunzelmann et al., 2018) and so on. However, the difference of sensitivity and accuracy of different platforms on the same data set is hardly known, and many of them generally still need quite a bit computational work like normalization, scaling, feature detection or peak resolution (Gorrochategui et al., 2016). Therefore a sensitive, efficient, automated and simultaneous non-targeted Quan/Qual analytical and data treatment strategy is demanded to be able to reveal what is in the sample.

The present study is aiming to make a comprehensive study for the development of both analytical and data treatment method. The study was carried out firstly by optimizing ion source parameters on the response factor in both ESI positive and negative mode. The effects of four key ion source parameters including nebulizer gas, capillary voltage, dry gas flow rate, in combination with gas temperatures in ESI and their interaction on an example of 34 common food contaminates and residues (pesticides, mycotoxins and veterinary drugs) were studied. Secondly, different data treatment strategies inspired from the field of metabolomics and suspected screening were compared and applied for food monitoring to enable peaks detection and discrimination. Settings in the software were optimized to minimize manual evaluation and irrelevant hits and to maximize relevant hit rate of the spiked compounds. The performance of

different data processing workflows was also evaluated to find out the most efficient and accurate approach for non-target analysis.

2. Methods and material

2.1 Chemicals and reagents

Propamocarb, Chloridazon, Pirimicarb-desmethyl, Clomazone, Napropamide, Phosphamidon, Imazaquin , aflatoxin B1, aflatoxin B2, Zearalenone, EPN, aflatoxin G1, Pencycuron, aflatoxin G2, Tebufenpyrad, Penthiopyrad, Isopyrazam, Pyridaben, DAS (diacetoxyscirpenol), Spiromesifen, Carbosulfan, Fluxapyroxad, Difloxacin, Benfuracarb, Indoxacarb, enniatin B, enniatin A1, enniatin B1, enniatin A, Erythromycin, Beaurericin, 3-O-acetyltylosin (all with a purity > 96%) were purchased from Dr. Ehrenstorfer (Augsburg, Germany). Milli-Q grade water was used for all analyses (Millipore Corporation, Molsheim, France). Acetonitrile (LC–MS grade), ammonium hydroxide solution (LC–MS grade), and formic acid (LC–MS grade) were purchased from Fluka (Steinheim, Germany), absorbent Bondesil-C18 was purchased from Agilent Technologies (Santa Clara, CA, USA), $MgSO_4$ was purchased from Merck (Darmstadt, Germany), and all other solvents used in this study were of analytical grade. The standards stock solutions of 1 mg mL^{-1} were prepared in toluene and stored at –18 °C. A standard mixture from 1 to 5 µg/ml in 80% acetonitrile was prepared from these stock solutions. Working solutions in the concentration range of 2 ng/mL to 40 ng/mL on 3 different concentrations were prepared by diluting the standard mixture with the extracts of blank samples as a matrix (free of studied compounds).

2.2 Samples and extraction

Blank rice reference samples (grown in the field produced in connection with the production of proficiency test material, SRM6) were used as the control samples. The extraction was based on the generic QuEChERS method developed by Dzuman's group (Dzuman et al., 2015), four grams of homogenized rice sample was weighed and extracted.

2.3 HPLC-ESI-QToF-MS.

Samples were analyzed on an Agilent 1200 HPLC (Agilent Technologies, CA, U.S.) coupled to a Bruker Daltonics micro-ToFq mass spectrometer (Bruker Daltonics, Bremen, Germany). The resolution of the Bruker micro-TOFq is approximately 10000. Samples were injected onto a

Nucleoshell C18 reversed-phase column (2.7 µm, 100 × 2 mm, MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). The mobile phase consisted of water (A) containing 2.5 mM ammonium formate with PH 3, and acetonitrile (B). The analysis was done using a gradient elution at a flow rate of 0.3 mL/min at 40 °C using the gradient: isocratic at 2% of B for 1.0 min, then increased to 100 % of B at 22.0 min, and maintained to 25 min, whereafter the gradient returned to starting conditions 2% of B at 22.1 min followed by 3 min equilibration with 2% of B to giving a total runtime of 28 min.

In data processing parts, ESI+ was used at the following settings: capillary voltage 3 kV, nebulizer pressure 20 psi, dry gas temperature 200 °C, and dry gas flow rate 8 L/min. Funnel radio frequency (RF) 400 V peak-to-peak (Vpp), multipole RF 200 Vpp. Spectra were collected from m/z 50 to 1100 Da in both positive and negative mode at a rate of 2 Hz. In the ESI optimization part, the settings used in both positive and negative ion mode (ESI+) for capillary voltage, nebulizer pressure, drying gas temperature, and drying gas flow, were demonstrated in Table S1 and Table S2. Mass calibration was performed in three steps according to the manufacturer's instructions, which has been reported previously (Wang et al., 2019).

2.4 Ion source optimization

In order to be able to make a general conclusion of methods applicability in more food crisis and for more unwanted contamination, not only focus on a specific scenario, 34 compounds used for optimization (Table 1) were selected to be as diverse as possible in terms of log*P* (-0.07~5.69), retention time (3.6~21.0 min), elemental composition (Cl, F, P, S), and molecular volume (127.82 ~ 772.53). There were in total 34 compounds were used for Analytical method optimization.

Table 1 Chemical formula, adducts, and molecular properties of 34 compounds for analytical method optimization

Name	Calculated	RT	log*P*	MV	Formula	ESI	Adducts
Propamocarb	189.1597	3.65	1.15	196.65	$C_9H_{20}N_2O_2$	+	M+H
Pirimicarb-desmethyl	225.1346	5.85	1.43	189.28	$C_{10}H_{16}N_4O_2$	+	M+H
Chloridazon	222.0429	6.9	0.92	155.44	$C_{10}H_8ClN_3O$	+/-	M+H
Difloxin	400.1467	6.96	-0.07	283.28	$C_{21}H_{19}F_2N_3O_3$	+	M+H
aflatoxin G2	331.0812	8.5	0.66	212.33	$C_{17}H_{14}O_7$	+	M+H
Phosphamidon	300.0762	8.92	0.98	245.81	$C_{10}H_{19}ClNO_5P$	+	M+H
Aflatoxin G1	329.0656	9	0.73	205.66	$C_{17}H_{12}O_7$	+	M+H
Aflatoxin B2	315.0863	9.03	0.92	206.26	$C_{17}H_{14}O_6$	+	M+H
Erythromycin	734.4685	9.28	2.44	229.45	$C_{37}H_{67}NO_{13}$	+	M+H
Imazaquin	312.1343	9.33	1.9	607.18	$C_{17}H_{17}N_3O_3$	+/-	M+H
DAS	384.2017	9.45	1.4	278.57	$C_{23}H_{30}O_4$	+	M+NH4
Aflatoxin B1	313.0707	9.56	1.08	199.58	$C_{17}H_{12}O_6$	+	M+H
3-o-acetyltylosin	958.537	10.14	2.77	772.53	$C_{48}H_{79}NO_{18}$	+	M+H
DNOC	199.0349	11.25	2.27	127.82	$C_7H_6N_2O_5$	-	M+H
Clomazone	240.0786	12.04	2.36	195.99	$C_{12}H_{14}ClNO_2$	+	M+H
Zearalenone	319.154	12.84	3.96	272.35	$C_{18}H_{22}O_5$	+/-	M+H
Fluxapyroxad	382.0968	13.55	3.5	267.86	$C_{19}H_{26}O_7$	+	M+NH4
Napropamide	272.1645	13.8	3.44	251.34	$C_{17}H_{21}NO_2$	+	M+H
Oryzacin	347.102	14.34	2.59	246.08	$C_{12}H_{18}N_4O_6S$	-	M+H
Penthiopyrad	360.1352	15.33	3.64	275.74	$C_{16}H_{20}F_3N_3OS$	+/-	M+H
Isopyrazam	360.1882	16.04	4.44	261.09	$C_{20}H_{23}F_2N_3O$	+/-	M+H
Pencycuron	329.1415	16.16	4.34	268.65	$C_{19}H_{21}ClN_2O$	+/-	M+H
EPN	324.0454	16.74	4.06	241.13	$C_{14}H_{14}NO_4PS$	+	M+H
Indoxacarb	528.078	16.75	3.7	343.31	$C_{22}H_{17}ClF_3N_3O_7$	+	M+H
Tebufenpyrad	334.1681	16.89	4.41	293.16	$C_{18}H_{24}ClN_3O$	+	M+H
Benfuracarb	411.1948	17.46	4.29	350	$C_{20}H_{30}N_2O_5S$	+	M+H
enniatin B	657.4433	18.65	3.42	617.25	$C_{33}H_{57}N_3O_9$	+	M+NH4
Pyridaben	365.1449	19.36	5.44	328.88	$C_{19}H_{25}ClN_2OS$	+	M+H
Spiromesifen	371.2217	19.38	5.25	650.26	$C_{20}H_{32}N_2O_3S$	+	M+H
enniatin A1	671.4589	19.39	4.21	323.53	$C_{35}H_{61}N_3O_9$	+	M+NH4
Beaurericin	801.4433	19.43	5.69	695.82	$C_{45}H_{57}N_3O_9$	+	M+NH4
enniatin B1	685.4746	20.1	3.81	633.75	$C_{34}H_{59}N_3O_9$	+	M+NH4
enniatin A	699.4902	20.76	4.67	666.77	$C_{36}H_{63}N_3O_9$	+	M+NH4
Carbosulfan	381.2206	21.01	5.2	346.14	$C_{18}H_{12}F_5N_3O$	+/-	M+H

A D-optimal custom screening design, was generated using JMP 13®. Four factors were entered as continuous 2-level parameters (level-1, level-0 and level 1) shown in Table S1. Due to the dry gas DryT of ion source is hard to change, all the combinations were separated into 5 different blocks based on DryT to avoid the frequent change of temperature (split plot). The design was then generated from a random seed with 10000 starts, resulting in a total of 20 experiments. The full design matrix is available as Table S2. Each experiment in the design was performed twice for each ESI polarity, where a full sequence was performed consequently in each ESI polarity. In total, two times 20 (40) experimental runs were performed. Each experimental condition was noted by a two digit number defined from the run order, e.g., 1-3 denotes the third experimental design run in the first block. All of the experimental conditions were considered unique combinations of factors. In between each block, a re-equilibration step (10 min) at the end of the gradient run was used to equilibrate the source parameters. Response factor of each compound was used in data analysis to evaluate the performance of ionization in various ion source settings.

2.5 Optimization and comparison of different data processing strategy for detection

Three different data treatment approaches including Suspected Screening (Method A), Metabolite Detect (Method B) and screening method based on XCMS package (Method C) (Giacomoni et al., 2014) were used to detect the compounds in a targeted or non-targeted way. Different data collection and processing strategies based on case and control samples in triplicate were developed and compared. Control samples were collected prior to the experiments, and reflected baseline chemical composition. Case samples were spiked control samples with 29 model compounds to mimic the unknown contamination at 3 concentration levels (Table 2), except for DAS (10 µg/kg, 50 µg/kg, 200 µg/kg) and Zearalenone (25 µg/kg, 125 µg/kg, 500 µg/kg), all the other model compounds were spiked at 5 µg/kg (level 1), 25 µg/kg (level 2) and 100 µg/kg (level 3). First different setting within each software were separately optimized or partially took the recommendation values from previous studies, such as smoothing width, various threshold and ratio, each of them can be an important factor for successful detection. Furthermore, the detection capacity and performance of various

processing approaches on the detection of true positive and false negative were also compared.

Table 2. LOD, levels and mass accuracy of different data processing methods for the detection of 29 comopounds

Name	LOD (μg/kg)	XCMS (ppm)	Metabolite Detect (ppm)	Suspected screening
Pirimicarb-desmethyl	3.05	-7.95	-7.92	-8.39
Chloridazon	2.91	-5.42	-4.35	-2.80
Difloxin	3.51	1.72	3.21	3.11
Aflatoxin G2	2.91	-1.91	0.05	-0.54
Phosphamidon	1.86	-3.04	-6.04	-2.70
Aflatoxin G1	2.59	0.78	-2.50	0.71
Aflatoxin B2	1.15	-2.92	-4.51	-1.97
Erythromycin	6.11	4.32	0.41	3.89
Imazaquin	1.97	-2.68	-3.52	-2.92
DAS*	4.59	0.92	0.00	2.46
Aflatoxin B1	1.92	-2.77	-2.91	-2.09
3-o-acetyltylosin	5.36	-0.18	-0.37	0.31
Clomazone	3.49	-4.10	-7.36	-4.63
Zearalenone*	18.25	4.01	3.39	0.45
Fluxapyroxad	2.87	2.73	1.41	3.26
Napropamide	1.46	-4.92	-7.59	-4.37
Penthiopyrad	1.80	-0.28	0.22	1.82
Isopyrazam	1.58	0.00	-1.33	0.74
Pencycuron	1.66	-1.74	-3.24	-0.37
EPN	7.72	-2.05	-7.79	0.53
Indoxacarb	3.13	5.19	4.54	5.09
Tebufenpyrad	3.44	1.59	1.95	3.42
enniatin B	1.26	3.33	4.75	4.43
Pyridaben	2.35	-0.59	-2.19	-0.06
Spiromesifen	5.09	-0.16	2.11	2.93
enniatin A1	2.09	3.34	2.32	3.77
Beaurericin	3.31	1.95	2.62	2.77
enniatin B1	0.92	2.24	1.44	3.65
enniatin A	0.95	3.15	2.49	3.83

*Note: *Spiked concentration levels of DAS are 200 µg/kg (level 3), 50 µg/kg (level 2), 10 µg/kg (level 1); Spiked concentration levels of Zearalenone are 500 µg/kg (level 3), 125 µg/kg (level 2), 25 µg/kg (level 1); Spiked concentration levels of all the other compounds are 100 µg/kg (level 3), 25 µg/kg (level 2), 5 µg/kg (level 1).*

Both Method 2 and Method 3 are non-targeted methods used in metabolomics (Smith et al., 2006; Antignac et al., 2011), but we applied them here for non-targeted food safety analysis. The goal was to detect as many as true positive, and as few as false negative as possible for the added model compounds, also to detect as many corresponding features (protonated ions, adducts, fragments, or any isotope of these) as possible and as few as irrelevant features originating from the model compounds. Apart from Target Screening, no databases or pre-selected lists of compounds were needed beforehand. Positive detection solely relied on detection of features/signals that were present in the spiked samples but not present in the control samples.

2.5.1 Method A: Suspected Screening

TargetAnalysis 1.3 (Method A) software from Bruker Daltonics was originally developed for pesticides. TargetAnalysis can screen up to 3000 compounds simultaneously, on the basis of mass accuracy, isotope fit, and retention time (Klitgaard et al., 2014). A database was created in csv. format including the chemical formula and names of all the compounds. For the successful detection, threshold of error, mSigma, whether smoothing or not and how much the values of them are all important factors for chromatographic resolution and the reliable detection (Dubbelman et al., 2018). The optimal settings after optimization were applied when analyzing the 29 compounds in the samples. The workflow and detailed setting of the software can be found in Figure 2-a and Table S3.

2.5.2 Method B: Metabolite Detect

Metabolite Detect 2.0 by Bruker Daltonics is a metabolite detection software for the identification and confirmation of expected and known metabolites as well as the annotation of unknowns. In this study, it was used for the non-targeted screening to mimic the screening of unknown food contamination (Figure 2-b). The eXpose algorithm, which is more robust compared with “simple difference” was applied in the processing, it masks the sample data with reference data without subtracting any intensities. After the complete calculation, a base peak

chromatogram of the difference is constructed. All the optimal settings after optimization in Metabolite Detect were shown in Table S4.

2.5.3 Method C: Non-targeted screening based on XCMS

This data processing procedure based on XCMS package with Workflow4Metabolomics was set up to be as automated as possible. The raw data file in .d format produced from Bruker micrOTOFq were converted to mzXML format by ProteoWizard firstly in peak picking mode with Vendor checked in order to centroid the data (Toolkit, 2012). Then the data files were uploaded to Workflow4Metabolomics platform, which is a virtual research environment built upon the Galaxy web-based platform (Giacomoni et al., 2014). Data matrix file in tsv. format was also created by this platform (Smith et al., 2006). The general workflow was shown in the Figure 2-c. The corresponding parameters settings for XCMS algorithm were inspired by the recommendation from Patti's study (Patti, Tautenhahn and Siuzdak, 2012). The full list of all these parameters settings for the MS data can be found in the supporting information Table S5. As the number of output ions generated by Method C is high (around 2930 entries in our case), it is necessary for the data matrices to filtrate and remove as many irrelevant ions as possible. Data filtration step is important to minimize relevant chemical information loss but also minimize false positive ions. Therefore, the ions that do not differ from the blanks or do not vary between samples were excluded in order to keep only suspect ions. The filtration of the data matrices (variable metadata and data matrix data) was done within three steps. In the pre-filtration step, the ions (around 2600 ions) show no significant difference between blank samples and any sample group (pairwise univariate T-test were used here to determine the significant difference for each ion between groups at a commonly used value of 0.05) were discarded after the first filtration (Li et al., 2014; Wang et al., 2015), there were 210 to 298 ions still left from groups level 1 to level 3. In the second filtration, the ions which exhibit a poor stability, i.e. the relative standard deviation (%RSD) of the peak area is higher than 100% in each sample group, the ions with intensity lower than control group were removed. In this step, around 184 to 237 ions were remained. In the last filtration, the "npeaks" value of ions with non-zero in control group were removed, then there were still 80 to 141 features left from level 1 to level 3 in the end.

3 Result and discussion

3.1 Ion source optimization

In order to investigate the source effects on the responses factors, a statistical factor analysis was performed to obtain a quantitative comparison. The experimental design space allowed full explorative analysis of 1st order factors and exploration on most of the 2nd order interactions in-between factors.

From the parameter estimates results, in the positive mode, 32 compounds were successfully detected among 34, taking the sum of the lowest 25% signals (min25%, S1+S2…+Sn, n = 8) as the optimized variable, Cap and DryF influence min25% the most than other factors with *t* Ratio of 0.02 and 0.005, respectively, and Cap and DryF also showed second-order interaction with a *t* Ratio of 0.02, indicating that min25%-s is optimal when Cap is at a high value or DryF is at relative low values without concern of interaction between these two factors. By using "maximize desirability" function in the model, the predicted maximum min25%-s value is 9169741 (Figure 1b), which has 40% increase compared with the predicted value of center points with a value of 6586388. The response surface can be found in Figure S2.

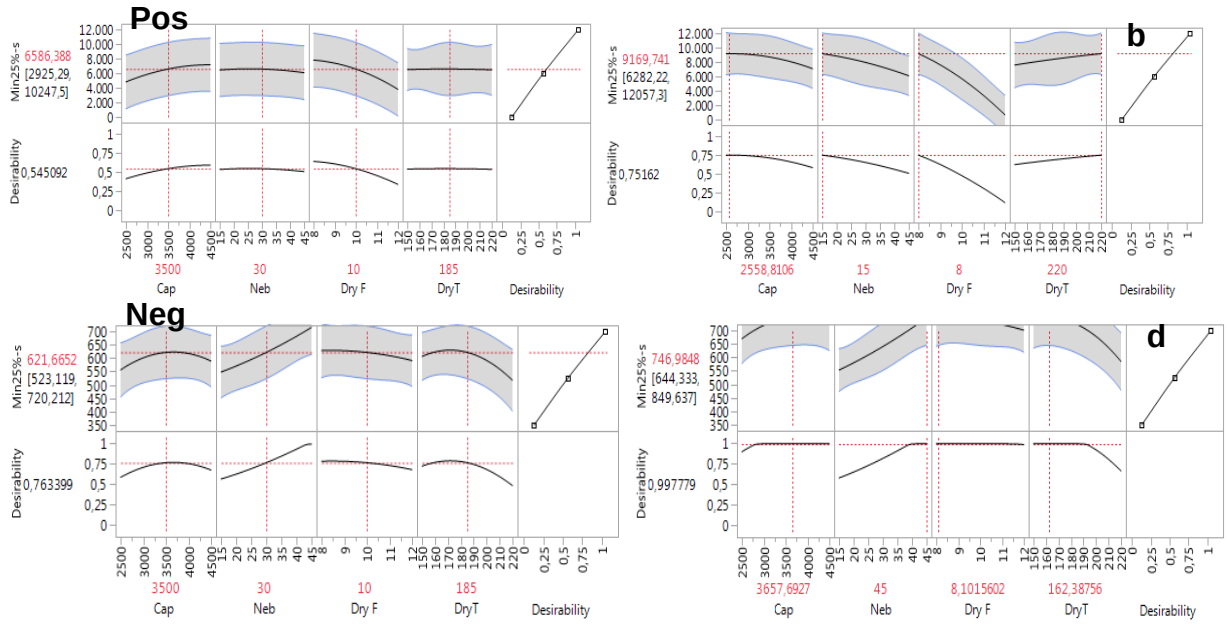


***Figure**.1. Prediction profiler of effects of source parameters on response a: effects of source parameters on response at center points in positive mode. b: effects of source parameters on response after optimization in positive mode. c: effects of source parameters effects on response in positive mode at center points in negative mode. d: effects of source parameters on response after optimization in negative mode*

In the negative mode, nine out of 34 compounds were successfully detected. It was found that Neb and DryT influence min25% more than other factors with *t* Ratio of 0.0004 and 0.0107, respectively, indicating that min25% is optimal when Neb is in at higher value and DryF is at low value, and there is no interaction between any two factors. By using “maximize desirability” function in the model, the predicted maximum min25% value is 7469848 (Figure 1d), which has 20% increase compared with the predicted value of center points. All the detailed information about effect summary, parameter estimate, fixed effect tests in both positive and negative mode are shown in Table S6 to Table S11.

3.2 Data processing

3.2.1 Method A: suspected screening

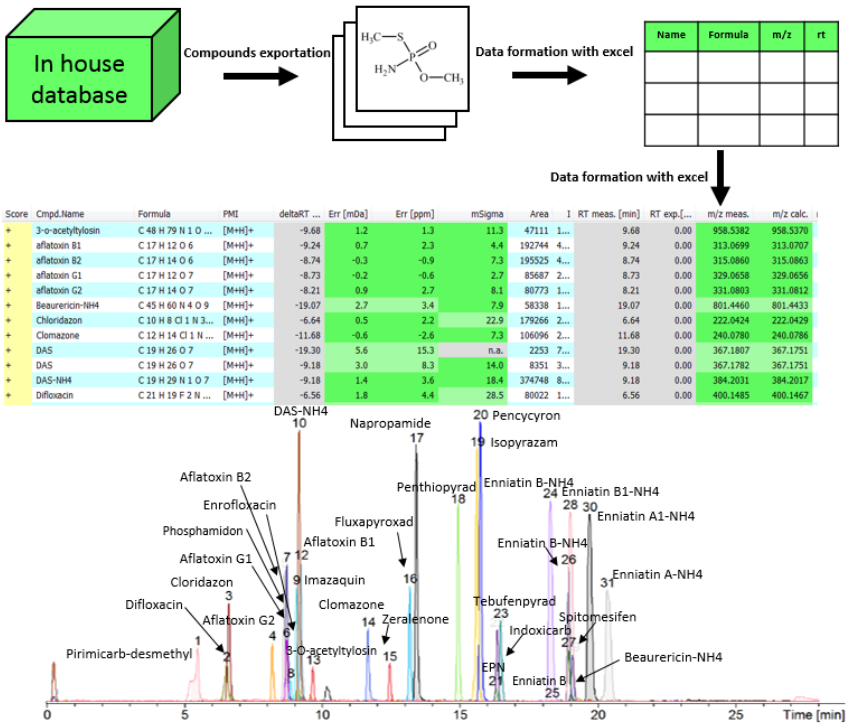
Suspected screening is a very common method used for food monitoring and were used here to screen the suspected compounds existence in rice samples. Apart from smoothing width, mSigma threshold is also important for the successful detection, which is the goodness of fit between measured and theoretical isotopic pattern. A lower mSigma value indicates a better fit, but it will also exclude some true positive, as we found in our case. mSigma value less than 100 allow the detection of most compounds (28 of 29), but the mSigma of EPN at 25 µg/kg is 418, with a mass error of 9 ppm. It has been reported that the mSigma value decreases with increasing concentration and also shows an increase with increasing matrix complexity (Peters, Bolck, Rutgers, Stolker, & Nielen, 2009). According to the results shown in Figure 2-a, the peaks in the samples have a good match with the database created. There was no problem to detect all the spiked compounds with high accuracy at the concentration of both level 2 and level 3, and even when the concentration was as low as 2 ng/mL (level 1, most of compounds at 5 µg/kg), All compounds above LOD were still successfully detected, which means the detection rate is 100% for all 3 concentration levels. The mean mass accuracy of all the compounds in all 3 levels is 2.69 ppm (Table 2). The mean mass errors from level 1 to 3 are 3.73 ppm, 2.84 ppm, and 3.04 ppm, respectively. It can be found that lower concentration tends to have higher mass error compared with high concentration.

3.2.2 Method B: Metabolite Detect

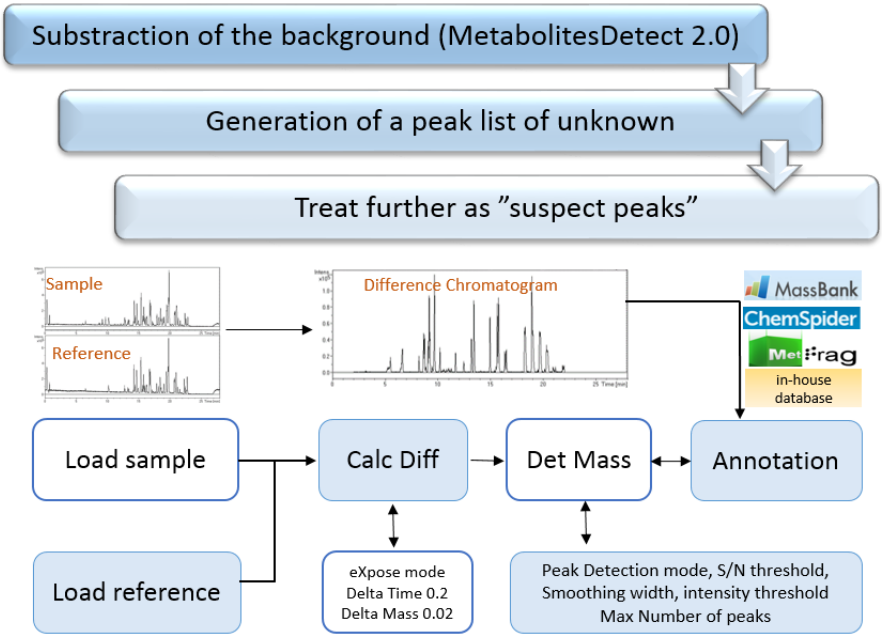
The most important factor which influence the detection capacity in Metabolite Detect is smoothing width. As shown in Figure S3, as the smoothing width increases from 1 to 5, the number of false negative increases from 12 to 22 for concentration level 1, from 2 to 7 for level 2, and from 1 to 5 for level 3, respectively. This strategy allowed for a rapid and substantial reduction in the number of suspect masses investigated.

Figure 2

a.



b.



c.

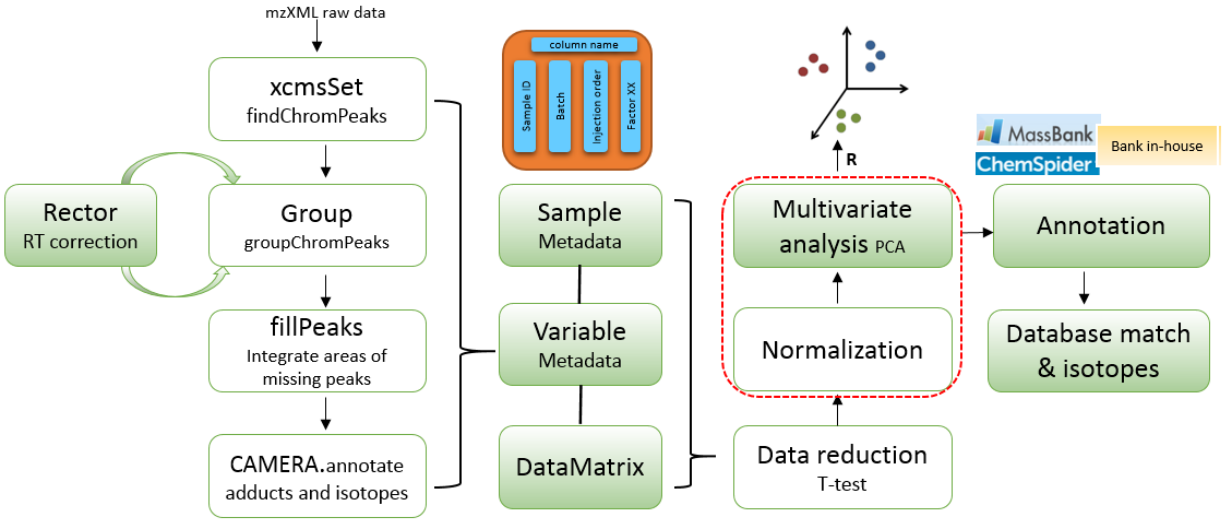


Figure 2. a. Example of workflow for screening of contaminants in rice with Target screening. Analysis via target screening yields both a graphical interpretation of the results and a table of the data. b. Example of workflow for detection of contaminates in rice with Merobolites Detect. c. Workflow developed based on XCMS for non-targeted contaminants detection at trace levels in rice.

From the results, there were around 282 to 216 features detected from the concentration level 1 to level 3, it had a good detection capacity at high concentration (level 3) with the average detection rate of 97% (Figure 4). Erythromycin (m/z: 734.4685) and beaurericin (m/z: 801.4433) were treated as false negatives in 2 of the 3 replicates. There were 2 and 10 compounds above LOD were detected as false negatives in level 2 and level 1, the detection rates were 93% and 66%, respectively, which were much lower than the detection rate of suspected screening screening. Many peak splits and redundant detection were observed for some compounds. The mass molecule weight of each compound with the highest intensity (mass difference within 0.01) value was used for the calculation of mass accuracy. The average MS accuracy of each compound at these three concentration with triplicate were taken. By using Metabolite Detect, the mean absolute mass error is 3.2 ppm. After the detection step, the suspected peaks can be treated further for annotation step by local or online database matching.

3.2.3 Method C: Non-targeted screening based on XCMS package

First of all, the mzXML file conversion is a critical step for the following data-processing, it is important to choose “peak picking” mode in the filters part. Although “Threshold Peak Filter” can be another option in ProteoWizard, it will generate much more irrelevant ions for the same compounds and more efforts need to be done in data cleaning. It is very common to get more than 12000 features for each sample in our cases by using “Threshold Peak Filter” even though we used the other settings in processing part. The mass accuracy of using “Threshold Peak Filter” is also much lower than “peak peaking” mode.

When we look into each group of the detection results, the detection rate of true positive is 97% for concentration level 3 and level 2, and 96% for group of level 1. It means that, even in the worst case that most compounds are in 5 µg/kg, Method C will still be able to successfully extract above 96% of ions (26 out of 27, Erythromycin and EPN were excluded here as LOD of them are higher than their level 1), which is higher than the study has been reported (75% overall) (Delaporte, Cladière, Bouveresse, et al., 2019). There are 2 ions (Erythromycin in level 2 and level 3, and tebufenpyrad in level 1) can not be correctly extracted from the raw data matrix file, for example, the feature of Erythromycin (m/z: 734.4685) was discarded arbitrarily by the T-test (extracted successfully in the raw data matrix file) in group level 2 and 3, and the *P* values of it in group level 3 and level 2 were 0.076 and 0.051, respectively. They are quite

close to 0.05 already, which means that the T-test method could be improved to retain these 2 ions or increase the *P* value limit, although this method has already a very high detection rate for all levels. Regarding the mass accuracy, this platform performs the best among three data processing method, with accuracies on average of 2.6 ppm for all the compounds (in the range of 0 to 7.95 ppm, Table 2).

These non-targeted approach can be used to detect the features that are normally neglected by the routine target screening. The theoretical elemental compositions were generated based on accurate mass, eg., $[M + H]^+$ or $[M + NH_4]^+$, and were prioritized for further investigation by selecting compounds with the smallest mass error and the highest MS/MS rank. Therefore, a high mass accuracy is critical as it allowed for tentative compound identification and also important for the correct interpretation. Even based on the same raw data, the mass accuracy is still different according to the algorithm of each processing method. Probable chemical structures can be generated for the highest priority elemental composition with the aid of ChemSpider, Massbank, MetFrag and local databases. However, the main purpose of this study was to focus on data processing, and how to get as many true positive as possible with as few as false negative. The compounds annotation part is out the scope of the present study and will not be included here.

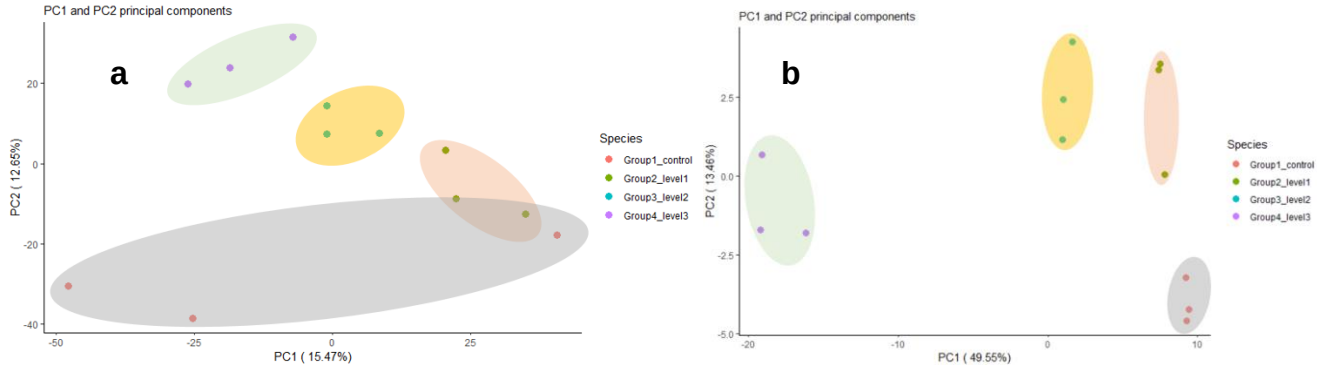


***Figure** 3. Example of Method C application. Principal component analysis plot of the features with an ANOVA p-value > 0.05 upon comparison between rice samples with different concentrations contaminant or control (rice blank). a: Before filtration. b: After filtration*

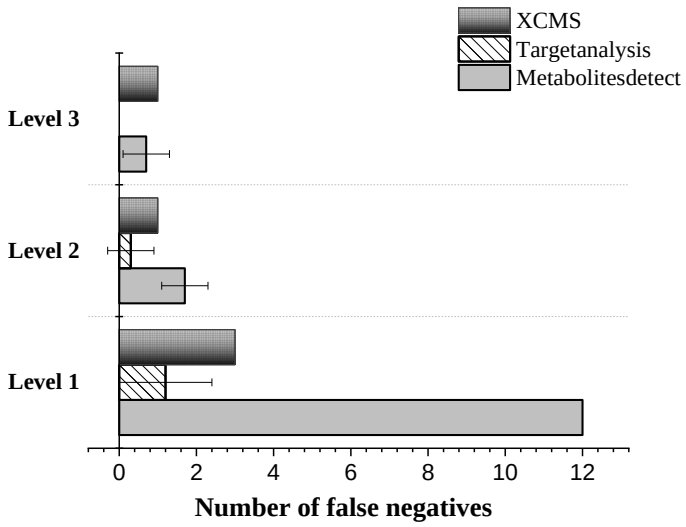


***Figure** 4. Number of false negatives in different level by using different data processing methods*

In order to discriminate between the groups and to evaluate the efficiency of a generic blind non-targeted analysis based on multivariate tools, Principal component analysis (PCA) was preliminarily applied to discriminate contaminated rice samples. The multivariate exploration was first applied without data filtration as shown in Figure 3-a (original 2930 features), it can be found that the control group (red dots) could not be discriminated from Group 2 (level 1, light grey dots). Figure 3-b shows the resulting PCA plot of the features with a *p*-value < 0.05 based on one-way ANOVA after data cleaning. A very clear distinction between the four groups can be found, even for the spiked sample group with lowest concentration of level 1 (5 µg/kg for most). This level is lower than the Maximum Residue Limits (MRL) for most pesticides (10 µg/kg), suggesting that this data processing method especially the cleaning step obviously increase the discrimination capacity of PCA. This metabolomics-like approach achieved discrimination of contaminated food samples with control sample on an even lower level than the lowest reported value (10 µg/kg) (Delaporte, Cladière, Bouveresse, et al., 2019), the explained variance of PC1 increased significantly from 15% to 50%. Partial least squares discriminant analysis (PLS-DA) was also performed to compare the results before and after data cleaning (Figure S4), the cleaning step significantly improve the model performance, PLS-DA was also able to separate the 4 groups with R2Y= 0.907 and Q2 = 0.575 showing a fairly robust model, also the ability to describe all the variables having an R2X = 0.947. Typically, Q^2

score >0.4 and R^2 > 0.5 indicates a robust model, and scores between 0.7 and 1.0 indicates a highly robust model (K. L. Nielsen et al., 2016). It obviously produce better model statistics after data cleaning to distinguish different sample groups.

4. Conclusion

This study developed and compared both analytical method and data treatment method. Regarding ion source optimization, Cap and DryF influence the chemical response most in positive mode. Neb and DryT influence the response most in the negative mode. A 40% increase in positive and 20% increase in the negative mode of response compared with center points can be achieved by optimization. Comparison of different data processing methods showed that, Method A has the highest mass accuracy, highest detection rate (100% for all 3 levels), while Method B has the lowest mass accuracy, lowest detection rate (66% for level 1) and gives the largest number of false negative. Method C performs better in mass accuracy, detection rate (96%) and lower the number of false negatives when compared to Method B in a non-targeted way, it was also possible to make a clear distinction between the control group and three spiked groups according to this approach by PCA and PLS-DA, even for the spiked samples with lowest concentration (5 µg/kg for most). Such approach as Method C could be a good complement to routine target analysis in a view of rapidly detecting potential contaminated food products without sacrificing too much accuracy and increasing the number of false negatives.

Acknowledgements

The authors acknowledge the funding from the Danish Food Authorities (Refs. Fødevareforlig 3).

.

Reference

Antignac, J., Courant, F., Pinel, G., Bichon, E., Monteau, F., Elliott, C., & Bizec, B. Le. (2011). Mass spectrometry-based metabolomics applied to the chemical safety of food. *Trends in Analytical Chemistry*, *30*(2), 292–301.

Bader, T., Schulz, W., Kümmerer, K., & Winzenbacher, R. (2016). General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Analytica Chimica Acta*, *935*, 173–186.

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

Blum, K. M., Andersson, P. L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., & Haglund, P. (2017). Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. *Science of the Total Environment*, *575*, 265–275.

Brug, F. J. Van De, Luijckx, N. B. L., Cnossen, H. J., & Houben, G. F. (2014). Early signals for emerging food safety risks : From past cases to future identification. *Food Control*, *39*, 75–86.

Chen, L., Song, F., Liu, Z., Zheng, Z., Xing, J., & Liu, S. (2014). Study of the ESI and APCI interfaces for the UPLC-MS/MS analysis of pesticides in traditional Chinese herbal medicine. *Analytical and Bioanalytical Chemistry*, *406*(5), 1481–1491 .

Coble, J. B., & Fraga, C. G. (2014). Comparative evaluation of preprocessing freeware on chromatography / mass spectrometry data for signature discovery. *Journal of Chromatography A*, *1358*, 155–164.

Creydt, M., & Fischer, M. (2017). Plant Metabolomics - Maximizing Metabolome Coverage by Optimizing Mobile Phase Additives for Non-Targeted Mass Spectrometry in Positive and Negative Electrospray Ionization Mode. *Analytical Chemistry*, *89*, 10474–10486.

Delaporte, G., Cladière, M., Bouveresse, D. J., & Camel, V. (2019). Untargeted food contaminant detection using UHPLC-HRMS combined with multivariate analysis : Feasibility study on tea. *Food Chemistry*, *277*, 54–62.

Dubbelman, A. C., Cuyckens, F., Dillen, L., Gross, G., Vreeken, R. J., & Hankemeier, T. (2018). Mass spectrometric recommendations for Quan/Qual analysis using liquid-chromatography coupled to quadrupole time-of-flight mass spectrometry. *Analytica Chimica Acta*, *1020*, 62–75.

Dzuman, Zbynek, Zachariasova, Milena, Veprikova, Z., Godula, M., & Hajslova, J. (2015). Multi-analyte high performance liquid chromatography coupled to high resolution tandem mass spectrometry method for control of pesticide residues, mycotoxins, and pyrrolizidine alkaloids. *Analytica Chimica Acta*, *863*(1), 29–40.

European Commission. (2007). https://ec.europa.eu/food/safety/chemical_safety.

Antignac, J., Courant, F., Pinel, G., Bichon, E., Monteau, F., Elliott, C., & Bizec, B. Le. (2011). Mass spectrometry-based metabolomics applied to the chemical safety of food. *Trends in Analytical Chemistry*, *30*(2), 292–301.

Bader, T., Schulz, W., Kümmerer, K., & Winzenbacher, R. (2016). General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Analytica Chimica Acta*, *935*, 173–186.

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

Blum, K. M., Andersson, P. L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., & Haglund, P. (2017). Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. *Science of the Total Environment*, *575*, 265–275.

Brug, F. J. Van De, Luijckx, N. B. L., Cnossen, H. J., & Houben, G. F. (2014). Early signals for emerging food safety risks : From past cases to future identification. *Food Control*, *39*, 75–86.

Chen, L., Song, F., Liu, Z., Zheng, Z., Xing, J., & Liu, S. (2014). Study of the ESI and APCI interfaces for the UPLC-MS/MS analysis of pesticides in traditional Chinese herbal medicine. *Analytical and Bioanalytical Chemistry*, *406*(5), 1481–1491 .

Coble, J. B., & Fraga, C. G. (2014). Comparative evaluation of preprocessing freeware on chromatography / mass spectrometry data for signature discovery. *Journal of Chromatography A*, *1358*, 155–164.

Creydt, M., & Fischer, M. (2017). Plant Metabolomics - Maximizing Metabolome Coverage by Optimizing Mobile Phase Additives for Non-Targeted Mass Spectrometry in Positive and Negative Electrospray Ionization Mode. *Analytical Chemistry*, *89*, 10474–10486.

Delaporte, G., Cladière, M., Bouveresse, D. J., & Camel, V. (2019). Untargeted food contaminant detection using UHPLC-HRMS combined with multivariate analysis : Feasibility study on tea. *Food Chemistry*, *277*, 54–62.

Dubbelman, A. C., Cuyckens, F., Dillen, L., Gross, G., Vreeken, R. J., & Hankemeier, T. (2018). Mass spectrometric recommendations for Quan/Qual analysis using liquid-chromatography coupled to quadrupole time-of-flight mass spectrometry. *Analytica Chimica Acta*, *1020*, 62–75.

Dzuman, Zbynek, Zachariasova, Milena, Veprikova, Z., Godula, M., & Hajslova, J. (2015). Multi-analyte high performance liquid chromatography coupled to high resolution tandem mass spectrometry method for control of pesticide residues, mycotoxins, and pyrrolizidine alkaloids. *Analytica Chimica Acta*, *863*(1), 29–40.

European Commission. (2007). https://ec.europa.eu/food/safety/chemical_safety.

Patti, G. J., Tautenhahn, R., & Siuzdak, G. (2012). Meta-analysis of untargeted metabolomic data from multiple profiling experiments. *Nature Protocols*, *7*(3), 508–516.

Peters, R. J. B., Bolck, Y. J. C., Rutgers, P., Stolker, A. A. M., & Nielen, M. W. F. (2009). Multi-residue screening of veterinary drugs in egg, fish and meat using high-resolution liquid chromatography accurate mass time-of-flight mass spectrometry. *Journal of Chromatography A*, *1216*(46), 8206–8216.

Smith, C. A., Want, E. J., Maille, G. O., Abagyan, R., & Siuzdak, G. (2006). XCMS : Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment , Matching , and Identification. *Anal Chem*, *78*(3), 779–787.

Stafford, E. G., Tell, L. A., Lin, Z., Davis, J. L., Vickroy, T. W., Riviere, J. E., … Service, I. (2018). FARAD Digest Consequences of fipronil exposure in egg-laying hens. *FARAD Digest*, *253*(1), 57–60.

Tengstrand, E., Rosén, J., Hellenäs, K. E., & Åberg, K. M. (2013). A concept study on non-targeted screening for chemical contaminants in food using liquid chromatography-mass spectrometry in combination with a metabolomics approach. *Analytical and Bioanalytical Chemistry*, *405*(4), 1237–1243.

Toolkit, P. (2012). Nature Biotechnology, *30*, 10-12.

Tyan, Y., Yang, M., & Jong, S. (2009). Melamine contamination, 729–735.

Vergeynst, L., Langenhove, H. Van, & Demeestere, K. (2015). Balancing the False Negative and Positive Rates in Suspect Screening with High-Resolution Orbitrap Mass Spectrometry Using Multivariate Statistics. *Anal. Chem.*, *87*, 2170–2177.

Wang, T., Frandsen, H. L., Christiansson, N. R., Rosendal, S. E., Pedersen, M., & Smedsgaard, J. (2019). Pyrrolizidine alkaloids in honey: Quantification with and without standards. *Food Control*, *98*, 227–237.

Wang, T., Li, X., Zhou, B., Li, H., Zeng, J., & Gao, W. (2015). Anti-diabetic activity in type 2 diabetic mice and α-glucosidase inhibitory, antioxidant and anti- inflammatory potential of chemically profiled pear peel and pulp extracts (Pyrus spp.). *Journal of Functional Food*, *13*, 276–288.

Wang, Y., Feng, R., Wang, R., Yang, F., Li, P., & Wan, J. B. (2017). Enhanced MS/MS coverage for metabolite identification in LC-MS-based untargeted metabolomics by target-directed data dependent acquisition with time-staggered precursor ion list. *Analytica Chimica Acta*, *992*, 67–75.

Yu, W., He, Z., Liu, J., & Zhao, H. (2008). Improving Mass Spectrometry Peak Detection Using Multiple Peak Alignment Results. *Journal of Proteome Research*, *7*, 123–129.

Zheng, H., Clausen, M. R., Dalsgaard, T. K., Mortensen, G., & Bertram, H. C. (2013). Time-saving design of experiment protocol for optimization of LC-MS data processing in metabolomic approaches. *Analytical Chemistry*, *85*(15), 7109–7116.

Manuscript 2: Targeted and untargeted semi-quantitative analysis

Food Control 98 (2019) 227–237

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Pyrrolizidine alkaloids in honey: Quantification with and without standards

Tingting Wang, Henrik Lauritz Frandsen, Natasha Roed Christiansson, Sofie Elisabeth Rosendal, Mikael Pedersen, Jørn Smedsgaard*

National Food Institute, Research Group for Analytical Food Chemistry, Technical University of Denmark, Kemitorvet Building 202, Kgs. Lyngby, DK-2800, Denmark

ARTICLE INFO

Keywords:
Pyrrolizidine alkaloids
Validation
Quantification without standards
Screening
Honey
HPLC-HRMS

ABSTRACT

Quantification of a large number of compounds in the absence of authentic standards is always a challenge. More than 600 pyrrolizidine alkaloids (PAs) have been found in plants however only limited number of PAs standards are commercially available. As PAs are the most widely distributed natural plant toxins with threat to human health, risk assessment calls for quantitative analytical methods with a wide scope including PAs without available standards. In this study, a method was developed that allows simultaneous quantification of 12 PAs in the honey samples by using HPLC-HRMS with authentic standards. This method was further extended to screen for other potential PAs in the honey using multi-target screening combined with a quantitative prediction model in the cases that authentic PAs were not available. The prediction model was subsequently validated by cross-validation and additional PAs standards which were not included in the model. The maximum concentration prediction error was 50.8%.

1. Introduction

Pyrrolizidine alkaloids (PAs) are a group of at least 600 different secondary metabolites that are produced by more than 6000 plant species worldwide. They are suspected to be hepatotoxic and genotoxic carcinogens in humans, since these PAs are activated to electrophilic compounds through the liver's metabolic enzymes (Bolechová; Čáslavský, Pospíchalová, & Kosubová, 2015:; Lucchetti et al., 2016). Single or continuous intake of 1, 2-unsaturated PAs can cause hemorrhagic necrosis and hepatic venous occlusion, and also increase the incidence of liver haemangiosarcoma for chronic risk assessment (EFSA, 2017). As approximately 3% of all flowering plants contain PAs compounds (Boppré, 2011), this PAs exposure occurs worldwide, and intoxication by contaminated herbal teas, cereals, herbal medicine, salads, and feed, have been reported (Culvenor, Edgar, Smith, Kumana, & Lin, 1986; Molyneux, Gardner, Colegate, & Edgar, 2011; Tandon, Tandon, & Narndranathan, 1976; Wiedenfeld, 2011). Honey is one of the food products that contain PAs. Bees collect pollen or floral nectar from a wide variety of plants and therefore diverse profiles of PAs are observed in honey (Boppré, Colegate, & Edgar, 2005; Lucchetti et al., 2016). Previous studies have shown a very high frequency of PA positive honey samples, e.g., up to 94.2% samples contain PAs in a study of Spanish raw honey (n = 103) (Martinello, Cristofoli, Gallina, & Mutinelli, 2014; Orantes-Bermejo, Serra Bonvehí, Gómez-Pajuelo, Megías, & Torres, 2013). Hence with regard to the increased awareness of this potential health concern, it is important to clarify the types and concentration of PAs in honey.

Based on the data collected by the European Food Safety Authority (EFSA), a total of 64 PAs were reported in food products (EFSA, 2011), while the lack of certified standards hampers an unequivocal quantification. This is, particularly, a problem in PA analysis, where a complex and variable profile of PAs can be expected. To date, very limited PA are commercially available and no more than approximately 34 different PAs in one method are reported, while isotopic labeling PAs are still not commercially available (Schenk, Siewert, Toff, & Drewe, 2017). This is unlikely to change in the near future since there is no facile chemical access, *in vitro* biosynthesis and isolation (EFSA, 2011). Therefore measuring the concentrations of a wide variety of PAs within a single experimental detection continues to remain challenging. Many studies about semi-quantification model have been reported, and the ionization efficiency were reported to be correlated to p*K*a, logP, molecular volume, molecular size, surface tension, or the WAPS value as developed by Kruve's group, (Alymatiri, Kouskoura, & Markopoulou, 2015; Hermans, Ongay, Markov, & Bischoff, 2017; Golubović, Birkemeyer, Protić, & Otašević, 2016; Raji, et al., 2009; Oss, Kruve, Herodes, & Leito, 2010; Kruve & Kaupmees, 2017). However, their models either used for specific compounds (such as amino acid, steroids, peptides and sartans) or challenged to calculate the relevant descriptors used in the model. To our knowledge, there are still no quantitative prediction model developed specific for PAs up to now.

* Corresponding author.
E-mail address: smeds@food.dtu.dk (J. Smedsgaard).

https://doi.org/10.1016/j.foodcont.2018.11.033
Received 17 August 2018; Received in revised form 19 October 2018; Accepted 12 November 2018
Available online 19 November 2018
0956-7135/

Table 1
Information of all the tested honey samples: places of production, origins and symbols used in the text.

No	Geographical origin	Botanical origin	PAs (μg/kg)
H1	Spain	Mandarin honey	–
H2	Romania	Creamy flowers honey	Echimidine (1.4)
H3	Mixture from outside the EU	Not clear	Intermedine (< LOQ), Rinderine//Echinatine/Indicine (ca. 6.1)
H4	Argentina and Ukraine	Not clear	Echimidine (3.4), Senecionine (< LOQ)
H5	China	Mountain honey (many wild flowers)	Integerrimine/Senecivernine (ca.1.6), Trichodesmine (ca. 33.9)
H6	Mexico	Not clear	Intermedine (3.5), Lycopsamine (1.7), Retrorsine (4.3), Rinderine//Echinatine/Indicine (ca.16.4)
H7	Brazil	Not clear	Echimidine (1.6), Intermedine (2.3), Lycopsamine (2.2), Retrorsine (1.6), Senecionine (2.4), Rinderine//Echinatine/Indicine (ca.16.5), Integerrimine/Senecivernine (ca.9.5)
H8	Denmark	Clover and rape	Crotananine (~51.0)
H9	Denmark	Heather flowers honey	Senecionine (< LOQ), Crotananine (ca.24.7), 3′-acetylindicine (ca.45.4), Acetyl-lycopsamine (ca.46.8)
H10	Denmark	Fruit trees (apple, plum, rape)	Crotananine (~14.2)
H11	Denmark	Summer flowers and white clover honey	Lycopsamine (< LOQ), Myoscorpine *N*-oxide/Echiumine *N*-oxide/Echiupinine-*N*-oxide (ca. 2.7–3.2), Rinderine//Echinatine/Indicine (ca.10.3), Crotananine (ca.7.8), 3′-acetylindicine (ca.8.8),Acetyl-lycopsamine (ca.6.7)
H12	Brazil	Not clear	Retrorsine-*N*-oxide (< LOQ)
H13	Nicaragua	Mountain honey (many wild flowers)	Crotananine (ca.5.2)
H14	Ukraine	Fruit trees	Echimidine (< LOQ)
H15	Mixture from countries inside and outside the EU	Not clear	Intermedine (8.4), Lycopsamine (2.0), Rinderine//Echinatine/Indicine (ca.56.4),
H16	Ukraine	Not clear	Echimidine (2.2)
H17	Mixture from outside the EU	Not clear	Crotananine (ca.3.4)
H18	South America	Not clear	Intermedine (8.0), Lycopsamine (0.8), Rinderine//Echinatine/Indicine (ca.31.7)
H19	Denmark	Fruit trees	Crotananine (ca.7.7), Rinderine//Echinatine/Indicine (ca.1.1)
H20	Ukraine	Not clear	Echimidine (2.0), Crotananine (~2.0)
H21	Mixture from countries inside and outside the EU	Acacia	Riddelline/Jacozine (ca.1.5–4.9)
H22	Mixture from outside the EU	Not clear	Retrorsine (8.2), Crotananine (ca. 2.6), Rinderine//Echinatine/Indicine (ca. 2.5)
H23	Mixture from outside the EU	Not clear	Intermedine (< LOQ), Retrorsine (8.2), Retrorsine-*N*-oxide (< LOQ), Crotananine (ca. 4.4), Rinderine//Echinatine/Indicine (ca. 3.6)
H24	Mixture from outside the EU	Acacia honey with hazelnuts	Crotananine (ca. 1.8)
H25	Argentina and Ukraine	Not clear	Echimidine (< LOQ), Lycopsamine (< LOQ),
H26	China	Acacia	–
H27	Mixture from outside the EU	Not clear	Intermedine (11.1), Lycopsamine (3.1)
H28	Denmark	Not clear	Crotananine (ca. 5.3), Riddelline/Jacozine (ca. 2.3–7.5), Rinderine//Echinatine/Indicine (ca. 1.5)
H29	Mixture from countries inside and outside the EU	Wild flower	Echimidine (2.7), Lycopsamine (< LOQ), Crotananine (ca. 2.1)
H30	Mixture from countries inside and outside the EU	Acacia	–
H31	Mixture from outside the EU	Not clear	Echimidine (1.6), Intermedine (1.4), Lycopsamine (0.7), Crotananine (ca. 12.8), Rinderine//Echinatine/Indicine (ca. 3.3)
H32	Mixture from outside the EU	Not clear	–

Solid-phase extraction (SPE), followed by HPLC-MS analysis is a commonly used sequence to analyze PAs in honey. However, the recovery of PAs from the samples are highly dependent on the types of cartridges, sample loading amount, and elutes used during the extraction procedures (Betteridge, Cao, & Colegate, 2005; Bodi et al., 2014; Chung & Lam, 2018; Colegate, Edgar, Knill, & Lee, 2005). The determination and quantification are generally carried out by a target analysis using GC-MS (with and without derivatization), HPLC-MS or HPLC(GC)-MS/MS (Crews, Berthiller, & Krska, 2010; Lucchetti et al., 2016; Molyneux et al., 2011). Latest LC-MS methods mostly relied on the high selectivity of mass spectrometry operated in the multiple reaction-monitoring (MRM) mode (Chung & Lam, 2018; Kowalczyk & Kwiatek, 2018), where the MRM method can only detect preselected PAs within the correct retention time windows and fragmentation patterns (These, Bodi, Ronczka, Lahrssen-wiederholt, & Preiss-weigert, 2013). Furthermore, the PAs not included in the specific methods, e.g. if authentic standards are not available, will remain undetected and therefore not be quantified.

A screening approach (multi-target screening) using high resolution/accurate LC-QTOF-MS with full scan acquisition can achieve both high sensitivity and selectivity. Full-scan MS offer the advantage to determine a virtually unlimited number of analytes simultaneously without preselection of precursor ions. The screening approach also allows a retrospective "targeted" evaluation of data by reconstructing any desired ion chromatogram to explore the data for presence of additional PAs (Heeft et al., 2009). These data are however qualitative in character as standards with known concentration is needed to determine the concentration of compounds.

The objective of this study is to develop a high-resolution analytical method to detect and estimate the concentrations of a wide range of PAs in honey, even for those without authentic PAs standards. Firstly, we developed a validated method that allows the simultaneous detection and quantification of 12 PAs in honey using HPLC-QTOF-MS with pure authentic standards. Secondly, the extracts of different honey samples were analyzed for occurrence of other PAs using a multi-target screening approach. Suspicious signals were verified by recording of product ions using a broadband fragmentation. Consequently, we proposed a semi-quantification strategy for those detected peaks without available standards through a quantitative prediction model (QPM). To evaluate this prediction model, feasibility and reliability have been validated by leave-one-out (LOO) cross validation. In addition, as a test, quantification of four unknown PAs, that were not included in the original dataset, has also been carried out. Finally, we estimated the contents of PAs in honey samples combining the available reference standards and the QPM.

2. Materials and methods

2.1. Chemicals and reagents

Milli-Q grade water was used for all analyses (Millipore Corporation, Molsheim, France). Methanol (LC–MS grade), acetonitrile (LC–MS grade), ammonium hydroxide solution (LC–MS grade), and formic acid (LC–MS grade) were purchased from Fluka (Steinheim, Germany). Pyrrolizidine alkaloids (PAs) (7-acetylintermedine, echimidine, heliotrine, heliotrine *N*-oxide, intermedine, intermedine *N*-oxide, lasiocarpine, lasiocarpine *N*-oxide, lycopsamine, lycopsamine *N*-oxide, retrorsine, retrorsine N-oxide, senecionine, senecionine *N*-oxide, seneciphylline, seneciphylline *N*-oxide) were purchased from PhytoLab (PhytoLab GmbH & Co. KG, Vestenbergsgreuth, Germany). Sulfuric acid (analytical grade, 95–97%) was obtained from Merck (Darmstadt, Germany). All other solvents used in this study were of analytical grade.

Each alkaloid standard was dissolved in either acetonitrile or acetonitrile/water (50/50 v/v) depending on their solubility to obtain stock solutions at a concentration of 0.5 mg/mL. All the stock solutions were stored at −20 °C. Stock standard mixtures (1 μg/mL) were prepared in acetonitrile/water (50/50 v/v) from the individual stock solutions, and used set up calibration curves.

2.2. Samples

Thirty-two honey samples were included in this study purchased from either local supermarkets or bee keepers, originating from at least 10 different countries. The honey samples were declared to be collected from summer flowers, white clover, wild flowers, creamy flowers, heather flowers, acacia, fruit trees (apple, plum, rape), and etc. The information of samples: names, geographical and botanical origin, and symbols are listed in Table 1.

2.3. Negative control and quality control samples

Honey sample previously analyzed and found not containing PA compounds, was used as blank sample (mountain honey coming from wild flowers of Nicaragua) for method validation. Quality control (QC) samples were made by spiking blank honey samples (25 μL stock standard mixtures to 1.0 g blank honey sample). QC samples were analyzed in each series of samples and were controlled to be in the range of ± 20% of the validation results. If not, the whole set of samples were reanalyzed.

2.4. Solid phase extraction

The strong cation exchange (SCX) SPE was found to be suitable for plant extracts and for a series of pyrrolizidine alkaloid/N-oxide (PAs/PANOs) representing the major hepatotoxic structural groups (monoesters, diesters and macrocyclic diesters of retronecine, heliotridine and otonecine bases), and also provided efficient capture from honey-based matrices and subsequent release into ammoniated methanol (Betteridge et al., 2005; Colegate et al., 2005). The SPE columns (150 mg/6 mL, Biotage Evolute Express CX) used in our study combines non-polar and SCX functionality for extraction of basic analytes from aqueous samples through a dual, mixed-mode mechanism. Also these specific SPE columns can be used without a preanalysis conditioning and equilibration steps without affecting the recovery significantly.

The extraction of PAs was based on a modified method from German Federal Institute of Risk Assessment (BfR) (BfR, 2013). One gram of honey was weighed into a 10 mL centrifuge tube and dissolved in 3 mL of 0.05 M H_2SO_4 in water at room temperature by shaking for 30 min. The samples were centrifuged at 2000 × *g* for 10 min at room temperature. The supernatant was transferred to a SPE cartridge by using a vacuum manifold. After loading the sample, the column was washed with 4 mL of 0.05 M H_2SO_4 followed by 4 mL of methanol, then the column was eluted with 5 mL of 2.5% (v/v) ammonium hydroxide in methanol at a rate of 0.5–1 mL/min. The eluate was dried under a stream of nitrogen at 45 °C. The residue was dissolved in 0.5 mL acetonitrile/water (5/95, v/v), transferred to an Eppendorf vial, and centrifuged at 10.000 × *g* for 10 min at room temperature. Then 200 μL of supernatant was transferred into a HPLC-vial with insert.

2.5. HPLC-ESI-QTOF-MS

Samples were analyzed on an Agilent 1200 HPLC (Agilent Technologies, CA, U.S.) coupled to a Bruker Daltonics micro-TOFq mass spectrometer (Bruker Daltonics, Bremen, Germany). The resolution of the Bruker micro-TOFq is approximately 10000. Samples were injected onto a Nucleoshell C18 reversed-phase column (2.7 μm, 100 × 2 mm, MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). The mobile phase consisted of water (A) containing 0.1% formic acid and 2.5 mM ammonium hydroxide, and acetonitrile (B). The analysis was done using a gradient elution at a flow rate of 0.3 mL/min at 40 °C using the gradient: isocratic at 2% of B for 1.0 min, then increased to 10% of B at 4.0 min, to 40% of B at 9 min, and finally to 95% of B at 10.0 min and maintained to 11 min, where after the gradient returned to starting conditions 2% of B at 11.2 min followed by 3.8 min equilibration with 2% of B to giving a total runtime of 15 min.

An electrospray ion source (ESI) was used in positive ion mode (ESI^+) with the following settings: capillary voltage 4 kV, nebulizer pressure 2 bar, dry gas temperature 180 °C, and dry gas flow rate 10 L/min. Spectra were collected from m/z 50–800 Da at a rate of 2 Hz. Sodium formate dissolved in 50% of 2-propanol was introduced into the ion source between 0.2 and 0.4 min for internal calibration of the data files. Hexakis (1H, 1H, 2H-perfluoroethoxy) phosphazene was used as the lock mass calibrant. To confirm the presence of pyrrolizidine alkaloids, LC-MS/MS analyses were conducted using broadband collision-induced dissociation (bbCID) with a fragmentation energy of 35 eV and nitrogen as the collision gas.

2.6. Design of validation

Validation parameters such as linearity, specificity, limits of detection (LOD), limit of quantification (LOQ), measurement uncertainty, recovery, and precision were determined for 12 PAs using the established method according to Nordic Committee on Food Analysis (NMKL) procedure No. 4 with minor modification (NMKL, 2009). The spiked samples were prepared according to the procedure described above and analyzed on three different days. All 12 PA compounds were spiked at 4 concentration levels: 2 ng/mL, 10 ng/mL, 50 ng/mL, and 150 ng/mL corresponding to 1 μg/kg, 5 μg/kg, 25 μg/kg, and 75 μg/kg in the honey samples. At each concentration level at least 10 replicates of the same spiked blank honey samples were analyzed (13 replicates at 10 ng/mL, and 10 replicates for the other 3 levels). Method precision was evaluated by intra- and inter-day analyses. The repeatability (intra-day variability) was tested within one day for each concentration, reproducibility (inter-day variability) was examined on three different days for each concentration (13 replicates for 5 μg/kg, 10 replicates for 1 μg/kg, 25 μg/kg and 75 μg/kg).

LOD and LOQ were calculated by using the standard deviation of 10 replicates at 2 ng/mL (blank honey spiked before SPE clean-up). LOD = 3 × SD/recovery of each compound, LOQ = 6 × SD/recovery of each compound. Matrix effects (ME) was estimated by adding different levels of standard solution to a sample of the blank honey extracts. Matrix match calibration curves were used to correct for matrix effects (ME), and were made by analysis of each analyte at 7 calibration levels from 2 to 150 ng/mL. The standard curves were the peak area versus the concentration of each analyte. ME (%) = (Slope of matrix match standard curve/Slope of standard curve without honey matrix) × 100%. ME of some other honey samples were also measured at 50 ng/mL spike level to compare the ME difference between different

honey samples. The recovery was determined by comparing the analytical results of the extracted PAs from spike honey samples with the results of standards added in the same concentrations to blank matrix (13 replicates for 5 μg/kg, 10 replicates for 1 μg/kg, 25 μg/kg and 75 μg/kg).

2.7. Quantification

All the quantification was performed by integration of peak area of extracted ion chromatograms using exact mass ± 5 mDa. The concentration of PAs in the samples was calculated by QuantAnalysis 2.2 (Bruker Daltonics, Bremen, Germany). All the values of PAs of samples were calculated by using the average value in duplicate samples of the same sequence.

2.8. Software

The Chemdraw Professional 15.0 was used for drawing and displaying chemical structures. R (version 3.3.3), R studio and Origin software were used for all statistical analysis. Jchem version 18.5.0 was used to predict the pK_a, pK_b, logP values, and other properties of molecules.

2.9. Statistical model evaluation and validation

Several studies have investigated the relationship between signal response in electrospray ionization-mass spectrometry (ESI-MS) and molecular descriptors (molecular volume, logP, pK_a, polarity) of analytes to estimate analyte concentration (Golubović, Birkemeyer, Protić, Otašević, & ečević, 2016; Kruve & Kaupmees, 2017; Oss et al., 2010). Seneciphylline was randomly chosen as the reference standard in the method to correct the daily variation. Multi-linear regression (MLR) analysis was carried out to obtain the QPM describing the relationship between relative calibration curves' slope (RS) corresponding to seneciphylline in honey matrix and molecular properties in the following formula:

$$\ln \mathrm{RS} = \ln\left(\frac{Slope\ of analytes}{Slope\ of\ Seneciphylline}\right) = \sum_{i=1}^{n} a_i X_i$$

where X_i is the physicochemical parameters or their combinations, and a_i is the coefficients of these parameters. Experimental values of several parameters such as logP and pK_a are not available for all compounds, therefore these molecular descriptors including Molecular volume, logP, pK_a (strongest acidic), pK_b (strongest basic), and polarizability were computed by using Physico-chemical property predictors plugin of JChem in excel (Szisz, 2017). All the chemical structures of PAs were transferred as the SMILES (Simplified molecular input line entry specification) obtained from PubChem. The molecular descriptors with correlation above 0.9 were removed to avoid the co-linearity in the model. X_i and a_i were obtained by MLR which was calculated in R. Twelve PAs compounds were used to develop the MLR model (exact mass is in the range from 299 to 427 Da, pK_a is in the range from 10.54 to 12.37, pK_b is in the range from 2.82 to 8.13, logP is in the range from −1.42 to 1.64). In the gradient elution, all the PAs were eluted from 10% to 37% organic phase, which is considered as a narrow range, thus the corrections with retention parameters or solvent change was ignored in this part.

Cross-validation method 'leave-one-out' (LOO) approach was used to validate the obtained results. LOO means that each compound was left out from the model fitting process once; thereafter, the model was used to predict the RS value of the compound not involved in the model development. LOO is equivalent to k-fold cross-validation where k = n (the number of observations in the data set), and LOO for linear regression is exactly equivalent to the PRESS (Predicted Residual Sum of Squares) method based on the leave-one-out or Jackknife technique, detail explanation can be found in Tarpey's study (Tarpey, 2000). Therefore in this study, PRESS and root mean square error of prediction (RMSEP) were used to describe the differences between predicted RS values and measured values.

Using the fact of the predicted value for y_i when the i^{th} compound is deleted from the MLR model is equal to

$$y_i - \hat{y}_{(i)} = \frac{\hat{e}_i}{(1-h_i)} = \hat{e}_{(i)}$$

Here $\hat{e}_i$ = *usual* i^{th} *residual* and h_i = potential or leverage value for the i^{th} compound, y_i and $\hat{y}_{(i)}$ are predicted and measured lnRS value in this study, respectively.

$$PRESS = \sum_{i=1}^{n} (y_i - \hat{y}_{(i)})^2$$

$$\mathrm{RMSEP} = \sqrt{\frac{PRESS}{n}}$$

The resulting model was also used to predict the slope of the calibration curves for 4 additional PAs which were not included in the original dataset but within the model coverage to verify the prediction quality. The predicted slopes were used to calculate the concentrations of the other PAs compounds detected in the collected samples.

$$C_{\mathrm{predicted}} = \frac{\mathrm{Response}_{from\ sample}}{\mathrm{Slope}_{predicted}}$$

$C_{predicted}$ is the predicted PA concentration, $Response_{from\ sample}$ is the peak area of the tentatively identified PAs in honey samples, $Slope_{predicted}$ is the slope of predicted calibration curve of the identified PAs, and it was calculated from the RS value corresponding to the slope value of reference standard (seneciphylline) in honey matrix, assuming that the standard curve pass through zero.

All statistical tests were performed at 95% significant level. The prediction error for additional 4 PAs out of the model was calculated with the following formula:

$$\text{Prediction error} = \left(\frac{\mathrm{RS}_{predicted} - \mathrm{RS}_{measured}}{\mathrm{RS}_{measured}}\right) \times 100\%$$

3. Results and discussion

3.1. PART 1. validation

3.1.1. Specificity

Blank honey samples and blank honey spiked with 5 ng/mL mix standard solution (corresponding to 2.5 μg/kg) were analyzed to examine if other compounds in the samples were interfering with these analytes. The HPLC shows good separation of all 12 PA compounds, and there are no apparent interferences observed from the matrix evaluated from the extracted ion chromatograms, thus confirming the specificity. The structures and chromatogram of 12 PAs are shown in Fig. 1.

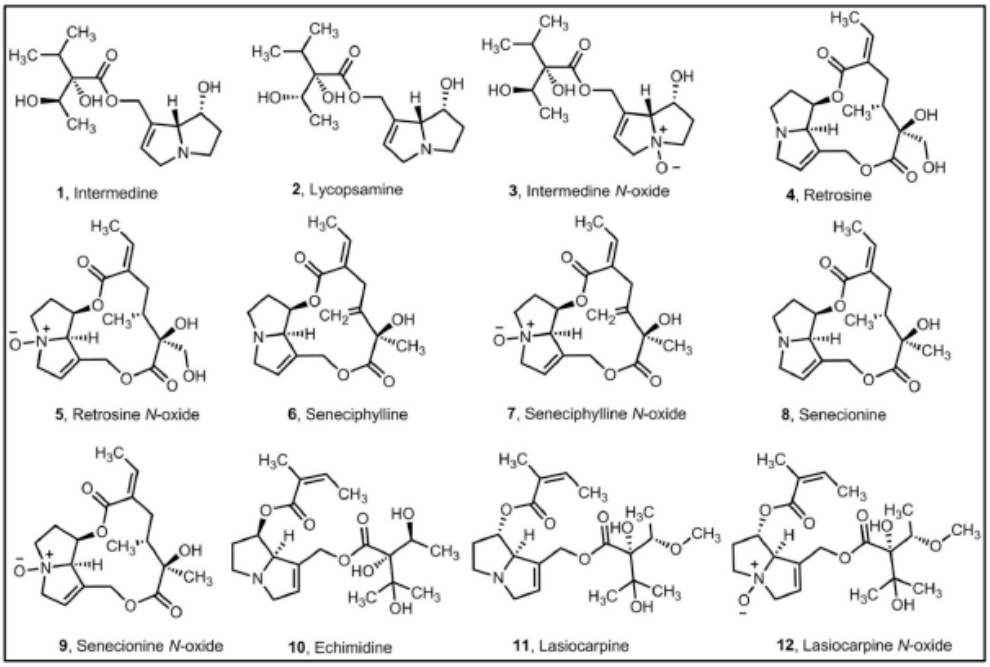
3.1.2. Linearity

A good linear relationship was observed based on correlation coefficients higher than 0.99 (from 2 ng/mL to 150 ng/mL). The average relative residual is less than 25% at all levels, and the intercepts obtained are close to the theoretical value zero, demonstrating good consistency. Information about residues plot can be found in Supporting Information Figure S1.

3.1.3. Precision

The standard deviations of all the PAs at different spiked levels were calculated. Results are shown in Table 2. Coefficient of Variation (CV%) for the reproducibility is in the range from 6.3% to 21.6%, and CV% for the repeatability is in the range from 2.6% to 18.9%, which is

(a)



(b)

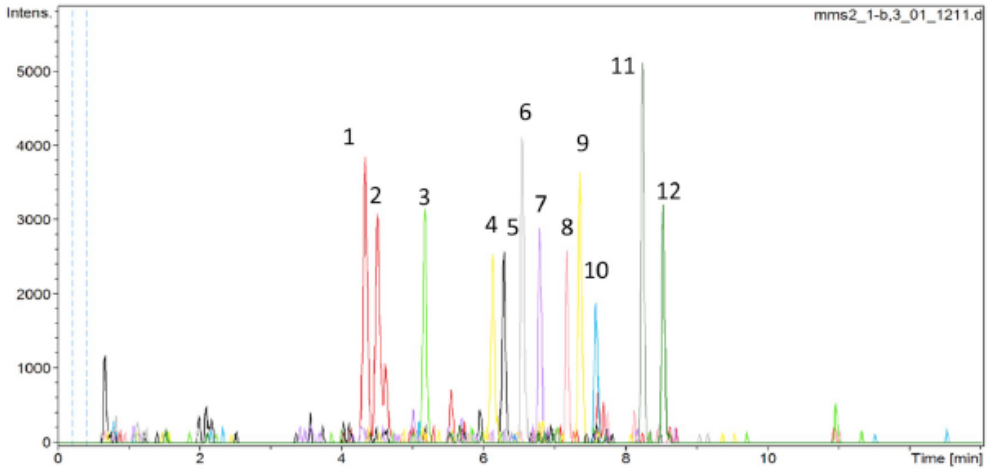


Fig. 1. Structures and chromatogram of 12 PAs (a) Chemical structures of 12 PAs. (b) Chromatogram of 12 PAs in 5 ng/mL spiked blank honey sample: 1. Intermedine, 2. Lycopsamine, 3. Intermedine *N*-oxide, 4. Retrosine, 5. Retrosine *N*-oxide, 6. Seneciphylline, 7. Seneciphylline *N*-oxide, 8. Senecionine, 9. Senecionine *N*-oxide, 10. Echimidine, 11. Lasiocarpine, 12. Lasiocarpine *N*-oxide.

considered acceptable.

3.1.4. Recovery

The average recovery is 89% for all the 12 PA compounds, ranging from 79.2% to 104.4% with a standard deviation lower than 7.6% (Table 2), thus confirming a good reproducibility of the method and sufficient for reliable analysis.

3.1.5. Limit of detection and limit of quantification

Results are shown in Table 2. LOD values for the 12 PA compounds are in the range from 0.2 μg/kg to 0.6 μg/kg, LOQ values of the 12 PA compounds are in the range from 0.5 μg/kg to 1.3 μg/kg, which are comparable to those reported in literature (Bolechová, Čáslavský, Pospíchalová, & Kosubová, 2015; EFSA, 2011). Although there are studies which present methods with lower LOQs (BfR, 2013; Kowalczyk & Kwiatek, 2018; Lucatello, Merlanti, Rossi, Montesissa, & Capolongo, 2016), they all rely on MRM mode rather than full scan detection. These MRM method are very specific however MRM are blind to other PAs that may occur in samples. **Figure** S2 shows samples spiked with 1 ng/mL (corresponding to −0.5 μg/kg in honey) with peaks still clearly visible for all analytes, indicating that the calculated limits are realistic to detect and quantify for the 12 PA compounds.

Table 2
Validation results.

Compounds	Conc		Inter-day reproducibility			Intra-day Repeatability			Recovery	LOD	LOQ	ME
			mean	std	CV %	mean	std	CV %	%	μg/kg		
Echimidine	1.00	μg/kg	0.89	0.19	21.6	0.89	0.17	18.8	93.5	0.54	1.07	101.92
	5.00	μg/kg	5.22	0.64	12.3	5.22	0.52	10.0	104.4			
	25.00	μg/kg	24.20	2.88	11.9	24.20	2.08	8.6	96.8			
	75.00	μg/kg	67.96	5.80	8.5	67.96	1.80	2.6	90.6			
Intermedine	1.00	μg/kg	0.81	0.08	10.3	0.81	0.07	8.0	89.4	0.41	0.81	90.50
	5.00	μg/kg	4.06	0.69	17.0	4.06	0.46	11.3	81.2			
	25.00	μg/kg	24.21	2.12	8.8	24.21	0.81	3.4	96.8			
	75.00	μg/kg	72.45	7.74	10.7	72.45	3.38	4.7	96.6			
Intermedine-*N*-oxide	1.00	μg/kg	0.96	0.16	16.9	0.96	0.16	16.9	95.6	0.52	1.04	94.05
	5.00	μg/kg	4.01	0.76	18.9	4.01	0.76	18.9	80.2			
	25.00	μg/kg	22.77	2.53	11.1	22.77	1.48	6.5	91.1			
	75.00	μg/kg	72.71	7.53	10.4	72.71	4.54	6.2	96.9			
Lasiocarpine	1.00	μg/kg	0.81	0.11	13.4	0.81	0.07	8.6	80.5	0.22	0.45	112.28
	5.00	μg/kg	4.19	0.59	14.1	4.19	0.24	5.8	83.8			
	25.00	μg/kg	23.57	3.20	13.6	23.57	1.30	5.5	94.3			
	75.00	μg/kg	68.47	8.95	13.1	68.47	3.65	5.3	91.3			
Lasiocarpine-*N*-oxide	1.00	μg/kg	0.81	0.15	18.6	0.81	0.15	18.6	81.4	0.60	1.20	97.10
	5.00	μg/kg	4.23	0.57	13.5	4.23	0.41	9.6	84.7			
	25.00	μg/kg	22.42	2.14	9.5	22.42	1.16	5.2	89.7			
	75.00	μg/kg	68.65	4.58	6.7	68.65	2.91	4.2	91.5			
Lycopsamine	1.00	μg/kg	0.83	0.10	12.3	0.83	0.08	9.4	82.7	0.28	0.56	90.68
	5.00	μg/kg	4.65	0.61	13.1	4.65	0.54	11.5	93.0			
	25.00	μg/kg	22.68	3.86	17.0	22.68	0.81	3.6	90.7			
	75.00	μg/kg	71.06	8.94	12.6	71.06	1.87	2.6	94.7			
Retrorsine	1.00	μg/kg	0.96	0.14	15.1	0.96	0.14	15.1	95.5	0.34	0.67	91.55
	5.00	μg/kg	4.66	0.93	20.0	4.66	0.62	13.3	93.1			
	25.00	μg/kg	22.94	3.05	13.3	22.94	2.59	11.3	91.8			
	75.00	μg/kg	69.08	6.10	8.8	69.08	2.35	3.4	92.1			
Retrorsine-*N*-oxide	1.00	μg/kg	0.79	0.10	12.9	0.79	0.10	12.9	79.2	0.64	1.27	81.01
	5.00	μg/kg	4.31	0.77	17.9	4.31	0.77	17.9	86.2			
	25.00	μg/kg	23.69	3.42	14.4	23.69	1.47	6.2	94.8			
	75.00	μg/kg	70.80	9.91	14.0	70.80	3.47	4.9	94.4			
Senecionine	1.00	μg/kg	0.81	0.13	16.5	0.81	0.13	16.5	81.4	0.45	0.91	82.95
	5.00	μg/kg	4.19	0.83	19.9	4.19	0.76	18.2	83.9			
	25.00	μg/kg	22.91	2.92	12.7	22.91	1.03	4.5	91.6			
	75.00	μg/kg	69.81	11.96	17.1	69.81	5.01	7.2	93.1			
Senecionine-*N*-oxide	1.00	μg/kg	0.80	0.14	17.8	0.80	0.14	17.8	79.8	0.42	0.84	99.93
	5.00	μg/kg	4.18	0.81	19.5	4.18	0.52	12.5	83.5			
	25.00	μg/kg	22.58	2.66	11.8	22.58	1.98	8.8	90.3			
	75.00	μg/kg	68.89	6.05	8.8	68.89	3.41	5.0	91.9			
Seneciphylline	1.00	μg/kg	0.79	0.05	6.5	0.79	0.05	6.5	79.2	0.23	0.45	81.55
	5.00	μg/kg	4.12	0.63	15.2	4.12	0.52	12.7	82.4			
	25.00	μg/kg	22.20	2.18	9.8	22.20	1.46	6.6	88.8			
	75.00	μg/kg	67.64	4.50	6.6	67.64	2.57	3.8	90.2			
Seneciphylline-*N*-oxide	1.00	μg/kg	0.84	0.15	17.5	0.84	0.13	15.4	84.4	0.60	1.19	88.53
	5.00	μg/kg	4.17	0.52	12.5	4.17	0.52	12.5	83.5			
	25.00	μg/kg	19.80	2.26	11.4	19.80	1.31	6.6	79.2			
	75.00	μg/kg	61.75	3.90	6.3	61.75	2.70	4.4	82.3			

3.1.6. Matrix effect

The matrix effect (ME) of spiked blank honey sample with 1 g sample load is in the range from 81.0% to 112.3%, indicating that the analytes' signals are only slightly influenced by the honey matrix (Table 2). In order to compare the difference of ME between different honey matrix, additional experiments were done to compare the ME. Examining these results showed a RSD% in the range from 4% to 12% from ME of 8 honey samples, which indicates that it is feasible to use one blank matrix to quantify all the other honey samples. Full data are shown in Supporting Information Table S1.

3.1.7. Measurement uncertainty

The measurement uncertainty (MU) consists of intra-laboratory reproducibility ("CVR"), bias ("bias") calculated as 100% recovery, standard deviation on bias ("CVRbias/sqrt(n)") calculated as intra-laboratory standard deviation divided by square root of number of results, and standard deviation on matrix effect ("CVmatrix effect") listed in Table S1 in the Supporting Information. The method uncertainty was calculated as:

$$MU = \sqrt{CVR^2 + bias^2 + \frac{CVRbias^2}{n} + CVRmatrixeffect^2}$$

Seneciphylline-N-oxide, with the highest bias (100–82.4% = 17.6%) had a MU of 22% ($\sqrt{11.9^2 + 17.6^2 + \frac{11.9^2}{43} + 7.1^2}$).

3.2. 2. PA detection and contents in honey samples

3.2.1. PA quantification and identification

The 12 PAs included as standards in this study were selected based on the ones reported most often and in highest concentrations in honey (EFSA, 2011), and they belong to 3 major PAs chemical groups found in honey:

- **Senecionine-type** PAs (Retrorsine, retrorsine *N*-oxide, senecionine, senecionine *N*-oxide, seneciphylline, seneciphylline *N*-oxide), which occur primarily in the *Senecioneae* spp. (*Asteraceae* family), but are also found in *Crotalaria* spp. (*Fabaceae* family);

Table 3
Molecular ions and selected MS/MS fragment ions of major PAs.

RT	PAs	Formula	$[M+H]^+$ calculated	$[M+H]^+$ experimental	Fragment pattern and abundancy	Error (ppm)
7.6	Echimidine	$C_{20}H_{31}NO_7$	398.2173	398.2170	120.0810 (48), 220.1301 (2)	−0.75
4.3	Intermedine	$C_{15}H_{25}NO_5$	300.1805	300.1800	138.0916 (33), 156.1021(100)	−1.66
5.2	Intermedine *N*-oxide	$C_{15}H_{25}NO_6$	316.1755	316.1751	138.0914(42), 172.0961(100)	−1.27
8.2	Lasiocarpine	$C_{21}H_{33}NO_7$	412.2330	412.2333	120.0815(100), 238.1427(5)	−0.73
8.5	Lasiocarpine *N*-oxide	$C_{21}H_{33}NO_8$	428.2279	428.2262	94.0651(19), 120.0806 (73), 138.0908 (51)	−3.96
4.5	Lycopsamine	$C_{15}H_{25}NO_5$	300.1805	300.1805	138.0917(28), 156.1020(100)	0
6.1	Retrorsine	$C_{18}H_{25}NO_6$	352.1755	352.1744	120.0822 (100), 138.0894 (58)	−3.12
6.3	Retrorsine-*N*-oxide	$C_{18}H_{25}NO_7$	368.1704	368.1707	118.0648(100), 136.0765 (52)	0.81
7.2	Senecionine	$C_{18}H_{25}NO_5$	336.1805	336.1792	120.0810 (98) 138,0906 (33)	−3.87
7.4	Senecionine *N*-oxide	$C_{18}H_{25}NO_6$	352.1755	352.1753	94.0645(26), 118.0649 (94), 136.0741(77)	−0.56
6.5	Seneciphylline	$C_{18}H_{23}NO_5$	334.1649	334.1655	120.1818(100), 138.0911(66)	1.79
6.8	Seneciphylline *N*-oxide	$C_{18}H_{23}NO_6$	350.1598	350.1595	120.0827 (100), 136.0748 (41)	−0.85
Tentative identified PAs						
7.8	3′-acetylindicine	$C_{17}H_{27}NO_6$	342.1911	342.1910	94.0699 (7), 121.0680 (81), 120.0813 (25)	−0.29
8.1	Acetyl-lycopsamine	$C_{17}H_{27}NO_6$	342.1911	342.1911	120.0791 (46), 156.0903 (8)	0
7.1	Crotanaline	$C_{17}H_{25}NO_5$	324.1805	324.1793	120.0768(92), 138.0907 (18)	−3.70
4.6	Echinatine/Indicine/Rinderine	$C_{15}H_{25}NO_5$	300.1805	300.1795	138.0912 (100). 156.0986 (28)	−1.67
7.1	Integerrimine/Senecivernine	$C_{18}H_{25}NO_5$	336.1805	336.1813	120.0796 (100), 138.0904 (54)	2.38
6.1	Jacozine/Riddelline	$C_{18}H_{23}NO_6$	350.1598	350.1605	120.0772 (33.9), 138.0963 (6)	2
7.5	Myoscorpine *N*-oxide/Echiumine *N*-oxide	$C_{20}H_{31}NO_7$	398.2173	398.2184	120.0807 (56), 136.0693(2), 172.1412 (15)	3.26
5.2	Trichodesmine	$C_{18}H_{27}NO_6$	354.1911	354.1911	94.0700(17), 120.0810 (100)	0

Note: the corresponding chromatograms of LC-MS/MS are shown in Figure S1 in supporting information.

- **Lycopsamine-type** PAs (Intermedine, intermedine *N*-oxide, lycopsamine, echimidine), which occur in the *Boraginaceae* family and in the *Eupatorieae* spp. (*Asteraceae* family);
- **Heliotrine-type** PAs (lasiocarpine, lasiocarpine *N*-oxide), which occur in *Heliotropium* spp. (*Boraginaceae* family) (Mulder, López, These, & Preiss-weigert, 2015).

The LC-MS/MS chromatograms of 12 validated PAs in honey samples were characterized by comparing the fragments pattern with authentic standards. The fragmentation pattern of selected PAs, C-1 (monoester or open-chained diester), and C-2 (*N*-oxide open-chained diester) are shown inFigure S3 (Avula et al., 2015; These et al., 2013). A compilation of chemical formulas, accurate mass, fragment ions, retention time, and mass error of HRMS detection for 12 reference PAs are shown in Table 3. The MS^2 broadband fragmentation spectrums show fragment ions at m/z 172, 156, 138, 136, 120, 118, and 94 which are the most common and characteristic fragments from these PAs. These fragment ions are also well known in unsaturated necine ester type PAs and are similar to fragments reported in the literature (Avula et al., 2015; Beales, Betteridge, Colegate, & Edgar, 2004; Schulz et al., 2015; These et al., 2013). Detailed information about fragment patterns of PAs found in this study are shown in supplementalFigure S4.

The analyses showed that one or more of the 12 PA compounds were found in 60% of these samples. The combined results are shown in Table 1. The concentration of each PA compound in the positive samples is in the range between 1.4 and 11.1 μg/kg honey and the total PAs contents are in the range of 1.4–14.2 μg/kg (the sum of the 12 PAs). The highest total PA content is found in honey H27.

The most frequently found PA is echimidine. Seven of the tested samples are positive with concentrations ranging from 1.4 to 3.4 μg/kg, followed by intermedine and senecionine. These findings are in accordance with the results reported by Lucatello (et al., 2016), most likely because plants from the genus *Echium*, namely *Echium italicum* L., *Echium vulgare* L., and *Echium plantagineum* L. are widely distributed in the European region (Lucatello, Merlanti, Rossi, Montesissa, 2016). For echimidine in H14, H25 intermedine in H3, H23, retrosine *N*-oxide in H12, H23, lycopsamine in H11, H16, H25, H29, and senecionine in H4 and H9, they are below the LOQ value but are still detected in LC-MS. Information about extracted ion chromatograms of these detected PAs are shown in Figure S5.

Although PAs are detected in almost half of the samples studied, they are not found in as high concentration as in other studies (Avula et al., 2015; Beales et al., 2004). The levels of the 12 validated and quantified PAs are so low that the honey can be accepted according to both $BMDL_{10}$ of EFSA to assess the carcinogenic risks of PAs (237 μg/kg bw per day for chronic toxicity) and also BFR recommendation (maximal intake of 0.007 μg/kg bw per day) if consumers only consume a limited quantity of honey (see Table 1) (Dusemund et al, 2018; EFSA, 2017; BfR, 2013). However, food safety issues frequently originate from compounds that are not included or quantified by the analytical method. Therefore, a multi-target screening method including more PAs compounds is particularly important to predict potential risk from PAs in honey.

3.2.2. PAs screening

A screening method monitoring 85 different PAs compounds based on elemental composition from the literature was used to detect if other PAs could be found in the honey samples. The PAs in the screening list were compiled from literature and only 1, 2-un-saturated compounds were taken into consideration. Then the suspicious signals with same molecule weight were further analyzed by LC-MS/MS. Tentative identification of the PAs was done by comparing the fragments with literature information (EFSA, 2011). Elemental composition and fragments ion of these tentative identified PAs from the screening method are shown in Table 3.

According to the screening results, several other PAs were found in the honey apart from the 12 validated PAs already quantified. Some of these PAs occur in relative high concentration. For example, a chromatographic peak just next to lycopsamine with the same protonated molecular ions (m/z 300.1805 $[M+H]^+$), same fragments pattern, and similar retention time (only 0.1 min difference) is found in 12 out of 32 honey samples. In order to clarify whether this compound is lycopsamine or not, we spiked one of these honey samples with this unknown peak. We proved that this large peak is not lycopsamine, but its peak area is much larger than that of lycopsamine. According to the previous reports and fragments information we obtained (m/z 94.0655, 138.0932. 156.1019), this unknown compound is most likely rinderine (other possibilities are echinatine or indicine), and all of them are

belong to lycopsamine-type PA stereoisomers, which include indicine, intermedine, lycopsamine, rinderine and echinatine (Beales et al., 2004; Kempf et al., 2011). All these five PAs are structurally closely related, thus a similar retention behaviour and mass spectrometric response can be assumed to be the same. There are also other large peaks appearing next to senecionine and echimidine again (0.1 min retention time difference) with the same molecule weight and fragment pattern, but not the same as verified by standard addition.

Several PAs have the similar molecular structures, therefore we will expect similar properties and fragments pattern in LC-MS/MS. For these types of compounds, it is difficult to predict the correct structures without further identification or authentic standard. Therefore, the data can only be seen as indicative. A list of considered PA compounds included in the screening, and their accurate mass data, are shown in Supporting Information Table S2.

To evaluate the concentrations of these other PAs, which were not included in the validation method, a semi-quantification approach was used to estimate their concentration as discussed in the following section.

3.3. 3. QPM and estimation of concentration of other PAs in honey samples

Estimation of the concentration for PAs where standards are unavailable are based on all the information we can obtain from direct measurement of available standards and honey samples, also model calculation. In the initial model, all the previously mentioned molecular properties (molecule volume, pK_a, pK_b, logP, relative slope, exact mass, and retention time of 12 PAs used in the model and tentative identified PAs in samples are shown in Table S3a) were included. While by backward MLR modelling, the best fit for data was achieved with a model containing pK_a, pK_b, logP, and exact mass. Several authors have found that hydrophobicity (logP, usually expressed as octanol/water partition coefficient) are important characteristics correlated to ESI ionization efficiency, and increasing the extent of nonpolar character in a compound leads to enhanced affinity for the surface phase and higher ESI response (Cech & Enke, 2000; Leito et al., 2008). p*K*a can be expressed as the ability to be protonated and become a cation, and it has been proved that p*K*a is highly correlated with ionization efficiency in ESI (Oss et al., 2010). Exact mass was characterized by the molecular size or molecular volume, which also has been reported as a parameter that can influence ionization efficiency (Hermans et al., 2017; Oss et al., 2010).

The formula calculated in R is shown in the following formula (Procedures of calculation is shown in Supporting Information).

$\ln(RS_{predicted}) = (27.0965 \pm 7.6792) + (1.0240 \pm 0.3192) \times \text{logP} - (1.4075 \pm 0.4115) \times \text{p}K_a - (0.3048 \pm 0.1010) \times \text{p}K_b - (0.0285 \pm 0.0073) \times \text{exact mass}$.

Regarding the calculation, eg., rinderine, $\ln(RS_{predicted})$ of rinderine = 27.0965 + 1.0240 × (−0.2900) −1.4075 × 11.3400 – 0.3048 × 7.8200 – 0.0285 × 299.1733 = −0.07149, therefore predicted relative slope of rinderine corresponding to seneciphylline is 0.9310.

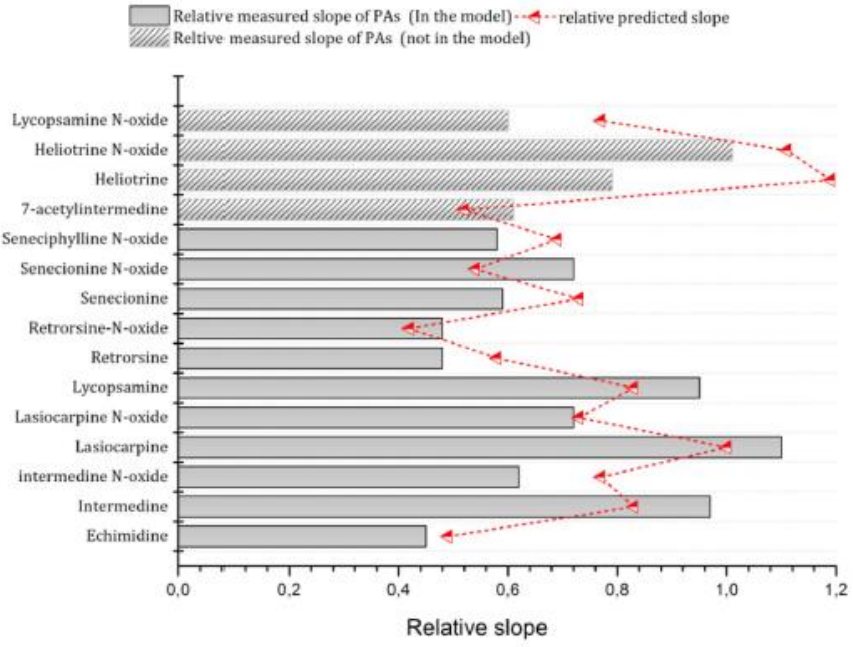
The correlation between predicted RS and measured RS is shown in Fig. 2-a. The R^2 for this model was 0.75 and the residual standard error was 0.21 logarithmic units. By LOO validation, the accuracy of the models can also be described with RMSEP, RMSEP = 0.28 lnRS units. This value shows that on average the mismatch between the predicted and measured RS from LOO is lower than 1.3 times. The prediction errors are in the range from −0.3% to 27.1% for the 12 PA compounds used to develop the model (Table S3-b in Supporting Information). To validate the method, the model was also used to quantify 4 additional PAs compounds available as pure standards but not included in the model above in a spiking experiment. The concentrations of these 4 authentic standard compounds were estimated using the model developed by the 12 PAs giving a prediction error in the range between 10.2% and 50.8%. Fig. 2-b shows the difference in standard curve slope using the predicted concentration and measured concentration of the PAs not included in the model (from 5 to 100 ng/mL). Therefore, a predicting model based on a 12 PAs calibration sets provides acceptable quantification results for a risk assessment. Despite some PAs being of different chemical structures, they share quite similar physicochemical properties, hence this will not have large effect on the predicted concentration. We also calculated the model based on the data we obtained from two different days; and the coefficients in the formula are quite similar as shown in Table S3-b. The difference between the two formulas obtained from two different experimental days was compared; the variation for all compounds is within a range of −24.5% to 10.9%. This indicates the feasibility for the future study, to use only one model and one reference PA standard (e.g., Seneciphylline in this study) to predict the concentrations of the other PAs within maximum 50.8% prediction error; as long as we have the structures information of these PAs.

In many cases, studies about PAs estimation may choose calibration alkaloids of similar structure using different calibration alkaloids (often an isomer) (Hoogenboom et al., 2011; These et al., 2013). Or just simply reported all alkaloids as equivalent of one alkaloid such as echimidine in Beales' study (Beales et al., 2004). In the approach of PAs estimation with different calibration alkaloids, serveral PAs standards are still necessary. They cover various PAs types, while in the approach of PAs estimation, with only one alkaloid, it is more unpersuasive due to the variation of PAs response which can be nearly 3 times based on our results. Hereby, the difference between the approach of PAs estimation, with only one alkaloid and our prediction model, were compared. It is shown in Table S4 in Supporting Information, 3 concentrations (5 ng/mL, 10 ng/mL, 25 ng/mL) were performed, and Max25% (the sum of 25% biggest absolute prediction error, e.g., absolute prediction error of 4 PAs were chosen among 16 PAs in our case) was used to estimate the performance of different approaches. In the approach simply reported every alkaloid as equivalent of one alkaloid, the Max25% values are in the range of 232.2%–603.1%, 202.5%–502.2%, 170.3%–621.2% in the concentration of 5, 10, 25 ng/mL, respectively. However, when we use the prediction model, the Max25% values are 196.50%, 142.16%, 162.26% separately, which is more accurate and controllable. It has been reported by the EFSA Scientific Panel on Contaminants in the Food Chain (CONTAM Panel) that they identify mainly 4 types of PAs of particular importance for food and feed (Senecionine-types, Lycopsamine-type, Heliotrine-type, Monocrotaline-type) (EFSA, 2011). Although there were only a limited number of PAs in our dataset when we developed the QPM, it covers 3 types of most common PAs which exist in honey. Therefore, this model can make relative accurate estimation of PAs covered by these 3 types, at least (only monocrotaline-type PAs was excluded). To make this model more generic, further study can be carried out to include more PAs, such as monocrotaline-type PAs in the model development.

The QPM was used for the estimation of other PAs based on the tentative identification, and prediction error of RS between measured and predicted value are shown in supplemental Table S3-b. The results from semi-quantification of these PAs are shown in Table 1. Riderine (or Echinatine/Indicine), which elutes just next to lycopsamine and shows the highest level among all the quantified PAs reaching an estimated concentration of about 56 μg/kg in H15. Crotananine shows the concentration of about 51 μg/kg in H8, followed by acetyl-lycopsamine and 3′-acetylindicine, with a concentration of 47 μg/kg and 45 μg/kg both detected in H9.

Although according to assessment of chronic risks referring to a $BMDL_{10}$ of 237 μg/kg bw per day established by EFSA (EFSA, 2017) or acute toxicity data reported by Dusemund et al., (2018), the detected PAs levels are acceptable. However, following the recommendations from the BFR as mentioned before, the oral intake of PAs should not exceed a daily dose of 0.007 μg/kg b.w (BfR, 2011). Assuming that a 60 kg body weight and a daily consumption of honey of 20 g, a honey sample should not contain more than 21 μg/kg of PAs. Our estimated concentration of PAs, including all PAs even those where standards are

(a)



(b)

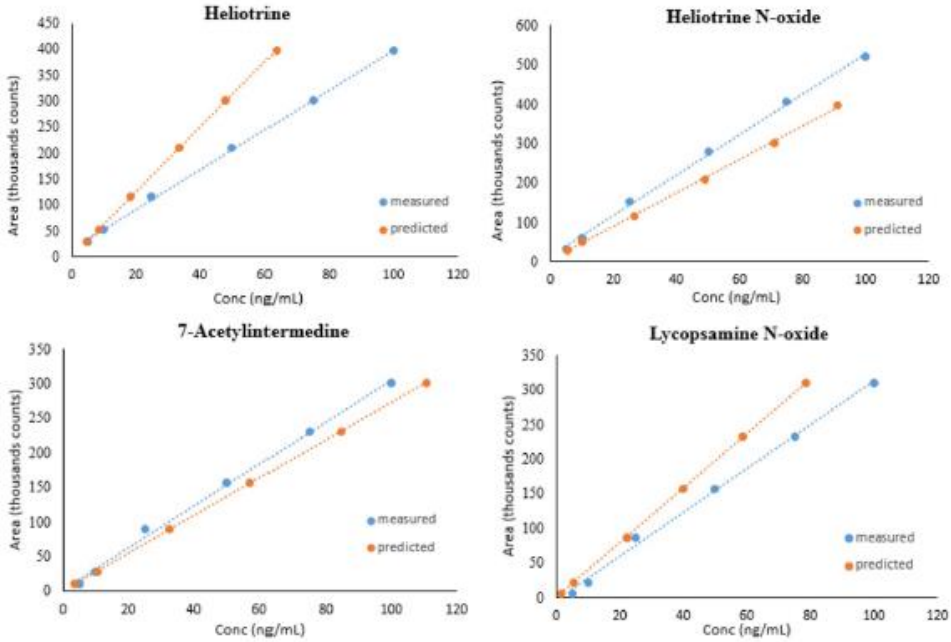


Fig. 2. (a) Correlation between relative predicted slope and measured slope in all PAs corresponding to Seneciphylline. (b) Correlation between predicted concentration and measured concentration in PAs (from 5 to 100 ng/mL) out of the model. Note: Lycopsamine *N*-oxide, heliotrine, heliotrine *N*-oxide, 7-acetylintermedine are 4 other PAs out of the 12 PAs in the model that used for verification of the prediction model.

not available, show the total PAs content in honey samples may be as high as: H6 (ca. 26 μg/kg), H7 (ca. 36 μg/kg), H5 (ca. 36 μg/kg), H11 (ca. 36 μg/kg), H18 (ca. 41 μg/kg), H8 (ca. 51 μg/kg), H15 (ca. 67 μg/kg) and H9 (ca.117 μg/kg) all exceeding 21 μg/kg. Even if we consider a quantification error of 51% the levels will still be of concern.

In conclusion, we proposed a method that can be extended to a general strategy to achieve estimation of quantitative data where authentic standards are not available. By using a prediction model based on physio-chemical parameters, analysis of similar standards, and literature relevant quantitative data for risk assessment can be estimated using HPLC-HR-MS/MS techniques. This approach was demonstrated by the simultaneous quantification of 12 PAs and also the further analysis of honey containing other high content PAs without standards. Future applications of this method could be the quantification or semi-quantification of PAs in other honey samples, and also other PA containing products, such as herbal tea, cereal, feed, and milk.

Acknowledgements

The authors acknowledge the funding from the Danish Food Authorities (Refs. Fødevareforlig 3) Helen Fodnæss is also acknowledged for the proof reading.

Abbreviations used

PAs	Pyrrolizidine alkaloids
GC-MS	Gas chromatography–mass spectrometry
LC-MS	Liquid chromatography–mass spectrometry
MRM	multiple reaction-monitoring
LC-QTOF	liquid chromatography-quadrupole-time-of-flight-mass spectrometry
QC	Quality control
LOD	limits of detection
LOQ	limit of quantification; pKa: pKa value of strongest acidic; pKb: pKa value of strongest basic
SMILE	simplified molecular-input line-entry system; CV%: Coefficient of Variation
ME	matrix effect; ESI: electrospray ion source
BMDL10	the lower confidence limit on the benchmark dose associated with a 10% response

Funding source

The study was financially supported by the Danish Food Authorities.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2018.11.033.

References

Alymatiri, C. M., Kouskoura, M. G., & Markopoulou, C. K. (2015). Decoding the signal response of steroids in electrospray ionization mode (ESI-MS). *Analytical Methods, 7*, 10433–10444.

Avula, B., Sagi, S., Wang, Y. H., Zweigenbaum, J., Wang, M., & Khan, I. A. (2015). Characterization and screening of pyrrolizidine alkaloids and N-oxides from botanicals and dietary supplements using UHPLC-high resolution mass spectrometry. *Food Chemistry, 178*, 136–148.

Beales, K. A., Betteridge, K., Colegate, S. M., & Edgar, J. A. (2004). Solid-phase extraction and LC-MS analysis of pyrrolizidine alkaloids in honeys. *Journal of Agricultural and Food Chemistry, 52*(21), 6664–6672.

Betteridge, K., Cao, Y., & Colegate, S. M. (2005). Improved method for extraction and LC-MS analysis of pyrrolizidine alkaloids and their N-oxides in honey: Application to Echium vulgare honeys. *Journal of Agricultural and Food Chemistry, 53*(6), 1894–1902.

BfR (2011). *Chemical analysis and toxicity of pyrrolizidine alkaloids and assessment of the health risks posed by their occurrence in honey.*

BfR (2013). *Determination of pyrrolizidine alkaloids.* (PA) in honey by SPE-LC-MS/MS.

Bodi, D., Ronczka, S., Gottschalk, C., Behr, N., Skibba, A., Wagner, M., et al. (2014). Determination of pyrrolizidine alkaloids in tea, herbal drugs and honey. *Food Additives & Contaminants: Part A, 31*(11), 1886–1895.

Bolechová, M., Čáslavský, J., Pospíchalová, M., & Kosubová, P. (2015). UPLC-MS/MS method for determination of selected pyrrolizidine alkaloids in feed. *Food Chemistry, 170*, 265–270.

Boppré, M. (2011). The ecological context of pyrrolizidine alkaloids in food, feed and forage: An overview. *Food Additives & Contaminants, 28*(3), 260–281.

Boppré, M., Colegate, S. M., & Edgar, J. A. (2005). Pyrrolizidine alkaloids of Echium vulgare honey found in pure pollen. *Journal of Agricultural and Food Chemistry, 53*, 594–600.

Cech, N. B., & Enke, C. G. (2000). Relating electrospray ionization response to nonpolar character of small peptides. *Analytical Chemistry, 72*, 2717–2723.

Chung, S. W. C., & Lam, C.-H. (2018). Development of an analytical method for analyzing pyrrolizidine alkaloids in different groups of food by UPLC-MS/MS. *Journal of Agricultural and Food Chemistry, 66*, 3009–3018.

Colegate, S. M., Edgar, J. A., Knill, A. M., & Lee, S. T. (2005). Solid-phase extraction and HPLC-MS profiling of pyrrolizidine alkaloids and their N-oxides: A case study of *Echium* plantagineum. *Phytochemical Analysis, 16*, 108–119.

Crews, C., Berthiller, F., & Krska, R. (2010). Update on analytical methods for toxic pyrrolizidine alkaloids. *Analytical and Bioanalytical Chemistry, 396*, 327–338.

Culvenor, C. C. J., Edgar, J. A., Smith, L. W., Kumana, C. R., & Lin, H. J. (1986). Heliotropium lasiocarpum Fisch and mey identified as cause of veno-occlusive disease due to a herbal tea. *The Lancet, 978*.

Dusemund, B., Nowak, N., Sommerfeld, C., Lindtner, O., Schäfer, B., & Lampen, A. (2018). Risk assessment of pyrrolizidine alkaloids in food of plant and animal origin. *Food and Chemical Toxicology, 115*, 63–72.

EFSA (2011). Scientific Opinion on Pyrrolizidine alkaloids in food and feed. *EFSA Journal, 9*(11), 2406.

EFSA (2017). Risks for human health related to the presence of pyrrolizidine alkaloids in honey, tea, herbal infusions and food supplements. *EFSA Journal, 15*(7), 4908.

Golubović, J., Birkemeyer, C., Protić, A., Otašević, B., & ečević, M. (2016). Structure – response relationship in electrospray ionization-mass spectrometry of sartans by artificial neural networks. *Journal of Chromatography A, 1438*, 123–132.

Heeft, E., Bolck, Y. J. C., Beumer, B., Nijroider, A. W. J. M., Stolker, A. A. M., & Nielen, M. W. (2009). Combined with time-of-flight and orbitrap. *Journal of the American Society for Mass Spectrometry, 20*, 451–463.

Hermans, J., Ongay, S., Markov, V., & Bischoff, R. (2017). Physicochemical parameters affecting the electrospray ionization efficiency of amino acids after acylation. *Analytical Chemistry, 89* 9166-9159.

Hoogenboom, L. A. P., Mulder, P. P. J., Zeilmaker, M. J., van den Top, H. J., Remmelink, G. J., Brandon, E. F. a, et al. (2011). Carry-over of pyrrolizidine alkaloids from feed to milk in dairy cows. *Food Additives & Contaminants: Part A, 28*(3), 359–372.

Kempf, M., Wittig, M., Reinhard, A., von der Ohe, K., Blacquiere, T., Raezke, K.-P., et al. (2011). Pyrrolizidine alkaloids in honey: Comparison of analytical methods. *Food Additives & Contaminants Part A-chemistry Analysts Control Exposure & Risk Assessment, 28*(3), 332–347.

Kowalczyk, E., & Kwiatek, K. (2018). Pyrrolizidine alkaloids in honey: Determination with liquid chromatography-mass spectrometry method. *Journal of Veterinary Research, 62*(2), 173–181.

Kruve, A., & Kaupmees, K. (2017). Predicting ESI/MS signal change for anions in different solvents. *Analytical Chemistry, 89*, 5079–5086.

Leito, I., Herodes, K., Huopolainen, M., Virro, K., Künnapas, A., Kruve, A., et al. (2008). Towards the electrospray ionization mass spectrometry ionization efficiency scale of organic compounds. *Rapid Communications in Mass Spectrometry, 22*, 379–384.

Lucatello, L., Merlanti, R., Rossi, A., Montesissa, C., & Capolongo, F. (2016). Evaluation of some pyrrolizidine alkaloids in honey samples from the veneto region (Italy) by LC-MS/MS. *Food Analytical Methods, 9*, 1825–1836.

Lucchetti, M. A., Glauser, G., Kilchenmann, V., Dübecke, A., Beckh, G., Praz, C., et al. (2016). Pyrrolizidine alkaloids from *Echium vulgare* in honey originate primarily from floral Nectar. *Journal of Agricultural and Food Chemistry, 64*, 5267–5273.

Martinello, M., Cristofoli, C., Gallina, A., & Mutinelli, F. (2014). Easy and rapid method for the quantitative determination of pyrrolizidine alkaloids in honey by ultra performance liquid chromatography-mass spectrometry: An evaluation in commercial honey. *Food Control, 37*(1), 146–152.

Molyneux, R. J., Gardner, D. L., Colegate, S. M., & Edgar, J. A. (2011). Pyrrolizidine alkaloid toxicity in livestock: A paradigm for human poisoning? *Food Additives & Contaminants: Part A, 28*(3), 293–307.

Mulder, P. P. J., Sánchez, L. P., These, A., & Preiss-weigert, A. (2015). *Occurrence of pyrrolizidine alkaloids in food.* EFSA Supporting Publication EN-859.

NMKL (2009). *NMKL Procedure No. 4. Validation of chemical analytical methods, Vols. 1–45*https://www.nmkl.org/index.php/en/.

Orantes-Bermejo, F. J., Serra Bonvehí, J., Gómez-Pajuelo, A., Megías, M., & Torres, C. (2013). Pyrrolizidine alkaloids: Their occurrence in Spanish honey collected from purple viper's bugloss (*Echium* spp.). *Food Additives & Contaminants: Part A, 30*, 1799–1806.

Oss, M., Kruve, A., Herodes, K., & Leito, I. (2010). Electrospray ionization efficiency scale of organic compound. *Analytical Chemistry, 82*, 2865–2872.

Raji, M. A., Fryčák, P., Temiyasathit, C., Kim, S. B., Mavromaras, G., Ahn, J.-M., et al. (2009). Using multivariate statistical methods to model the electrospray ionization response of GXG tripeptides based on multiple physicochemical parameters. *Rapid Communications in Mass Spectrometry, 23*, 2221–2232.

Schenk, A., Siewert, B., Toff, S., & Drewe, J. (2017). Determination of 34 pyrrolizidine alkaloids (PA) as contaminants in various plant extracts using UHPLC-ToF-HRMS. *Planta Medica International Open, 4*, S1–S202. https://doi.org/10.1055/s-0037-1608497.

Schulz, M., Meins, J., Diemert, S., Zagermann-Muncke, P., Goebel, R., Schrenk, D., et al. (2015). Detection of pyrrolizidine alkaloids in German licensed herbal medicinal

teas. *Phytomedicine, 22*, 648–656.
Szisz, D. (2017). *pKa Plugin*. http://www.weightless.org/about/spectrum-for-weightless.
Tandon, R. K., Tandon, H. D., & Narndranathan, M. (1976). An epidemic of veno-occlusive disease of liver in central Asia. *The Lancet*, 271–272.
Tarpey, T. (2000). A note on the prediction sum of squares statistic for restricted least squares. *The American Statistician, 54*(2), 116–118.
These, A., Bodi, D., Ronczka, S., Lahrssen-wiederholt, M., & Preiss-weigert, A. (2013). Structural screening by multiple reaction monitoring as a new approach for tandem mass spectrometry : Presented for the determination of pyrrolizidine alkaloids in plants. *Analytical and Bioanalytical Chemistry, 405*, 9375–9383.
Wiedenfeld, H. (2011). Plants containing pyrrolizidine alkaloids: Toxicity and problems. *Food Additives & Contaminants, 28*(3), 282–292.

Manuscript 3: Semi-quantification model application in cereals

Supporting information of manuscript 3 is available at https://doi.org/10.1038/s41598-020-62573-z

SCIENTIFIC REPORTS
natureresearch

OPEN

Quantification for non-targeted LC/MS screening without standard substances

Jaanus Liigand[1], Tingting Wang[2], Joshua Kellogg[3], Jørn Smedsgaard[2], Nadja Cech[3] & Anneli Kruve[1,4*]

Non-targeted and suspect analyses with liquid chromatography/electrospray/high-resolution mass spectrometry (LC/ESI/HRMS) are gaining importance as they enable identification of hundreds or even thousands of compounds in a single sample. Here, we present an approach to address the challenge to quantify compounds identified from LC/HRMS data without authentic standards. The approach uses random forest regression to predict the response of the compounds in ESI/HRMS with a mean error of 2.2 and 2.0 times for ESI positive and negative mode, respectively. We observe that the predicted responses can be transferred between different instruments via a regression approach. Furthermore, we applied the predicted responses to estimate the concentration of the compounds without the standard substances. The approach was validated by quantifying pesticides and mycotoxins in six different cereal samples. For applicability, the accuracy of the concentration prediction needs to be compatible with the effect (e.g. toxicology) predictions. We achieved the average quantification error of 5.4 times, which is well compatible with the accuracy of the toxicology predictions.

Liquid chromatography mass spectrometry (LC/MS) has become the most versatile analytical tool to discover and detect metabolites[1], pharmaceuticals and their transformation products[2], environmental contaminants[3], and food contaminants[4] with non-targeted[5] analysis. To better understand the chemical and mechanistic dynamics of a system, a quantitative approach is preferred, which requires two main elements: identification and quantification of each metabolite (determining the concentration of compounds within the dataset). Currently, accurate mass measurements from high-resolution mass spectrometry (HRMS), together with relevant data analysis[6,7], are increasingly able to assign putative structures to the detected features[8]. Quantifying, however, remains a primary challenge. For example, out of 114 100 compounds in the Human Metabolome Database, only ca. 21 000 have been detected and identified[9]. Currently, the ability to obtain quantitative information from LC/MS is almost exclusively limited by the availability of standard substances, as different compounds ionize to different extents in an electrospray (ESI) source. The response of the compounds in LC/MS is influenced by the properties of the compound, eluent composition, and instrument. Thus, quantifying all detected compounds with a targeted analysis is exceedingly difficult, as standard reference materials (to match retention time, mass fragmentation pattern, and provide a calibration curve) are not available for the majority of compounds.

Additionally, the results of most LC/MS analyses conducted in different laboratories can currently only be compared based on qualitative data, as the measurement conditions and instruments used vary strongly and quantitative data are not available[5]. The lack of facile quantification also represents an obstacle to longitudinal studies, as samples collected over a long period of time must be stored and analysed all together in the same laboratory with the same methods. This raises concerns about sample preservation and stability, and delays in information dissemination, especially in cases where fast interventions may be crucial.

Several groups have studied ionization efficiencies in electrospray sources to understand which factors contribute to the ionization process[10–15]. Several ionization efficiency prediction models have been proposed for small sets of compounds, mainly based on multilinear regression algorithms (see Table S1 in Supporting Information

[1]Institute of Chemistry, Faculty of Science and Technology, University of Tartu, Ravila 14A, 50411, Tartu, Estonia. [2]National Food Institute, Research Group for Analytical Food Chemistry, Technical University of Denmark, Kemitorvet Building 202, Kgs, Lyngby, DK-2800, Denmark. [3]Department of Chemistry & Biochemistry, University of North Carolina at Greensboro, Greensboro, North Carolina, 27412, United States. [4]Department of Environmental Science and Analytical Chemistry, Stockholm University, Svante Arrhenius väg 16, 106 91, Stockholm, Sweden. *email: anneli.kruve@su.se

for a detailed overview). In general, researchers have focussed on finding a suitable predictive model for a specific compound group (drugs, metabolites, steroids) in studies that normally cover a few tens of compounds measured on a single instrument, and most have been carried out in ESI positive mode with one eluent composition in infusion or flow injection mode. More recently, some groups have used LC separation with gradient elution[16]. The latter point is important to effectively model the influence of both compound properties and eluent properties. However, a general model that would fit a wide range of compounds, LC conditions, instruments, and would also allow quantitation is lacking.

Based on the promising results obtained for predicting ionization efficiencies for specific compound classes, we propose a general approach for quantifying compounds using their predicted ionization efficiency values. We demonstrate that this approach can be transferred between different eluents, LC setups, and instruments and can be used for the analysis of real samples. The approach incorporates both positive and negative ionization modes under more than 100 eluent compositions and covers over 450 compounds. We validated this approach by predicting the concentrations of 35 compounds, including pesticides and mycotoxins, in cereal samples. The quantification for the samples was conducted on an instrument that was not involved in the ionization efficiency model development. The results presented herein demonstrate that our approach for calculating ionization efficiencies has the potential for making non-targeted analysis (semi-)quantitative.

Experimental

Ionization efficiency data: training and test set. Previously measured ionization efficiency values were collected in positive[17–24] and negative[24–28] mode. The ionization efficiency values for both modes and all studied eluent compositions are presented in Table S2. We measured an additional 165 compounds with diverse chemical properties in both the main eluent composition (acetonitrile/0.1% formic acid (aq) 80/20) and ca. 1000 new analyte-eluent combinations. For ESI positive mode, a total of 3139 ionization efficiency values were measured and collected from our previous works[17–21]. These data belong to 353 unique compounds and 106 different eluent compositions (Table S3 in SI). For ESI negative mode, an additional 1286 ionization efficiency values were collected from our previous works[25,26], including 33 eluent compositions Table S4), and 101 unique compounds.

The compounds covered by the model include protonated and intrinsically charged compounds in positive mode and deprotonated compounds in the negative mode. In ESI positive mode both nitrogen and a significant amount of oxygen bases could be measured. In ESI negative mode compounds with significant acidic moieties could be detected as $[M\text{-}H]^-$, including amino acids, benzoic acids and derivatives, phenols, amines, heterocycles, guanidines, and diazines (see Table S5 and S6). As we focus on $[M]^+$, $[M+H]^+$ and $[M\text{-}H]^-$, some of the compound groups are better represented then some of the other groups. For example, the lipids and sugars form primarily sodiated ions in ESI positive mode and are, therefore, not included in the scope of this approach.

From an application perspective, the training and test set compounds fall into a diverse array of categories, including drugs or drug-like compounds (e.g. terfenadine, ketoconazole, lidocaine), metabolites (e.g. acetylcholine, dopamine, thiamine), amino acids, small organic precursors (amines primarily), lipids (e.g. myristic acid, progesterone, glyceryl tributyrate) as well as industrial dyes (e.g. sudan II, sudan IV). The chemical space covered by these compounds was compared to the chemical databases Drugbank, NORMAN, and the Human Metabolome Database (HMDB), and is presented at Figs. S1 and S2.

Validation set. The proposed method was validated on a set of 35 pesticides and mycotoxins (Table S7), 28 of which had not been included in the training or test set. The chemical space covered by the validation compounds can be seen from Fig. S3. The compounds were measured at 10 concentration levels in solvent (acetonitrile), oat, barley, rye, wheat, rice, and maize with Agilent 6495 triple quadrupole instrument. The concentrations of the compounds ranged over 5 orders of magnitude from 3.6 nM to 0.35 mM. Altogether, 2233 data points (pesticide, matrix, and concentration combinations) were measured and corresponding concentrations were predicted. The validation set was used to evaluate the applicability of the ionization efficiency predictions for compounds not included in the training or test set. All measurements were also done under gradient separation and on an instrument that was not the primary instrument for ionization efficiency measurements.

Ionization efficiency measurements. For the evaluation of log*IE* values, the responses of $[M+H]^+$ and $[M\text{-}H]^-$ were recorded in an MS full scan mode corresponding to the polarity mode. In case in-source fragmentation of the compounds occurred, the intensities of the fragment ions peaks were summed with the intensity of the molecular ion peak. Six dilutions of the stock solutions were made (1-, 1.25-, 1.67-, 2-, 2.5-, and 5 times) with the corresponding eluent by the autosampler and delivered to MS in flow injection mode. The injection volume was 10 µL, and the eluent flow rate was 0.2 mL/min. Concentrations in the injected solutions ranged from 10^{-4} M to 10^{-9} M and measurements were conducted in the linear range. Most of the measurements of log*IE* values were carried out using an Agilent XCT ion trap mass spectrometer. Additionally, six mass spectrometers from five vendors were used in three labs around the world (Tartu/Estonia, Lyon/France, and Beerse/Belgium, see Table S8). On all instruments, the default MS and ESI parameters were used. In the case of the ion trap instrument, only the optimized Target Mass (TM) parameter was used[29].

The absolute ionization efficiency values vary significantly depending on the ionization source geometry, ion optics, day, cleanliness of the ionization source, etc. Therefore, we measured the relative ionization efficiency (*RIE*) of a compound M_1 relative to anchor compound (M_2) according to the following equation[26]:

$$RIE(M_1/M_2) = \frac{slope(M_1) \cdot IC(M_1)}{slope(M_2) \cdot IC(M_2)} \quad (1)$$

where the slope of the signal versus concentration is estimated via linear regression in the linear range of the signal-concentration plot and the *IC* is the sum of relative abundances of isotopologues where highest abundance is taken equal to 100. All measurements were made relative to tetraethylammonium (positive mode) and benzoic acid (negative mode). To make the data easier to present and analyze, the logarithmic scale (log*IE*) was used. The scale in negative mode was anchored to the log*IE* of benzoic acid, taken as 0 in the 0.1% ammonia/acetonitrile 20/80 mixture and the scale in positive mode was anchored to log*IE* of tetraethylammonium, taken as 3.95 in the 0.1% formic acid/acetonitrile 20/80. The log$IE = 0$ corresponds to methyl benzoate in the 0.1% formic acid/acetonitrile 20/80. The log*IE* values for each compound were obtained based on the *RIE* and log*IE* value of the anchor compound.

$$\log IE_{M_1} = \log RIE(M_1/M_2) + \log IE_{\text{anchor}} \tag{2}$$

To minimize the influence of possible differences in conditions when measuring M_1 and anchor compound, two steps were taken: (1) each compound was measured on at least three different runs (on three different days) and the results were averaged, and (2) anchor compound was measured in the beginning and the end of each run on each day. To anchor the scales of other eluent compositions, the MS signal intensities of anchor compound in all eluent compositions were measured in the same day and the log*IE* value of anchor compound in an eluent composition b was calculated using Eq:

$$\log IE_{Eb} = \log\left(IE_{Ea} \cdot \frac{Signal_{Eb}}{c_{Eb}} \cdot \frac{c_{Ea}}{Signal_{Ea}}\right) \tag{3}$$

where the $Signal_{Eb}$ and $Signal_{Ea}$ are the signal intensities in eluent composition b and a and c_{Ea} and c_{Eb} are the corresponding concentrations of benzoic acid in the respective eluent compositions.

To develop the model, log*IE* values measured on different instruments were transformed to unified log*IE* values. The unification was performed using the intersection of compounds measured on two instruments in the eluent composition acetonitrile/0.1% formic acid (aq) 80/20[22,23]. To unify the data, linear regression between log*IE* values measured on two instruments was performed (see SI).

Eluent effect of ionization efficiency. PCA analysis of the physicochemical properties (Fig. S4) of the training and test set compounds (n = 353) was conducted in order to incorporate different eluent compositions to the ionization efficiency predictive model without measuring all of the compounds in all eluents. Next, random sets of 40 compounds were sampled, and the set representing the widest chemical space was chosen for studying the solvent effects. As a result, 20 additional eluent compositions were studied (Table S3), yielding 106 different solvent compositions together with the previously measured eluent conditions. Both acetonitrile and methanol were studied as an organic modifier and the organic modifier percentages studied were 0%, 20%, 50%, 80% and 100%. Water phase additives included formic acid, trifluoroacetic acid, ammonia, ammonium acetate, ammonium formate, ammonium bicarbonate, and ammonium fluoride (pH = 2.0–10.3).

Model development. *Data preprocessing.* For model development, PaDEL descriptors[30] (1444 descriptors) were calculated using ChemDes calculator[31] for every compound. As PaDEL descriptor calculation of some descriptors fails for some compounds, the descriptors with NA (not available values) were removed from the dataset. Next, all the descriptors with the same value for >95% of compounds were eliminated from the dataset. As the third cleaning step, the pairwise correlation of descriptors was considered. If the R^2 was higher than 0.8 the former descriptor was removed from the dataset. After data pre-processing for ESI positive mode 1086 descriptors were left in the dataset and for ESI negative mode 822 descriptors were left in the dataset. Additionally, five empirical eluent descriptors (viscosity[32], surface tension[33], polarity index[34], pH, NH_4 content (yes/no)) were added to the dataset.

Algorithm selection and model development. Different machine learning algorithms (multilinear regression, Ridge regression, support vector machine regression, artificial neural networks and random forest regression) were tested for model development. For each algorithm, a suitable R package (glmnet[35], e1071[36], h2o[37], RRF[38]) was used and the parameters were optimized.

Regularized random forest regression algorithm[38] from library RRF in R yielded the best prediction accuracy. For ESI negative and positive mode, individual models were developed. For model development, the order of the dataset was randomized and split into two sets. 80% of observations were used for developing the model and 20% of observations were used as a validation set. The number of trees used in the random forest was optimized, with the optimal number was 100 decision trees. The regularization isotherm selected 450 significant descriptors (Table S9) in ESI positive mode and 145 significant descriptors (Table S10) in ESI negative mode.

The code used for training the model is available from Code S1.

Concentration from predicted ionization efficiency. As the model output is in universal ionization efficiency values and not instrumentation specific, a set of compounds with known concentrations is used to transform the universal predicted values to instrumentation specific values.

For transforming the predicted ionization efficiency values a set of compounds with known concentration either spiked to the matrix or as a standard solution was measured in dynamic range with the same method as compounds of interest. Sets of 6, 10, 15, and 31 compounds were tested; see discussion below. From the analysis results logarithmic response factors (*RF*) were calculated:

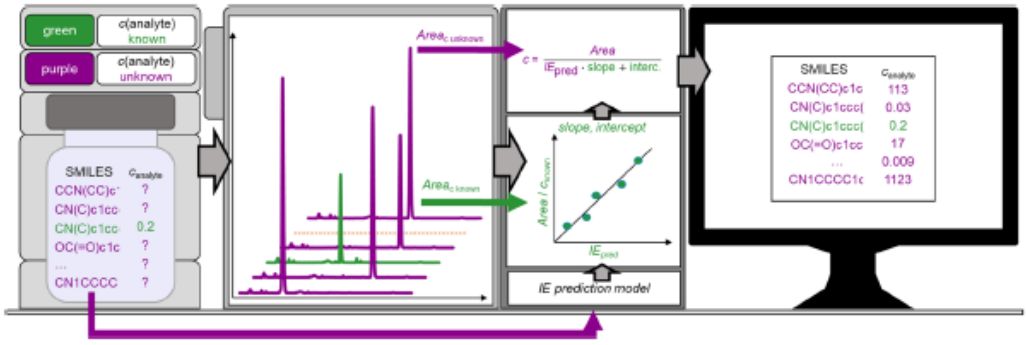


Figure 1. Flow chart of the developed approach to apply ionization efficiency prediction to estimate concentration. Purple is used for compounds of interest and green is used for compounds with known concentration; the latter are used to account for instrument-specific effects in the prediction model.

$$\log RF = \frac{Signal}{concentration} \tag{4}$$

and correlated with the predicted ionization efficiency values $\log IE_{pred}$:

$$\log RF_{pred} = slope \cdot \log IE_{pred} + intercept \tag{5}$$

where the signal is the MS1 peak area corrected with isotope distribution. Molar concentrations were used for the calculation of the response factor. Logarithmic response factors were correlated with predicted universal ionization efficiency values to obtain parameters necessary for transforming the predicted ionization efficiency values. The process of concentration prediction is visualized in Fig. 1. For studied pesticides and mycotoxins in cereals, the obtained parameters were used to transform predicted ionization efficiency values to logarithmic response factors. Based on the MS1 peak areas of compounds of interest and predicted response factors it is possible to predict the concentration. Slope and intercept values in Eq. 5 were calculated based on the coefficients of the linear regression curve between $\log RF$ and $\log IE_{pred}$ values in the transformation set. In order to validate the obtained results, the prediction errors between real concentration and predicted concentration were calculated.

Measuring the prediction accuracy. We express the accuracy of ionization efficiency prediction as a prediction error:

$$\text{prediction error} = \max\left\{\frac{predicted\ IE}{measured\ IE}, \frac{measured\ IE}{predicted\ IE}\right\} \tag{6}$$

The lowest possible prediction error is 1.0 times and values close to 1.0 times are desirable. In order to illustrate the general prediction accuracy, we use the average error. The accuracy of concentration predictions was estimated with a similar comparison.

Sample preparation for cereal samples. All of the cereal samples were prepared according to the QuEChERS extraction method, described elsewhere[39]. In short, the blank samples (pesticides free) were obtained from proficiency test material for the six European Union Proficiency Test EUPTs: EU-PT-CF8 (wheat), C3 (oat), CF10 (rye) and C6 (barley), CF9 (maize), SRM6 (rice). Two grams of homogenized cereal samples were soaked with 10 mL acidified Milli-Q water containing 0.2% formic acid. Then, the sample was extracted with 10 mL of acetonitrile. Thereafter, 4 g of magnesium sulfate and 1 g of sodium chloride were added, and the tube was shaken for 1 min followed by centrifugation. The organic upper layer (2 mL) was removed and shaken with 0.1 g of Bondesil-C18 and 0.3 g of magnesium sulphate for 2 min followed by centrifugation. Then 1.5 mL of purified extract was removed into a vial with insert and spiked with different concentration of tested compounds prior to injection on the LC/MS system.

LC/MS analysis of cereal samples. Samples were analyzed on an Agilent 1290 ultrahigh performance liquid chromatograph (Agilent Technologies, CA, U.S.) coupled to an Agilent 6495 triple quadrupole instrument (QQQ) at the University of Tartu (UT). Samples were injected onto an Agilent Zorbax RRHD SB-C18 reversed-phase column (1.8 µm, 2.1 × 50 mm). The mobile phase consisted of (A) water containing 0.1% formic acid and (B) acetonitrile. The analysis was done using a gradient elution at a flow rate of 0.3 mL/min at 30 °C. The gradient was from 5% to 100% B in 7 min, then maintained at 100% for 2 min and returned back to 5% B in 2 min, and maintaining starting conditions at 5% B for 2 min equilibration with 5% B to yield a total runtime of 13 min.

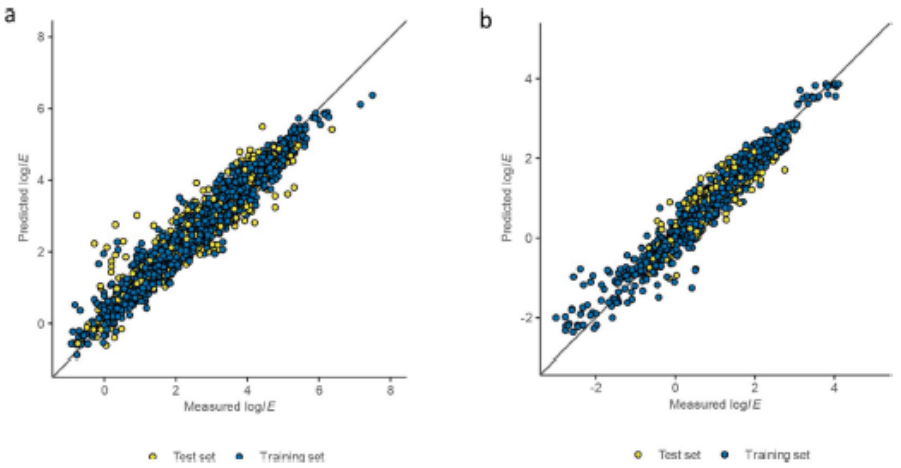


Figure 2. Performance of ionization efficiency prediction models. Black line denotes ideal fit. (**a**) ESI positive mode with 3139 datapoints. (**b**) ESI negative mode with 1286 datapoints.

An Agilent's Jet Stream electrospray ionization (ESI) interface was used in positive ion mode (ESI+) with the following settings: capillary voltage 3 kV, nebulizer pressure 20 psi, sheath gas flow rate 11 L/min, sheath gas temperature 350 °C, dry gas temperature 250 °C, and dry gas flow rate 14 L/min. Spectra were collected from m/z 100 to 1100 Da. The injection volume was 1 µL. Peaks were manually integrated, and peak areas were used for all calculations.

Results

Predicting ionization efficiencies. In order to quantify the compounds based on the ionization efficiency values we (1) measured *IE* values for a wide set of compounds, (2) used the measured *IE* values as well as compound and eluent descriptors for developing the model that would allow predicting ionization efficiencies, and (3) validated the approach by using the predicted *IE* to quantify a set of compounds in cereal samples. All of the *IE* values have been measured as relative values; the *IE* of methyl benzoate and benzoic acid have been arbitrarily taken as $IE = 1$ ($\log IE = 0$) in ESI positive and negative mode, respectively.

Based on these ionization efficiencies, predictive models for positive and negative mode were developed. The overall root mean squared prediction error was 2.2 times (training set 1.9 and test set 3.0 times) (Fig. 2a and Table S11). This means that if the ionization efficiency of compound A is predicted to be 100 times higher than the ionization efficiency of the methyl benzoate the actual ionization efficiency would be 45 to 220 higher than that of methyl benzoate ($\log IE = 2.00 \pm 0.34$).

In negative mode, the best performing model was obtained also with random forest regression with 145 significant descriptors and 100 regression trees. The regression model explained 93% of the variation in ionization efficiency values. The overall root mean squared error was 2.0 times (training set 2.0 and test set 2.3 times, Fig. 2b).

Upon closer examination of the ionization efficiency prediction model in ESI positive mode, it was observed that the model performed universally well for different organic modifier percentages (Fig. S5). The lowest prediction error, 1.4 times, was observed for eluents containing 20% of organic modifier and the highest prediction error, 1.9 times, for eluent containing 90% organic modifier. This is expected, as eluents containing 20% of organic modifier had the highest number of data points, which improved prediction accuracy. Additionally, the model was well-performing for both methanol as well as for acetonitrile containing eluent compositions (Figs. S5 and S6). Based on the pH of the eluent, basic conditions had the highest prediction error; 2.5 times and 3.7 times for the training and test set, respectively.

Similar trends were observed for ESI negative mode; the prediction accuracy for the pure organic modifier was the lowest (prediction error of 4.1 times, Figs. S7 and S8). Regarding the pH, no significant differences in the prediction accuracy were observed (Fig. S7).

The most influential parameters (Table S12) were a number of structural parameters and also solvent parameters. PaDEL descriptors used in the Random Forest regression are 2D structural parameters and are, therefore, not directly linked to classical physicochemical parameters; however, some of the most influential parameters can be interpreted. From structure-related parameters, number of hydrogen atoms and number of nitrogen atoms was important in the model in ESI positive mode. The number of hydrogen atoms is associated with the size and hydrophobicity of the compounds while the number of nitrogen atoms is associated with the basicity of the compound. Also, the mobile phase parameters pH, viscosity, and presence of ammonium ions in the mobile phase were significant. Previous studies have shown that these factors can influence the response of compounds by orders of magnitude[18–20]. Here, the model confirms the previous empirical findings.

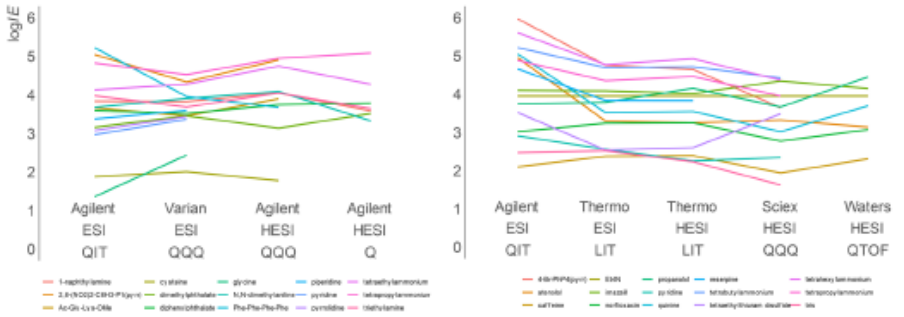


Figure 3. Measured ionization efficiencies on different instruments in ESI positive mode based on two subsets of druglike compounds measured on different instruments in the same eluent composition (acetonitrile/0.1% formic acid(aq) 80/20). Full data are shown in Table S2. ESI denotes conventional pneumatically assisted electrospray ionization source, HESI heated electrospray ionization source, QIT quadrupole ion trap, QQQ triple quadrupole, LIT linear iontrap, QTOF quadrupole time-of flight. Left and right correspond to two datasets.

Compatibility between different instruments. It is known that a compound's response varies from one LC/MS instrument to another; however, we have previously shown that the relative log*IE* values measured on different instruments are in good correlation to each other (Figs. 3 and S9)[22,23]. The ionization efficiency model was primarily developed on an ion trap instrument; however, for this study, we expanded the types of LC/MS instruments used for measuring ionization efficiency values. Altogether seven instruments and 9 different instrument-ionization source combinations (Table S8) and different mass analyser types from labs around the world have been included in studying the ionization efficiency. On all instruments, generic default settings have been used. It is known that some instruments (e.g. QQQ) may have lower mass cut-offs. Here, these effects are incorporated in the agreement of ionization efficiency values.

The comparison of the ionization efficiency values for compounds measured on a number of instruments reveals good intra-instrumental consistency (Figs. 3 and S9). The compounds being highest responders on one instrument are also generally top performers on the other instruments. Still, some instruments compress the ionization efficiency values more than other instruments. However, these effects could not be associated with a known lower mass cut-off of mass analysers or ionization source type (ESI vs HESI). Therefore, a more complicated relationship between the source design and ion optics design is likely to be important.

At the same time, it is not reasonable to build time and data-intensive prediction algorithms on each instrument. Therefore, it is necessary to have a robust approach to transfer the predicted values between different instruments. In order to achieve this, we use a linear correlation observed between the ionization efficiency values measured on different instruments (see SI for more details).

The feasibility of using linear correlation for transferring the predicted ionization efficiency values can be estimated by comparing the prediction error of ionization efficiency values for different instruments. We observed a good consistency between the prediction errors from instrument to instrument (Fig. S10). The highest average prediction error was 2.8 times (Waters Synapt G.2 Z-spray) and the lowest was 1.5 times (Thermo LTQ ESI). Both of these values are close to the repeatability of the log*IE* values. Therefore, it can be concluded that ionization efficiency values can be transferred between instruments with reasonable accuracy.

Additionally, we developed a model to predict ionization efficiency values with PaDEL descriptors and random forest regression based on data from a single instrument. For this, we chose data from an Agilent XCT instrument, which had been used to measure the largest number of ionization efficiency values (1816 data points). The mean prediction error for a single instrument-based model (ESI + mode) was 1.8 and 3.0 times for training and test set, respectively. The achieved prediction error values for a single instrument-based model were not significantly different from the generic model incorporating data from seven different instruments. Therefore, the developed model is robust for variations caused by different instruments and can be applied to predict ionization efficiency values measured on different mass spectrometers.

Validation: ionization efficiency and quantification of pesticides and mycotoxins in cereals. The proposed was validated based on 35 pesticides and mycotoxins (Table S7) measured under gradient separation on a triple quadrupole instrument. The list included 28 compounds not included in the training or test set. The log*IE* values were measured from the analysis of standard solutions of the 35 validation compounds and the values range from 1.60 to 4.11 with the average of 3.10 and median 3.29. The predicted log*IE* values take into account the eluent composition at retention time. The log*IE* prediction with the random forest model showed reasonable accuracy (RMSE 0.64, Fig. S11).

The predicted ionization efficiency was further used to predict the concentrations of the compounds detected assuming that the structure is known. For this, the compounds were spiked into blank oat, barley, rye, wheat, rice, and maize and analysed. For each compound, the ionization efficiency was predicted in ESI positive mode. To transform the predicted ionization efficiencies to instrument-specific response factors, a set of 31 compounds (Table S7) was used. Thereafter, the instrument-specific response factors were used to convert LC/MS signals into concentration (Eq.).

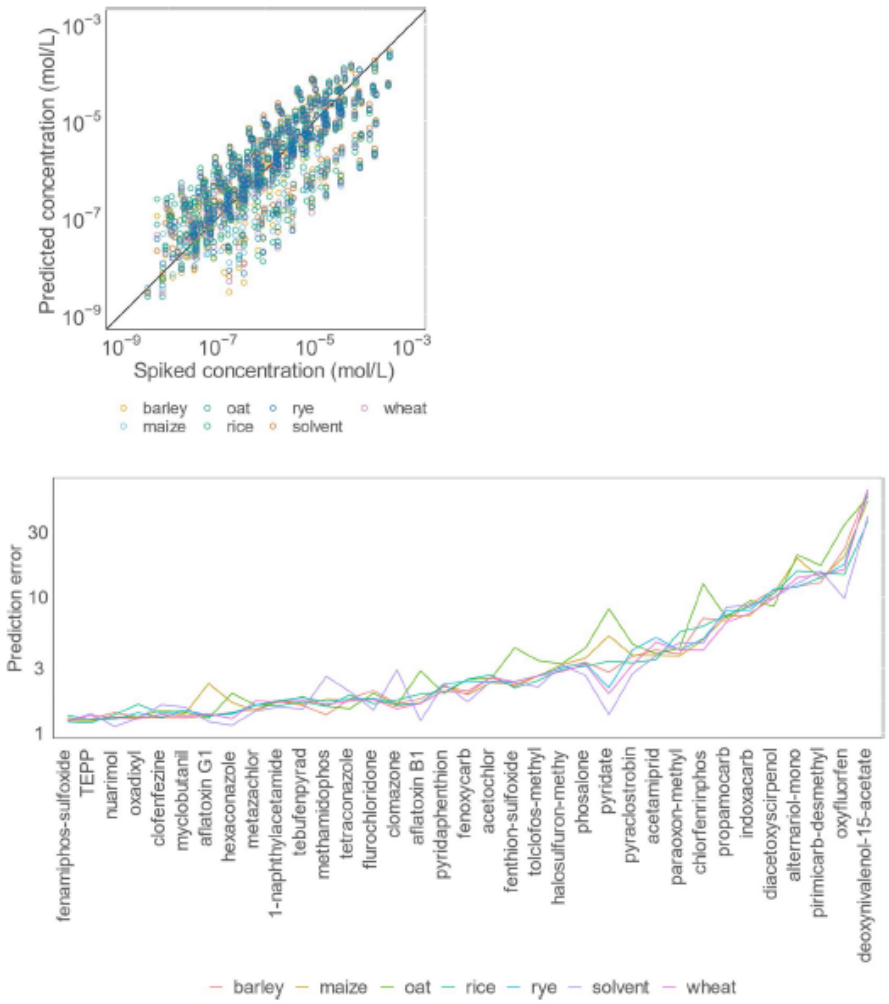


Figure 4. Performance of concentration prediction in the example of pesticides in cereals. above: concentration prediction of pesticides in cereal samples. below: prediction error of pesticide concentration in cereal samples, y-axis in logarithmic scale.

$$c = \frac{Signal}{10^{\log RF_{pred}}} \tag{7}$$

On average, the concentrations were predicted with the prediction error of 5.4 times (Fig. 4a, compound separated graphs on Fig. S12, Table S13). This means that if the pesticide concentration is estimated to be 1 ppm it would actually lie between 0.2 and 6 ppm. Compared to the conventional approach of assuming the equal response to all compounds detected (average prediction error of 526 times, Table S14), the developed approach improved prediction accuracy around 10 times and significantly reduces the width of the confidence interval. Consequently, more effective decision making can be made based on the predicted concentrations.

The lowest error observed was 1.2 times for TEPP in rice matrix, while the largest error, 62 times was observed for deoxynivalenol-15-acetate in barley matrix. For 88% of the compounds, the prediction error was lower than 10 times, for 76% of the compounds it was lower than 5 times, and for 48% of the compounds, it was lower than 2 times. The three compounds with the highest prediction error were deoxynivalenol-15-acetate (52 times), oxyfluorfen (19 times), and pirimicarb-desmethyl (15 times). The three best-performing compounds are fenamiphos-sulfoxide (1.2 times), TEPP (1.3 times), and nuarimol (1.3 times). These prediction errors are especially promising for mycotoxins for which the standard substances are either rarely available or very expensive. The large errors could not be associated with specific compound properties.

Additionally, we tested sets of 15, 10 and 6 compounds for transforming the predicted ionization efficiency values to instrument-specific response factors. For this random sets of 15, 10 or 6 compounds were sampled from the 31 compounds. On average, sets of 15, 10, and 6 compounds resulted in concentration prediction errors of 5.0, 5.0, and 5.0 times, respectively, which were not significantly different from the average prediction error observed for 31 compounds. Therefore, comparable prediction accuracies can be achieved with a smaller set of compounds. However, it needs to be considered that the compounds used should possess a wide range of chemical properties as well as a wide range of ionization efficiency values to allow for the establishment of a reasonable regression approach. Additionally, uniform elution distribution during the chromatographic run is beneficial.

Different matrices. Any model that aims at providing quantitative information needs to be applicable in a variety of matrices in order to be truly useful to researchers. The importance of matrices with LC/MS is even more relevant than for other analytical techniques, due to the possibility of significant matrix effects in the ESI source. A matrix effect is the suppression of ionization of a compound due to co-eluting compound(s). Previously, it has been qualitatively observed that the matrix effect and ionization efficiencies are influenced by the physicochemical properties of the compound[40]. Therefore, while applying the ionization efficiency predictions for concentration estimations, it was assumed that a model that considers ionization efficiencies also helps to account for matrix effect. This assumption was based on the fact that a small set of compounds with known concentrations were spiked into every sample; this helped to account for the differences arising from the instrument and also for matrix effects (Eq. 7).

Regarding different matrices, the prediction accuracy for all cereals was very similar. The lowest prediction error was observed for wheat and rice, 4.8 times, and highest for oat, 6.3 times (Table S15). This is expected, as oat samples possess a high content of polar lipids and free fatty acids[41] which possess high surface affinity and are, therefore, expected to cause ionization suppression[42]. Also, the mean prediction error for solvent (average 4.6 times) and all studied cereals (5.5 times) was comparable.

Moreover, the matrix effect is expected to vary strongly from sample to sample even for the same food commodity. The compounds that performed worst in one matrix performed also performed poorly in other matrices, and the best performers were among the top for all matrices (Fig. 4b). The consistency of the prediction error form one matrix to another indicates the insignificance of matrix effect and strongly indicates that the developed approach using ionization efficiency predictions, together with the transformation, helped to account for matrix effects.

Application area. The application range of all models depends on the data that have been used to train the model. The compounds covered by the model include protonated $[M+H]^+$ and intrinsically charged compounds $[M]^+$ in positive mode and deprotonated $[M-H]^-$ compounds in the negative mode. This defines the limitations of the application range of the model. For examples, compounds forming primarily sodium or ammonium adducts and no protonated ions are not within the scope of this approach.

We were additionally interested in comparing the compounds used in this study to compounds detectable with LC/MS based non-targeted screening. For this purpose, we compared the properties of the compounds used in this study with some of the most common databases; namely, NORMAN database[43], HMDB[9], and DrugBank[44]. NORMAN database is a database compiled by NORMAN network members and contains compounds that are important for various application areas from solvents to industrial chemicals, personal care products to metabolites. HMDB contains information about metabolites with different origins and DrugBank contains approved small molecule drugs, approved biologics (proteins, peptides, vaccines, and allergenics), nutraceuticals and experimental (discovery-phase) drugs. It is important to consider that LC/MS is unable to analyse all of these compounds due to limitations in retention and ionization. For HMDB and NORMAN it was possible to extract only the compounds that had been analysed with LC/MS. The compounds included in this study were compared with the databases based on the principal component analysis of PaDEL descriptors. It was observed that compounds included in this study covered a very large part on the chemical space of interest in HMDB and NOMRAN (Figs. S2 and S3). The coverage was somewhat less for DrugBank. This was expected, as the DrugBank includes also large biomolecules out of the scope of the model. Therefore, it is important to keep in mind the limitations of the initial data used in the modelling while applying the model and we hope to widen the model application are further in the future by incorporating (1) compounds forming adducts and (2) larger compounds, such as peptides.

Discussion

For the case study presented here, quantifying a variety of pesticides and mycotoxins in cereal matrices, we have shown that ionization efficiencies can be used to improve the accuracy of concentration predictions for small molecules that lack standard substances. In general, to obtain a reliable model, it is suggested to include several compounds that have been analysed together with the sample and quantified. Such compounds could be compounds confirmed with the aid of standard substances or compounds from quality control samples. This makes full scan LC/HRMS extremely appealing, as a combined targeted and non-targeted analysis method can be used.

Additionally, for risk assessment of contaminants in food and environmental samples, an estimated concentration of a compound is required to evaluate the potential risk of the exposure. In most cases, the error associated with other parts of the risk assessment (like intake and toxicology) is even larger than the error in concentration prediction[45]. Therefore, the ionization efficiency based quantification approach has high potential to complement non-targeted analysis and aid decision making based on the analysis results. The tool is made available at app.quantem.co. The approach has been developed on several instruments and validated on the example of five complicated matrices. Therefore, the quantification accuracy also represents the effects arising from mass discrimination and matrix effect. Validation for other applications is though, desirable.

Data availability

All data used for model development and validation are available as supporting information.

Received: 20 June 2019; Accepted: 16 March 2020;
Published online: 02 April 2020

References

1. Cajka, T. & Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **88**, 524–545 (2016).
2. Xu, R. N., Fan, L., Rieser, M. J. & El-Shourbagy, T. A. Recent advances in high-throughput quantitative bioanalysis by LC–MS/MS. *J. Pharm. Biomed. Anal.* **44**, 342–355 (2007).
3. Alder, L., Greulich, K., Kempe, G. & Vieth, B. Residue analysis of 500 high priority pesticides: Better by GC–MS or LC–MS/MS? *Mass Spectrom. Rev.* **25**, 838–865 (2006).
4. Malik, A. K., Blasco, C. & Picó, Y. Liquid chromatography–mass spectrometry in food safety. *J. Chromatogr. A* **1217**, 4018–4040 (2010).
5. Schymanski, E. L. *et al.* Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem.* **407**, 6237–6255 (2015).
6. Allen, F., Pon, A., Wilson, M., Greiner, R. & Wishart, D. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* **42**, W94–W99 (2014).
7. Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J. & Neumann, S. MetFrag relaunched: incorporating strategies beyond *in silico* fragmentation. *J. Cheminformatics* **8**, 3 (2016).
8. Djoumbou-Feunang, Y. *et al.* BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification. *J. Cheminformatics* **11**, 2 (2019).
9. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
10. Chalcraft, K. R., Lee, R., Mills, C. & Britz-McKibbin, P. Virtual Quantification of Metabolites by Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry: Predicting Ionization Efficiency Without Chemical Standards. *Anal. Chem.* **81**, 2506–2515 (2009).
11. Cech, N. B. & Enke, C. G. Relating Electrospray Ionization Response to Nonpolar Character of Small Peptides. *Anal. Chem.* **72**, 2717–2723 (2000).
12. Wu, L. *et al.* Quantitative structure–ion intensity relationship strategy to the prediction of absolute levels without authentic standards. *Anal. Chim. Acta* **794**, 67–75 (2013).
13. Huffman, B. A., Poltash, M. L. & Hughey, C. A. Effect of Polar Protic and Polar Aprotic Solvents on Negative-Ion Electrospray Ionization and Chromatographic Separation of Small Acidic Molecules. *Anal. Chem.* **84**, 9942–9950 (2012).
14. Henriksen, T., Juhler, R. K., Svensmark, B. & Cech, N. B. The relative influences of acidity and polarity on responsiveness of small organic molecules to analysis with negative ion electrospray ionization mass spectrometry (ESI-MS). *J. Am. Soc. Mass Spectrom.* **16**, 446–455 (2005).
15. Ehrmann, B. M., Henriksen, T. & Cech, N. B. Relative importance of basicity in the gas phase and in solution for determining selectivity in electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **19**, 719–728 (2008).
16. Hermans, J., Ongay, S., Markov, V. & Bischoff, R. Physicochemical Parameters Affecting the Electrospray Ionization Efficiency of Amino Acids after Acylation. *Anal. Chem.* **89**, 9159–9166 (2017).
17. Oss, M., Kruve, A., Herodes, K. & Leito, I. Electrospray Ionization Efficiency Scale of Organic Compounds. *Anal. Chem.* **82**, 2865–2872 (2010).
18. Liigand, J., Kruve, A., Leito, I., Girod, M. & Antoine, R. Effect of Mobile Phase on Electrospray Ionization Efficiency. *J. Am. Soc. Mass Spectrom.* **25**, 1853–1861 (2014).
19. Liigand, J., Laaniste, A. & Kruve, A. pH Effects on Electrospray Ionization Efficiency. *J. Am. Soc. Mass Spectrom.* **28**, 461–469 (2017).
20. Ojakivi, M., Liigand, J. & Kruve, A. Modifying the Acidity of Charged Droplets. *ChemistrySelect* **3**, 335–338 (2018).
21. Kruve, A. & Kaupmees, K. Adduct Formation in ESI/MS by Mobile Phase Additives. *J. Am. Soc. Mass Spectrom.* **28**, 887–894 (2017).
22. Liigand, J., Vries, Rde & Cuyckens, F. Optimization of flow splitting and make-up flow conditions in liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **33**, 314–322 (2019).
23. Liigand, J. *et al.* Transferability of the Electrospray Ionization Efficiency Scale between Different Instruments. *J. Am. Soc. Mass Spectrom.* **26**, 1923–1930 (2015).
24. Liigand, P. *et al.* Think Negative: Finding the Best Electrospray Ionization/MS Mode for Your Analyte. *Anal. Chem.* **89**, 5665–5668 (2017).
25. Kruve, A. & Kaupmees, K. Predicting ESI/MS Signal Change for Anions in Different Solvents. *Anal. Chem.* **89**, 5079–5086 (2017).
26. Kruve, A., Kaupmees, K., Liigand, J. & Leito, I. Negative Electrospray Ionization via Deprotonation: Predicting the Ionization Efficiency. *Anal. Chem.* **86**, 4822–4830 (2014).
27. Kruve, A. Influence of mobile phase, source parameters and source type on electrospray ionization efficiency in negative ion mode: Influence of mobile phase in ESI/MS. *J. Mass Spectrom.* **51**, 596–601 (2016).
28. Liigand, P., Liigand, J., Cuyckens, F., Vreeken, R. J. & Kruve, A. Ionisation efficiencies can be predicted in complicated biological matrices: A proof of concept. *Anal. Chim. Acta* **1032**, 68–74 (2018).
29. Kruve, A., Kaupmees, K., Liigand, J., Oss, M. & Leito, I. Sodium adduct formation efficiency in ESI source: Sodium adduct formation efficiency in ESI source. *J. Mass Spectrom.* **48**, 695–702 (2013).
30. Yap, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **32**, 1466–1474 (2011).
31. Dong, J. *et al.* ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminformatics* **7** (2015).
32. Snyder, L. R., Kirkland, J. J. & Dolan, J. W. *Introduction to Modern Liquid Chromatography*. (John Wiley & Sons, Inc. (2009).
33. Rudakov, O. B., Belyaev, D. S., Khorokhordina, E. A. & Podolina, E. A. Surface tension of binary mobile phases for liquid chromatography. *Russ. J. Phys. Chem. A* **81**, 366–369 (2007).
34. Katz, E., Eksteen, R., Schoenmakers, P. & Miller, N. *Handbook of HPLC*. (M. Dekker (1998).
35. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).
36. Meyer, D. Support Vector Machines. https://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf.
37. *h2oai/h2o-3*. (H2O.ai (2019).
38. Deng, H. & Runger, G. Feature Selection via Regularized Trees. *ArXiv12011587 Cs Stat* (2012).
39. Dzuman, Z., Zachariasova, M., Veprikova, Z., Godula, M. & Hajslova, J. Multi-analyte high performance liquid chromatography coupled to high resolution tandem mass spectrometry method for control of pesticide residues, mycotoxins, and pyrrolizidine alkaloids. *Anal. Chim. Acta* **863**, 29–40 (2015).

40. Trufelli, H., Palma, P., Famiglini, G. & Cappiello, A. An overview of matrix effects in liquid chromatography–mass spectrometry. *Mass Spectrom. Rev.* **30**, 491–509 (2011).
41. Krasilnikov, V. N., Batalova, G. A., Popov, V. S. & Sergeyeva, S. S. Fatty Acid Composition of Lipids in Naked Oat Grain of Domestic Varieties. *Russ. Agric. Sci.* **44**, 406–408 (2018).
42. Ismaiel, O. A., Halquist, M. S., Elmamly, M. Y., Shalaby, A. & Thomas Karnes, H. Monitoring phospholipids for assessment of ion enhancement and ion suppression in ESI and APCI LC/MS/MS for chlorpheniramine in human plasma and the importance of multiple source matrix effect evaluations. *J. Chromatogr. B* **875**, 333–343 (2008).
43. Stravs, M. A., Schymanski, E. L., Singer, H. P. & Hollender, J. Automatic recalibration and processing of tandem mass spectra using formula annotation: Recalibration and processing of MS/MS spectra. *J. Mass Spectrom.* **48**, 89–99 (2013).
44. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
45. Blacquière, T., Smagghe, G., van Gestel, C. A. M. & Mommaerts, V. Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. *Ecotoxicology* **21**, 973–992 (2012).

Acknowledgements

This work was financially supported by Smart specialization doctoral stipend, PRG300 from Estonian Research Council, Otto Mønsted Foundation, and Fødevareforlig 3. To all financing sources the authors are greatly indebted. We thank Stefan Schoder and Asko Laaniste for fruitful discussions. Open access funding provided by Stockholm University.

Author contributions

J.L. carried out most of the ionization efficiency measurements, all of the modelling, and wrote significant part of the manuscript. T.W. carried out all of the measurements for validation, including concentration prediction of pesticides. J.S. helped to plan and analyse data related to pesticide analysis. J.K. and N.C. helped in concept verification. A.K. designed the project, helped to plan the experiments and modelling and wrote significant part of the manuscript. All authors contributed significantly to writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-62573-z.

Correspondence and requests for materials should be addressed to A.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Manuscript 4: Semi-quantification application between labs

Food Chemistry 318 (2020) 126460

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Standard substances free quantification makes LC/ESI/MS non-targeted screening of pesticides in cereals comparable between labs

Tingting Wang[a,1], Jaanus Liigand[b,1], Henrik Lauritz Frandsen[a], Jørn Smedsgaard[a], Anneli Kruve[b,c,*]

[a] National Food Institute, Research Group for Analytical Food Chemistry, Technical University of Denmark, Kemitorvet Building 202, Kgs. Lyngby, DK-2800, Denmark
[b] University of Tartu, Institute of Chemistry, Ravila 14a, Tartu 50411, Estonia
[c] Department of Environmental Science and Analytical Chemistry, Stockholm University, Svante Arrhenius väg 16, 106 91 Stockholm, Sweden

ARTICLE INFO

Keywords:
Quantification without standards
Pesticides
LC/ESI/MS
Food monitoring
Cereal

ABSTRACT

LC/ESI/MS is the technique of choice for qualitative and quantitative food monitoring; however, analysis of a large number of compounds is challenged by the availability of standard substances. The impediment of detection of food contaminants has been overcome by the suspect and non-targeted screening. Still, the results from one laboratory cannot be compared with the results of another laboratory as quantitative results are required for this purpose. Here we show that the results of the suspect and non-targeted screening for pesticides can be made quantitative with the aid of in silico predicted electrospray ionization efficiencies and this allows direct comparison of the results obtained in two different laboratories. For this purpose, six cereal matrices were spiked with 134 pesticides and analysed in two independent labs; a high correlation for the results with the R^2 of 0.85.

1. Introduction

LC/ESI/MS is widely used for qualitative and quantitative analysis in food monitoring. This technique allows for the detection of a wide range of compounds such as pesticides, mycotoxins, veterinary drugs, other contaminants, and their metabolites in various food products (Dzuman, Zachariasova, Veprikova, Godula, & Hajslova, 2015; Jamin, Bonvallot, & Tremblay-franco, 2014; Mol et al., 2008). In ESI source, different compounds possess vastly different ionization efficiencies, the number of gas-phase ions produced at the same molar concentration; therefore, the quantification of these compounds requires corresponding standard substances to establish the relation between MS response and analyte concentration (Hermans, Ongay, Markov, & Bischoff, 2017; Kruve, Kaupmees, Liigand, & Leito, 2014). For example, more than 1000 different compounds against pests are in use worldwide (EU – Pesticides database; Li & Jennings, 2017). Thus, it is common that pesticide residue analysis includes 500 or more analytes in the routine scope and the same number of standard substances needs to be purchased and included in the sequence. Unfortunately, standard substances are often unavailable for various reasons and it is even more of a predicament when new active substances are put on the market, residue definitions are changed to include metabolites of compounds or when compounds that are newly discovered or unstable.

Non-targeted analysis and suspect screening earn much attention in food safety monitoring in these years, as these allow identifying compounds present in the sample from MS data without the need for authentic standard substances (Fraser et al., 2014; Hird, Lau, Schuhmacher, & Krska, 2014). However, for quantification purposes, standard substances are still required. As numerous tentatively identified compounds are often found it becomes either impossible or too expensive to purchase all these tentatively identified compounds. Furthermore, the peak areas obtained for different samples can be compared only if run in the same sequence and no matrix effect is expected. Comparison between different days, instruments or laboratories is impossible. Therefore, the ability to obtain quantitatively comparable results that do not depend on the instrument and lab is of utmost interest and importance (Kruve & Kaupmees, 2017).

To overcome this problem some semi-quantification approaches have been proposed. An approach based on chemical similarity was proposed by, eg., Kalogiouri, Aalizadeh, and Thomaidis (2017) who selected several most appropriate standard substances for semi-quantification according to the similarity calculation of molecular descriptors between tested molecules and quantification standard substances (14 phenolic compounds). Pieke, Granby, Trier, and

* Corresponding author at: Department of Environmental Science and Analytical Chemistry, Stockholm University, Svante Arrhenius väg 16, 106 91 Stockholm, Sweden.
E-mail address: anneli.kruve@su.se (A. Kruve).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.foodchem.2020.126460
Received 19 August 2019; Received in revised form 28 January 2020; Accepted 19 February 2020
Available online 20 February 2020

Smedsgaard (2017) demonstrated a proof of concept strategy for semi-quantification using LC/ESI/MS combining generic ESI ion source optimization using 18 quantification markers where the marker with the smallest difference in retention time was chosen to give the best average prediction of the concentration of the analytes. However, due to the limited number of test compounds and lack of validation with more compounds, the prediction quality is still unclear.

The prediction of ionization efficiency (IE) has the potential to offer higher accuracy and a wider application range. Recent attempts to predict ESI response have considered multiple physiochemical parameters simultaneously, and a correlation between ionization efficiency and different physicochemical properties (e.g., gas-phase basicity, log*P*, polarity, pK_a, surface area, molar volume, etc.) has been observed (Alymatiri, Kouskoura, & Markopoulou, 2015; Cramer, Johnson, & Kamel, 2017; Fusaro, Mani, Mesirov, & Carr, 2009; Golubović, Birkemeyer, Protić, Otašević, & Zečević, 2016; Hermans et al., 2017; Huffma, Poltash, & Hughey, 2012; Oss, Kruve, Herodes, & Leito, 2010; Wang et al., 2019; Wu et al., 2013). The application scopes have been confined to the specific type of compounds such as drug-like molecules (Cramer et al., 2017), amino acids (Hermans et al., 2017), peptides (Raji et al., 2009), steroids (Alymatiri et al., 2015), or organic acid et al (Wu et al., 2013). Unfortunately, these approaches have been hardly used for quantification. The solvent also has an important effect on ionization efficiency and needs to be taken into account, which is often neglected.

Our group (Kruve & Kaupmees, 2017; Kruve et al., 2014; P; Liigand, Liigand, Cuyckens, Vreeken, & Kruve, 2018; Oss et al., 2010) has previously proposed a more generic model to predict ionization efficiencies and the ionization efficiency predictions have been used to obtain the concentrations of small molecules analysis with gradient elution in ESI negative mode (Kruve & Kaupmees, 2017) for the analysis of standard substances. Additionally, we have shown that ionization efficiencies can be predicted for compounds in case of the analysis of complicated matrices (blood, urine, etc.) with flow injection analysis (Liigand et al., 2018). Very recently we have developed a fast and easy Quantem approach to predict the ionization efficiencies for a wide range of compounds and to use it for estimating the concentration of the compounds in LC/HRMS non-targeted screening (Liigand, Wang, Kellogg, Smedsgaard, Cech, & Kruve, 2019).

Based on the previously obtained promising results we aim to evaluate whether ionization efficiency prediction model can be used to reach quantitatively comparable results for suspect screening between labs for the analysis in food samples. Here we focus on the analysis of pesticides in cereals as cereals are one of the largest foods commodity and consumed worldwide (González-Curbelo, Herrera-Herrera, Ravelo-Pérez, & Hernández-Borges, 2012), there is an intense attention to the occurrence of pesticides and mycotoxins exposure from cereal to maintain a high food safety standard. We, firstly, apply the Quantem model to predict the ionization efficiencies of detected pesticides to obtain the concentrations (using 4 orders of magnitudes spiking levels) in six common cereal matrices (oat, barley, rice, rye, wheat, and maize) (Liigand et al., 2019). Secondly, we evaluate the robustness of the concentration predictions of 134 common pesticides by comparing the results obtained from two laboratories in Denmark and Estonia on two different mass spectrometric setups (Agilent 6495 triple quadrupole and Bruker Daltonics micro-TOFq).

2. Materials and methods

2.1. Chemicals and reagents

Milli-Q grade water was used for all analyses (18.2 MΩ·cm, TOC 1–2 ppb, prepared by a Millipore Milli-Q Advantage A10 water-purification system). Acetonitrile (LC-MS grade), and formic acid (LC–MS grade) were purchased from Sigma-Aldrich (Steinheim, Germany). Magnesium sulfate (purity > 99%) was purchased from Lach:ner (Neratovice, Czech Republic), and absorbent Bondesil-C18 was purchased from Agilent Technologies (Santa Clara, CA, USA). All other solvents used in this study were of analytical grade.

Thirty standard substances that were used to transfer the model from one lab to another were purchased from Sigma-Aldrich (Steinheim, Germany) or synthesized in-house (Liigand et al., 2019). All the pesticide standard substances (all with a purity greater than 96%) were purchased from Dr. Ehrenstorfer (Augsburg, Germany). The pesticide standard substances stock solutions of 1 mg/mL were prepared in toluene and stored at −18 °C. A standard mixture of 10 μg/mL in 80% acetonitrile was prepared from these stock solutions. Working solutions in the concentration range of 4 nM to 35 μM on 6 different concentration levels were prepared by diluting the standard mixture with the extracts of blank cereal as a matrix (free of pesticide residues). Due to the instability of some of the compounds the same spiked samples could not be analyzed in two labs. To overcome this obstacle, the same sample preparation was carried out in both labs and the extracts were spiked. The small differences in the spiking experiments were determined by weighting and corrected thereafter. A full list of all the pesticides and compounds used for transferring the model (structures, retention time, isotope distribution, concentration, et al.) are shown in Table S1 in Supporting information. The extracts used for the matrix matching were obtained by the extraction and clean-up procedure described below.

2.2. Samples collection and sample extraction

The blank samples (free of pesticides) of cereals specifically grown for preparation of European Union proficiency test material (EUPT) for the six EUPTs: EUPT-CF8 (wheat), EUPT-C3 (oat), EUPT-CF10 (rye) and EUPT-C6 (barley), EUPT-CF9 (maize), EUPT-SRM6 (rice).

The extraction was based on the generic QuEChERS method developed by Dzuman's group (Dzuman et al., 2015). Two grams of homogenized cereal samples (wheat, oat, rye, rice, barley, maize) were weighed into a 50 mL polypropylene centrifugation tube, and 10 mL acidified Milli-Q water containing 0.2% formic acid was added and left to soak the matrix for 0.5 h. Then, 10 mL acetonitrile was added, and the samples were extracted for 0.5 h using a laboratory shaker (Type 358S, Lubawa, Poland). Thereafter 4 g of magnesium sulfate and 1 g of sodium chloride were added, and the tube was shaken for 1 min followed by centrifugation (10 min, 4500 RPM; Centrifuge 5430R, Eppendorf AG, Hamburg, Germany). The organic upper layer (2 mL) was collected into a clean tube and shaken with 0.1 g of Bondesil-C18 and 0.3 g of magnesium sulfate for 2 min followed by centrifugation (10 min, 4500 RPM). Then 1.5 mL of purified supernatant was transferred into an autosampler vial with insert and spiked with different concentration of tested compounds prior to injection on LC-MS system.

2.3. Instrumental parameters

2.3.1. Agilent 6495 triple quadrupole

Samples were analyzed on an Agilent 1290 ultrahigh performance liquid chromatograph (Agilent Technologies, CA, U.S.) coupled to an Agilent 6495 triple quadrupole instrument (QQQ) in University of Tartu (UT). Samples were injected onto an Agilent Zorbax RRHD SB-C18 reversed-phase column (1.8 μm, 2.1 × 50 mm). The mobile phase consisted of water (A) containing 0.1% formic acid and acetonitrile (B). The analysis was done using a gradient elution at a flow rate of 0.3 mL/min at 30 °C. The gradient was from 5% to 100% of acetonitrile in 7 min, then acetonitrile was maintained at 100% for 2 min and returned back to 5% of acetonitrile in 2 min, and maintaining starting conditions at 5% of B for 2 min equilibration with 5% of B to giving a total runtime of 13 min.

An Agilent's Jet Stream electrospray ionization (ESI) interface was used in positive ion mode (ESI+) with the following settings: capillary voltage 3 kV, nebulizer pressure 20 psi, sheath gas flow rate 11 L/min,

sheath gas temperature 350 °C, dry gas temperature 250 °C, and dry gas flow rate 14 L/min. Spectra were collected from m/z 100 to 1100 Da. The injection volume was 1 µL. Peaks were manually integrated, and peak areas were used for all calculations.

2.3.2. Bruker Daltonics micro-TOFq

Samples were also analyzed on an Agilent 1200 HPLC (Agilent Technologies, CA, U.S.) coupled to a Bruker Daltonics micro-TOFq (QToF) mass spectrometer (Bruker Daltonics, Bremen, Germany) in the lab of Technical University of Denmark (DTU). Samples were injected onto a Nucleoshell C18 reversed-phase column (2.7 µm, 100 × 2 mm, MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). The mobile phase consisted of water (A) containing 2.5 mM ammonium formate at pH = 3.0, and acetonitrile (B). The analysis was done using a gradient elution at a flow rate of 0.3 mL/min at 40 °C: isocratic at 2% of B for 1.0 min, then increased to 100% of B at 22.0 min, and maintained to 25 min, whereafter the gradient returned to starting conditions 2% of B at 25.1 min followed by 2.9 min equilibration with 2% of B to giving a total runtime of 28 min.

ESI+ was used in the following settings: capillary voltage 3 kV, nebulizer pressure 20 psi, dry gas temperature 200 °C, and dry gas flow rate 8 L/min. Funnel radio frequency (RF) 400 Vpp, multipole RF 200 Vpp. Spectra were collected from m/z 100 to 1100 Da at a rate of 2 Hz.

2.4. Matrix effects, LoD, LoQ

The matrix effect (ME) was calculated with the following formula in Eq. (1):

$$\%ME = \left(\frac{Slope_{matrices}}{Slope_{solvent}}\right) \times 100\% \quad (1)$$

The calibration curves were prepared as the peak area versus the concentration of each analyte. $Slope_{matrices}$ is the slope of the calibration curves of analytes spiked into the cereal samples, $Slope_{solvent}$ is the slope of the calibration curves of analytes into the solvent.

$$LoD = 3.3 \times \frac{S_{residuals}}{slope} \quad (2)$$

$$LoQ = 10 \times \frac{S_{residuals}}{slope} \quad (3)$$

Detection of limit (LoD) and detection of quantification (LoQ) were calculated based formula Eqs. (2) and (3) (Evard, Kruve, & Leito, 2016), $S_{residuals}$ is the residual standard deviation of the calibration curve of each compound in the LoD region, and the *slope* is the slope value of the calibration curve of the analytes.

2.5. Model application

Previously developed Quantem model for predicting electrospray ionization efficiencies (log*IE*, predictions are made in the logarithmic scale) together with the transformation approach (Liigand et al., 2019) was used to estimate the concentration of the compounds for the results of both labs. For this empirical solvent descriptors (viscosity, surface tension, polarity index, pH, NH_4 presence) and 1444 PaDEL molecular descriptors were used for the model development (Yap, 2011). Solvent descriptors were calculated based on the mobile phase used in the study and the retention time of each compound. PaDEL descriptor calculator (Dong et al., 2015) was used to calculate the molecular descriptors of all the compounds (http://www.scbdd.com/padel_desc/index/). For the transformation, the same compounds as described previously (Liigand et al., 2019) were used.

For transferring the predicted ionization efficiency values, a set of compounds (either spiked to the matrix or as a standard solution) was measured in the linear range at the beginning and the end of the sequence with the same method as analytes. From the analysis results, logarithmic response factors (log*RF*) were calculated for transformation compounds as Eq. (4):

$$\log RF = \log\frac{Signal \cdot IC}{c_{spiked}} \quad (4)$$

where *Signal* is the MS1 peak area and *IC* is an isotopic correction, which is the sum of the abundances of all isotopomers for each compound. Molar concentrations are used for calculation of response factor.

Next ionization efficiencies were predicted using the Quantem model (Liigand et al., 2019). The generic ionization efficiency values predicted with Quantem model need to be transferred to the instrument and method-specific response factors. To obtain this the logarithmic response factors of transformation compounds were correlated with predicted ionization efficiencies of the same compounds. The predicted logarithmic response factors of analytes were obtained using Eq. (5):

$$\log RF_{pred} = Slope \times \log IE_{pred} + intercept \quad (5)$$

where slope and intercept values are determined from the above-mentioned correlation graph of logarithmic response factors of transformation compounds vs their predicted ionization efficiencies. As analysts are mainly interested in concentrations, the concentrations are calculated from predicted response factors of analytes according to Eq. (6):

$$c_{pred} = \frac{Signal \cdot IC}{10^{\log RF_{pred}}} \quad (6)$$

where *Signal* is MS1 peak area, and *IC* is an isotopic correction, which is the sum of the abundances of all isotopomers for each compound. As a result, molar concentrations are obtained.

To estimate the accuracy of prediction the absolute times differences of spiked concentration and predicted concentration in both cereal matrices as well as in pure solvent were calculated as Eq. (7).

$$\text{Prediction error} = \max\left\{\frac{c_{pred}}{c_{spiked}}, \frac{c_{spiked}}{c_{pred}}\right\} \quad (7)$$

3. Results and discussion

3.1. Predicting ionization efficiency values

The Quantem model (Liigand et al., 2019) developed previously, was used to predict the ionization efficiencies of the studied compounds. These values were transferred to instrument-specific response factors (log*RF*) with the aid of Eq. (4) and good transformation accuracy was observed both on QQQ in UT (R^2 = 0.77) and QToF in DTU (R^2 = 0.66).

The transformation compounds are used for transferring the ionization efficiency values into instrument-specific values and should be measured under conditions as close as possible to the actual sample analysis. Therefore, it is of interest if the transformation compounds need to be spiked into the sample or can be analyzed as a separate mixture within the same analysis sequence. To test this transformation compounds were spiked into two matrices (barley and oat) and were run as a separate standard solution. The concentration prediction accuracy for 109 randomly selected pesticides were compared. It was seen that if the model transferring compounds were spiked into the matrices, the mean errors changed only slightly from 3.8 to 3.9 times, thus there is hardly any difference (Table S2). Therefore, it is feasible to transform the predicted ionization efficiency values by measuring the transformation compounds in the standard solution instead of spiking into specific matrices.

3.2. Ionization efficiency and concentration prediction in solvent and matrices.

3.2.1. Concentration prediction in a solvent on QQQ

In order to determine the applicability of this Quantem model for a wider range of compounds, a set of pesticides with varying chemical nature was chosen and measured. The studied set on QQQ includes 139 common pesticides from different compound classes, such as amide, anilide, ether, pyrimidine, ketone. Each pesticide was studied at 6 concentration levels covering more than 4 orders of magnitude in both matrices and solvent. The spiked concentrations ranged from 10 nM to 35 μM. The Maximum Residue Levels (MRL) for pesticides in cereal samples is 0.01–0.02 ppm. This corresponds to a pesticide concentration of ~100 nM, which is in the lower end of the studied concentration range.

The predicted log*IE* values (Table S1) for all the tested compounds varied from 1.3 to 4.5. From the prediction of concentration in pure solvents, we found that for 48% of compounds, the concentration prediction error is better than 2 times, for 87% of compounds better than 5 times, and for 98% of studied compounds better than 10 times. The highest prediction error of concentration was 16 times. The 3 best-performing pesticides were imazamox, tetraconazole, and metrafenone (1.1 times for all). It is evident that the model is universal for both nitrogen as well as for oxygen bases, as among the top performers are both nitrogen and oxygen bases. Unsurprisingly, compounds (altogether 4 compounds, e.g., propiconazole) included previously in developing the log*IE* prediction model show high performance (Liigand et al., 2019). The correlation between measured and predicted concentration values in the solvent is presented in Fig. 1; it can be seen that the agreement between measured and predicted concentration of test compounds by the model was good, the average accuracy was 3.0 times.

Comparing the different concentration levels, the accuracy was consistent except for the lowest concentration level, where it is only slightly lower (4.1 times).

3.2.2. Concentration prediction in matrices on QQQ

The performance of Quantem model for prediction of log*IE* values for 6 cereal matrices was evaluated for the analyses of the same compounds at the same concentration levels as in solvent. Altogether, 6

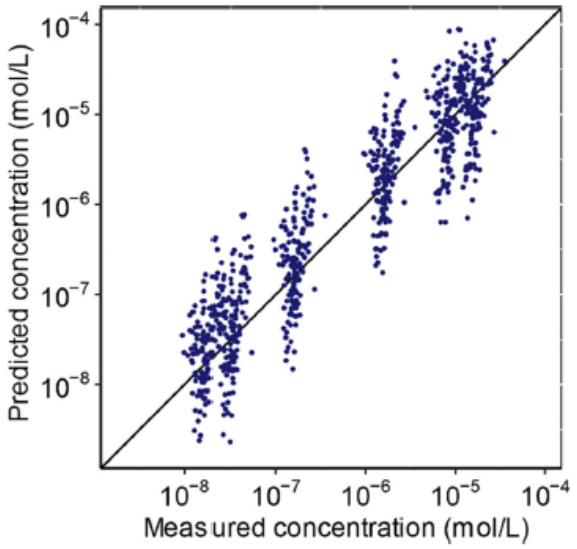


Fig. 1. Correlation between the predicted and measured concentrations for the pesticides in solvent (QQQ in UT).

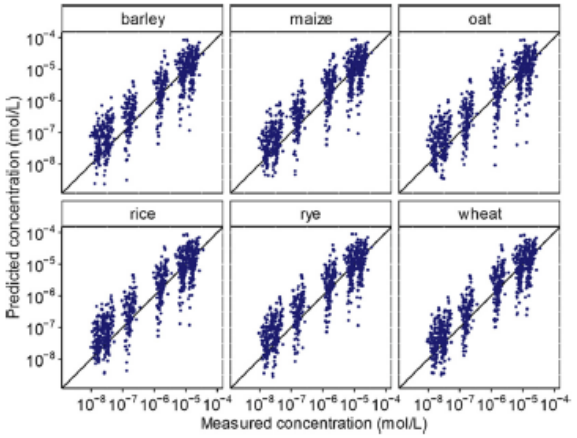


Fig. 2. Correlation between the predicted and measured concentrations for pesticides in 6 cereal matric extracts. The measurements were performed on the QQQ instrument in UT.

concentration levels from 10 nM to 35 μM were used. The detail prediction error results in different matrices were shown in Table S3. It was found the mean errors of concentration mismatch between predicted and measured values in all cereal matrices was 3.8 times (Fig. 2), the average concentration prediction error for each cereal matrix varied from 3.3 times to 4.6 times. In general, higher prediction accuracy was observed for wheat than for the other matrices (prediction error of 3.3 times), oat performed worst (4.6 times). Comparing all different concentration levels (Table S5), the prediction error of the lowest concentration level was slightly higher (4.4 times). For 46% of compounds, the concentration prediction error was less than 2 times, for 81% of compounds less than 5 times, and for 94% of compounds less than 10 times.

The concentration prediction accuracy of carbosulfan was most severely influence by matrices (prediction error 64.6 times), which had the highest difference while compared to prediction accuracy in solvent (3.9 times). The reason of carbosulfan's poor performance could be that it contains the N-S-N moiety which was not present in the training set of the Quantem model. On the other hand, carbosulfan is unstable, it decomposes slowly at room temperature especially in acidic organic solvents (Barros & Barros, 1995; Leppert, Markle, Helt, & Fujie, 1983). The QuEChERS sample preparation used in this study uses 0.2% formic acid in the sample extraction procedure. We also observed that the degradation of carbosulfan results in poor linearity in matrices. It can be found that carbosulfan in oat matrice results in the largest concentration prediction error of 173.8 times. Carbosulfan was followed by fluometuron (15.9 times), pirimicarb (15.7 times), pyridaben (15.4 times), and pyridate (14.1 times) on average in all matrices.

3.2.3. Matrix effect

The results are influenced by the matrix effect: the ionization suppression due to the matrix compounds co-eluting with the analyte. The most severe matrix effect was observed in oat matrix, the matrix effect was in the range of 2% (suppressed by 98%) to 109% (enhanced by 9%) for all compounds relative to the pure solvent (Table S7). In comparison to other cereals, oat is a rich source of polar lipids and free fatty acids. It has been discussed that phospholipids and free fatty acids will cause matrix effects in ESI (Krasilnikov, Batalova, Popov, & Sergeyeva, 2018; Pucci, Di, Alfieri, Bonelli, & Monteagudo, 2009), which can explain the severe matrix effect found in oat. The matrix effect may vary with analyte concentration (Furey, Moriarty, Bane, Kinsella, & Lehane, 2013; Kruve, Auling, Herodes, & Leito, 2011).

The concentration prediction errors were in general consistent for

most compounds in both solvent and matrices, which means that the matrix effect does not influence the prediction accuracy significantly and the transformation step help to account for the matrix effects. For example, fluometuron, pyridaben, pirimicarb, pirimicarb-desmethyl, EPN, and thiamethoxam, had low prediction accuracy in both solvent and matrices. For some compounds, specific matrices caused a notable decrease in prediction accuracy. Carbosulfan and pyridate suffered the most severe matrix effects, carbosulfan was suppressed 98% and 95% in oat and maize matrices (concentration prediction error of 173.8 times and 55.5 times), respectively. Signal of pyridate was suppressed 87% and 73% in oat and maize matrices (concentration prediction error of 20.3 times and 24.1 times), respectively. Signal of benfuracarb was suppressed 58% in the oat matrix, which results in 5.9 times concentration prediction error. Spirodiclofen and pyridaben were also suppressed 53% and 47% in the matrices, resulted in concentration prediction error of 10.3 times and 15.4 times on average.

3.2.4. Concentration prediction in matrices on QToF

In order to test the robustness of the Quantem model and the transferability of the ionization efficiency prediction model between labs, the same matrices spiked at the 6 concentration levels from 4 nM to 11 μM with QToF in DTU. The ionization efficiency prediction and transformation approach were the same as in the case of QQQ instrument. There were 134 pesticides in common that could be detected and quantified both in DTU and UT. After model transformation, the predicted log*IE* values (Table S4) for all the tested compounds on QToF varied from 1.6 to 4.5.

As shown in Fig. 3, the concentration prediction accuracy was even better on QToF in DTU. The average concentration prediction errors for all compounds was 2.4 times. For 61% of the compounds, the prediction error was better than 2 times and for 94% of compounds better than 5 times. These values are reasonably close to the ones obtained on the QQQ instrument in UT and described above. The average concentration prediction errors were 2.3 times in solvent and 2.6 times in all matrices. The prediction errors in different matrices were in the range from 2.5 to 3 times, which is quite closed. Rice and rye performed slightly better than other matrices. The concentration prediction errors were still in general consistent for most compounds in solvent and matrices. As the QToF instrument used for analysis in DTU is high-resolution mass spectrometry, the ability to distinguish the molecules with close MS weight is better compared to QQQ; also, the total running time for each sample in DTU is longer than the running time in UT which leads to a better separation on HPLC part. As a result, the integration of the compounds on QToF was influenced less by other compounds and the matrices background than QQQ. This might explain the slightly better performance of the concentration prediction for the data in DTU than UT.

On the whole, the quantitation of pesticides in cereal matrices with LC/ESI/MS without standard substances is feasible to achieve the prediction error of less than 4 times on average both on QToF in DTU and QQQ in UT.

3.2.5. Comparison of the predicted concentrations between labs

Moreover, after applying Quantem model we can also directly compare the concentrations obtained both in UT and DTU. The methods applied in UT and DTU are intrinsically different. These use slightly different chromatography, different mobile phase, different ESI source as well as different mass analyzers and vendors. As a result, the chromatograms obtained in both labs are not comparable. However, as the Quantem model translate these chromatograms into quantitative data for pesticides and the results can be numerically compared. Here we observed a very high correlation for the results from UT and DTU (R^2 = 0.85). Also, the slope and intercept of the correlation line were very close to the ideal fit, the slope was 0.95 while the ideal would be 1. The results are shown in Fig. 4. From this, it is obvious that for the first time the measurement results from two different laboratories can be made quantitatively comparable for suspect screening without the necessity of standard substances. This finding is extremely valuable for applications in both food monitoring and environmental areas.

3.2.6. LoD and LoQ

In spite of the generally good correlation between the predicted concentrations for the analysis of two different labs (UT and DTU), some of the compounds were detected only by one of the labs. In the case of interlaboratory comparisons, such differences often occur and commonly indicate different LoD and LoQ values. Therefore, we compared the estimated concentrations of the compounds detected by both labs with the concentrations of the compounds detected only in DTU with QToF instrument. The comparison revealed that most of the compounds not detected by UT were having among some of the lowest concentrations (Fig. 5). Therefore, the comparison for the concentration determined from different labs has the potential to bring clarity to the

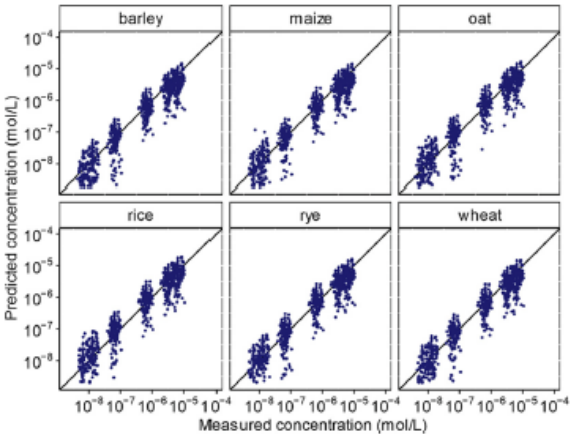


Fig. 3. Correlation between the predicted and measured concentrations for pesticides in 6 cereal matric extracts. The measurements were performed on QToF instrument in DTU.

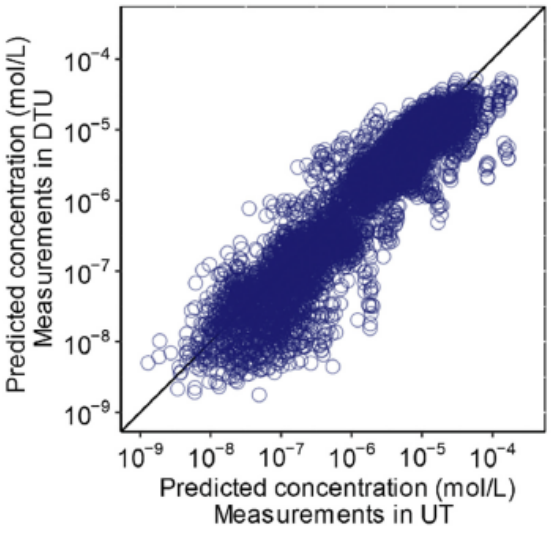


Fig. 4. Correlation between the predicted concentrations for pesticides in 6 cereal matrices measured on QQQ at UT and on QToF at DTU.

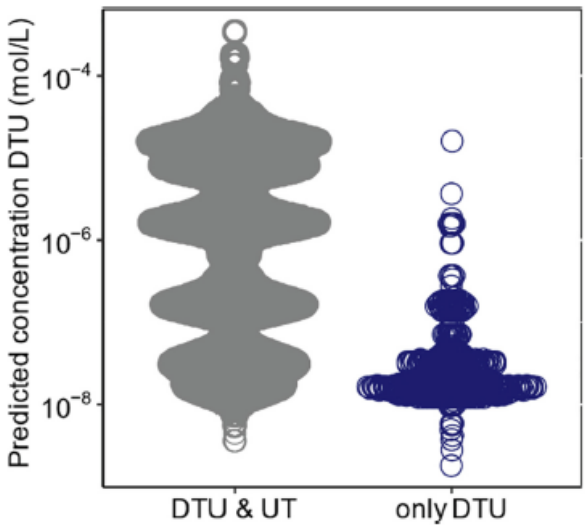


Fig. 5. The predicted concentrations of pesticides for compounds detected in both laboratories vs detected only in DTU. The results are summarized over all studied matrices.

discrepancies often observed in collaborative trials of non-targeted screening.

To further validate the conclusion, we determine the LoD values of all of the studied compounds. This was possible as the concentrations of the pesticides were known from the spiking experiments. A conservative LoD determination strategy was used (Evard et al., 2016). It was observed that the LoD values ranged between 0.1 nM and 1 μM, while most values were between 1 nM and 0.1 μM. The LoD and LoQ values are brought in Table S9. The comparison of the determined LoD values for UT and DTU, in Fig. S1, revealed that different compounds had varying LoD values between different labs. Sometimes the LoD was lower in UT, sometimes in DTU.

4. Conclusion

Here we have shown, for the first time that results obtained from suspect screening can be made quantitatively comparable between different laboratories using different instruments as well as different analysis methods. To achieve this, we use the ionization efficiency prediction with Quantem model to transfer the obtained signals into concentrations. In contrast to signals, the concentrations are comparable between different instruments and labs. The consistency between the results obtained independently in two labs was very high (R^2 = 0.85) in spite of the differences in the instrumentation and methods.

Also, the independent results show high accuracy. The quantitation of pesticides in different cereals matrices without standard substances is feasible with the accuracy of less than 4 times on average for chromatographic analyses. It can serve as a useful tool for preliminary quantitation in various fields such as food control, metabolomics, and environmental research, and there is a significant improvement compared to the conventional approach. Also, the obtained accuracy is sufficient for using the standard substance-free quantitation as the first step in food monitoring.

CRediT authorship contribution statement

Tingting Wang: Investigation, Formal analysis, Validation, Writing - original draft. **Jaanus Liigand:** Data curation, Formal analysis, Software, Writing - original draft. **Henrik Lauritz Frandsen:** Supervision. **Jørn Smedsgaard:** Supervision, Writing - review & editing. **Anneli Kruve:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was financially supported by Smart specialization doctoral stipend, PRG300 from Estonian Research Council, Otto Mønsted Foundation, and Fødevareforlig 3. To all financing sources the authors are greatly indebted.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2020.126460.

References

Alymatiri, C. M., Kouskoura, M. G., & Markopoulou, C. K. (2015). Decoding the signal response of steroids in electrospray ionization mode (ESI-MS). *Analytical Methods, 7*, 10433–10444. https://doi.org/10.1039/C5AY02839F.

Barros, M. W., & Barros, A. (1995). Determination of Carbosulfan in Oranges by High-performance Liquid Chromatography With Post-col umn Fluorescence, 120, pp. 2479–2481.

Cramer, C. J., Johnson, J. L., & Kamel, A. M. (2017). Prediction of mass spectral response factors from predicted chemometric data for druglike molecules. *Journal of the American Society for Mass Spectrometry, 28*, 278–285. https://doi.org/10.1007/s13361-016-1536-4.

Dong, J., Cao, D. S., Miao, H. Y., Liu, S., Deng, B. C., Yun, Y. H., ... Chen, A. F. (2015). ChemDes: An integrated web - based platform for molecular descriptor and fingerprint computation. *Journal of Cheminformatics, 7*(60), 1–10. https://doi.org/10.1186/s13321-015-0109-z.

Dzuman, Zbynek, Zachariasova, Milena, Veprikova, Z., Godula, M., & Hajslova, J. (2015). Multi-analyte high performance liquid chromatography coupled to high resolution tandem mass spectrometry method for control of pesticide residues, mycotoxins, and pyrrolizidine alkaloids. *Analytica Chimica Acta, 863*(1), 29–40. https://doi.org/10.1016/j.aca.2015.01.021.

EU - Pesticides database. (n.d.). Retrieved from http://ec.europa.eu/food/plant/pesticides/eu-pesti.

Evard, H., Kruve, A., & Leito, I. (2016). Tutorial on estimating the limit of detection using LC-MS analysis, part I: Theoretical review. *Analytica Chimica Acta, 942*, 23–39. https://doi.org/10.1016/j.aca.2016.08.043.

Fraser, K., Lane, G. A., Otter, D. E., Harrison, S. J., Quek, S. Y., Hemar, Y., & Rasmussen, S. (2014). Non-targeted analysis by LC-MS of major metabolite changes during the oolong tea manufacturing in New Zealand. *Food Chemistry, 151*, 394–403. https://doi.org/10.1016/j.foodchem.2013.11.054.

Furey, A., Moriarty, M., Bane, V., Kinsella, B., & Lehane, M. (2013). Ion suppression; A critical review on causes, evaluation, prevention and applications. *Talanta, 115*, 104–122. https://doi.org/10.1016/j.talanta.2013.03.048.

Fusaro, V. A., Mani, D. R., Mesirov, J. P., & Carr, S. A. (2009). Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nature Biotechnology, 27*(2), 190–198. https://doi.org/10.1038/nbt.1524.

Golubović, J., Birkemeyer, C., Protić, A., Otašević, B., & Zečević, M. (2016). Structure – response relationship in electrospray ionization-mass spectrometry of sartans by artificial neural networks. *Journal of Chromatography A, 1438*, 123–132. https://doi.org/10.1016/j.chroma.2016.02.021.

González-Curbelo, M.Á., Herrera-Herrera, A. V., Ravelo-Pérez, L. M., & Hernández-Borges, J. (2012). Sample-preparation methods for pesticide-residue analysis in cereals and derivatives. *Trends in Analytical Chemistry, 38*, 32–51. https://doi.org/10.1016/j.trac.2012.04.010.

Hermans, J., Ongay, S., Markov, V., & Bischoff, R. (2017). Physicochemical parameters affecting the electrospray ionization efficiency of amino acids after acylation. *Analytical Chemistry, 89*, 9159–9166. https://doi.org/10.1021/acs.analchem.7b01899.

Hird, S. J., Lau, B. P.-Y., Schuhmacher, R., & Krska, R. (2014). Liquid chromatography-mass spectrometry for the determination of chemical contaminants in food. *TrAC Trends in Analytical Chemistry, 59*, 59–72. https://doi.org/10.1016/j.trac.2014.04.005.

Huffma, B. A., Poltash, M. L., & Hughey, C. A. (2012). Effect of polar protic and polar aprotic solvents on negative-ion electrospray ionization and chromatographic separation of small acidic molecules. *Analytical Chemistry, 84*, 9942–9950. https://doi.org/10.1021/ac302397b.

Jamin, E. L., Bonvallot, N., & Tremblay-franco, M. (2014). Untargeted profiling of pesticide metabolites by LC – HRMS: An exposomics tool for human exposure evaluation. *Analytical and Bioanalytical Chemistry, 406*, 1149–1161. https://doi.org/10.1007/s00216-013-7136-2.

Kalogiouri, N. P., Aalizadeh, R., & Thomaidis, N. S. (2017). Investigating the organic and conventional production type of olive oil with target and suspect screening by LC-QTOF-MS, a novel semi-quantification method using chemical similarity and advanced chemometrics. *Analytical and Bioanalytical Chemistry, 49*(3), 5413–5426. https://doi.org/10.1007/s00216-017-0395-6.

Krasilnikov, V. N., Batalova, G. A., Popov, V. S., & Sergeyeva, S. S. (2018). Fatty acid composition of lipids in naked oat grain of domestic varieties. *Russian Agricultural Sciences, 44*(5), 406–408. https://doi.org/10.3103/S1068367418050117.

Kruve, A., Auling, R., Herodes, K., & Leito, I. (2011). Study of liquid chromatography/electrospray ionization mass spectrometry matrix effect on the example of glyphosate analysis from cereals. *Rapid Commun. Mass Spectrom, 25*, 3252–3258. https://doi.org/10.1002/rcm.5222.

Kruve, A., & Kaupmees, K. (2017). Predicting ESI/MS signal change for anions in different solvents. *Analytical Chemistry, 89*, 5079–5086. https://doi.org/10.1021/acs.analchem.7b00595.

Kruve, A., Kaupmees, K., Liigand, J., & Leito, I. (2014). Negative electrospray ionization via deprotonation: Predicting the ionization efficiency. *Analytical Chemistry, 86*(10), 4822–4830. https://doi.org/10.1021/ac404066v.

Leppert, B. C., Markle, J. C., Helt, R. C., & Fujie, G. H. (1983). Determination of carbosulfan and carbofuran residues in plants, soil, and water by gas chromatography. *Journal of Agricultural and Food Chemistry, 31*, 220–223. https://doi.org/10.1021/jf00116a009.

Li, Z., & Jennings, A. (2017). Worldwide regulations of standard values of pesticides for human health risk control: A review. *International Journal of Environmental Research and Public Health, 14*, 826. https://doi.org/10.3390/ijerph14070826.

Liigand, J., Wang, T., Kellogg, J., Smedsgaard, J., Cech, N., & Kruve, A. (2019). Quantifying the unquantifiable: Quantification for non-targeted LC/MS screening without standards, submitted.

Liigand, P., Liigand, J., Cuyckens, F., Vreeken, R. J., & Kruve, A. (2018). Ionisation efficiencies can be predicted in complicated biological matrices: A proof of concept. *Analytica Chimica Acta, 1032*, 68–74. https://doi.org/10.1016/j.aca.2018.05.072.

Mol, H. G. J., Plaza-Bolaños, P., Zomer, P., De Rijk, T. C., Stolker, A. A. M., & Mulder, P. P. J. (2008). Toward a generic extraction method for simultaneous determination of pesticides, mycotoxins, plant toxins, and veterinary drugs in feed and food matrixes. *Analytical Chemistry, 80*(24), 9450–9459. https://doi.org/10.1021/ac801557f.

Oss, M., Kruve, A., Herodes, K., & Leito, I. (2010). Electrospray ionization efficiency scale of organic compound. *Analytical Chemistry, 82*, 2865–2872. https://doi.org/10.1021/ac902856t.

Pieke, E. N., Granby, K., Trier, X., & Smedsgaard, J. (2017). A framework to estimate concentrations of potentially unknown substances by semi-quantification in liquid chromatography electrospray ionization mass spectrometry. *Analytica Chimica Acta, 975*, 30–41. https://doi.org/10.1016/j.aca.2017.03.054.

Pucci, V., Di, S., Alfieri, A., Bonelli, F., & Monteagudo, E. (2009). A novel strategy for reducing phospholipids-based matrix effect in LC–ESI-MS bioanalysis by means of HybridSPE. *Journal of Pharmaceutical and Biomedical Analysis, 50*, 867–871. https://doi.org/10.1016/j.jpba.2009.05.037.

Raji, M. A., Fryčák, P., Temiyasathit, C., Kim, S. B., Mavromaras, G., Ahn, J., Schug, K., & Using, A. (2009). multivariate statistical methods to model the electrospray ionization response of GXG tripeptides based on multiple physicochemical parameters. *Rapid Communications in Mass Spectrometry, 23*, 2221–2232. https://doi.org/10.1002/rcm.

Wang, T., Frandsen, H. L., Christiansson, N. R., Rosendal, S. E., Pedersen, M., & Smedsgaard, J. (2019). Pyrrolizidine alkaloids in honey: Quantification with and without standards. *Food Control, 98*, 227–237. https://doi.org/10.1016/j.foodcont.2018.11.033.

Wu, L., Wu, Y., Shen, H., Gong, P., Cao, L., Wang, G., & Hao, H. (2013). Quantitative structure-ion intensity relationship strategy to the prediction of absolute levels without authentic standards. *Analytica Chimica Acta, 794*, 67–75. https://doi.org/10.1016/j.aca.2013.07.034.

Yap, C. W. (2011). Software news and update PaDEL-Descriptor : An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry, 32*(7), 1466–1474. https://doi.org/10.1002/jcc.

References

Agilent Technologies. (2014). *Agilent 6200 Series TOF and 6500 Series Q-TOF LC / MS System Concepts Guide The Big Picture*.

Aguilera-Luiz, M. M., Vidal, J. L. M., Romero-González, R., & Frenich, A. G. (2008). Multi-residue determination of veterinary drugs in milk by ultra-high-pressure liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A*, *1205*(1–2), 10–16. https://doi.org/10.1016/j.chroma.2008.07.066

Al-Alam, J., Fajloun, Z., Chbani, A., & Millet, M. (2017). A multiresidue method for the analysis of 90 pesticides, 16 PAHs, and 22 PCBs in honey using QuEChERS–SPME. *Analytical and Bioanalytical Chemistry*, *409*(21), 5157–5169. https://doi.org/10.1007/s00216-017-0463-y

Alder, L., Greulich, K., Kempe, G., & Vieth, B. (2006). Residue analysis of 500 high priority pesticides: better by GC–MS or LC–MS/MS. *Mass Spectrometry Reviews*, (25), 838–865. https://doi.org/10.1002/mas

Alldrick, A. J. (2017). *Food Safety Aspects of Grain and Cereal Product Quality*. *Cereal Grains: Assessing and Managing Quality: Second Edition*. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100719-8.00015-2

Alymatiri, C. M., Kouskoura, M. G., & Markopoulou, C. K. (2015). Decoding the signal response of steroids in electrospray ionization mode (ESI-MS). *Analytical Methods*, *7*, 10433–10444. https://doi.org/10.1039/C5AY02839F

Amad, M. H., Cech, N. B., Jackson, G. S., & Enke, C. G. (2000). Importance of gas-phase proton affinities in determining the electrospray ionization response for analytes and solvents. *Journal of Mass Spectrometry*. https://doi.org/10.1002/1096-9888(200007)35:7<784::AID-JMS17>3.0.CO;2-Q

Andersen, A. J. C., Hansen, P. J., Jørgensen, K., & Nielsen, K. F. (2016). Dynamic Cluster Analysis: An Unbiased Method for Identifying A + 2 Element Containing Compounds in Liquid Chromatographic High-Resolution Time-of-Flight Mass Spectrometric Data. *Analytical Chemistry*, *88*(24), 12461–12469. https://doi.org/10.1021/acs.analchem.6b03902

Annunziata, L., Stramenga, A., Visciano, P., Schirone, M., De Colli, L., Colagrande, M. N., … Scortichini, G. (2017). Simultaneous determination of aflatoxins, T-2 and HT-2 toxins, and fumonisins in cereal-derived products by QuEChERS extraction coupled with LC-MS/MS. *Analytical and Bioanalytical Chemistry*, *409*, 5143–5155. https://doi.org/10.1007/s00216-017-0462-z

Antignac, J. P., Courant, F., Pinel, G., Bichon, E., Monteau, F., Elliott, C., & Le Bizec, B. (2011). Mass spectrometry-based metabolomics applied to the chemical safety of food. *TrAC - Trends in Analytical Chemistry*, *30*(2), 292–301. https://doi.org/10.1016/j.trac.2010.11.003

AOAC. (2002). *AOAC Guidelines for Single Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals*. https://doi.org/10.1063/1.4915424

Avula, B., Sagi, S., Wang, Y. H., Zweigenbaum, J., Wang, M., & Khan, I. A. (2015). Characterization and screening of pyrrolizidine alkaloids and N-oxides from botanicals and dietary supplements using UHPLC-high resolution mass spectrometry. *Food Chemistry*, *178*, 136–148. https://doi.org/10.1016/j.foodchem.2015.01.053

Awad, H., Khamis, M. M., & El-Aneed, A. (2015). Mass spectrometry, review of the basics: Ionization. *Applied Spectroscopy Reviews*, *50*(2), 158–175. https://doi.org/10.1080/05704928.2014.954046

Banerjee, S., & Mazumdar, S. (2012). Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry*, *2012*, 1–40. https://doi.org/10.1155/2012/282574

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*,

17(3), 166–173. https://doi.org/10.1002/cem.785
Basiri, B., Murph, M. M., & Bartlett, M. G. (2017). Assessing the Interplay between the Physicochemical Parameters of Ion-Pairing Reagents and the Analyte Sequence on the Electrospray Desorption Process for Oligonucleotides. *Journal of the American Society for Mass Spectrometry*, *28*(8), 1647–1656. https://doi.org/10.1007/s13361-017-1671-6
Beach, D. G., & Gabryelski, W. (2013). Linear and nonlinear regimes of electrospray signal response in analysis of urine by electrospray ionization-high field asymmetric waveform ion mobility spectrometry-MS and implications for nontarget quantification. *Analytical Chemistry*, *85*(4), 2127–2134. https://doi.org/10.1021/ac3027542
Beccari, G., Caproni, L., Tini, F., Uhlig, S., & Covarelli, L. (2016). Presence of Fusarium Species and Other Toxigenic Fungi in Malting Barley and Multi-Mycotoxin Analysis By Liquid Chromatography-High Resolution Mass Spectrometry. *Journal of Agricultural and Food Chemistry*, acs.jafc.6b00702. https://doi.org/10.1021/acs.jafc.6b00702
BfR. (2011). *Chemical analysis and toxicity of pyrrolizidine alkaloids and assessment of the health risks posed by their occurrence in honey*.
Biognosys. (2016). Generating high-quality spectral libraries for targeted analysis of data generated with data-independent acquisition. Retrieved from https://www.researchgate.net/institution/Biognosys/post/577a675eed99e18ccd4c5667_Generating_high-quality_spectral_libraries_for_targeted_analysis_of_data_generated_with_data-independent_acquisition
Blanz, J., Williams, G., Dayer, J., Délémonté, T., Gertsch, W., Ramstein, P., … Pearson, D. (2017). Evaluation of relative MS response factors of drug metabolites for semi-quantitative assessment of chemical liabilities in drug discovery. *Journal of Mass Spectrometry*, *52*(4), 210–217. https://doi.org/10.1002/jms.3918
Bletsou, A. A., Jeon, J., Hollender, J., Archontaki, E., & Thomaidis, N. S. (2015). Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. *TrAC - Trends in Analytical Chemistry*, *66*, 32–44. https://doi.org/10.1016/j.trac.2014.11.009
Bolechová, M., Čáslavský, J., Pospíchalová, M., & Kosubová, P. (2015). UPLC-MS/MS method for determination of selected pyrrolizidine alkaloids in feed. *Food Chemistry*, *170*, 265–270. https://doi.org/10.1016/j.foodchem.2014.08.072
Bolton, E. E., Wang, Y., Thiessen, P. A., & Bryant, S. H. (2008). *Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities*. *Annual Reports in Computational Chemistry* (Vol. 4). Elsevier Masson SAS. https://doi.org/10.1016/S1574-1400(08)00012-1
Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
Brereton, R.G. (2003). *Chemometrics. Data analysis for the laboratory and chemical plant*. Wiley.
Brereton, Richard G. (2009). *Chemometrics for Pattern Recognition*. *Chemometrics for Pattern Recognition*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470746462
Broeckling, C. D., Hoyes, E., Richardson, K., Brown, J. M., & Prenni, J. E. (2018). Comprehensive Tandem-Mass-Spectrometry Coverage of Complex Samples Enabled by Data-Set-Dependent Acquisition. *Analytical Chemistry*, *90*(13), 8020–8027. https://doi.org/10.1021/acs.analchem.8b00929
Bruins, A. P. (1991). Mass spectrometry with ion sources operating at atmospheric pressure. *Mass Spectrometry Reviews*, *10*(1), 53–77. https://doi.org/10.1002/mas.1280100104
Caetano, S., Delæ , caestecker, T., Put, R., Daszykowski, M., Van Bocxlaer, J., & Vander Heyden, Y. (2005). Exploring and modelling the responses of electrospray and atmospheric pressure chemical ionization techniques based on molecular descriptors. *Analytica Chimica Acta*, *550*(1–2), 92–106. https://doi.org/10.1016/j.aca.2005.06.069

Cartwright, H. (2015). Artificial Neural Networks. *Methods in Molecular Biology*, *1260*. https://doi.org/10.1007/978-1-4939-2239_0

Cech, N. B., & Enke, C. G. (2000). Relating electrospray ionization response to nonpolar character of small peptides. *Analytical Chemistry*, *72*(13), 2717–2723. https://doi.org/10.1021/ac9914869

Chalcraft, K. R., Lee, R., Mills, C., & Britz-McKibbin, P. (2009). Virtual Quantifcation of Metabolites by Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry: Predicting Ionization Ef ciency Without Chemical Standards. *Analytical Chemistry*, *81*(7), 2506–2515. https://doi.org/10.1021/ac802272u

Cheli, F., Battaglia, D., Gallo, R., & Dell'Orto, V. (2014). EU legislation on cereal safety: An update with a focus on mycotoxins. *Food Control*, *37*(1), 315–325. https://doi.org/10.1016/j.foodcont.2013.09.059

Chen, L., Song, F., Liu, Z., Zheng, Z., Xing, J., & Liu, S. (2014). Study of the ESI and APCI interfaces for the UPLC-MS/MS analysis of pesticides in traditional Chinese herbal medicine. *Analytical and Bioanalytical Chemistry*, *406*(5), 1481–1491. https://doi.org/10.1007/s00216-013-7508-7

Chung, S. W. C., & Lam, C.-H. (2018). Development of an Analytical Method for Analyzing Pyrrolizidine Alkaloids in Different Groups of Food by UPLC-MS/MS. *Journal of Agricultural and Food Chemistry*, *66*, 3009–3018. https://doi.org/10.1021/acs.jafc.7b06118

Cífková, E., Holčapek, M., Lísa, M., Ovčačíková, M., Lyčka, A., Lynen, F., & Sandra, P. (2012). Nontargeted quantitation of lipid classes using hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry with single internal standard and response factor approach. *Analytical Chemistry*, *84*(22), 10064–10070. https://doi.org/10.1021/ac3024476

Cigić, I. K., & Prosen, H. (2009). An overview of conventional and emerging analytical methods for the determination of mycotoxins. *International Journal of Molecular Sciences*, *10*(1), 62–115. https://doi.org/10.3390/ijms10010062

Coceral. (2019). Balance sheet for cereals 2018/12-2019/12.

Cole, R. B. (2000). Some tenets pertaining to electrospray ionization mass spectrometry. *Journal of Mass Spectrometry*, *772*(35), 763–772.

Colin A Smith, O'Maille, G., Want, E. J., Qin, C., Sunia A Trauger, Theodore R Brandon, Darlene E Custodio, R. A., & Siuzdak, G. (2005). METLIN: A Metabolite Mass Spectral Database. *Ther Drug Monit*, *27*(6), 747–751.

Cramer, C. J., Johnson, J. L., & Kamel, A. M. (2017). Prediction of Mass Spectral Response Factors from Predicted Chemometric Data for Druglike Molecules. *Journal of the American Society for Mass Spectrometry*, *28*, 278–285. https://doi.org/10.1007/s13361-016-1536-4

Cramer, L., Schiebel, H. M., Ernst, L., & Beuerle, T. (2013). Pyrrolizidine alkaloids in the food chain: Development, validation, and application of a new HPLC-ESI-MS/MS sum parameter method. *Journal of Agricultural and Food Chemistry*, *61*(47), 11382–11391. https://doi.org/10.1021/jf403647u

Creydt, M., & Fischer, M. (2017). Plant Metabolomics - Maximizing Metabolome Coverage by Optimizing Mobile Phase Additives for Non-Targeted Mass Spectrometry in Positive and Negative Electrospray Ionization Mode. *Analytical Chemistry*, *89*, 10474–10486. https://doi.org/10.1021/acs.analchem.7b02592

Croley, T. R., White, K. D., Callahan, J. H., & Musser, S. M. (2012). The chromatographic role in high resolution mass spectrometry for non-targeted analysis. *Journal of the American Society for Mass Spectrometry*, *23*(9), 1569–1578. https://doi.org/10.1007/s13361-012-0392-0

Crotti, S., Isak, I., & Traldi, P. (2017). *Advanced spectroscopic detectors for identification and quantification: Mass spectrometry*. *Liquid Chromatography: Fundamentals and Instrumentation: Second Edition* (Second Edi, Vol. 1). Elsevier Inc. https://doi.org/10.1016/B978-0-12-805393-5.00018-X

Cunha, S. C., & Fernandes, J. O. (2010). Development and validation of a method based on a

QuEChERS procedure and heart-cutting GC-MS for determination of five mycotoxins in cereal products. *Journal of Separation Science*, *33*(4–5), 600–609. https://doi.org/10.1002/jssc.200900695

Cunsolo, V., Muccilli, V., Saletti, R., & Foti, S. (2014). Mass spectrometry in food proteomics: A tutorial. *Journal of Mass Spectrometry*, *49*(9), 768–784. https://doi.org/10.1002/jms.3374

Daglia, M., Antiochia, R., Sobolev, A. P., & Mannina, L. (2014). Untargeted and targeted methodologies in the study of tea (Camellia sinensis L .). *FRIN*, *63*, 275–289. https://doi.org/10.1016/j.foodres.2014.03.070

de Aguiar, P. F., Bourguignon, B., Khots, M. S., Massart, D. L., & Phan-Than-Luu, R. (1995). D-optimal designs. *Chemometrics and Intelligent Laboratory Systems*, *30*(2), 199–210. https://doi.org/10.1016/0169-7439(94)00076-X

Delaporte, G., Cladière, M., Bouveresse, D. J., & Camel, V. (2019). Untargeted food contaminant detection using UHPLC-HRMS combined with multivariate analysis : Feasibility study on tea. *Food Chemistry*, *277*, 54–62. https://doi.org/10.1016/j.foodchem.2018.10.089

Delaporte, G., Cladière, M., & Camel, V. (2019). Untargeted food chemical safety assessment : A proof-of-concept on two analytical platforms and contamination scenarios of tea. *Food Control*, *98*, 510–519. https://doi.org/10.1016/j.foodcont.2018.12.004

Delatour, T., Racault, L., Bessaire, T., & Desmarchelier, A. (2018). Screening of veterinary drug residues in food by LC-MS/MS. Background and challenges. *Food Additives & Contaminants: Part A ISSN:*, *35*(4), 632–645. https://doi.org/10.1080/19440049.2018.1426890

Díaz, R., Ibáñez, M., Sancho, J. V., & Hernández, F. (2012). Target and non-target screening strategies for organic contaminants, residues and illicit substances in food, environmental and human biological samples by UHPLC-QTOF-MS. *Analytical Methods*, *4*(1), 196. https://doi.org/10.1039/c1ay05385j

Du, F., Liu, T., Shen, T., Zhu, F., & Xing, J. (2012). Qualitative-(semi)quantitative data acquisition of artemisinin and its metabolites in rat plasma using an LTQ/Orbitrap mass spectrometer. *Journal of Mass Spectrometry*, *47*(2), 246–252. https://doi.org/10.1002/jms.2958

Dzuman, Z., Zachariasova, M., Veprikova, Z., Godula, M., & Hajslova, J. (2015). Multi-analyte high performance liquid chromatography coupled to high resolution tandem mass spectrometry method for control of pesticide residues, mycotoxins, and pyrrolizidine alkaloids. *Analytica Chimica Acta*, *863*(1), 29–40. https://doi.org/10.1016/j.aca.2015.01.021

Edmands, W. M. B., Petrick, L. M., Barupal, D. K., Scalbert, A., Wilson, M., Wickliffe, J., & Rappaport, S. M. (2017). compMS2Miner: an automatable metabolite identification, visualization and data-sharing R package for high-resolution LC-MS datasets. *Analytical Chemistry*, acs.analchem.6b02394. https://doi.org/10.1021/acs.analchem.6b02394

EFSA. (2011). Scientific Opinion on Pyrrolizidine alkaloids in food and feed. *EFSA Journal 2011*, *9*(11), 2406. https://doi.org/10.2903/j.efsa.

EFSA. (2016). *Scientific Opinion on Tropane alkaloids in food and feed* (Vol. 11). https://doi.org/10.2903/j.efsa.2013.3386

EFSA. (2017). Risks for human health related to the presence of pyrrolizidine alkaloids in honey, tea, herbal infusions and food supplements. *EFSA Journal*, *15*(7), 4908. https://doi.org/10.2903/j.efsa.2017.4908

Ehrmann, B. M., Henriksen, T., & Cech, N. B. (2008). Relative Importance of Basicity in the Gas Phase. *Journal of American Society for Mass Spectrometry*, *19*, 719–728. https://doi.org/10.1021/jasms.8b03186

El-Aneed, A., Cohen, A., & Banoub, J. (2009). Mass spectrometry, review of the basics: Electrospray, MALDI, and commonly used mass analyzers. *Applied Spectroscopy Reviews*, *44*(3), 210–230. https://doi.org/10.1080/05704920902717872

Eliasson, M., Rännar, S., Madsen, R., Donten, M. A., Marsden-Edwards, E., Moritz, T., … Trygg, J.

(2012). Strategy for optimizing LC-MS data processing in metabolomics: A design of experiments approach. *Analytical Chemistry*, *84*(15), 6869–6876. https://doi.org/10.1021/ac301482k

Empereur-Mot, C., Zagury, J.-F., & Montes, M. (2016). Screening Explorer–An Interactive Tool for the Analysis of Screening Results. *Journal of Chemical Information and Modeling*, *56*(12), 2281–2286. https://doi.org/10.1021/acs.jcim.6b00283

EU - Pesticides database. (n.d.). Retrieved from http://ec.europa.eu/food/plant/pesticides/eu-pesti

European Commission. (2001). *Setting maximum levels for certain contaminants in foodstuffs*. *COMMISSION REGULATION (EC) No 466/2001 of 8 March 2001* (Vol. R0466).

European Commission. (2020). Chemical safety.

FAO. (2020a). Food safety and quality.

FAO. (2020b). *National food control systems : Core elements and functions*.

FAO & WHO. (2010). *FAO/WHO Framework for Developing National Food Safety Emergency Response Plans*.

FDA. (1994). *Validation of chromatographic methods*. https://doi.org/10.1201/9781315118024-31

Fernández Pierna, J. A., Vincke, D., Baeten, V., Grelet, C., Dehareng, F., & Dardenne, P. (2016). Use of a multivariate moving window PCA for the untargeted detection of contaminants in agro-food products, as exemplified by the detection of melamine levels in milk using vibrational spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, *152*, 157–162. https://doi.org/10.1016/j.chemolab.2015.10.016

Freeman. (2013). *quantative analysis*. *Journal of Chemical Information and Modeling* (Vol. 53). https://doi.org/10.1017/CBO9781107415324.004

Frenich, A. G., Romero-González, R., Aguilera-Luiz, M. del M., & Research. (2014). Comprehensive analysis of toxics (pesticides, veterinary drugs and mycotoxins) in food by UHPLC-MS. *TrAC - Trends in Analytical Chemistry*, *63*, 158–169. https://doi.org/10.1016/j.trac.2014.06.020

García-Reyes, J. F., Ferrer, I., Thurman, E. M., Molina-Díaz, A., & Fernández-Alba, A. R. (2005). Searching for non-target chlorinated pesticides in food by liquid chromatography/time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, *19*(19), 2780–2788. https://doi.org/10.1002/rcm.2127

Garcı, J. F., Hernando, M. D., & Molina-dı, A. (2007). Comprehensive screening of target , non-target and unknown pesticides in food by LC-TOF-MS, *26*(8). https://doi.org/10.1016/j.trac.2007.06.006

Gerbig, S., Stern, G., Brunn, H. E., Düring, R.-A., Spengler, B., & Schulz, S. (2016). Method development towards qualitative and semi-quantitative analysis of multiple pesticides from food surfaces and extracts by desorption electrospray ionization mass spectrometry as a preselective tool for food control. *Analytical and Bioanalytical Chemistry*, *409*, 2107–2117. https://doi.org/10.1007/s00216-016-0157-x

Ghosh, B., & Jones, A. D. (2015). Dependence of negative-mode electrospray ionization response factors on mobile phase composition and molecular structure for newly-authenticated neutral acylsucrose metabolites. *Analyst*, *140*(19), 6522–6531. https://doi.org/10.1039/c4an02124j

Giacomoni, F., Corguillé, G. Le, Monsoor, M., Landi, M., Peri-, P., Pétéra, M., … Caron, C. (2014). Workflow4Metabolomics : A collaborative research infrastructure for computational metabolomics. *Bioinformatics*, *31*(9), 1–2.

Gioumouxouzis, C. I., Kouskoura, M. G., & Markopoulou, C. K. (2015). Negative electrospray ionization mode in mass spectrometry: A new perspective via modeling. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *998–999*, 97–105. https://doi.org/10.1016/j.jchromb.2015.06.009

Golubović, J., Birkemeyer, C., Protić, A., Otašević, B., & M., Z. (2016). Structure – response relationship in electrospray ionization-mass spectrometry of sartans by artificial neural networks. *Journal of Chromatography A*, *1438*, 123–132. https://doi.org/10.1016/j.chroma.2016.02.021

Gong, Y., Chen, J., Shi, Y., Lim, H.-K., Weng, N., & Salter, R. (2017). Standard-Free Bioanalytical

Approach for Absolute Quantitation of Drug Metabolites Utilizing Biosynthesis of Reciprocal Radio and Stable Isotopologues and Its Application. *Analytical Chemistry*, acs.analchem.7b01830. https://doi.org/10.1021/acs.analchem.7b01830

González-Curbelo, M. Á., Herrera-Herrera, A. V, Ravelo-Pérez, L. M., & Hernández-Borges, J. (2012). Sample-preparation methods for pesticide-residue analysis in cereals and derivatives. *Trends in Analytical Chemistry*, *38*, 32–51. https://doi.org/10.1016/j.trac.2012.04.010

Griffin, C. T., Danaher, M., Elliott, C. T., Glenn Kennedy, D., & Furey, A. (2013). Detection of pyrrolizidine alkaloids in commercial honey using liquid chromatography-ion trap mass spectrometry. *Food Chemistry*, *136*(3–4), 1577–1583. https://doi.org/10.1016/j.foodchem.2012.02.112

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. *The Elements of Statistical Learning, Second Edition*. https://doi.org/10.1007/b94608

Hatsis, P., Waters, N. J., & Argikar, U. A. (2017). Implications for metabolite quantification by mass spectrometry in the absence of authentic standards. *Drug Metabolism and Disposition*, *45*(5), 492–496. https://doi.org/10.1124/dmd.117.075259

Henriksen, T., & Juhler, R. K. (2005). The Relative Influences of Acidity and Polarity on Responsiveness of Small Organic Molecules to Analysis with Negative Ion Electrospray Ionization Mass Spectrometry. *Journal of American Society for Mass Spectrometry*, *16*, 446–455. https://doi.org/10.1016/j.jasms.2004.11.021

Hermans, J., Ongay, S., Markov, V., & Bischoff, R. (2017). Physicochemical Parameters Affecting the Electrospray Ionization Efficiency of Amino Acids after Acylation. *Analytical Chemistry*, *89*, 9159−9166. https://doi.org/10.1021/acs.analchem.7b01899

Herodes, K. (2020). LC-MS method validation.

Hird, S. J., Lau, B. P. Y., Schuhmacher, R., & Krska, R. (2014). Liquid chromatography-mass spectrometry for the determination of chemical contaminants in food. *TrAC-Trends in Analytical Chemistry*, *59*, 59–72. https://doi.org/10.1016/j.trac.2014.04.005

Ho, T. K. (1995). Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, *1*, 278–282. https://doi.org/10.1109/ICDAR.1995.598994

Hoffmann, E. de. (2005). Mass spectrometry 1.

Hoffmann, E. de, & Stroobant, V. (2004). *Mass Spectrometry-Principles and Applications-Third Edition*.

Holčapek, M., Jirásko, R., & Lísa, M. (2012). Recent developments in liquid chromatography-mass spectrometry and related techniques. *Journal of Chromatography A*, *1259*, 3–15. https://doi.org/10.1016/j.chroma.2012.08.072

Huffma, B. A., Poltash, M. L., & Hughey, C. A. (2012). Effect of Polar Protic and Polar Aprotic Solvents on Negative-Ion Electrospray Ionization and Chromatographic Separation of Small Acidic Molecules. *Analytical Chemistry*, *84*, 9942–9950. https://doi.org/10.1021/ac302397b

Iribarne, J. V., & Thomson, B. A. (1976). On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*, *64*(6), 2287–2294. https://doi.org/10.1063/1.432536

Jackson, L. S. (2009). Chemical Food Safety Issues in the United States: Past, Present, and Future. *Journal of Agricultural and Food Chemistry*, *57*(18), 8161–8170. https://doi.org/10.1021/jf900628u

Jaeger, C., Hoffmann, F., Schmitt, C. A., & Lisec, J. (2016). Automated Annotation and Evaluation of In-Source Mass Spectra in GC/Atmospheric Pressure Chemical Ionization-MS-Based Metabolomics. *Analytical Chemistry*, *88*(19), 9386–9390. https://doi.org/10.1021/acs.analchem.6b02743

Jamin, E. L., Bonvallot, N., & Tremblay-franco, M. (2014). Untargeted profiling of pesticide metabolites by LC – HRMS : an exposomics tool for human exposure evaluation. *Analytical and Bioanalytical Chemistry*, *406*, 1149–1161. https://doi.org/10.1007/s00216-013-7136-2

Jayaraj, R., Megha, P., & Sreedev, P. (2016). Review Article. Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. *Interdisciplinary Toxicology*, *9*(3–4), 90–100. https://doi.org/10.1515/intox-2016-0012

Jin, P., Xu, S., Xu, W., He, X., Kuang, Y., & Hu, X. (2020). Screening and quantification of fourteen synthetic antidiabetic adulterants in herbal pharmaceuticals and health foods by HPLC and confirmation by LC-Q-TOF-MS/MS. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment*, *37*(1), 11–18. https://doi.org/10.1080/19440049.2019.1675910

Kalogiouri, N. P., Aalizadeh, R., & Thomaidis, N. S. (2017). Investigating the organic and conventional production type of olive oil with target and suspect screening by LC-QTOF-MS, a novel semi-quantification method using chemical similarity and advanced chemometrics. *Analytical and Bioanalytical Chemistry*, *49*(3), 5413–5426. https://doi.org/10.1007/s00216-017-0395-6

Kaltner, F., Stiglbauer, B., Rychlik, M., Gareis, M., & Gottschalk, C. (2019). Development of a sensitive analytical method for determining 44 pyrrolizidine alkaloids in teas and herbal teas via LC-ESI-MS/MS. *Analytical and Bioanalytical Chemistry*, *411*(27), 7233–7249. https://doi.org/10.1007/s00216-019-02117-1

Katajamaa, M., & Orešič, M. (2005). Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics*, *6*, 179. https://doi.org/10.1186/1471-2105-6-179

Kaufmann, A. (2012). The current role of high-resolution mass spectrometry in food analysis. *Analytical and Bioanalytical Chemistry*, *403*(5), 1233–1249. https://doi.org/10.1007/s00216-011-5629-4

Khatal, L., Gaur, A., Naphade, A., Kandikere, V., & Mookhtiar, K. (2016). Impact of APCI ionization source in liquid chromatography tandem mass spectrometry based tissue distribution studies. *Biomedical Chromatography*, *30*(10), 1676–1685. https://doi.org/10.1002/bmc.3740

Kiontke, A., Oliveira-Birkmeier, A., Opitz, A., & Birkemeyer, C. (2016). Electrospray ionization efficiency is dependent on different molecular descriptors with respect to solvent pH and instrumental configuration. *PLoS ONE*, *11*(12), 1–16. https://doi.org/10.1371/journal.pone.0167502

Knolhoff, A. M., & Croley, T. R. (2016). Non-targeted screening approaches for contaminants and adulterants in food using liquid chromatography hyphenated to high resolution mass spectrometry. *Journal of Chromatography A*, *1428*, 86–96. https://doi.org/10.1016/j.chroma.2015.08.059

Koopmans, F., Ho, J. T. C., Smit, A. B., & Li, K. W. (2018). Comparative Analyses of Data Independent Acquisition Mass Spectrometric Approaches: DIA, WiSIM-DIA, and Untargeted DIA. *Proteomics*, *18*(1), 1–6. https://doi.org/10.1002/pmic.201700304

Kruve, A. (2016). Influence of mobile phase , source parameters and source type on electrospray ionization efficiency in negative ion mode. *Journal of Mass Spectrometry*, *51*, 596–601. https://doi.org/10.1002/jms.3790

Kruve, A. (2019). Semi-quantitative non-target analysis of water with liquid chromatography/high-resolution mass spectrometry: How far are we? *Rapid Communications in Mass Spectrometry*, *33*(S3), 54–63. https://doi.org/10.1002/rcm.8208

Kruve, A. (2020). Strategies for drawing quantitative conclusions from non-targeted liquid chromatography high-resolution mass spectrometry analysis. *Analytical Chemistry*, In Press. https://doi.org/10.1021/acs.analchem.9b03481

Kruve, A., & Kaupmees, K. (2017). Predicting ESI/MS signal change for anions in different solvents. *Analytical Chemistry*, *89*, 5079−5086. https://doi.org/10.1021/acs.analchem.7b00595

Kruve, A., Kaupmees, K., Liigand, J., & Leito, I. (2014). Negative electrospray ionization via deprotonation: Predicting the ionization efficiency. *Analytical Chemistry*, *86*(10), 4822–4830. https://doi.org/10.1021/ac404066v

Kunzelmann, M., Winter, M., Åberg, M., Hellenäs, K. E., & Rosén, J. (2018). Non-targeted analysis of unexpected food contaminants using LC-HRMS. *Analytical and Bioanalytical Chemistry*, *410*, 5593–5602. https://doi.org/10.1007/s00216-018-1028-4

Leitner, A., Emmert, J., Boerner, K., & Lindner, W. (2007). Influence of solvent additive composition on chromatographic separation and sodium adduct formation of peptides in HPLC-ESI MS. *Chromatographia*, *65*(11–12), 649–653. https://doi.org/10.1365/s10337-007-0219-5

Leito, I., Herodes, K., Huopolainen, M., Virro, K., Ku¨nnapas, A., Kruve, A., & Tanner, R. (2008). Towards the electrospray ionization mass spectrometry ionization efficiency scale of organic compounds. *Rapid Communications in Mass Spectrometry*, *22*, 3567–3577. https://doi.org/10.1002/rcm

León, N., Pastor, A., & Yusà, V. (2016). Target analysis and retrospective screening of veterinary drugs, ergot alkaloids, plant toxins and other undesirable substances in feed using liquid chromatography-high resolution mass spectrometry. *Talanta*, *149*, 43–52. https://doi.org/10.1016/j.talanta.2015.11.032

Li, X., Wang, T., Zhou, B., Gao, W., Cao, J., & Huang, L. (2014). Chemical composition and antioxidant and anti-inflammatory potential of peels and flesh from 10 different pear varieties (Pyrus spp.). *Food Chemistry*, *152*, 531–538. https://doi.org/10.1016/j.foodchem.2013.12.010

Li, Z., Li, Y., Chen, W., Cao, Q., Guo, Y., Wan, N., … Shui, W. (2017). Integrating MS1 and MS2 Scans in High-Resolution Parallel Reaction Monitoring Assays for Targeted Metabolite Quantification and Dynamic ^{13} C-Labeling Metabolism Analysis. *Analytical Chemistry*, *89*(1), 877–885. https://doi.org/10.1021/acs.analchem.6b03947

Liigand, J., Kruve, A., Leito, I., Girod, M., & Antoine, R. (2014). Effect of mobile phase on electrospray ionization efficiency. *Journal of the American Society for Mass Spectrometry*, *25*(11), 1853–1861. https://doi.org/10.1007/s13361-014-0969-x

Liigand, J., Laaniste, A., & Kruve, A. (2017). pH Effects on Electrospray Ionization Efficiency. *Journal of the American Society for Mass Spectrometry*, *28*(3), 461–469. https://doi.org/10.1007/s13361-016-1563-1

Liigand, P., Heering, A., Kaupmees, K., Leito, I., Girod, M., Antoine, R., & Kruve, A. (2017). The Evolution of Electrospray Generated Droplets is Not Affected by Ionization Mode. *Journal of The American Society for Mass Spectrometry*, 14–17. https://doi.org/10.1007/s13361-017-1737-5

Liigand, P., Liigand, J., Cuyckens, F., Vreeken, R. J., & Kruve, A. (2018). Ionisation efficiencies can be predicted in complicated biological matrices: A proof of concept. *Analytica Chimica Acta*, *1032*, 68–74. https://doi.org/10.1016/j.aca.2018.05.072

Liu, D. Q., & Hop, C. E. C. A. (2005). Strategies for characterization of drug metabolites using liquid chromatography-tandem mass spectrometry in conjunction with chemical derivatization and on-line H/D exchange approaches. *Journal of Pharmaceutical and Biomedical Analysis*, *37*(1), 1–18. https://doi.org/10.1016/j.jpba.2004.09.003

Luttrell, C. B. (2002). *Outlook for Agriculture*. *Review* (Vol. 58). https://doi.org/10.20955/r.58.11-16.qnt

Mandra, V. J., Kouskoura, M. G., & Markopoulou, C. K. (2015). Using the partial least squares method to model the electrospray ionization response produced by small pharmaceutical molecules in positive mode. *Rapid Communications in Mass Spectrometry*, *29*, 1661–1675. https://doi.org/10.1002/rcm.7263

Maragou, N. C., Thomaidis, N. S., & Koupparis, M. A. (2011). Optimization and comparison of ESI and APCI LC-MS/MS methods: A case study of Irgarol 1051, Diuron, and their degradation products in environmental samples. *Journal of the American Society for Mass Spectrometry*, *22*(10), 1826–1838. https://doi.org/10.1007/s13361-011-0191-z

Martens, H., & Næs, T. (1989). *Multivariate calibration*. Wiley.

Martos, P. A., Thompson, W., & Diaz, G. J. (2010). Multiresidue mycotoxin analysis in wheat, barley, oats, rye and maize grain by high-performance liquid chromatography-tandem mass spectrometry. *World Mycotoxin Journal*, *3*(3), 205–223. https://doi.org/10.3920/WMJ2010.1212

Masiá, A., Suarez-Varela, M. M., Llopis-Gonzalez, A., & Picó, Y. (2016). Determination of pesticides and veterinary drug residues in food by liquid chromatography-mass spectrometry: A review. *Analytica Chimica Acta*, *936*, 40–61. https://doi.org/10.1016/j.aca.2016.07.023

Mauser, H., Roche, O., Stahl, M., & Müller, S. (2005). Prediction of UV and ESI - MS signal intensities. *Journal of Chemical Information and Modeling*, *45*(4), 1039–1046.

https://doi.org/10.1021/ci0496548
Mehta, N., Porterfield, M., Struwe, W. B., Heiss, C., Azadi, P., Rudd, P. M., … Aoki, K. (2016). Mass Spectrometric Quantification of N-Linked Glycans by Reference to Exogenous Standards. *Journal of Proteome Research*, *15*(9), 2969–2980. https://doi.org/10.1021/acs.jproteome.6b00132
MetaboAnalyst. (n.d.). Retrieved from https://www.metaboanalyst.ca/MetaboAnalyst/home.xhtml
Mie, A., Laursen, K. H., Aberg, K. M., Forshed, J., Lindahl, A., Thorup-Kristensen, K., … Husted, S. (2014). Discrimination of conventional and organic white cabbage from a long-term field trial study using untargeted LC-MS-based metabolomics. *Analytical and Bioanalytical Chemistry*, *406*(12), 2885–2897. https://doi.org/10.1007/s00216-014-7704-0
Mol, H. G. J., Plaza-Bolaños, P., Zomer, P., De Rijk, T. C., Stolker, A. A. M., & Mulder, P. P. J. (2008). Toward a generic extraction method for simultaneous determination of pesticides, mycotoxins, plant toxins, and veterinary drugs in feed and food matrixes. *Analytical Chemistry*, *80*(24), 9450–9459. https://doi.org/10.1021/ac801557f
Mollerup, C. B., Dalsgaard, P. W., Mardal, M., & Linnet, K. (2017). Targeted and non-targeted drug screening in whole blood by UHPLC-TOF-MS with data-independent acquisition. *Drug Testing and Analysis*, *9*(7), 1052–1061. https://doi.org/10.1002/dta.2120
Mulder, P. P. J., López, P., Castelari, M., Bodi, D., Ronczka, S., Preiss-Weigert, A., & These, A. (2018). Occurrence of pyrrolizidine alkaloids in animal- and plant-derived food: results of a survey across Europe. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment*, *35*(1), 118–133. https://doi.org/10.1080/19440049.2017.1382726
Nguyen, T. B., Nizkorodov, S. A., Laskinc, A., & Laskin, J. (2013). An approach toward quantification of organic compounds in complex environmental samples using high-resolution electrospray ionization mass spectrometry. *Analytical Methods*, *5*(5), 72. https://doi.org/10.1039/c2ay25682g
Nielsen, K. F., & Smedsgaard, J. (2003). Fungal metabolite screening: Database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography-UV-mass spectrometry methodology. *Journal of Chromatography A*, *1002*(1–2), 111–136. https://doi.org/10.1016/S0021-9673(03)00490-4
Nielsen, K. L., Telving, R., Andreasen, M. F., Hasselstrøm, J. B., & Johannsen, M. (2016). A Metabolomics Study of Retrospective Forensic Data from Whole Blood Samples of Humans Exposed to 3,4-Methylenedioxymethamphetamine: A New Approach for Identifying Drug Metabolites and Changes in Metabolism Related to Drug Consumption. *Journal of Proteome Research*, *15*(2), 619–627. https://doi.org/10.1021/acs.jproteome.5b01023
O'Haver, T. C. (2019). Smoothing.
Ortiz, X., Korenkova, E., Jobst, K. J., MacPherson, K. A., & Reiner, E. J. (2017). A high throughput targeted and non-targeted method for the analysis of microcystins and anatoxin-A using on-line solid phase extraction coupled to liquid chromatography–quadrupole time-of-flight high resolution mass spectrometry. *Analytical and Bioanalytical Chemistry*, *409*, 4959–4969. https://doi.org/10.1007/s00216-017-0437-0
Oss, M., Kruve, A., Herodes, K., & Leito, I. (2010). Electrospray ionization efficiency scale of organic compound. *Analytical Chemistry*, *82*, 2865–2872. https://doi.org/10.1021/ac902856t
Pattison, M. (2019). The Role of Methanol and Acetonitrile as Organic Modifiers in Reversed-phase Liquid Chromatography. *Chromatography Today*, (2), 24–26.
Pence, H. E., & Williams, A. (2010). Chemspider: An online chemical information resource. *Journal of Chemical Education*, *87*(11), 1123–1124. https://doi.org/10.1021/ed100697w
Pereira, V. L., Fernandes, J. O., & Cunha, S. C. (2014). Mycotoxins in cereals and related foodstuffs: A review on occurrence and recent methods of analysis. *Trends in Food Science and Technology*, *36*, 96–136. https://doi.org/10.1016/j.tifs.2014.01.005
Pérez-Ortega, P., Lara-Ortega, F. J., García-Reyes, J. F., Gilbert-López, B., Trojanowicz, M., & Molina-Díaz, A. (2016). A feasibility study of UHPLC-HRMS accurate-mass screening methods for

multiclass testing of organic contaminants in food. *Talanta*, *160*, 704–712. https://doi.org/10.1016/j.talanta.2016.08.002

Peters, R. J. B., Bolck, Y. J. C., Rutgers, P., Stolker, A. A. M., & Nielen, M. W. F. (2009). Multi-residue screening of veterinary drugs in egg, fish and meat using high-resolution liquid chromatography accurate mass time-of-flight mass spectrometry. *Journal of Chromatography A*, *1216*(46), 8206–8216. https://doi.org/10.1016/j.chroma.2009.04.027

Picó, Y., Blasco, C., & Font, G. (2004). Environmental and food applications of LC-tandem mass spectrometry in pesticide-residue analysis: An overview. *Mass Spectrometry Reviews*, *23*(1), 45–85. https://doi.org/10.1002/mas.10071

Pieke, E. N., Granby, K., Trier, X., & Smedsgaard, J. (2017). A framework to estimate concentrations of potentially unknown substances by semi-quantification in liquid chromatography electrospray ionization mass spectrometry. *Analytica Chimica Acta*, *975*, 30–41. https://doi.org/10.1016/j.aca.2017.03.054

Pieke, E. N., Smedsgaard, J., & Granby, K. (2018). Exploring the chemistry of complex samples by tentative identification and semiquantification: A food contact material case. *Journal of Mass Spectrometry*. https://doi.org/10.1002/jms.3995

Raji, M. A., Fryčák, P., Temiyasathit, C., Kim, S. B., Mavromaras, G., Ahn, J. -., & Schug, K. A. (2009). Using multivariate statistical methods to model the electrospray ionization response of GXG tripeptides based on multiple physicochemical parameters. *Rapid Communications in Mass Spectrometry*, *23*, 2221–2232. https://doi.org/10.1002/rcm

Rebane, R., Kruve, A., Liigand, P., Liigand, J., Herodes, K., & Leito, I. (2016). Establishing Atmospheric Pressure Chemical Ionization Efficiency Scale. *Analytical Chemistry*, *88*(7), 3435–3439. https://doi.org/10.1021/acs.analchem.5b04852

Reinholds, I., Bartkevics, V., Silvis, I. C. J., van Ruth, S. M., & Esslinger, S. (2015). Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: A review. *Journal of Food Composition and Analysis*, *44*, 56–72. https://doi.org/10.1016/j.jfca.2015.05.004

Renaud, J. B., Sabourin, L., Topp, E., & Sumarah, M. W. (2017). A spectral counting approach to measure selectivity of high resolution LC-MS methods for environmental analysis. *Analytical Chemistry*, *89*, 2747–2754. https://doi.org/10.1021/acs.analchem.6b03475

Ridgway, K., Lalljie, S. P. D., & Smith, R. M. (2007). Sample preparation techniques for the determination of trace residues and contaminants in foods. *Journal of Chromatography A*, *1153*(1–2), 36–53. https://doi.org/10.1016/j.chroma.2007.01.134

Rochat, B. (2016). From targeted quantification to untargeted metabolomics: Why LC-high-resolution-MS will become a key instrument in clinical labs. *TrAC - Trends in Analytical Chemistry*. https://doi.org/10.1016/j.trac.2016.02.009

Skinner, C. G., Thomas, J. D., & Osterloh, J. D. (2010). Melamine toxicity. *Journal of Medical Toxicology*, *6*(1), 50–55. https://doi.org/10.1007/s13181-010-0038-1

Smith, C. A., Want, E. J., Maille, G. O., Abagyan, R., & Siuzdak, G. (2006). XCMS : Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment , Matching , and Identification. *Analytical Chemistry*, *78*(3), 779–787.

Sulyok, M., Krska, R., & Schuhmacher, R. (2010). Application of an LC-MS/MS based multi-mycotoxin method for the semi-quantitative determination of mycotoxins occurring in different types of food infected by moulds. *Food Chemistry*, *119*(1), 408–416. https://doi.org/10.1016/j.foodchem.2009.07.042

Tang, K., Page, J. S., Kelly, R. T., & Marginean, I. (2016). *Electrospray ionization in mass spectrometry*. *Encyclopedia of Spectroscopy and Spectrometry* (3rd ed.). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-803224-4.00319-8

Tang, Y., Lu, L., Zhao, W., & Wang, J. (2009). Rapid detection techniques for biological and chemical

contamination in food: A review. *International Journal of Food Engineering*, *5*(5). https://doi.org/10.2202/1556-3758.1744

Tengstrand, E., Lindberg, J., & Åberg, K. M. (2014). TracMass 2-A modular suite of tools for processing chromatography-full scan mass spectrometry data. *Analytical Chemistry*, *86*(7), 3435–3442. https://doi.org/10.1021/ac403905h

Tengstrand, E., Rosén, J., Hellenäs, K. E., & Åberg, K. M. (2013). A concept study on non-targeted screening for chemical contaminants in food using liquid chromatography-mass spectrometry in combination with a metabolomics approach. *Analytical and Bioanalytical Chemistry*, *405*(4), 1237–1243. https://doi.org/10.1007/s00216-012-6506-5

Thompson, J. W., Kaiser, T. J., & Jorgenson, J. W. (2006). Viscosity measurements of methanol-water and acetonitrile-water mixtures at pressures up to 3500 bar using a novel capillary time-of-flight viscometer. *Journal of Chromatography A*, *1134*(1–2), 201–209. https://doi.org/10.1016/j.chroma.2006.09.006

Thurman, E. M., Ferrer, I., & Barceló, D. (2001). Choosing between atmospheric pressure chemical ionization and electrospray ionization interfaces for the HPLC/MS analysis of pesticides. *Analytical Chemistry*, *73*(22), 5441–5449. https://doi.org/10.1021/ac010506f

Toolkit, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, *30*, 10–12. https://doi.org/10.1038/nbt.2377

Trygg, J., & Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*, *16*(3), 119–128. https://doi.org/10.1002/cem.695

Tsou, C. C., Tsai, C. F., Teo, G. C., Chen, Y. J., & Nesvizhskii, A. I. (2016). Untargeted, spectral library-free analysis of data-independent acquisition proteomics data generated using Orbitrap mass spectrometers. *Proteomics*. https://doi.org/10.1002/pmic.201500526

Tu, J., Yin, Y., Xu, M., Wang, R., & Zhu, Z. J. (2018). Absolute quantitative lipidomics reveals lipidome-wide alterations in aging brain. *Metabolomics*, *14*, 1–11. https://doi.org/10.1007/s11306-017-1304-x

United States Environmental Protection Agency. (2019). Retrieved from https://www.epa.gov/pesticide-worker-safety

van Egmond, H. P., & Jonker, M. A. (2005). Worldwide regulations for mycotoxins in food and feed in 2003: Summary of study. *Food and Nutrition Paper*, 9–28.

Veenaas, C., & Haglund, P. (2017). Methodology for non-target screening of sewage sludge using comprehensive two-dimensional gas chromatography coupled to high-resolution mass spectrometry. *Analytical and Bioanalytical Chemistry*, 4867–4883. https://doi.org/10.1007/s00216-017-0429-0

Wang, T., Frandsen, H. L., Christiansson, N. R., Rosendal, S. E., Pedersen, M., & Smedsgaard, J. (2019). Pyrrolizidine alkaloids in honey: Quantification with and without standards. *Food Control*, *98*, 227–237. https://doi.org/10.1016/j.foodcont.2018.11.033

Wang, T., Li, X., Zhou, B., Li, H., Zeng, J., & Gao, W. (2015). Anti-diabetic activity in type 2 diabetic mice and α-glucosidase inhibitory, antioxidant and anti- inflammatory potential of chemically profiled pear peel and pulp extracts (Pyrus spp.). *Journal of Functional Food*, *13*, 276–288.

Wang, X., Wang, S., & Cai, Z. (2013). The latest developments and applications of mass spectrometry in food-safety and quality analysis. *TrAC - Trends in Analytical Chemistry*, *52*, 170–185. https://doi.org/10.1016/j.trac.2013.08.005

Watrous, J. D., Henglin, M., Claggett, B., Lehmann, K. A., Larson, M. G., Cheng, S., & Jain, M. (2017). Visualization, Quantification, and Alignment of Spectral Drift in Population Scale Untargeted Metabolomics Data. *Analytical Chemistry*, *89*(3), 1399–1404. https://doi.org/10.1021/acs.analchem.6b04337

WHO. (2017). Health Topics: Pesticides. Retrieved from http://www.who.int/topics/ pesticides/en/

WHO & FAO. (2009). Monitoring for Chemicals in Foods. INFOSAN Information Note No. 1/2009.

Retrieved from https://www.who.int/foodsafety/fs_management/No_01_ Chem_Mar09_en.pdf
WHO & FAO. (2019). *Introduction and glossary-Food control system assessment tool*.
Wolf, S., Schmidt, S., Müller-Hannemann, M., & Neumann, S. (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics*, *11*. https://doi.org/10.1186/1471-2105-11-148
Wu, L., Wu, Y., Shen, H., Gong, P., Cao, L., Wang, G., & Hao, H. (2013). Quantitative structure-ion intensity relationship strategy to the prediction of absolute levels without authentic standards. *Analytica Chimica Acta*, *794*, 67–75. https://doi.org/10.1016/j.aca.2013.07.034
Yang, J., Wang, A. Q., Li, X. J., Fan, X., Yin, S. S., & Lan, K. (2016). A chemical profiling strategy for semi-quantitative analysis of flavonoids in Ginkgo extracts. *Journal of Pharmaceutical and Biomedical Analysis*, *123*, 147–154. https://doi.org/10.1016/j.jpba.2016.02.017
Yang, X., Qu, Y., Yuan, Q., Wan, P., Du, Z., Chen, D., & Wong, C. (2013). Effect of ammonium on liquid- and gas-phase protonation and deprotonation in electrospray ionization mass spectrometry. *Analyst*, *138*(2), 659–665. https://doi.org/10.1039/c2an36022e
Yap, C. W. (2011). Software News and Update PaDEL-Descriptor : An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *Journal of Computational Chemistry*, *32*(7), 1466–1474. https://doi.org/10.1002/jcc
Zendong, Z., Sibat, M., Herrenknecht, C., Hess, P., & McCarron, P. (2017). Relative molar response of lipophilic marine algal toxins in liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry*, *31*(17), 1453–1461. https://doi.org/10.1002/rcm.7918
Zheng, H., Clausen, M. R., Dalsgaard, T. K., Mortensen, G., & Bertram, H. C. (2013). Time-saving design of experiment protocol for optimization of LC-MS data processing in metabolomic approaches. *Analytical Chemistry*, *85*(15), 7109–7116. https://doi.org/10.1021/ac4020325
Zhu, X., Chen, Y., & Subramanian, R. (2014). Comparison of information-dependent acquisition, SWATH, and MS All techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Analytical Chemistry*, *86*(2), 1202–1209. https://doi.org/10.1021/ac403385y

Appendix

Appendix contents:

Appendix I

I. Manuscript 1 supplementary information (163)

Supporting information

Non-targeted food unexpected compounds analysis by LC/HRMS: feasibility study on rice

Tingting Wang, * Lene Duedahl-Olesen, Henrik Lauritz Frandsen,

National Food Institute, Research Group for Analytical Food Chemistry, Technical University of Denmark, Kemitorvet Building 202, Kgs. Lyngby, DK-2800, Denmark

Supporting information

Table S1. D optimal design, four factors

Parameters	Term	Level-1	Level_0	Level+1
Capillary voltage	Cap	2500	3500	4500
Nebulizer	Neb	15	30	45
Dry Gas Flow rate	Dry F	8	10	12
Dry Temperature	DryT	150	185	220

Table S2. D optimal design with 10000 starts, four factors, split plot design.

Whole Plots	Cap	Neb	Dry F	DryT
1	4500	15	12	220
1	2500	15	8	220
1	2500	45	12	220
1	4500	45	8	220
2	4500	45	12	220
2	2500	45	8	220
2	4500	15	8	220
2	2500	15	12	220
3	4500	45	10	150
3	2500	45	8	150
3	3500	15	8	150
3	2500	30	12	150
4	3500	30	10	185
4	4500	45	12	185
4	2500	15	12	185
4	4500	15	8	185
5	3470	45	12	150
5	2500	15	10	150
5	4500	15	12	150
5	4500	30	8	150

TableS3 Parameters for Target analysis

parameters	value
Detection parameters	
Overal retention time range	0.4-25 min
Retention time tolerance	± 25 min
Extracted ion chromatogram width	± 5 mDa
Creat chromatogram slices	± 0.5 min
Identification parameters	
Identification m/z tolerance	± 10 mDa
mSigma threshold	500
Smoothing of chromatogram	
Smoothing algrithm	Gauss
Smoothing width	1
cycle	1
Baseline substraction of chromatograms	
Smoothing of mass spectra	
Smoothing algrithm	Gauss
Smoothing width	0.5
cycle	1
Baseline substraction of mass spectra	0.8

TableS4 Parameters for Metabolitedetect

parameters	value
Calculate difference	
mode	eXpose
ratio	3
DeltaTime	0.5
DeltaMass	0.02
interval	2-21min
Det.Mass	
Peak detection Chromatogram	Base peak
Peak detection Mode	with peak intergration
Trace width	± 0.02
Delta RT	0.2
S/N	3
Skim ratio	3
Smoothing width	1
Intensity threshold	5%
Max Number of Peaks	8

TableS5 Full parameters for peak extraction using XCMS based approach

Step	Parameter	Value
xcmset	Extraction method for peaks detection	centWave
	ppm	30
	peakwidth	10-60
	mzdiff	-0.001
	snthresh	10
	integrate	1
	prefilter	0,0
	Noise filter	0
group - 1	Method to use for grouping	PeakDensity
	bw	2
	minFraction	0.5
	minSamples	1
	binSize	0.025
	maxFeatures	50
retcor	Method to use for retention time correction	peakgroups
	smooth	loess
	minFraction	0.9
	extraPeaks	1
	span	0.2
	family	gaussian
group - 2	Method to use for grouping	PeakDensity
	bw	2
	minFraction	0.5
	minSamples	1
	binSize	0.025
	maxFeatures	50
fillPeaks	convertRTMinute	YES
	decimal places for MS	4
	decimal places for RT	2
	Replace the remain NA by 0	YES
	intval	INTO
	sigma	6
	perfwhm	0.6
	ppm	30
	mzabs	0.025
	maxcharge	1
	maxiso	4
	minfrac	0.5
	minfrac	0.5
	convertRTMinute	FALSE
	numDigitsMZ	4

Table S6. Pos-Effect summary

Source	LogWorth	PValue
Dry F(8,12)	2,315	0,00484
Cap*Dry F	1,682	0,02077
Cap(2500,4500)	1,647	0,02253
Cap*Neb	1,031	0,09321
Neb*Dry F	0,727	0,18749
Dry F*DryT	0,720	0,19066
Dry F*Dry F	0,426	0,37540
Cap*Cap	0,307	0,49319
Neb(15,45)	0,259	0,55066
Cap*DryT	0,215	0,60893
Neb*DryT	0,199	0,63177
Neb*Neb	0,155	0,69955
DryT*DryT	0,025	0,94365
DryT(150,220)	0,012	0,97174

Table S7. Pos-Parameter estimate

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	6586.3883	1348.101	4.239	4.89	0.0070*
Cap(2500,4500)	1179.7121	282.239	3.159	4.18	0.0225*
Neb(15,45)	-188.0057	282.1773	3.161	-0.67	0.5507
Dry F(8,12)	-2003.184	282.6968	3.164	-7.09	0.0048*
DryT(150,220)	-21.21279	535.3189	2.169	-0.04	0.9717
Cap*Cap	-589,7753	759.1832	3.04	-0.78	0.4932
Cap*Neb	724.40818	299.5505	3.039	2.42	0.0932
Neb*Neb	-321.5699	757.7865	3.04	-0.42	0.6996
Cap*Dry F	1326.0949	299.1792	3.039	4.43	0.0208*
Neb*Dry F	507.70096	299.5361	3.039	1.69	0.1875
Dry F*Dry F	-786.4884	759.0233	3.04	-1.04	0.3754
Cap*DryT	-168.3032	297.0694	3.147	-0.57	0.6089
Neb*DryT	-157.3058	297.299	3.149	-0.53	0.6318
Dry F*DryT	-494.7107	297.603	3.151	-1.66	0.1907
DryT*DryT	-93.82311	1176.617	2.016	-0.08	0.9437

Table S8. Pos-Fixed effect tests

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Cap(2500,4500)	1	1	3.159	17.4710	0.0225*
Neb(15,45)	1	1	3.161	0.4439	0.5507
Dry F(8,12)	1	1	3.164	50.2111	0.0048*
DryT(150,220)	1	1	2.169	0.0016	0.9717

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Cap*Cap	1	1	3.04	0.6035	0.4932
Cap*Neb	1	1	3.039	5.8483	0.0932
Neb*Neb	1	1	3.04	0.1801	0.6996
Cap*Dry F	1	1	3.039	19.6466	0.0208*
Neb*Dry F	1	1	3.039	2.8729	0.1875
Dry F*Dry F	1	1	3.04	1.0737	0.3754
Cap*DryT	1	1	3.147	0.3210	0.6089
Neb*DryT	1	1	3.149	0.2800	0.6318
Dry F*DryT	1	1	3.151	2.7633	0.1907
DryT*DryT	1	1	2.016	0.0064	0.9437

Table S9. Neg-effect summary

Source	LogWorth	PValue
Neb(15,45)	3.365	0.00043
DryT(150,220)	1.970	0.01072
DryT*DryT	1.280	0.05251
Neb*DryT	1.181	0.06595
Dry F(8,12)	0.902	0.12521
Cap*Cap	0.830	0.14799
Cap(2500,4500)	0.784	0.16442
Cap*DryT	0.776	0.16752
Cap*Neb	0.491	0.32303
Cap*Dry F	0.432	0.36963
Dry F*DryT	0.180	0.66012
Dry F*Dry F	0.160	0.69202
Neb*Neb	0.128	0.74420
Neb*Dry F	0.023	0.94781

Table S10. Neg-Parameter estimate

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	621.66522	38.33627	16.22	<.0001*
Cap(2500,4500)	16.491786	10.12921	1.63	0.1644
Neb(15,45)	83.299139	10.1211	8.23	0.0004*
Dry F(8,12)	-18.62396	10.12388	-1.84	0.1252
DryT(150,220)	-44.65511	11.27098	-3.96	0.0107*
Cap*Cap	-49.09068	28.71048	-1.71	0.1480
Cap*Neb	-12.42247	11.33415	-1.10	0.3230
Neb*Neb	9.8858916	28.66219	0.34	0.7442
Cap*Dry F	-11.15746	11.32148	-0.99	0.3696
Neb*Dry F	0.7799545	11.33462	0.07	0.9478
Dry F*Dry F	-12.05198	28.70496	-0.42	0.6920
Cap*DryT	-17.29739	10.71923	-1.61	0.1675
Neb*DryT	-25.13614	10.71815	-2.35	0.0660
Dry F*DryT	5.0059304	10.71861	0.47	0.6601
DryT*DryT	-59.792	23.62973	-2.53	0.0525

Table S11. Neg-Fixed effect tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Cap(2500,4500)	1	1	4186.55	2.6508	0.1644
Neb(15,45)	1	1	106978.79	67.7369	0.0004*
Dry F(8,12)	1	1	5344.69	3.3842	0.1252
DryT(150,220)	1	1	24790.83	15.6971	0.0107*
Cap*Cap	1	1	4617.32	2.9236	0.1480
Cap*Neb	1	1	1897.19	1.2013	0.3230
Neb*Neb	1	1	187.88	0.1190	0.7442
Cap*Dry F	1	1	1533.90	0.9712	0.3696
Neb*Dry F	1	1	7.48	0.0047	0.9478
Dry F*Dry F	1	1	278.40	0.1763	0.6920
Cap*DryT	1	1	4112.51	2.6040	0.1675
Neb*DryT	1	1	8686.19	5.4999	0.0660
Dry F*DryT	1	1	344.48	0.2181	0.6601
DryT*DryT	1	1	10112.10	6.4028	0.0525

Figure S1.

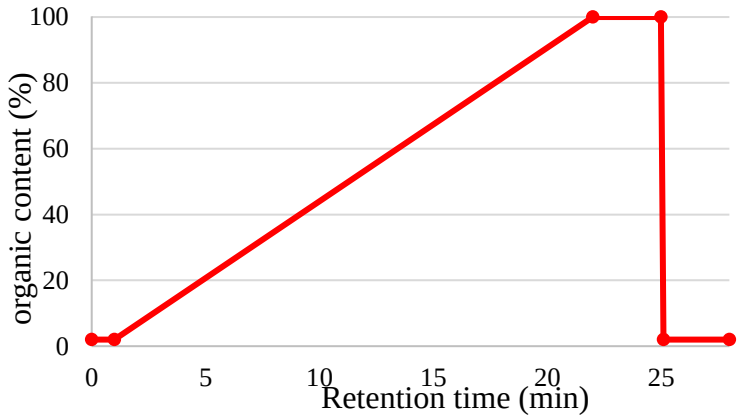


Figure S2.

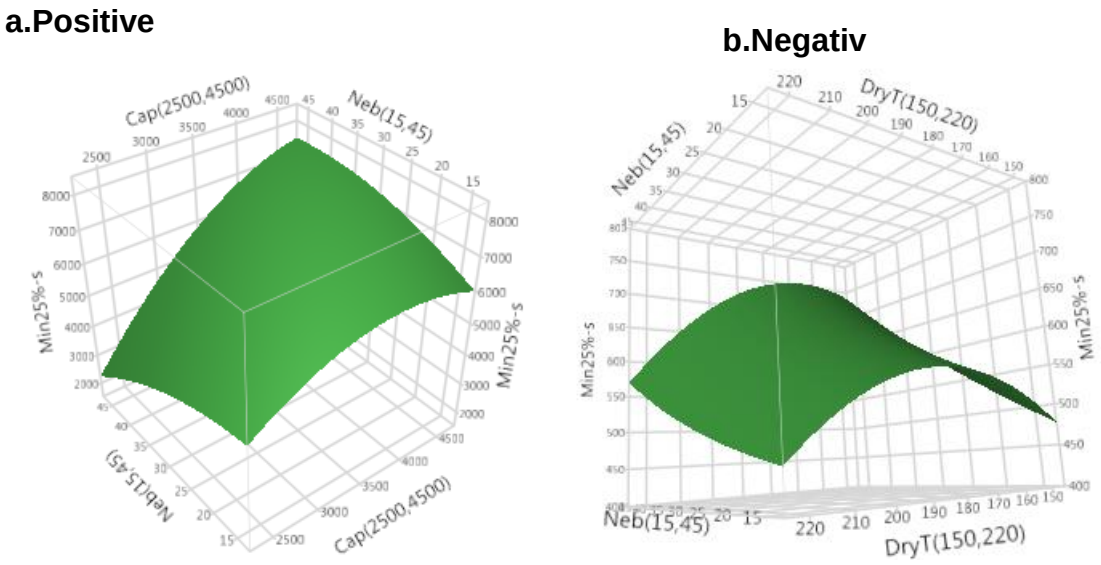


Figure S3

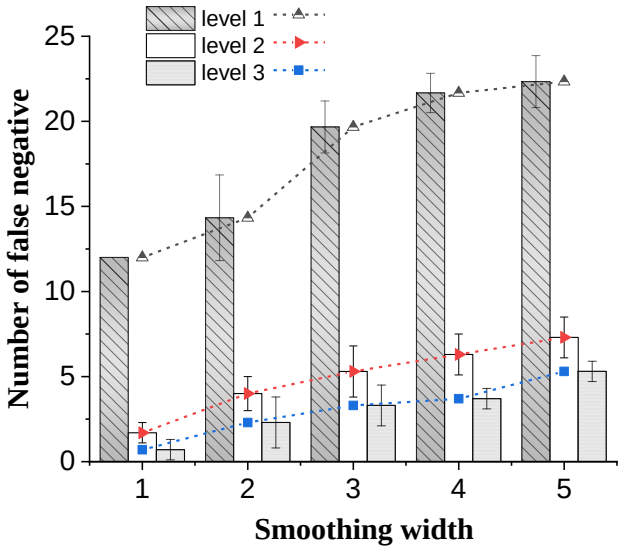
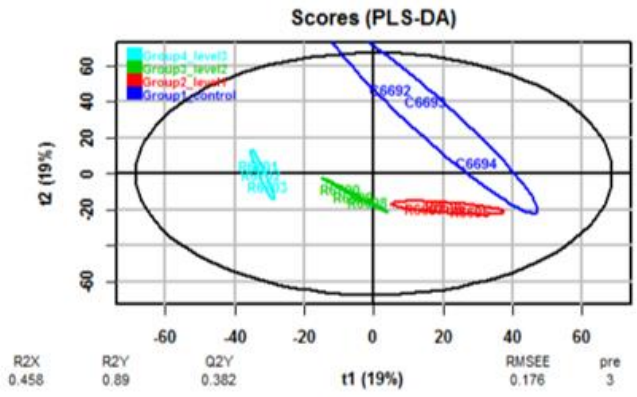
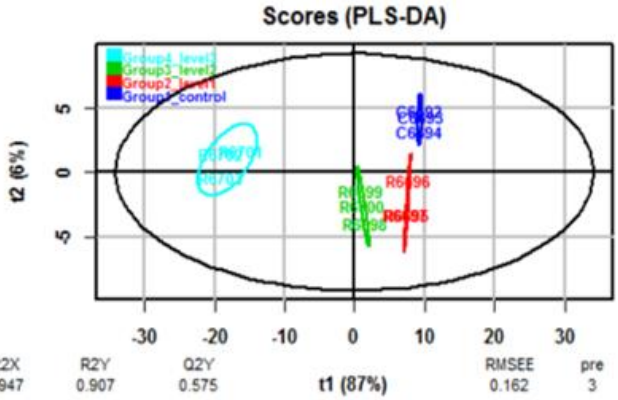


Figure S4.

a **b**


Scores (PLS-DA)
Group4_level3
Group3_level2
Group2_level1
Group1_control
C6692
C6693
C6694
t2 (19%)
t1 (19%)
R2X
0.458
R2Y
0.89
Q2Y
0.382
RMSEE
0.176
pre
3


Scores (PLS-DA)
Group4_level3
Group3_level2
Group2_level1
Group1_control
C6694
R6696
t2 (6%)
t1 (87%)
R2X
0.947
R2Y
0.907
Q2Y
0.575
RMSEE
0.162
pre
3

Appendix II

II. Manuscript 2 supplementary information (174)

Supporting information
Pyrrolizidine alkaloids in honey: Quantification with and without standards

Tingting Wang, Henrik Lauritz Frandsen, Natasha Roed Christiansson, Sofie Elisabeth Rosendal, Mikael Pedersen, Jørn Smedsgaard*

National Food Institute, Division of Analytical Food Chemistry, Technical University of Denmark, Kemitorvet Building 202, DK 2800 Lyngby, Denmark

---

* Corresponding author: Jørn Smedsgaard
E-mail address: smeds@food.dtu.dk

Supporting Information
Tables

Table of Contents:

Supporting Information
Figures

Table of Contents:

Figure S1. Residues plot

Accuracy residues (%) = Accuracy (%)-100%

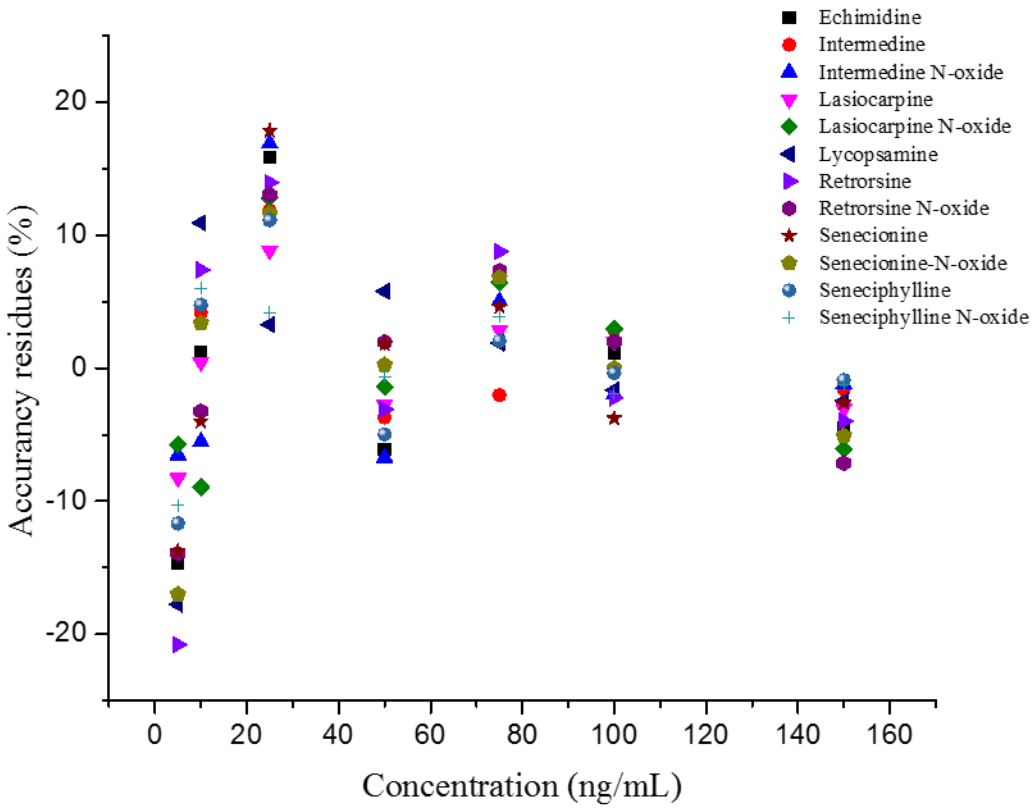


Figure S2. **Spiked blank honey sample with 1 ng/ml (corresponding to 0.5 µg/kg).** 1. Intermedine, 2. Lycopsamine, 3. Intermedine N-oxide, 4. Retrosine, 5. Retrosine N-oxide, 6. Seneciphylline, 7. Seneciphylline N-oxide, 8. Senecionine, 9. Senecionine N-oxide, 10. Echimidine, 11. Lasiocarpine, 12. Lasiocarpine N-oxide.

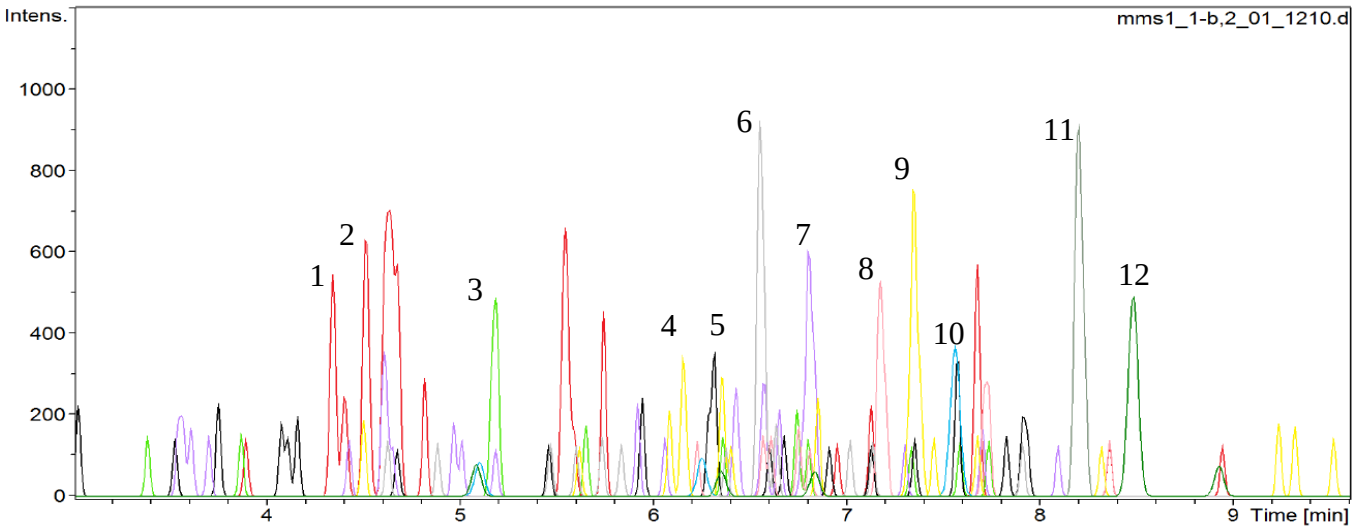


Figure S3. Fragmentation pattern of selected PAs. C-1, monoester and open-chained diester; C-2, *N*-oxide open-chained diester.

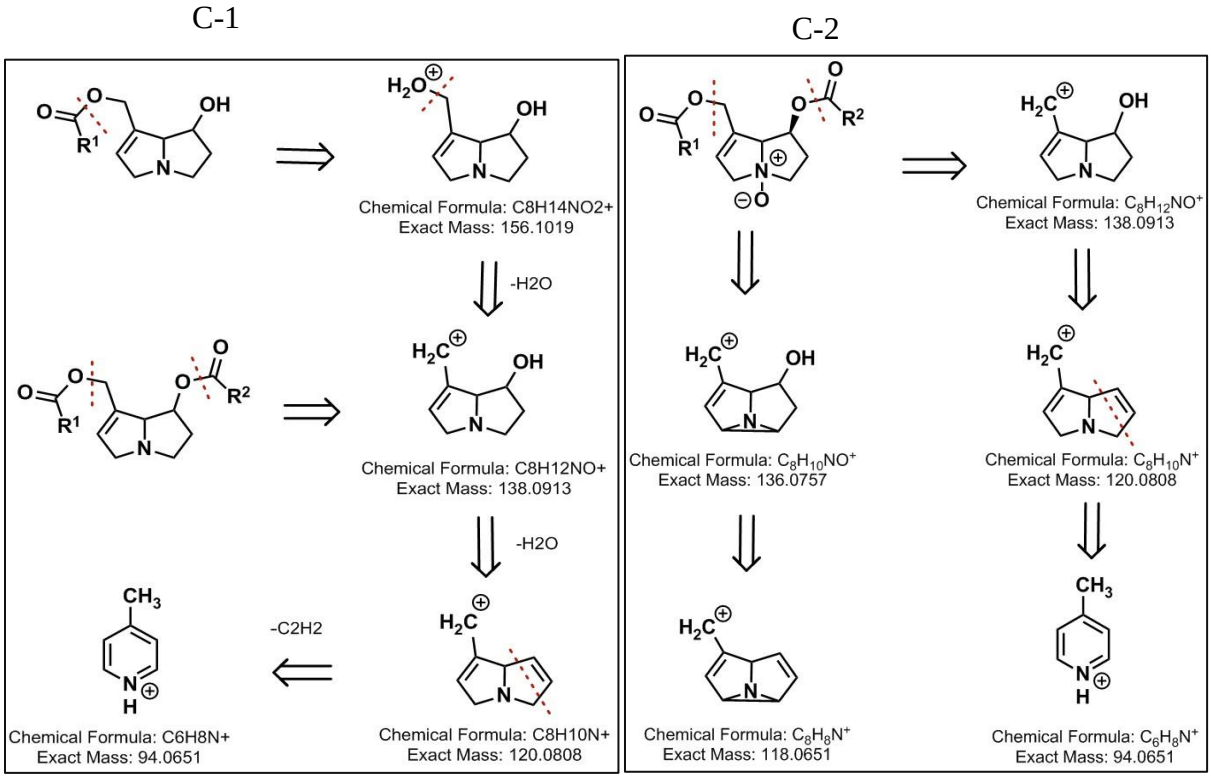
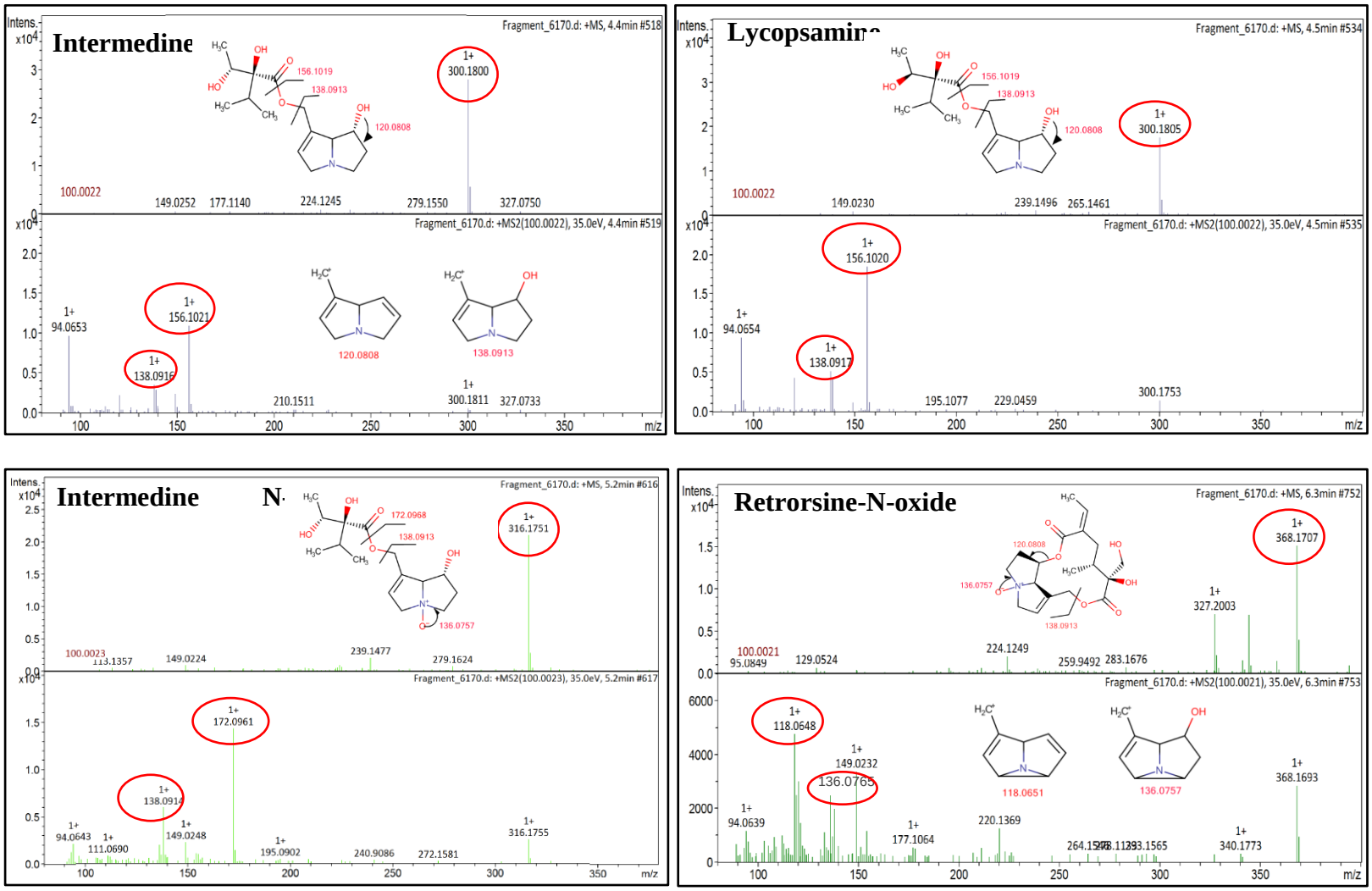
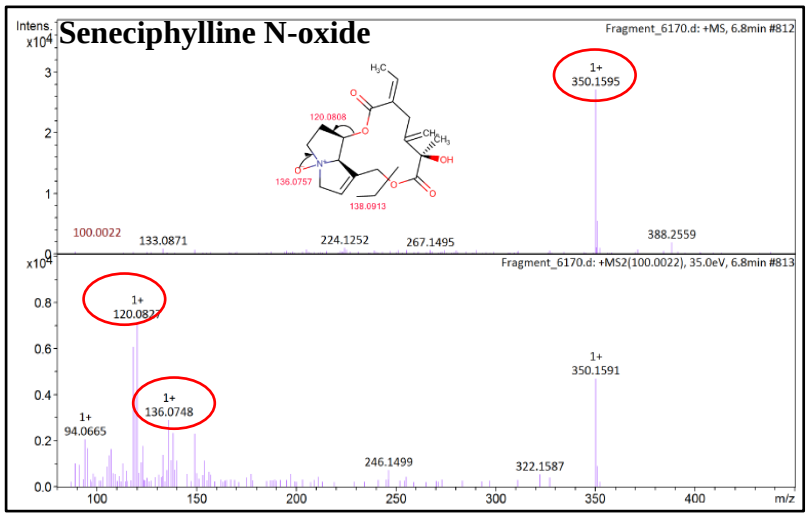
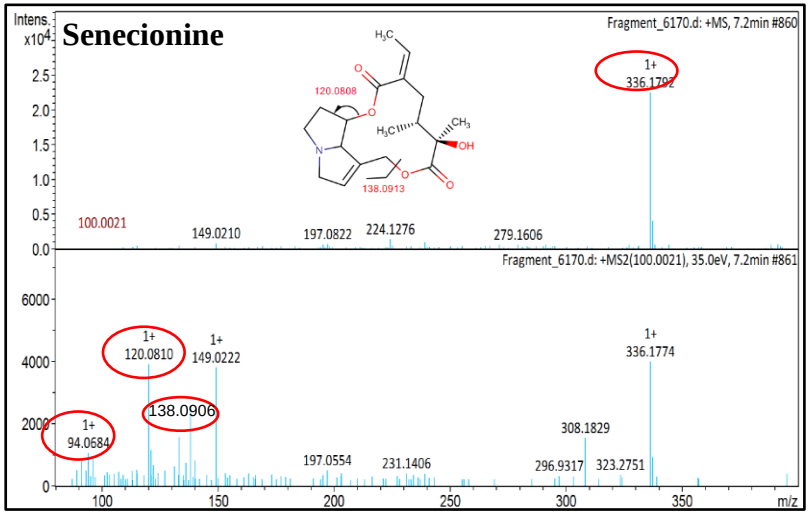
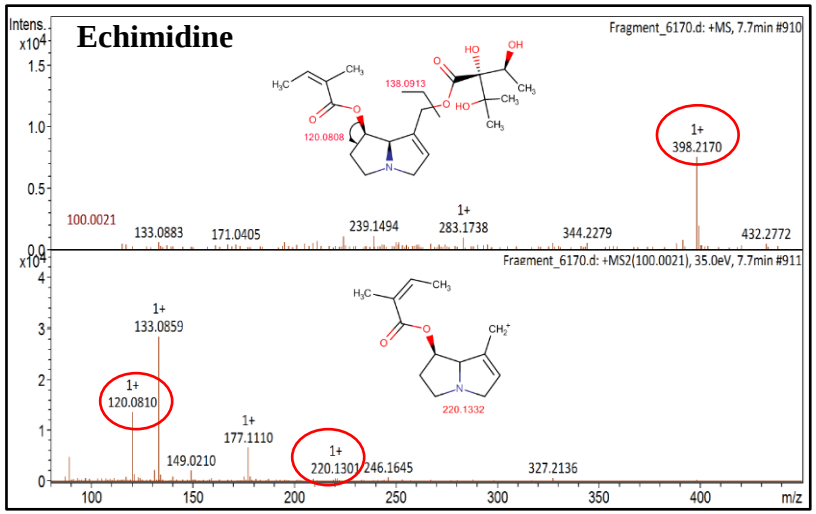


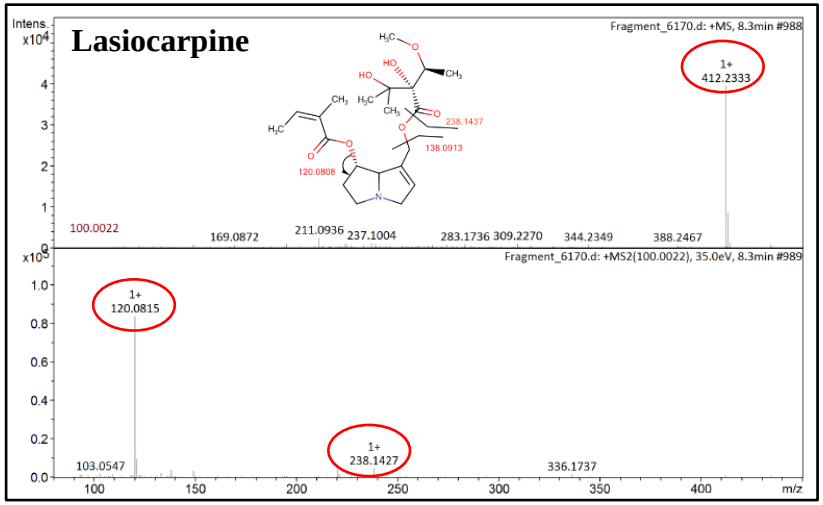
Figure S4. 12 PAs with authentic standards and other tentatively identified PAs found in honey

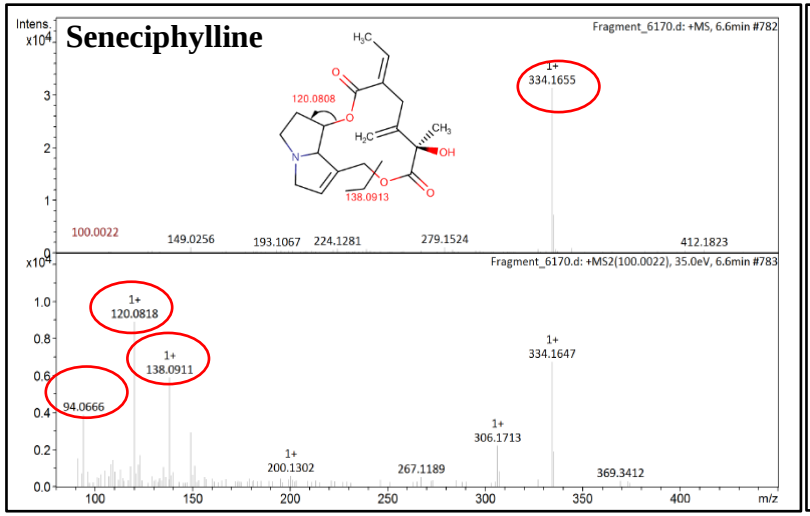


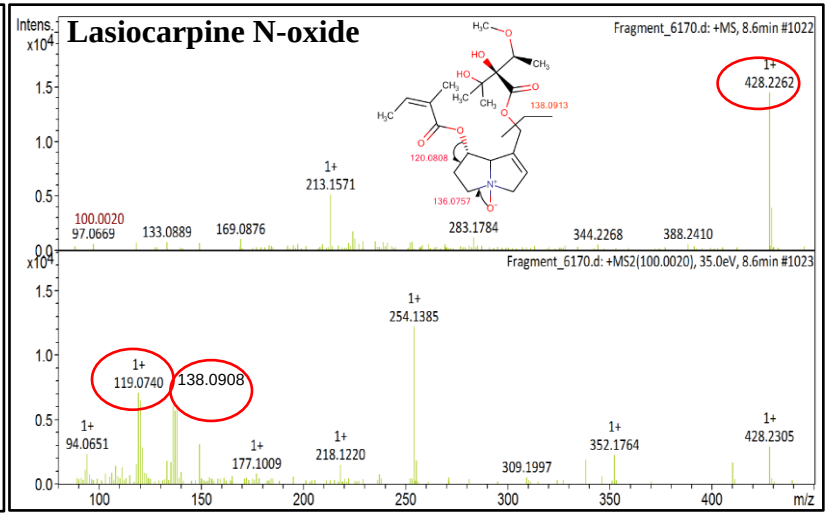

Seneciphylline N-oxide
Fragment_6170.d: +MS, 6.8min #812
1+ 350.1595
Fragment_6170.d: +MS2(100.0022), 35.0eV, 6.8min #813
1+ 120.0810
1+ 136.0748
1+ 94.0665
1+ 350.1591

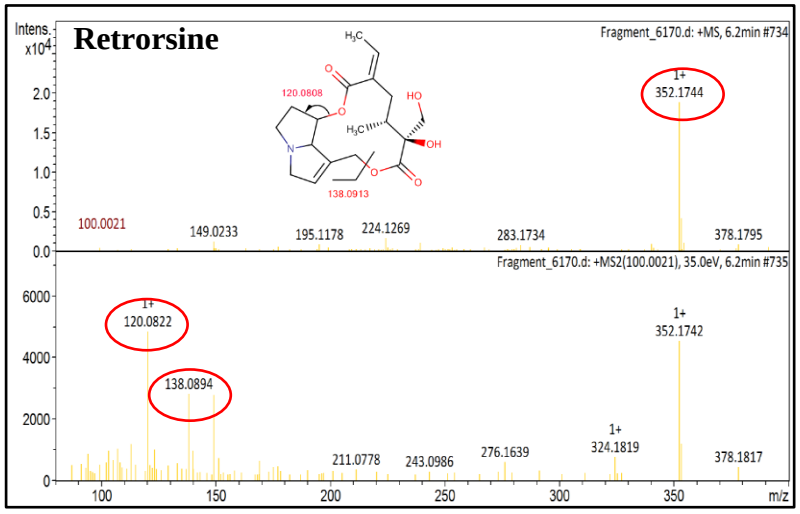

Senecionine
Fragment_6170.d: +MS, 7.2min #860
1+ 336.1792
Fragment_6170.d: +MS2(100.0021), 35.0eV, 7.2min #861
1+ 120.0810
1+ 149.0222
138.0906
1+ 94.0684
1+ 336.1774

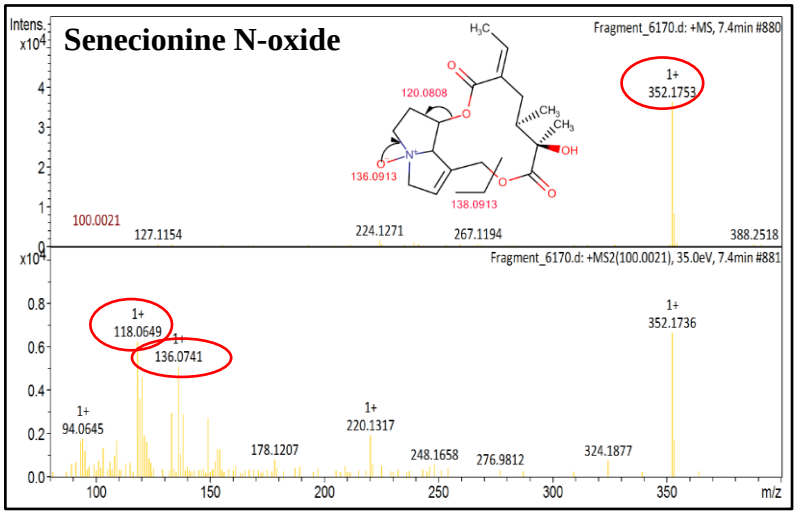

Echimidine
Fragment_6170.d: +MS, 7.7min #910
1+ 398.2170
Fragment_6170.d: +MS2(100.0021), 35.0eV, 7.7min #911
1+ 133.0859
1+ 120.0810
1+ 177.1110
1+ 220.1301

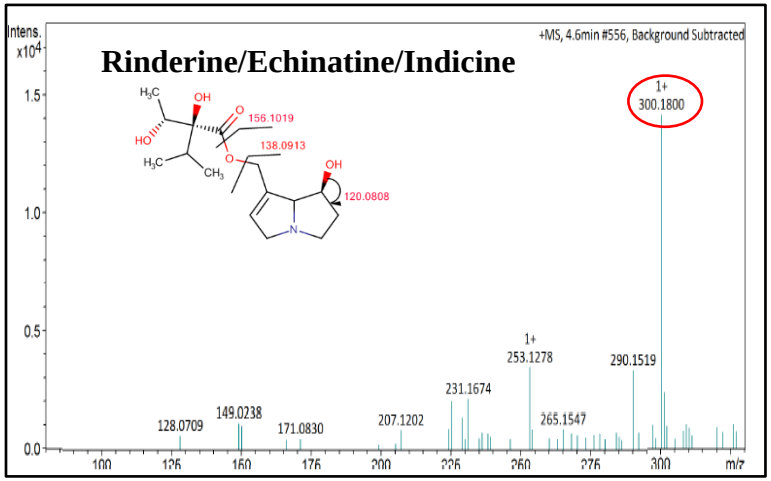

Lasiocarpine
Fragment_6170.d: +MS, 8.3min #988
1+ 412.2333
Fragment_6170.d: +MS2(100.0022), 35.0eV, 8.3min #989
1+ 120.0815
1+ 238.1427

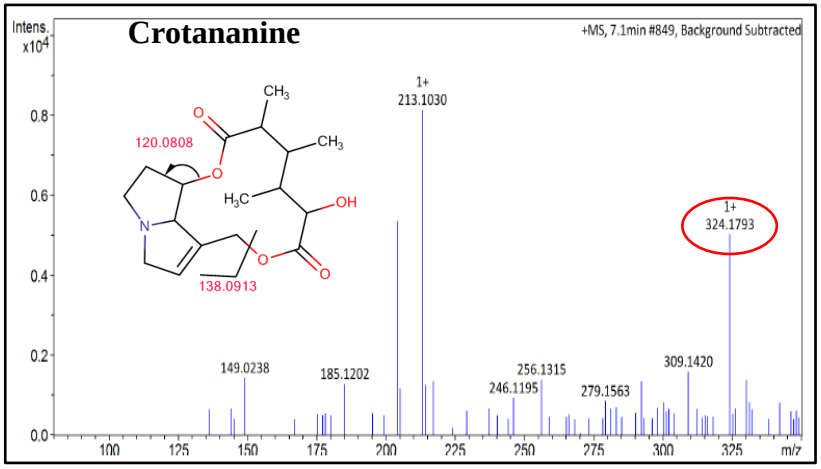

Seneciphylline
Fragment_6170.d: +MS, 6.6min #782
1+ 334.1655
120.0808
138.0913
100.0022
149.0256
193.1067
224.1281
279.1524
412.1823
Fragment_6170.d: +MS2(100.0022), 35.0eV, 6.6min #783
1+ 120.0818
1+ 138.0911
94.0666
1+ 334.1647
1+ 306.1713
1+ 200.1302
267.1189
369.3412
m/z

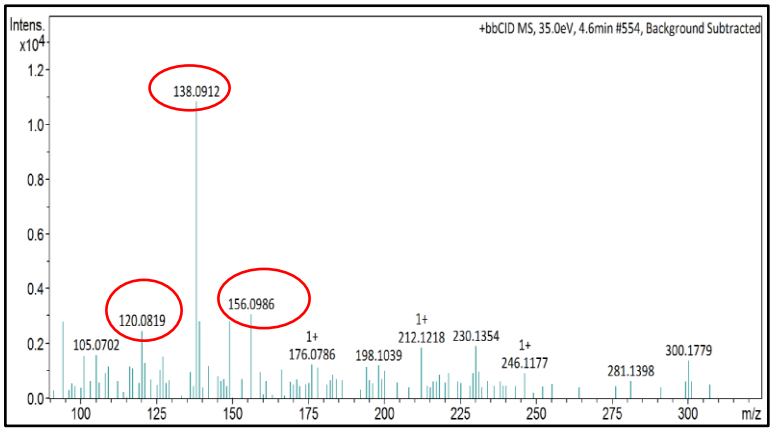

Lasiocarpine N-oxide
Fragment_6170.d: +MS, 8.6min #1022
1+ 428.2262
138.0913
120.0808
136.0757
1+ 213.1571
100.0020
97.0669
133.0889
169.0876
283.1784
344.2268
388.2410
Fragment_6170.d: +MS2(100.0020), 35.0eV, 8.6min #1023
1+ 254.1385
1+ 119.0740
138.0908
1+ 94.0651
1+ 177.1009
1+ 218.1220
309.1997
1+ 352.1764
1+ 428.2305
m/z

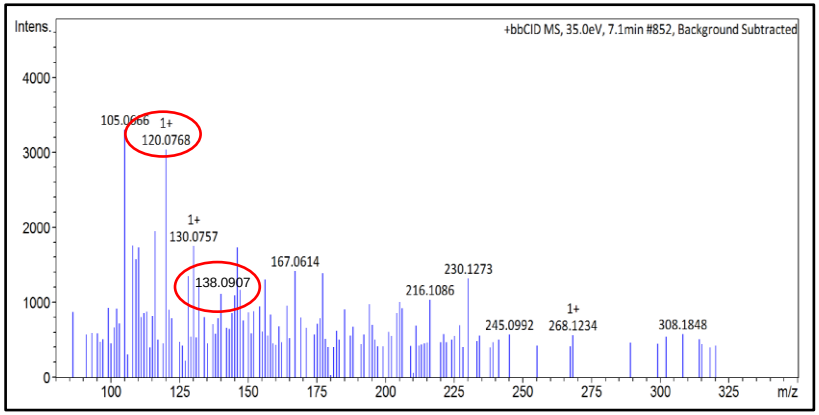

Retrorsine
Fragment_6170.d: +MS, 6.2min #734
1+ 352.1744
120.0808
138.0913
100.0021
149.0233
195.1178
224.1269
283.1734
378.1795
Fragment_6170.d: +MS2(100.0021), 35.0eV, 6.2min #735
1+ 120.0822
1+ 352.1742
138.0894
211.0778
243.0986
276.1639
1+ 324.1819
378.1817
m/z

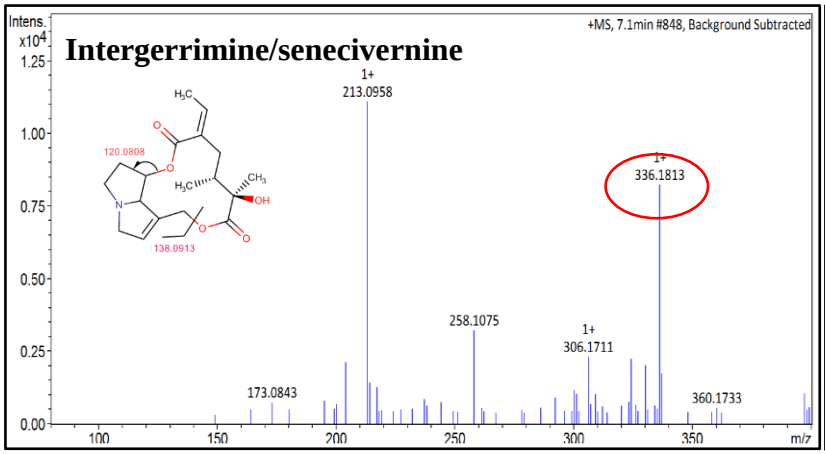

Senecionine N-oxide
Fragment_6170.d: +MS, 7.4min #880
1+ 352.1753
120.0808
136.0913
138.0913
100.0021
127.1154
224.1271
267.1194
388.2518
Fragment_6170.d: +MS2(100.0021), 35.0eV, 7.4min #881
1+ 118.0649
1+ 136.0741
1+ 352.1736
1+ 94.0645
1+ 220.1317
178.1207
248.1658
276.9812
324.1877
m/z

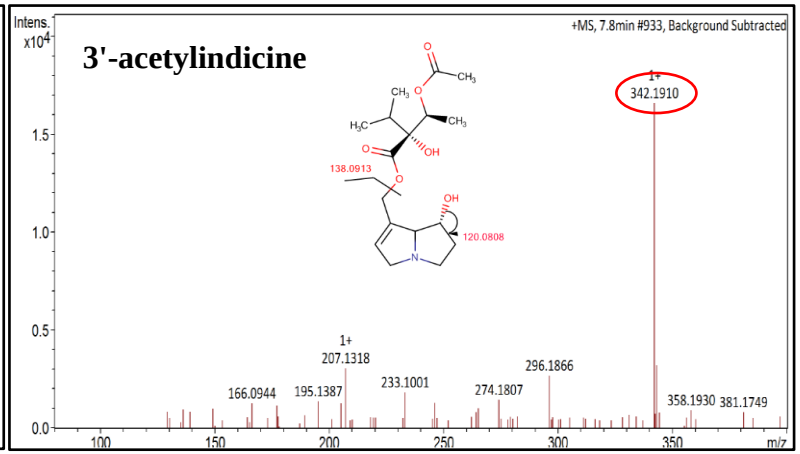

Rinderine/Echinatine/Indicine
+MS, 4.6min #556, Background Subtracted
Intens.
x10^4
1+
300.1800
156.1019
138.0913
120.0808
1+
253.1278
290.1519
231.1674
149.0238
128.0709
171.0830
207.1202
265.1547
m/z

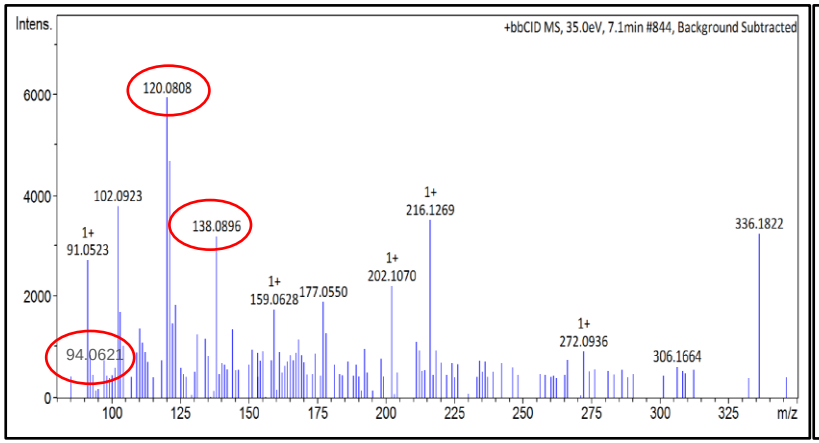

Crotananine
+MS, 7.1min #849, Background Subtracted
Intens.
x10^4
120.0808
138.0913
1+
213.1030
1+
324.1793
149.0238
185.1202
256.1315
246.1195
279.1563
309.1420
m/z

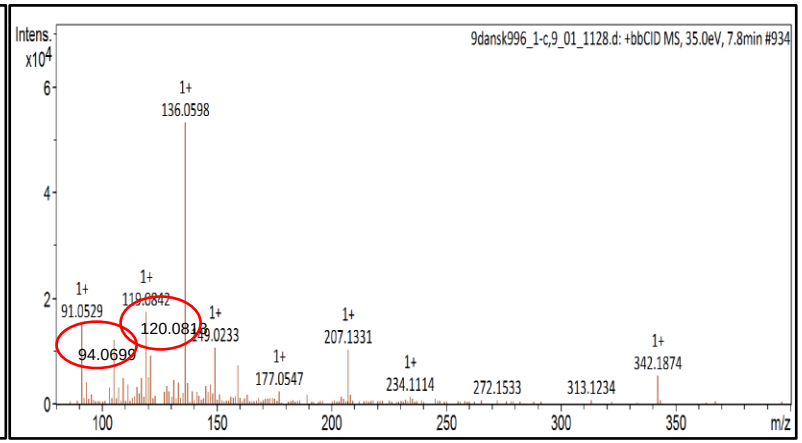

+bbCID MS, 35.0eV, 4.6min #554, Background Subtracted
Intens.
x10^4
138.0912
120.0819
156.0986
105.0702
1+
176.0786
198.1039
1+
212.1218
230.1354
1+
246.1177
281.1398
300.1779
m/z

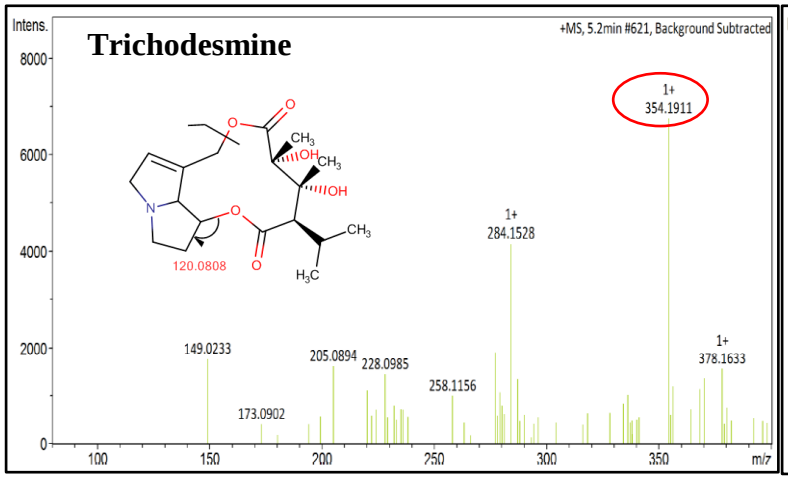

+bbCID MS, 35.0eV, 7.1min #852, Background Subtracted
Intens.
105.0666
1+
120.0768
1+
130.0757
138.0907
167.0614
230.1273
216.1086
245.0992
1+
268.1234
308.1848
m/z

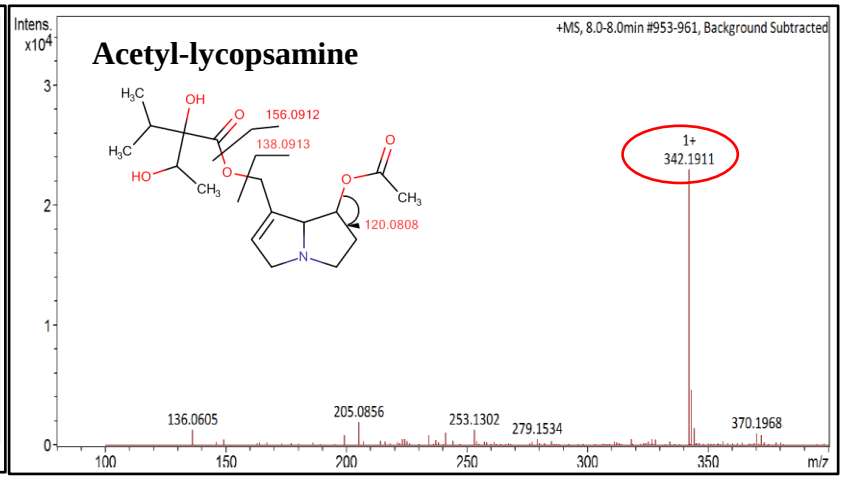

Intergerrimine/senecivernine
+MS, 7.1min #848, Background Subtracted
Intens.
x10^4
1+
213.0958
1+
336.1813
258.1075
1+
306.1711
173.0843
360.1733
120.0808
138.0913
m/z

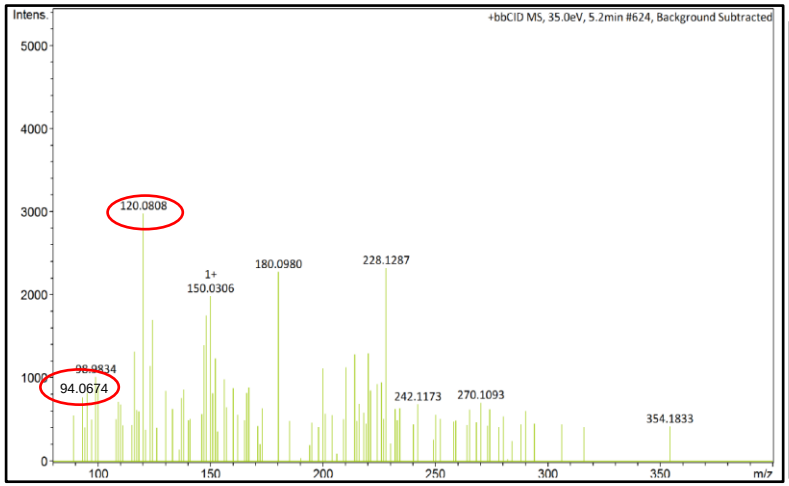

3'-acetylindicine
+MS, 7.8min #933, Background Subtracted
Intens.
x10^4
1+
342.1910
1+
207.1318
233.1001
296.1866
166.0944
195.1387
274.1807
358.1930
381.1749
138.0913
120.0808
m/z

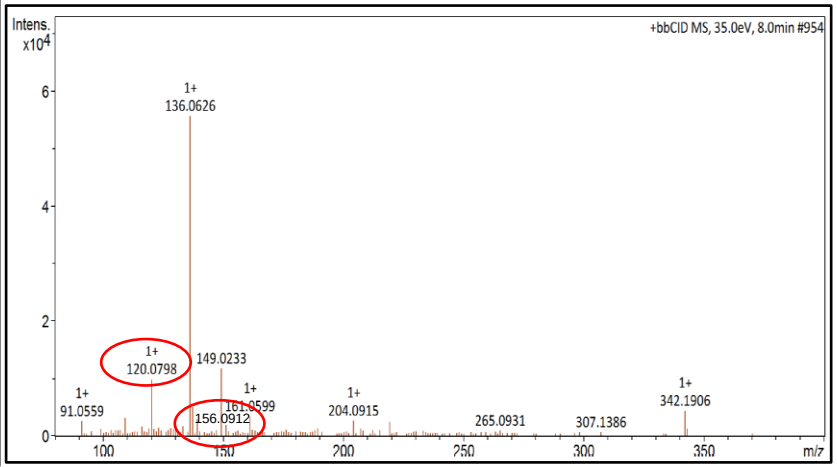

+bbCID MS, 35.0eV, 7.1min #844, Background Subtracted
Intens.
120.0808
102.0923
138.0896
1+
91.0523
1+
216.1269
1+
202.1070
1+
159.0628
177.0550
336.1822
1+
272.0936
306.1664
94.0621
m/z

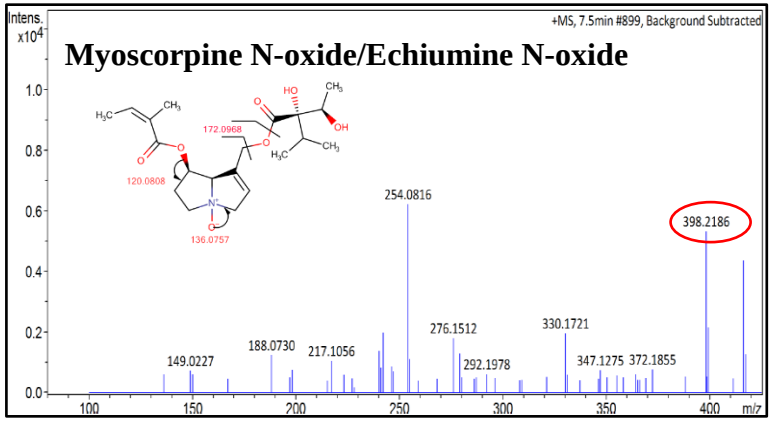

9dansk996_1-c,9_01_1128.d: +bbCID MS, 35.0eV, 7.8min #934
Intens.
x10^4
1+
136.0598
1+
91.0529
1+
118.0842
120.0818
1+
149.0233
94.0699
1+
177.0547
1+
207.1331
1+
234.1114
272.1533
313.1234
1+
342.1874
m/z

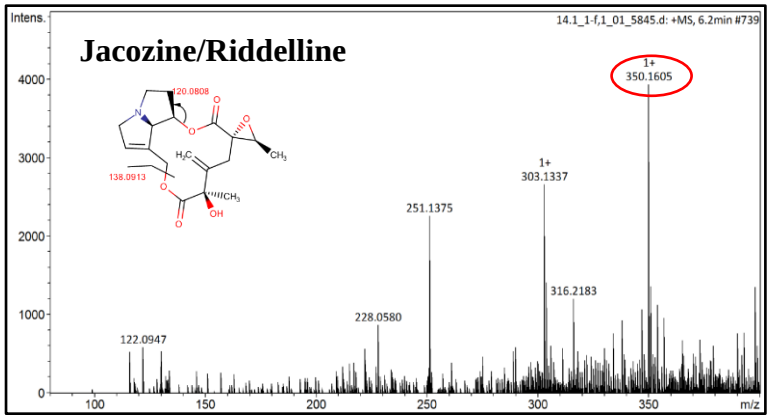

Trichodesmine
+MS, 5.2min #621, Background Subtracted
Intens.
1+
354.1911
1+
284.1528
1+
378.1633
149.0233
205.0894
228.0985
258.1156
173.0902
120.0808
m/z

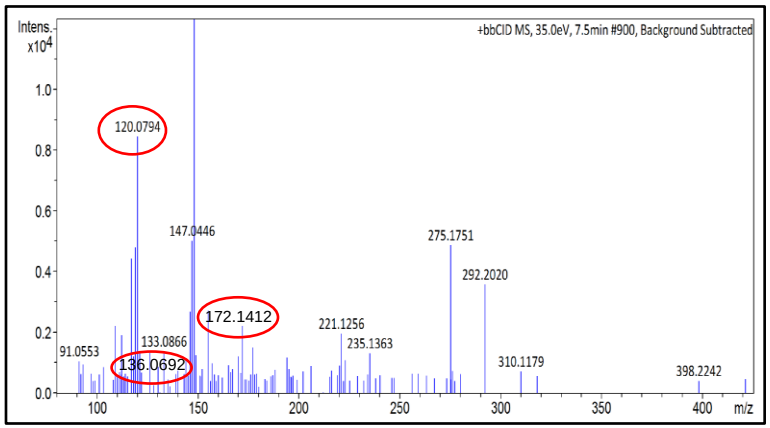

Acetyl-lycopsamine
+MS, 8.0-8.0min #953-961, Background Subtracted
Intens.
x10^4
1+
342.1911
156.0912
138.0913
120.0808
136.0605
205.0856
253.1302
279.1534
370.1968
m/z


+bbCID MS, 35.0eV, 5.2min #624, Background Subtracted
Intens.
120.0808
1+
150.0306
180.0980
228.1287
94.0674
242.1173
270.1093
354.1833
m/z


+bbCID MS, 35.0eV, 8.0min #954
Intens.
x10^4
1+
136.0626
1+
120.0798
149.0233
1+
91.0559
156.0912
1+
204.0915
265.0931
307.1386
1+
342.1906
m/z


Myoscorpine N-oxide/Echiumine N-oxide
+MS, 7.5min #899, Background Subtracted
Intens.
x10^4
120.0808
172.0968
136.0757
254.0816
398.2186
276.1512
330.1721
188.0730
217.1056
149.0227
292.1978
347.1275
372.1855
m/z


Jacozine/Riddelline
14.1_1-f,1_01_5845.d: +MS, 6.2min #739
Intens.
120.0808
138.0913
1+
350.1605
1+
303.1337
251.1375
316.2183
228.0580
122.0947
m/z


+bbCID MS, 35.0eV, 7.5min #900, Background Subtracted
Intens.
x10^4
120.0794
147.0446
275.1751
292.2020
172.1412
221.1256
133.0866
235.1363
91.0553
136.0692
310.1179
398.2242
m/z

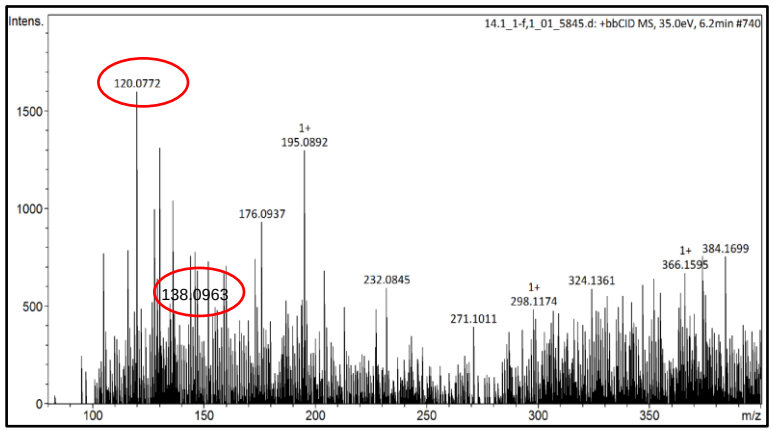
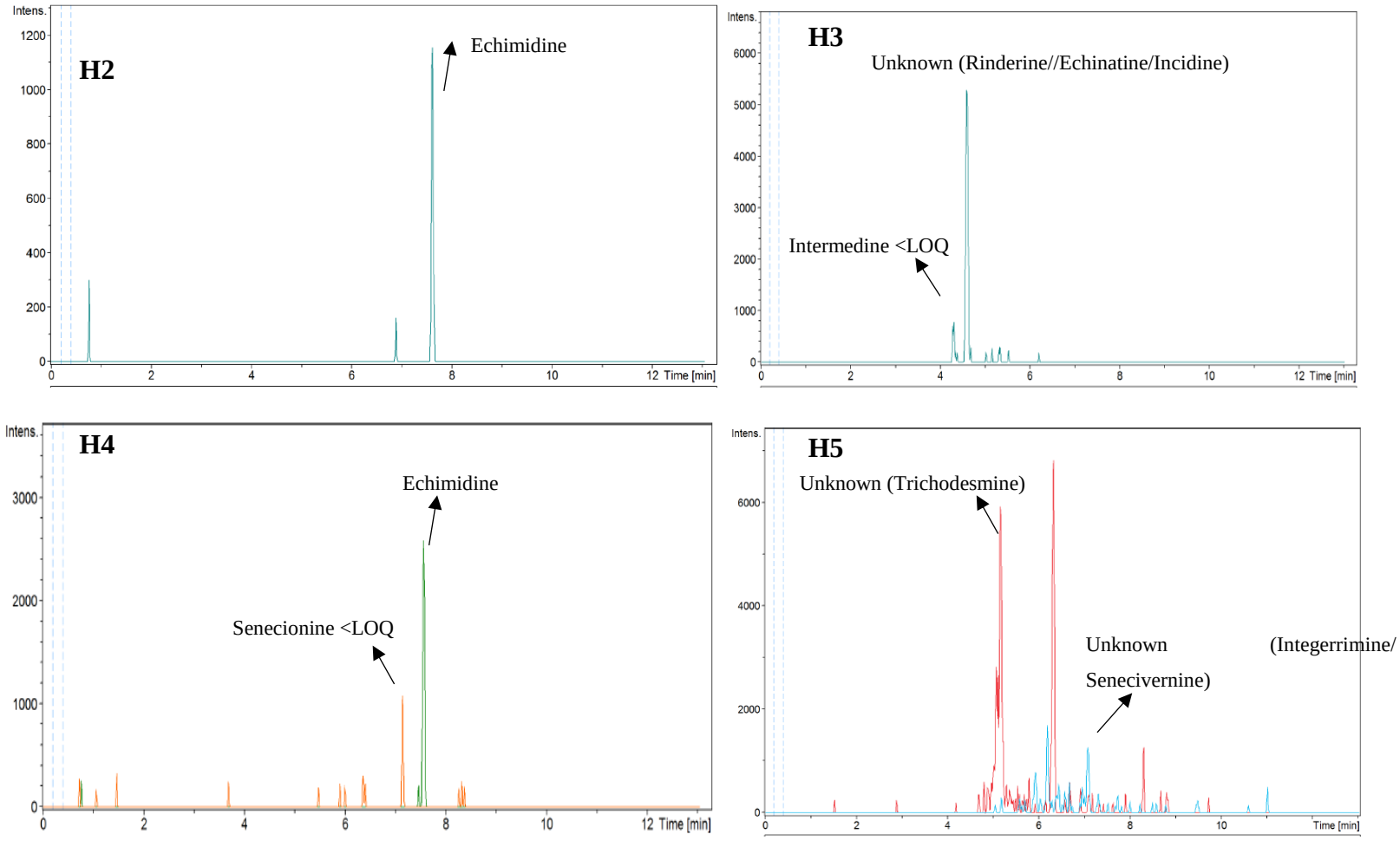
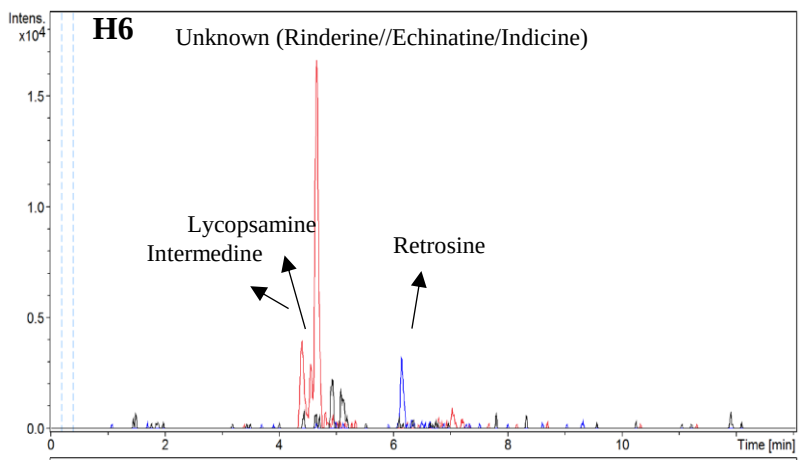
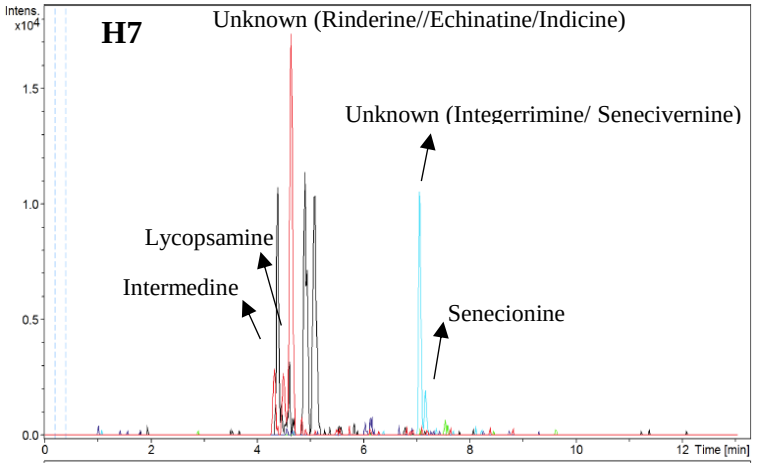
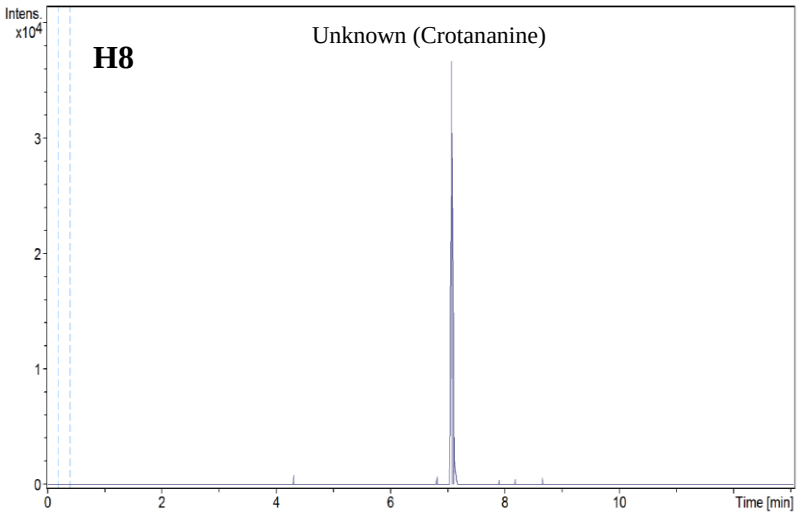

14.1_1-f,1_01_5845.d: +bbCID MS, 35.0eV, 6.2min #740
Intens.
120.0772
1+
195.0892
176.0937
232.0845
138.0963
1+
298.1174
271.1011
324.1361
1+
366.1595
384.1699
m/z

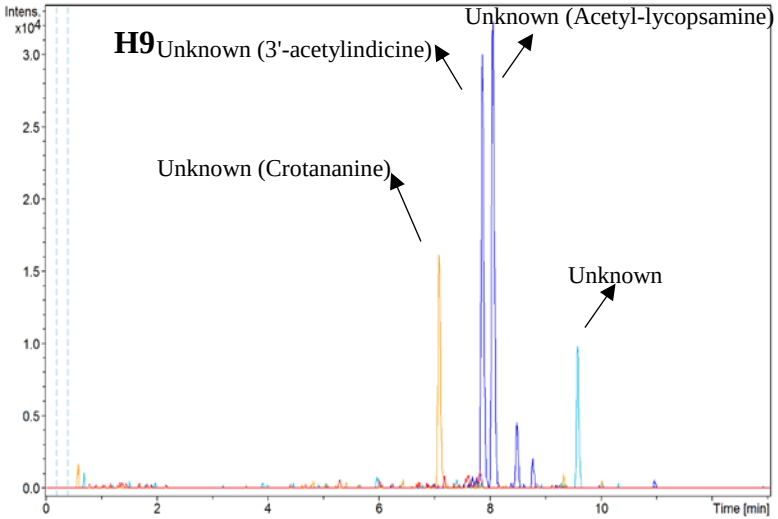
Figure S5. Extracted ion chromatograms of honey samples

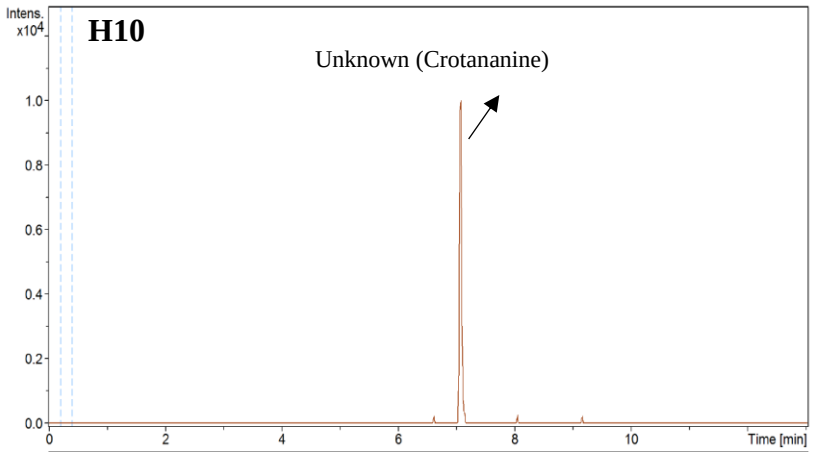


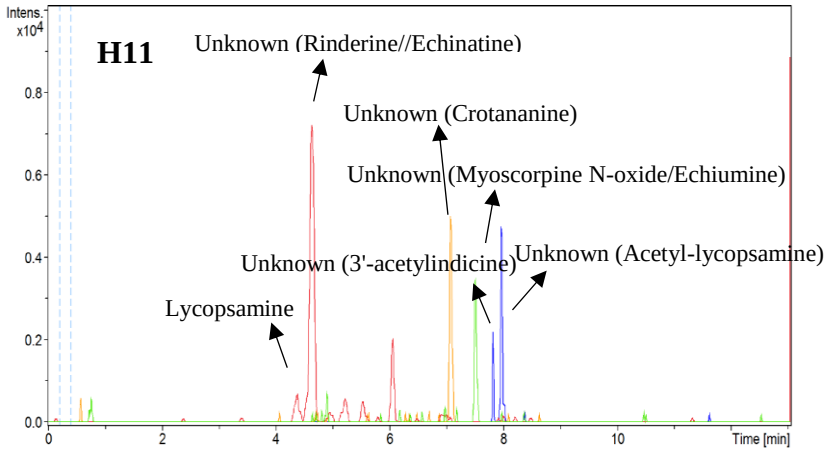

H6
Unknown (Rinderine//Echinatine/Indicine)
Lycopsamine
Intermedine
Retrosine
Intens.
x10^4
Time [min]

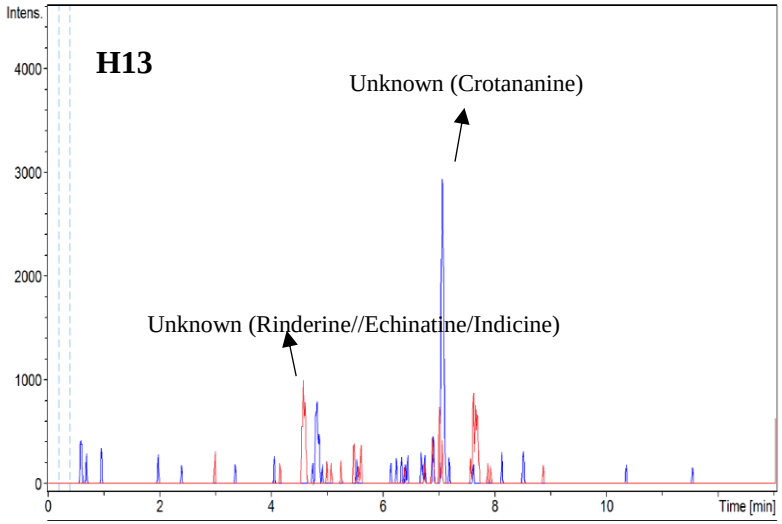

H7
Unknown (Rinderine//Echinatine/Indicine)
Unknown (Integerrimine/ Senecivernine)
Lycopsamine
Intermedine
Senecionine
Intens.
x10^4
Time [min]

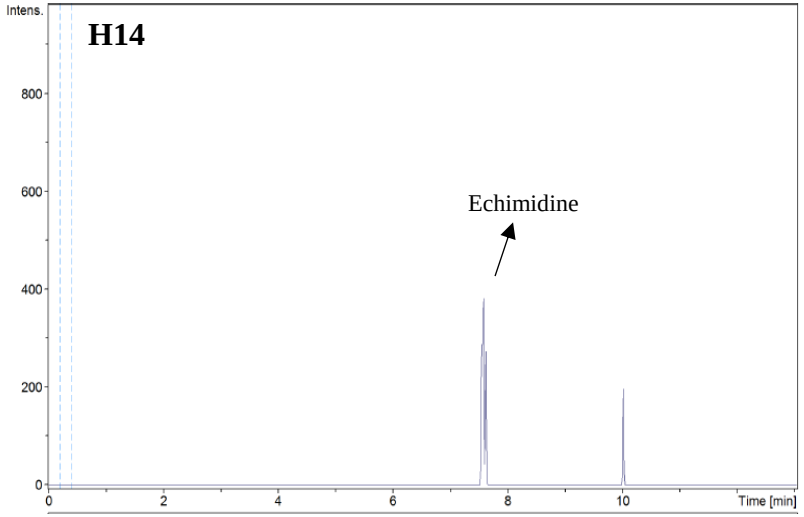

H8
Unknown (Crotananine)
Intens.
x10^4
Time [min]

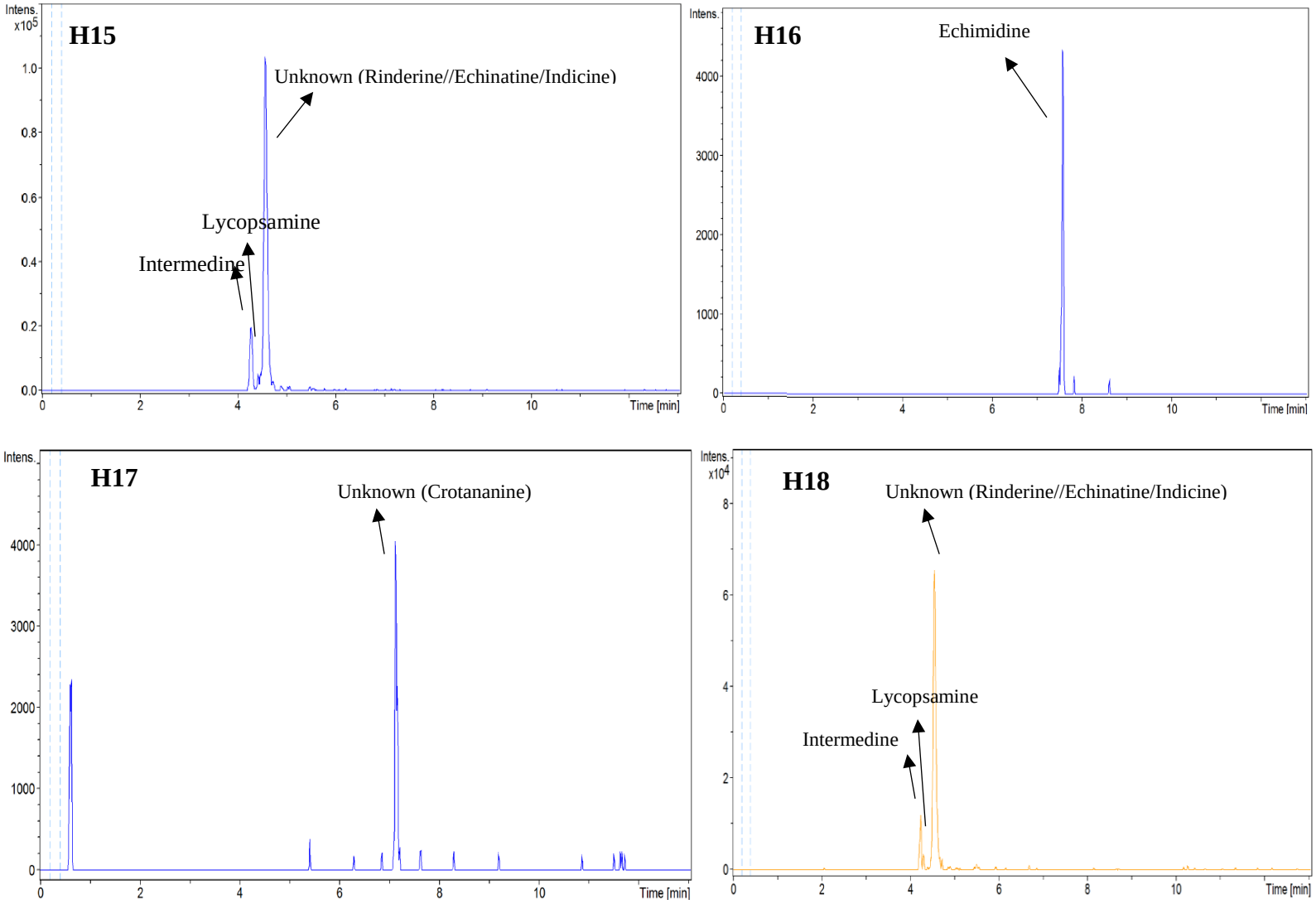

H9
Unknown (3'-acetylindicine)
Unknown (Acetyl-lycopsamine)
Unknown (Crotananine)
Unknown
Intens.
x10^4
Time [min]

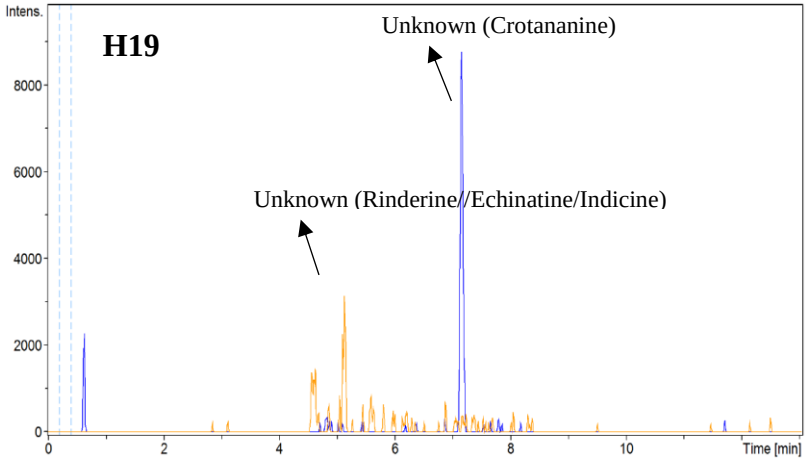

H10
Intens.
x10^4
Unknown (Crotananine)
Time [min]

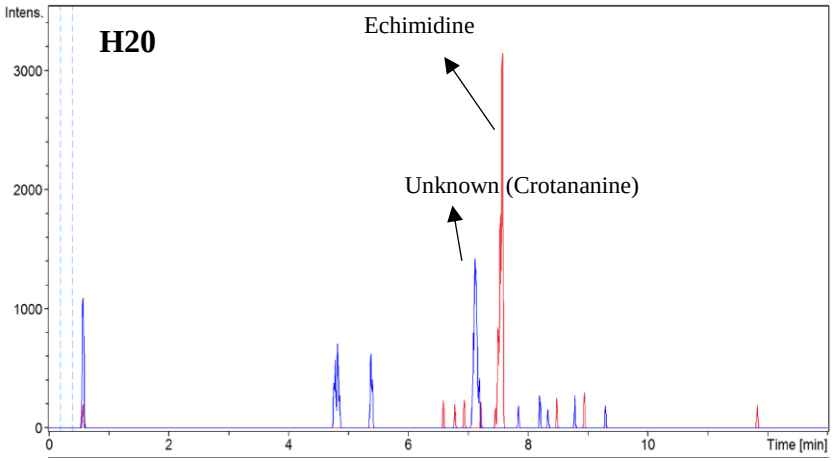

H11
Intens.
x10^4
Unknown (Rinderine//Echinatine)
Unknown (Crotananine)
Unknown (Myoscorpine N-oxide/Echiumine)
Unknown (Acetyl-lycopsamine)
Unknown (3'-acetylindicine)
Lycopsamine
Time [min]

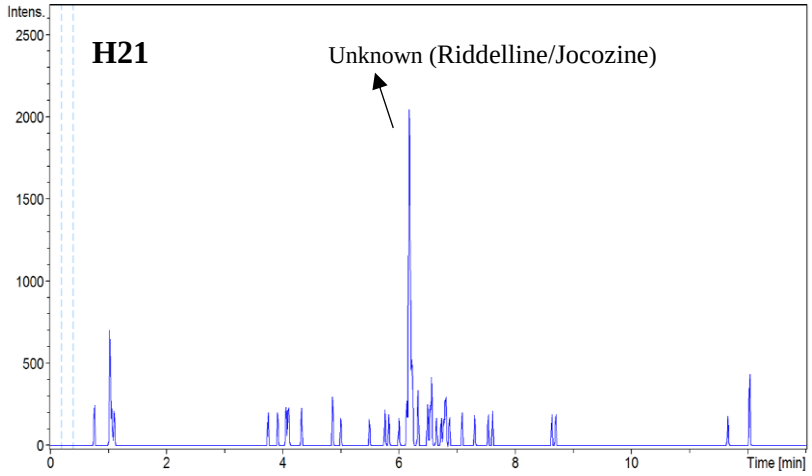

H13
Intens.
Unknown (Crotananine)
Unknown (Rinderine//Echinatine/Indicine)
Time [min]

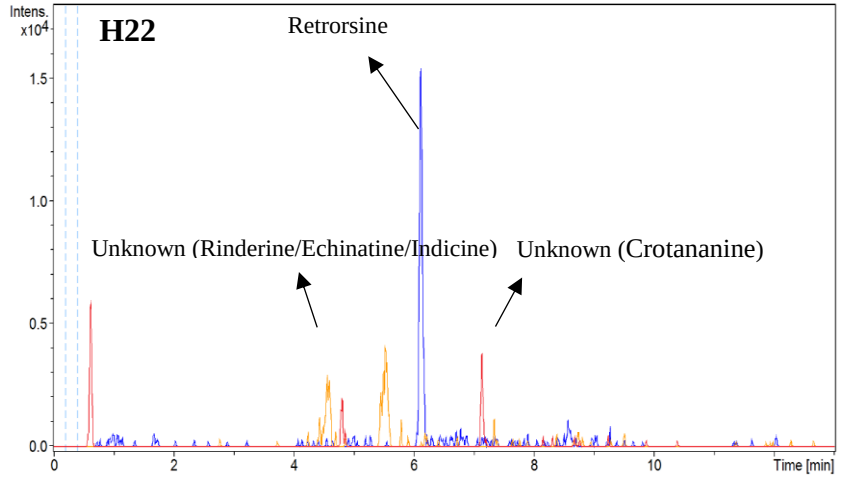

H14
Intens.
Echimidine
Time [min]

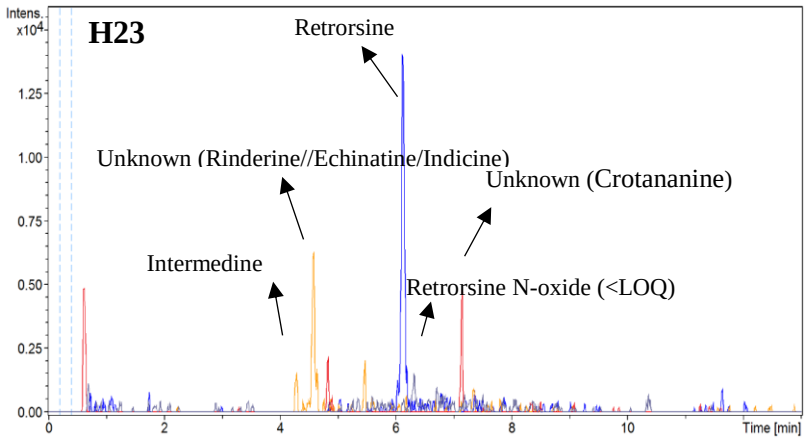

H15
Intens.
x10^5
1.0
0.8
0.6
0.4
0.2
0.0
Unknown (Rinderine//Echinatine/Indicine)
Lycopsamine
Intermedine
0
2
4
6
8
10
Time [min]
H16
Intens.
4000
3000
2000
1000
0
Echimidine
0
2
4
6
8
10
Time [min]
H17
Intens.
4000
3000
2000
1000
0
Unknown (Crotananine)
0
2
4
6
8
10
Time [min]
H18
Intens.
x10^4
8
6
4
2
0
Unknown (Rinderine//Echinatine/Indicine)
Lycopsamine
Intermedine
0
2
4
6
8
10
Time [min]

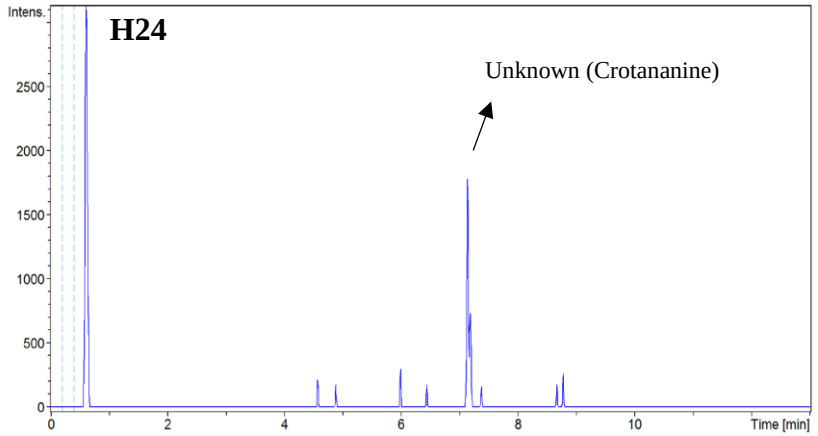

H19
Unknown (Crotananine)
Unknown (Rinderine//Echinatine/Indicine)
Intens.
Time [min]

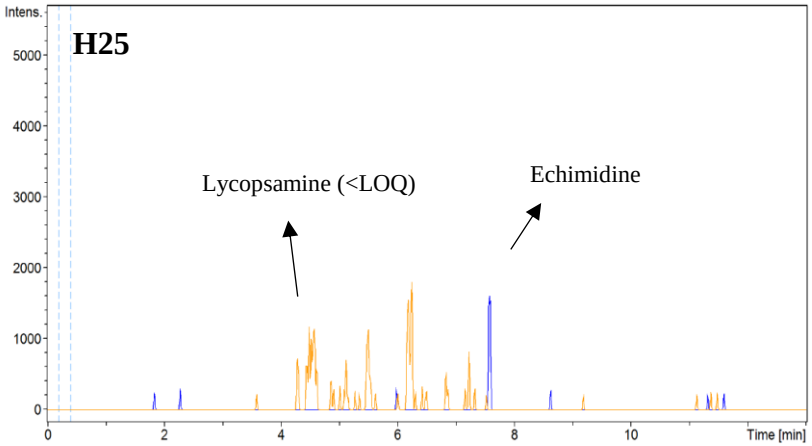

H20
Echimidine
Unknown (Crotananine)
Intens.
Time [min]

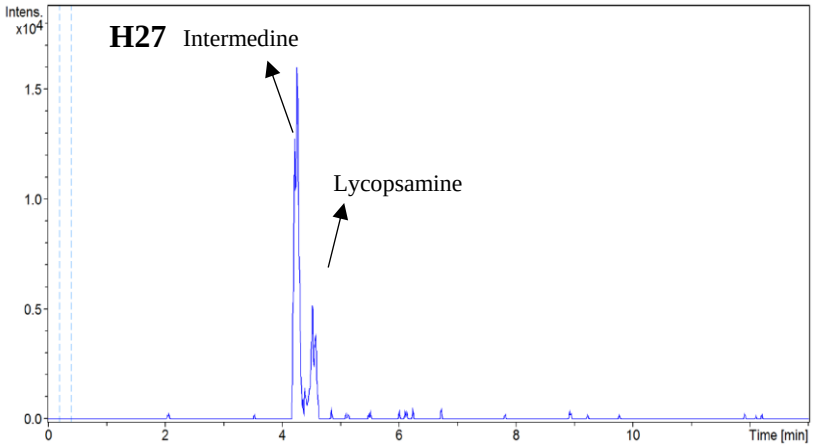

H21
Unknown (Riddelline/Jocozine)
Intens.
Time [min]

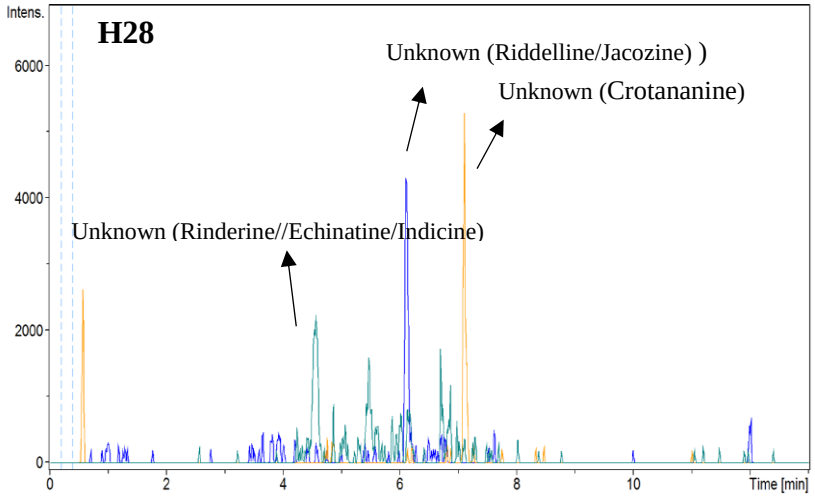

H22
Retrorsine
Unknown (Rinderine/Echinatine/Indicine)
Unknown (Crotananine)
Intens.
Time [min]

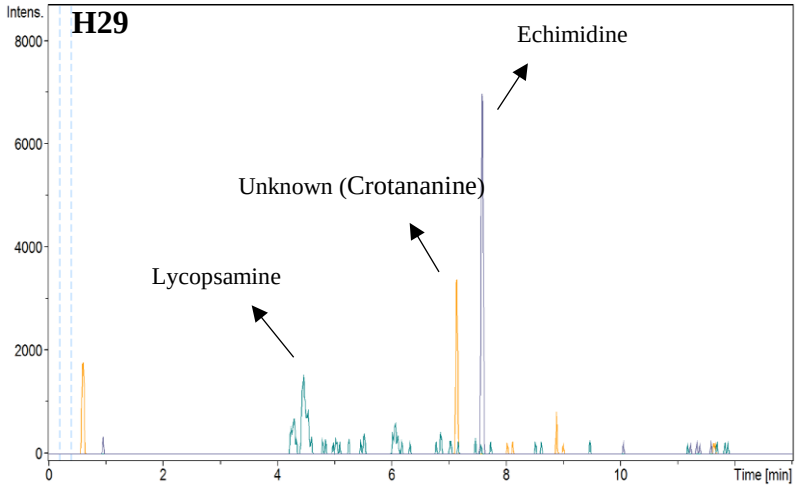

H23
Retrorsine
Unknown (Rinderine//Echinatine/Indicine)
Unknown (Crotananine)
Intermedine
Retrorsine N-oxide (<LOQ)
Intens.
x10^4
Time [min]


H24
Unknown (Crotananine)
Intens.
Time [min]


H25
Lycopsamine (<LOQ)
Echimidine
Intens.
Time [min]


H27
Intermedine
Lycopsamine
Intens.
x10^4
Time [min]


H28
Intens.
6000
4000
2000
0
Unknown (Riddelline/Jacozine))
Unknown (Crotananine)
Unknown (Rinderine//Echinatine/Indicine)
0
2
4
6
8
10
Time [min]


H29
Intens.
8000
6000
4000
2000
0
Echimidine
Unknown (Crotananine)
Lycopsamine
0
2
4
6
8
10
Time [min]

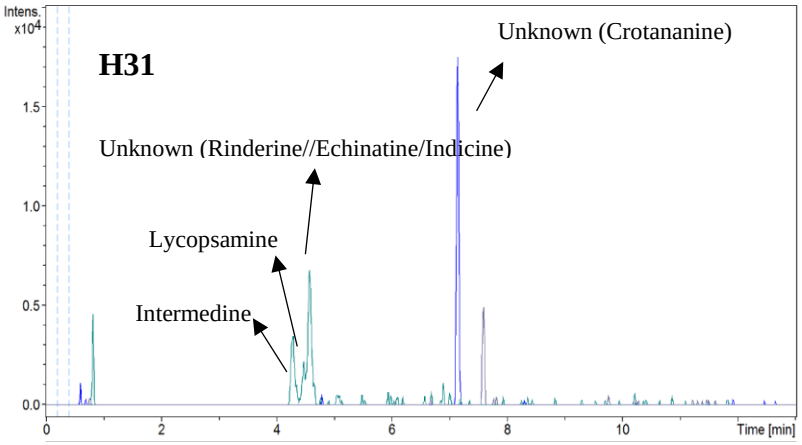

Intens.
x10^4
1.5
1.0
0.5
0.0
H31
Unknown (Crotananine)
Unknown (Rinderine//Echinatine/Indicine)
Lycopsamine
Intermedine
0
2
4
6
8
10
Time [min]

Table S1. Matrix effect of other different honey measured at 50 ng/ml spiking level to give a honey concentration of 25 μg/kg with same procedure as method description

Compounds	H1	H2	H4	H5	H6	H7	H11	H13	RSD (%)
Echimidine	101.09	112.97	122.91	112.03	111.41	107.63	118	127.09	7.3
Intermedine	88.92	98.79	108.97	94.56	104.91	96.04	100.37	123.86	10.54
Intermedine-N-oxide	92.06	96.99	96.43	94.75	95.31	89.39	93.23	102.34	4.04
Lasiocarpine	95.67	83.26	93.09	90.39	93.16	93.19	90.89	102.74	5.87
Lasiocarpine-N-oxide	101.47	97.37	95.75	98.94	100.06	99.9	102.14	113	5.19
Lycopsamine	97.81	90.25	103.92	88.92	97.33	85.9	89.37	103.1	7.27
Retrorsine	93.81	89.75	97.84	88.7	104.08	70.95	95.93	105.75	11.68
Retrorsine-N-oxide	86.7	83.26	91.02	89.57	83.65	80.75	91.1	99.32	6.71
Senecionine	97.62	95.16	92.92	87.62	92.42	81.85	91.38	101.65	6.53
Senecionine-N-oxide	94.02	83.1	89.27	92.19	80.79	82.27	92.61	103.15	8.35
Seneciphylline	90.63	89.42	94.15	93.88	80.74	76.56	80.09	100.51	9.42
Seneciphylline-N-oxide	101.03	85.07	92.91	97.17	96.14	88.22	96.43	106.49	7.13

Botanical origin. H1: Mandarin honey; H2: Creamy flowers honey; H4; not clear; H5: Mountain honey; H6: not Clear; H7; not clear; H11: Summer flowers; and white clover honey; H13: Mountain honey
Note: The matrix effects of these 8 samples were calculated by the following formula.
ME (%) = (response of spiking in 50 ng mL^{-1}/ response of standards mixture in 50 ng mL^{-1} without matrix)*100%].

Table S2. PAs screening list

Formula	Name	M+H
$C_{19}H_{28}NO_7$	18-hydroxysenkirkine	383.1933
$C_{22}H_{33}NO_8$	3'-Acetylechimidine	440.2279
$C_{17}H_{27}NO_6$	3'-acetylindicine	342.1911
$C_{17}H_{27}NO_6$	7-acetylintermedine	342.1911
$C_{13}H_{19}NO_3$	7-O- angeloylheliotridine	238.1438
C13H19NO4	9-Angeloylretronecine N-oxide	254.1386
C13H19NO3	9-Tigloylretronecine	238.1437
$C_{22}H_{33}NO_8$	Acetylechimidine	440.2279
$C_{22}H_{33}NO_9$	Acetylechimidine N-oxide	456.2228
$C_{20}H_{25}NO_8$	Acetylerucifoline N-oxide	408.1652

$C_{17}H_{27}NO_6$	Acetyl-lycopsamine	342.1911
$C_{20}H_{25}NO_6$	Acetylseneciphylline	376.1755
$C_{20}H_{25}NO_7$	Acetylseneciphylline N-oxide	392.1704
C20H25NO7	Acetylerucifoline	392.1703
$C_{15}H_{25}NO_4$	Amabiline	284.1856
$C_{15}H_{25}NO_5$	Amabiline N-oxide	300.1805
$C_{21}H_{27}NO_7$	Clivorine	406.186
$C_{16}H_{23}NO_6$	Crotaline	326.1598
$C_{17}H_{25}NO_5$	Crotananine	324.1805
$C_{15}H_{27}NO_4$	Cynaustine	286.2013
$C_{18}H_{24}ClNO_6$	Dehydrojaconine	386.1365
$C_{19}H_{28}ClNO_7$	Desacetyldoronine	418.1627
$C_{21}H_{30}ClNO_8$	Doronine	460.1733
$C_{20}H_{31}NO_8$	Echimidine-N-oxide	414.2122
$C_{15}H_{25}NO_5$	Echinatine	300.1805
$C_{20}H_{31}NO_6$	Echiumine	382.2224
$C_{20}H_{31}NO_7$	Echiumine N-oxide	398.2173
$C_{20}H_{31}NO_7$	Echiupinine-N-oxide	398.2173
$C_{20}H_{31}NO_6$	Echiuplatine	382.2224
$C_{20}H_{31}NO_7$	Echiuplatine-N-oxide	398.2173
$C_{25}H_{37}NO_8$	Echivulgarine	480.2592
$C_{25}H_{37}NO_9$	Echivulgarine N-Oxide	496.2541
$C_{18}H_{23}NO_7$	Erucifoline N-oxide	366.1547
$C_{16}H_{27}NO_6$	Europine	330.1911
$C_{16}H_{27}NO_7$	Europine N-oxide	346.186
$C_{19}H_{27}NO_7$	Fukinotoxin	382.186
$C_{16}H_{23}NO_5$	Fulvine	310.1649
$C_{16}H_{27}NO_4$	Heleurine	298.2013
$C_{16}H_{27}NO_5$	Heleurine N-oxide	314.1962
$C_{20}H_{31}NO_7$	Heliosupine	398.2173
$C_8H_{13}NO_2$	Heliotridine	156.1019
$C_{16}H_{27}NO_5$	Heliotrine	314.1962
$C_{16}H_{27}NO_6$	Heliotrine N-oxide	330.1911
$C_{15}H_{25}NO_5$	Indicine	300.1805
$C_{15}H_{25}NO_6$	Indicine N-oxide	316.1755
$C_{18}H_{25}NO_5$	Integerrimine	336.1805
$C_{18}H_{25}NO_6$	Integerrimine N-oxide	352.1755
$C_{18}H_{25}NO_6$	Jacobine	352.1755
$C_{18}H_{25}NO_7$	Jacobine N-oxide	368.1704
$C_{18}H_{27}NO_7$	Jacoline	370.186
$C_{18}H_{27}NO_8$	Jacoline N-oxide	386.1809
$C_{18}H_{26}ClNO_6$	Jaconine	388.1521

$C_{18}H_{26}ClNO_7$	Jaconine N-oxide	404.147
$C_{18}H_{23}NO_6$	Jacozine	350.1598
$C_{18}H_{23}NO_7$	Jacozine-N-oxide	368.1704
$C_{19}H_{29}NO_7$	Junceine	384.2017
$C_{20}H_{27}NO_7$	Latifoline	394.186
$C_{21}H_{29}NO_7$	Ligularidine	408.2017
$C_{15}H_{25}NO_6$	Lycopsamine-N-oxide	316.1755
$C_{16}H_{23}NO_7$	Monocrotaline N-oxide	342.1547
C20H31NO6	Myoscorpine	382.2224
$C_{20}H_{31}NO_7$	Myoscorpine N-oxide	398.2173
$C_{22}H_{33}NO_8$	Parsonsine	440.2279
C22H33NO9	Parsonsine 4-oxide	456.2228
$C_{18}H_{27}NO_5$	Platyphylline	338.1962
$C_8H_{13}NO_2$	Retronecine	156.1019
$C_{19}H_{26}NO_7$	Retusamine	381.1776
$C_{18}H_{23}NO_6$	Riddelline	350.1598
$C_{18}H_{23}NO_7$	Riddelline N-oxide	366.1547
$C_{15}H_{25}NO_5$	Rinderine	300.1805
$C_{18}H_{25}NO_5$	Senecivernine	336.1805
$C_{18}H_{25}NO_6$	Senecivernine N-oxide	352.1755
$C_{19}H_{27}NO_6$	Senkirkine	366.1911
$C_{18}H_{23}NO_5$	Spartioidine	334.1649
$C_{18}H_{25}NO_7$	Spectabiline	368.1704
$C_{15}H_{25}NO_4$	Supinine	284.1856
$C_{20}H_{31}NO_6$	Symphytine	382.2224
$C_{21}H_{31}NO_8$	Triacetylindicine	426.2122
$C_{18}H_{27}NO_6$	Trichodesmine	354.1911
$C_{18}H_{27}NO_7$	Trichodesmine-N-oxide	370.186
$C_{17}H_{27}NO_7$	Uplandicine	358.186
$C_{17}H_{27}NO_8$	Uplandicine N-oxide	374.1809
$C_{20}H_{31}NO_7$	Vulgarine	398.2173
$C_{20}H_{31}NO_8$	Vulgarine- N-oxide	414.2122
$C_{20}H_{25}NO_7$	(Z)-Acetylerucifoline	392.1704

Table S3-a. Molecular properties of related PAs.

Molecule	Exact mass	Formula	pK_b	pK_a	RT	logP	MolecularVolume
Echimidine	397.2101	$C_{20}H_{31}NO_7$	8.13	10.55	7.5	0.68	374.47
Intermedine	299.1733	$C_{15}H_{25}NO_5$	7.82	11.34	4.3	-0.29	285.69
Intermedine N-oxide	315.1682	$C_{15}H_{25}NO_6$	2.83	11.33	5.1	-1.42	294.97
Lasiocarpine	411.2257	$C_{21}H_{33}NO_7$	8.13	10.54	8.2	1.33	392.36
Lasiocarpine N-oxide	427.2206	$C_{21}H_{33}NO_8$	2.82	10.54	8.5	0.20	402.02

Lycopsamine	299.1733	$C_{15}H_{25}NO_5$	7.82	11.34	4.5	-0.29	285.69
Retrorsine	351.1682	$C_{18}H_{25}NO_6$	7.14	11.44	6.1	0.60	320.40
Retrorsine-N-oxide	367.1631	$C_{18}H_{25}NO_7$	2.82	11.44	6.3	-0.53	329.94
Senecionine	335.1733	$C1_8H_{25}NO_5$	7.14	12.37	7.2	1.64	312.03
Senecionine N-oxide	351.1682	$C_{18}H_{25}NO_6$	2.82	12.37	7.4	0.52	321.19
Seneciphylline	333.1576	$C_{18}H_{23}NO_5$	7.14	12.04	6.6	1.37	303.84
Seneciphylline N-oxide	349.1525	$C_{18}H_{23}NO_6$	2.82	12.04	6.8	0.25	313.05
7-acetylintermedine	341.1838	$C_{20}H_{31}NO_7$	7.22	11.34	6.6	0.15	322.19
3'-acetylindicine	341.1838	$C_{17}H_{27}NO_6$	7.82	11.31	7.8	0.15	322.32
Crotananine	323.1733	$C_{17}H_{25}NO_5$	7.08	12.47	7.1	0.97	302.71
Heliotrine	313.1889	$C_{16}H_{27}NO_5$	7.82	11.33	6.4	0.35	303.54
Heliotrine N-oxide	329.1838	$C_{16}H_{27}NO_6$	2.83	11.33	6.6	-0.77	312.67
Jacozine	350.1598	$C_{18}H_{23}NO_6$	6.92	12.04	6.1	0.40	309.56
Lycopsamine N-oxide	315.1682	$C_{15}H_{25}NO_6$	2.83	11.33	5.3	-1.42	294.97
Myoscorpine N-oxide	397.2101	$C_{20}H_{31}NO_7$	2.82	11.34	7.5	0.80	374.94
Echiumine N-oxide	397.2101	$C_{20}H_{31}NO_7$	2.82	11.34	7.5	0.80	374.94
Echiupinine-N-oxide	397.2101	$C_{20}H_{31}NO_7$	2.82	11.34	7.5	0.65	374.94
Indicine	299.1733	$C_{15}H_{25}NO_5$	7.82	11.34	4.6	-0.29	285.69
Rinderine	299.1733	$C_{15}H_{25}NO_5$	7.82	11.34	4.6	-0.29	285.69
Riddelline	350.1598	$C_{18}H_{23}NO_6$	7.14	11.11	6.1	0.33	312.29
Echinatine	299.1733	$C_{15}H_{25}NO_5$	7.82	11.34	4.6	-0.29	285.69
Integerrimine	335.1733	$C_{18}H_{25}NO_5$	7.14	12.37	7.1	1.64	312.03
Trichodesmine	353.1838	$C_{18}H_{27}NO_6$	8.08	11.49	7.8	0.40	329.00
Senecivernine	335.1733	$C_{18}H_{25}NO_5$	7.14	12.37	7.1	1.64	312.03

Table S3-b. Prediction error of relative slope (RS)

Molecule	RSM	RSP	Prediction error	RSM	RSP	Prediction error	Prediction error difference
		a			b		Use b
Echimidine	0.45	0.49	8.6%	0.38	0.42	9.9%	-14.1%
Intermedine	0.97	0.83	-14.1%	1.04	0.92	-11.7%	10.9%
Intermedine N-oxide	0.62	0.77	24.6%	0.68	0.85	24.8%	10.4%
Lasiocarpine	1.10	1.00	-9.8%	0.80	0.75	-6.1%	-24.5%
Lasiocarpine N-oxide	0.72	0.73	0.3%	0.58	0.56	-2.7%	-23.0%
Lycopsamine	0.95	0.83	-12.0%	1.06	0.92	-13.2%	10.9%
Retrorsine	0.48	0.58	20.4%	0.48	0.56	15.9%	-4.0%
Retrorsine-N-oxide	0.48	0.42	-11.5%	0.47	0.41	-12.7%	-2.1%
Senecionine	0.59	0.73	22.7%	0.54	0.69	27.1%	-5.6%
Senecionine N-oxide	0.72	0.54	-25.7%	0.68	0.52	-23.4%	-3.8%
Seneciphylline	1.00	-	-	1.00	-	-	-
Seneciphylline N-oxide	0.58	0.69	18.8%	0.55	0.66	20.5%	-4.5%
7- acetylintermedine	0.61	0.52	-14.7%	–	–	–	1.5%
Heliotrine	0.79	1.19	50.8%	–	–	–	0.7%
Heliotrine N-oxide	1.01	1.11	10.2%	–	–	–	0.3%
Lycopsamine N-oxide	0.60	0.77	29.7%	–	–	–	10.4%
3’-acetylindicine	–	–	–	–	0.46	–	
Echiupinine-N-oxide					0.67		
Crotananine	–	–	–	–	0.43	–	
Myoscorpine N-oxide	–	–	–	–	0.78	–	
Rinderine/Echinatine/indicine	–	–	–	–	0.93	–	
Integerrimine	–	–	–	–	0.69	–	
Riddelline/Jacozine					0.71/0.22		
Senecivernine	–	–	–	–	0.63	–	
Trichodesmine	–	–	–	–	0.30	–	
Formula	= (28.8671 ± 7.7165) + (1.1691± 0.3107) × logP - (1.5495±0.4134) × pK_a - (0.3378±0.1015) × pK_b - (0.0281±0.0.0074) × exact.mass			= (27.096527 ± 7.6793) + (1.024035 ± 0.3092) × logP - (1.407454±0.4115) × pK_a - (0.304834±0.1010) × pK_b - (0.02853±0.0.0073) × exact.mass			

Note: New standards curves for 12 PAs were remade in every sequence, “a” and “b” means the results obtained from difference sequence on different days. RSM, relative measured slope corresponding to Seneciphylline; RSP, relative predicted slope corresponding to Seneciphylline; Prediction error was calculated separately

Appendix III

III. Manuscript 3. supplementary information (198)

Supporting information

Standard substances free quantification makes LC/ESI/MS non-targeted screening of pesticides in cereals comparable between labs

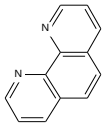
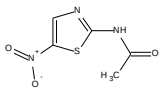
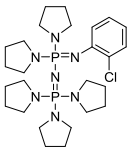
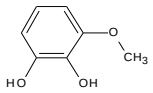
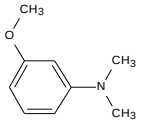
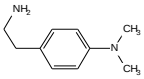
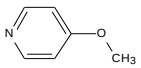
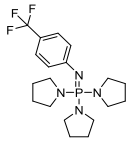
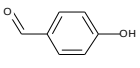
Tingting Wang,[†,1] Jaanus Liigand,[‡,1] Henrik Lauritz Frandsen, [†] Jørn Smedsgaard,[†] Anneli Kruve[‡,$,*]

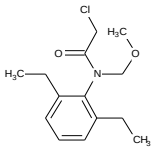
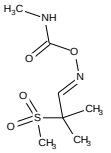
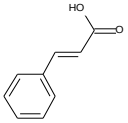
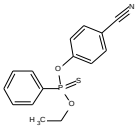
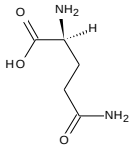
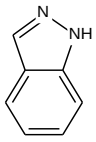
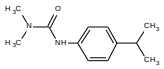
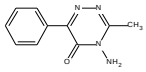
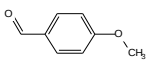
† National Food Institute, Research Group for Analytical Food Chemistry, Technical University of Denmark, Kemitor-vet Building 202, Kgs. Lyngby, DK-2800, Denmark
‡ University of Tartu, Institute of Chemistry, Ravila 14a Tartu, 50411, Estonia
$ Department of Environmental Science and Analytical Chemistry, Stockholm University, Svante Arrhenius väg 16, 106 91 Stockholm

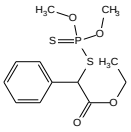
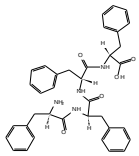
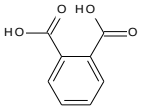
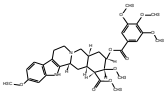
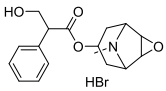
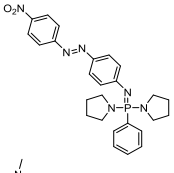
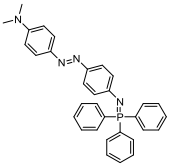
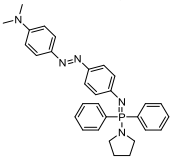
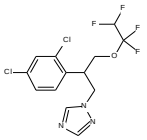
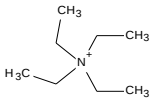
Supporting Information

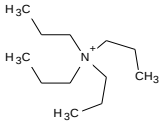
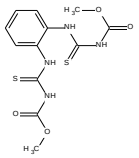
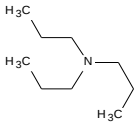
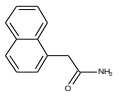
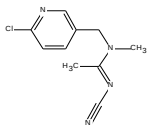
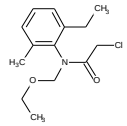
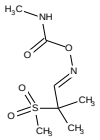
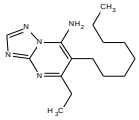
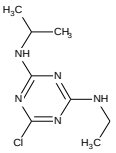
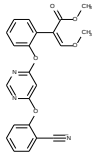
Table of Contents:

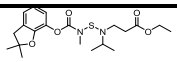
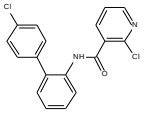
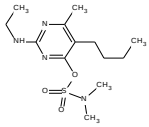
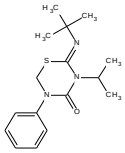
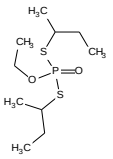
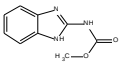
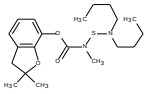
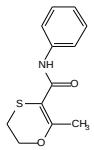
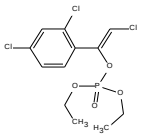
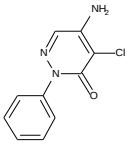
Table S1: Information of all compounds used in calibrants set and validation set, C denotes the used molar concentration range for each compound.

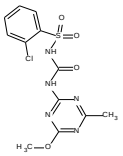
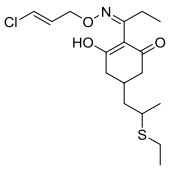
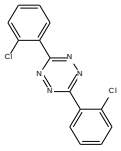
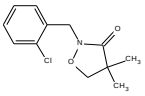
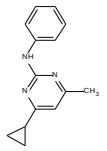
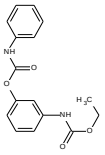
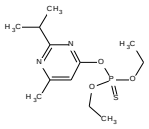
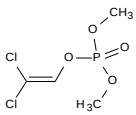
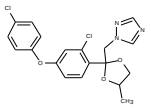
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Transformation compounds									
1,10-phenanthroline		$C_{12}H_8N_2$	114.7	181.1	1.6	1.5E-07~3.0E-05	2.88	15.16	15.16
2-acetamido-5-nitrothiazole		$C_5H_5N_3O_3S$	112.9	188.0	2.8	6.0E-07~1.2E-04	0.39	13.18	13.85
(2-Cl-C6H4-P2(pyrr)		$C_{26}H_{44}ClN_7P_2$	180.7	552.3	7.0	2.5E-10~5.0E-08	6.02	17.67	16.59
3-methoxycatechol		$C_7H_8O_3$	108.6	141.1	2.0	5.5E-06~1.1E-03	0.47	13.24	12.97
3-methoxy-N,N-dimethylaniline		$C_9H_{13}NO$	110.9	152.1	1.8	5.6E-08~1.1E-05	3.40	15.58	15.37
4-dimethylamino-N,N-dimethylaniline		$C_{10}H_{16}N_2$	112.4	165.1	1.6	3.4E-08~6.8E-06	2.68	15.01	14.93
4-methoxypyridine		C_6H_7NO	107.4	110.1	0.6	9.2E-09~1.8E-06	3.07	15.32	17.06
4-CF3-C6H4-P(pyrr)		$C_{19}H_{28}F_3N_4P$	125.0	401.2	4.5	5.4E-10~1.1E-07	4.22	16.24	16.41
4-OH benzaldehyde		C_7H_6O2	108.3	123.0	2.2	4.1E-07~8.2E-05	0.90	13.59	13.55

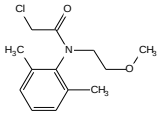
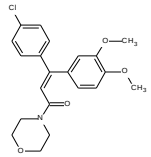
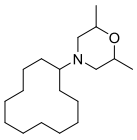
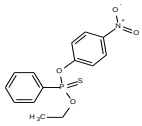
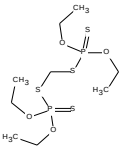
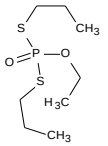
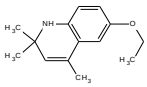
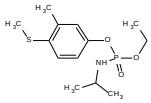
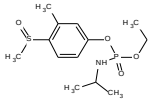
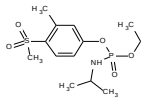
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
alachlor		$C_{14}H_{20}ClNO_2$	155.6	270.1	5.4	2.3E-08~4.5E-06	2.05	14.50	14.65
aldicarb-sulfone		$C_7H_{14}N_2O_4S$	115.5	223.1	1.8	1.1E-07~2.3E-05	1.62	14.16	14.36
cinnamic acid		$C_9H_8O_2$	110.8	149.1	2.0	3.2E-07~6.3E-05	0.69	13.42	14.10
cyanophenphos		$C_{15}H_{14}NO_2PS$	124.8	304.1	6.0	2.1E-07~4.2E-05	2.07	14.52	14.34
glutamine		$C_5H_{10}N_2O_3$	107.1	147.1	0.5	5.4E-07~1.1E-04	1.43	14.01	13.92
indazole		$C_7H_6N_2$	108.7	119.1	2.9	1.7E-07~3.5E-05	2.33	14.72	14.89
isoproturon		$C_{12}H_{18}N_2O$	115.1	207.1	4.1	1.2E-08~2.3E-06	3.35	15.54	16.14
metamitrone		$C_{10}H_{10}N_4O$	113.3	203.1	2.5	1.4E-08~2.8E-06	2.65	14.98	15.47
p-anisaldehyde		C_8H_8O2	109.6	137.1	3.4	7.1E-06~1.4E-03	1.11	13.75	12.60

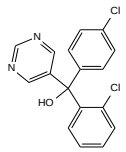
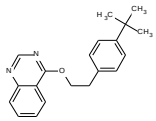
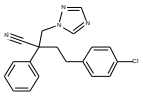
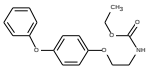
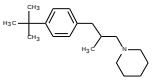
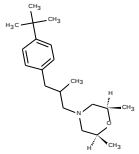
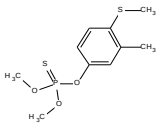
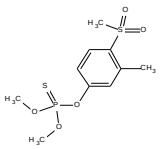
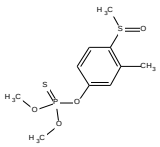
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
phenthoate		$C_{12}H_{17}O_4PS_2$	127.2	321.0	5.8	5.9E-08~1.2E-05	1.95	14.42	13.74
Phe-Phe-Phe-Phe		$C_{36}H_{38}N_4O_5$	152.1	607.3	3.6	1.1E-08~2.2E-06	3.82	15.91	14.96
phthalic acid		$C_8H_6O_4$	110.1	167.0	1.9	5.8E-07~1.2E-04	1.06	13.71	12.93
reserpine		$C_{33}H_{40}N_2O_9$	147.5	609.3	3.9	2.2E-09~4.5E-07	3.98	16.05	15.15
scopolamine HBr		$C_{17}H_{21}NO_4$	122.0	304.2	1.7	6.4E-09~1.3E-06	3.30	15.50	15.57
Sigrid VII		$C_{26}H_{29}N_6O_2P$	136.2	489.2	5.4	2.6E-10~5.2E-08	4.93	16.80	16.50
Sigrid X		$C_{32}H_{29}N_4P$	141.2	501.2	5.3	4.4E-10~8.8E-08	5.33	17.12	17.54
Sigrid XII		$C_{30}H_{32}N_5P$	138.7	494.2	5.4	1.1E-09~2.3E-07	5.04	16.89	17.40
tetraconazole		$C_{13}H_{11}Cl_2F_4N_3O$	204.3	372.0	5.1	8.2E-09~1.6E-06	3.32	15.52	15.42
tetraethylammonium		$C_8H_{20}N$	109.6	130.1	0.6	9.6E-09~1.9E-06	3.61	15.75	15.62

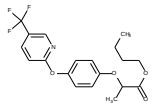
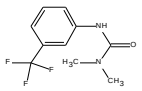
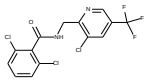
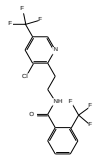
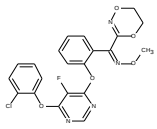
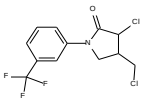
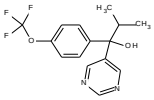
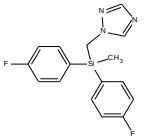
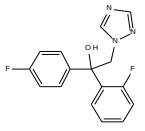
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
tetrapropylammonium		$C_{12}H_{27}N$	114.6	186.2	1.9	3.6E-08~7.2E-06	3.65	15.78	15.97
thiophanate-methyl		$C_{12}H_{14}N_4O_4S_2$	129.0	343.0	3.7	2.7E-07~5.4E-05	2.68	15.01	14.74
tripropylamine		$C_9H_{21}N$	110.8	144.2	1.4	8.1E-08~1.6E-05	3.55	15.70	15.65
Analyzed pesticides									
1-Naphtylacetamide		$C_{12}H_{11}NO$	114.6	186.1	3.1	2.7E-08~2.7E-05 (1.1E-08~1.1E-05)	3.06	15.36	14.93
Acetamiprid		$C_{10}H_{11}ClN_4$	149.1	223.1	2.8	2.3E-08~2.3E-05 (9.0E-09~9.0E-06)	2.5	14.89	15.17
Acetochlor		$C_{14}H_{20}ClNO_2$	154.6	270.1	5.3	1.9E-08~1.9E-05 (7.4E-09~7.4E-06)	2.82	15.16	14.49
Aldicarb-sulfone		$C_7H_{14}N_2O_4S$	115.6	223.1	1.9	2.3E-08~2.3E-05 (9.0E-09~9.0E-06)	1.31	13.90	14.64
Ametoctradin		$C_{15}H_{25}N_5$	120.0	276.2	5.4	1.8E-08~1.8E-05 (7.2E-09~7.2E-06)	3.49	15.72	15.97
Atrazine		$C_8H_{14}ClN_5$	146.6	216.1	4.0	2.3E-08~2.3E-05 (9.3E-09~9.3E-06)	3.52	15.74	15.83
Azoxystrobin		$C_{22}H_{17}N_3O_5$	129.9	404.1	4.9	1.2E-08~1.2E-05 (4.9E-09~4.9E-06)	3.17	15.45	15.72

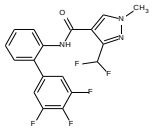
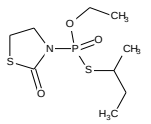
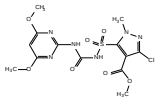
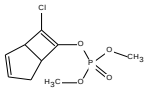
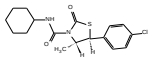
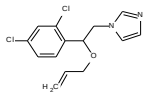
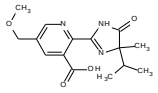
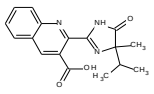
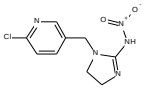
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Benfuracarb		$C_{20}H_{30}N_2O_5S$	133.3	411.2	6.4	1.2E-08~1.2E-05 (4.9E-09~4.9E-06)	3.09	15.38	15.75
Boscalid		$C_{18}H_{12}Cl_2N_2O$	211.4	343.0	5.0	1.5E-08~1.5E-05 (5.8E-09~5.8E-06)	3.38	15.63	14.77
Bupirimate		$C_{13}H_{24}N_4O_3S$	124.0	317.2	4.5	1.6E-08~1.6E-05 6.3E-09~6.3E-06	4.11	16.23	16.27
Buprofezine		C16H23N3OS	127.0	306.2	5.6	1.6E-08~1.6E-05 (6.5E-09~6.5E-06)	3.53	15.75	16.09
Cadusafos		$C_{10}H_{23}O_2PS_2$	124.0	271.1	5.8	1.9E-08~1.9E-05 (7.4E-09~7.4E-06)	3.16	15.44	15.38
Carbendazim		$C_9H_9N_3O_2$	112.0	192.1	1.5	2.6E-08~2.6E-05 (1.0E-08~1.0E-05)	2.87	15.20	15.62
Carbosulfan		$C_{20}H_{32}N_2O_3S$	132.8	381.2	7.6	1.3E-08~1.3E-05 (5.2E-09~5.2E-06)	3.92	16.07	15.46
Carboxin		$C_{12}H_{13}NO_2S$	121.0	236.1	4.1	2.1E-08~2.1E-05 (8.5E-09~8.5E-06)	2.9	15.23	15.26
Chlorfenvinphos		$C_{12}H_{14}Cl_3O_4P$	260.3	359.0	5.5	1.4E-08~1.4E-05 (5.6E-09~5.6E-06)	2.46	14.86	15.02
Chloridazon		$C_{10}H_8ClN_3O$	149.1	222.0	2.6	2.3E-08~2.3E-05 (9.0E-09~9.0E-06)	2.46	14.86	15.26

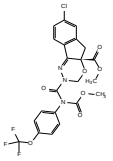
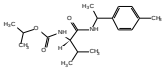
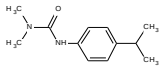
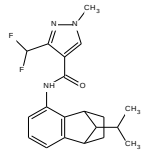
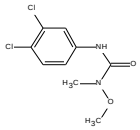
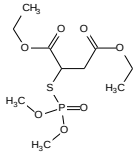
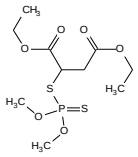
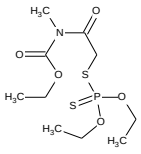
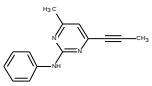
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Chlorsulfuron		$C_{12}H_{12}ClN_5O_4S$	162.5	358.0	3.9	1.4E-08~1.4E-05 (5.6E-09~5.6E-06)	2.46	14.86	14.76
Clethodim		$C_{17}H_{26}ClNO_3S$	169.4	360.1	6.3	1.4E-08~1.4E-05 (5.6E-09~5.6E-06)	3.22	15.49	15.18
Clofentezine		$C_{14}H_8Cl_2N_4$	205.6	303.0	5.9	1.7E-08~1.7E-05 (6.6E-09~6.6E-06)	2.71	15.07	14.32
Clomazone		$C_{12}H_{14}ClNO_2$	151.7	240.1	4.5	2.1E-08~2.1E-05 (8.3E-09~8.3E-06)	3.13	15.42	15.37
Cyprodinil		$C_{14}H_{15}N_3$	117.7	226.1	4.5	2.2E-08~2.2E-05 (8.8E-09~8.8E-06)	3.55	15.77	16.27
Desmedipham		$C_{16}H_{16}N_2O_4$	121.0	301.1	4.7	1.7E-08~1.7E-05 (6.6E-09~6.6E-06)	2.94	15.26	14.57
Diazinon		$C_{12}H_{21}N_2O_3PS$	121.7	305.1	5.8	1.6E-08~1.6E-05 (6.6E-09~6.6E-06)	3.14	15.43	16.22
Dichlorvos		$C_4H_7Cl_2O_4P$	183.2	221.0	3.5	2.3E-08~2.3E-05 (9.1E-09~9.1E-06)	1.58	14.13	14.38
Difenoconazole		$C_{19}H_{17}Cl_2N_3O_3$	217.8	406.1	5.7	1.2E-08~1.2E-05 (4.9E-09~4.9E-06)	3.92	16.07	15.63

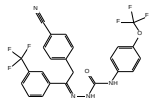
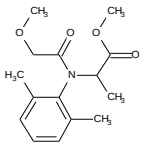
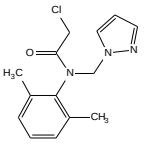
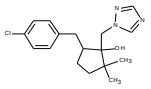
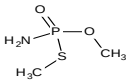
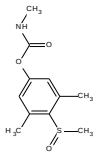
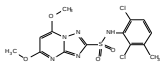
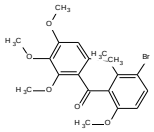
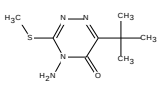
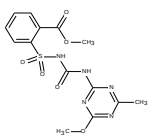
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Dimethachlor		$C_{13}H_{18}ClNO_2$	153.4	256.1	4.4	2.0E-08~2.0E-05 (7.8E-09~7.8E-06)	2.89	15.22	15.39
Dimethomorph		$C_{21}H_{22}ClNO_4$	168.0	388.1	4.6	1.3E-08~1.3E-05 (5.2E-09~5.2E-06)	3.6	15.81	15.03
Dodemorph		$C_{18}H_{35}NO$	122.6	282.3	3.8	1.8E-08~1.8E-05 (7.1E-09~7.1E-06)	4.48	16.54	16.06
EPN		$C_{14}H_{14}NO_4PS$	124.1	324.0	6.1	1.5E-08~1.5E-05 (6.2E-09~6.2E-06)	2.8	15.14	14.04
Ethion		$C_9H_{22}O_4P_2S_4$	136.9	385.0	6.6	1.3E-08~1.3E-05 (5.2E-09~5.2E-06)	3.78	15.96	15.04
Ethoprophos		$C_8H_{19}O_2PS_2$	121.6	243.1	5.0	2.1E-08~2.1E-05 (8.2E-09~8.2E-06)	2.82	15.16	15.54
Ethoxyquin		$C_{14}H_{19}NO$	117.2	218.2	3.7	2.3E-08~2.3E-05 (9.2E-09~9.2E-06)	3.66	15.86	15.88
Fenamiphos		$C_{13}H_{22}NO_3PS$	122.6	304.1	4.8	1.6E-08~1.6E-05 (6.6E-09~6.6E-06)	3.15	15.43	15.41
Fenamiphos-sulfoxide		$C_{13}H_{22}NO_4PS$	122.9	320.1	3.1	1.6E-08~1.6E-05 (6.2E-09~6.2E-06)	3.14	15.43	15.30
Fenamiphos-sulphone		$C_{13}H_{22}NO_5PS$	123.2	336.1	5.9	1.5E-08~1.5E-05 (6.0E-09~6.0E-06)	2.99	15.30	14.72

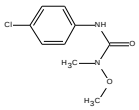
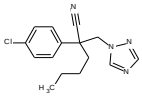
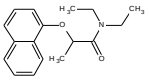
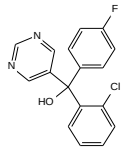
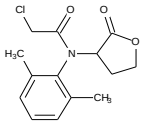
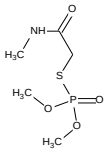
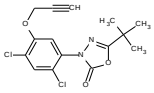
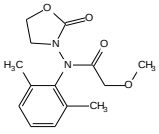
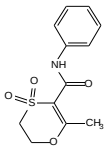
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Fenarimol		$C_{17}H_{12}Cl_2N_2O$	211.3	331.0	5.0	1.5E-08~1.5E-05 (6.0E-09~6.0E-06)	3.24	15.51	15.03
Fenazaquin		$C_{20}H_{22}N_2O$	125.5	307.2	7.1	1.6E-08~1.6E-05 (6.5E-09~6.5E-06)	3.92	16.07	16.29
Fenbuconazole		$C_{19}H_{17}ClN_4$	164.6	337.1	5.2	1.5E-08~1.5E-05 (5.9E-09~5.9E-06)	3.6	15.81	15.29
Fenoxycarb		$C_{17}H_{19}NO_4$	121.9	302.1	5.3	1.7E-08~1.7E-05 (6.6E-09~6.6E-06)	3.01	15.32	14.87
Fenpropidin		$C_{19}H_{31}N$	123.6	274.3	4.0	1.8E-08~1.8E-05 (7.3E-09~7.3E-06)	4.28	16.37	16.44
Fenpropimorph		$C_{20}H_{33}NO$	125.2	304.3	4.0	1.6E-08~1.6E-05 (6.6E-09~6.6E-06)	4.32	16.41	16.06
Fenthion		$C_{10}H_{15}O_3PS_2$	124.5	279.0	2.5	1.8E-08~1.8E-05 (7.2E-09~7.2E-06)	2.87	15.20	15.41
Fenthion Sulfone		$C_{10}H_{15}O_5PS_2$	125.1	311.0	4.4	1.6E-08~1.6E-05 (6.4E-09~6.4E-06)	2.48	14.88	14.06
Fenthion -sulfoxid		$C_{10}H_{15}O_4PS_2$	124.8	295.0	3.9	1.7E-08~1.7E-05 (6.8E-09~6.8E-06)	2.72	15.08	15.20

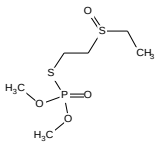
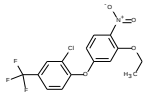
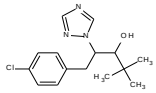
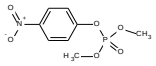
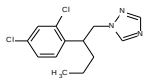
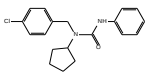
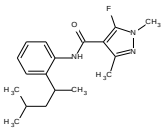
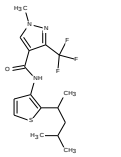
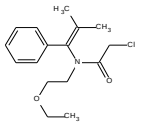
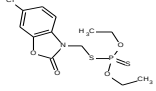
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Fluazifop-p-buthyl		$C_{19}H_{20}F_3NO4$	124.6	384.1	6.5	1.3E-08~1.3E-05 (5.2E-09~5.2E-06)	2.93	15.25	15.84
Fluometuron		$C_{10}H_{11}F_3N_2O$	112.5	233.1	3.9	2.2E-08~2.2E-05 (8.6E-09~8.6E-06)	2.02	14.49	15.60
Fluopicolide		$C_{14}H_8Cl_3F_3N_2O$	269.6	383.0	5.1	1.3E-08~1.3E-05 (5.2E-09~5.2E-06)	2.22	14.66	15.03
Fluopyram		$C_{16}H_{11}ClF_6N_2O$	158.5	397.1	5.1	1.3E-08~1.3E-05 (5.0E-09 5.0E-06)	2.67	15.03	15.50
Fluoxastrobin		$C_{21}H_{16}ClFN_4O_5$	170.1	459.1	5.4	1.1E-08~1.1E-05 (4.4E-09~4.4E-06)	3.18	15.46	15.62
Flurochloridone		$C_{12}H_{10}Cl_2F_3NO$	199.5	312.0	5.4	1.6E-08~1.6E-05 (6.4E-09~6.4E-06)	2.03	14.50	14.19
Flurprimidole		$C_{15}H_{15}F_3N_2O_2$	119.0	313.1	4.7	1.6E-08~1.6E-05 (6.4E-09~6.4E-06)	2.66	15.03	15.32
Flusilazole		$C_{16}H_{15}F_2N_3Si$	130.4	316.1	5.2	1.6E-08~1.6E-05 (6.3E-09~6.3E-06)	3.48	15.71	15.77
Flutriafol		$C_{16}H_{13}F_2N_3O$	120.6	302.1	3.9	1.7E-08~1.7E-05 (6.6E-09~6.6E-06)	3.15	15.43	15.41

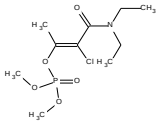
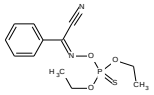
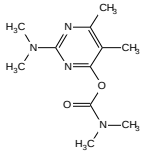
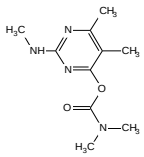
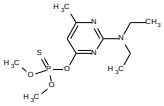
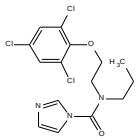
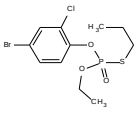
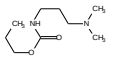
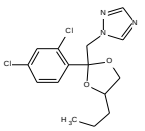
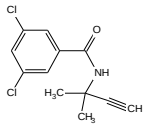
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Fluxapyroxad		$C_{18}H_{12}F_5N_3O$	123.2	382.1	5.0	1.3E-08~1.3E-05 (5.2E-09~5.2E-06)	2.81	15.15	15.16
Fosthiazate		$C_9H_{18}NO_3PS_2$	123.7	284.1	3.9	1.8E-08~1.8E-05 (7.0E-09~7.0E-06)	2.78	15.13	15.22
Halosulfuron methyl		$C_{13}H_{15}ClN_6O_7S$	166.4	435.0	4.9	1.2E-08~1.2E-05 (4.6E-09~4.6E-06)	2.47	14.87	15.21
Heptenophos		$C_9H_{12}ClO_4P$	147.0	251.0	4.2	2.0E-08~2.0E-05 (8.0E-09~8.0E-06)	2.5	14.89	15.00
Hexythiazox		$C_{17}H_{21}ClN_2O_2S$	169.2	353.1	6.6	1.4E-08~1.4E-05 (5.7E-09~5.7E-06)	3.01	15.32	14.72
imazalil		$C_{14}H_{14}Cl_2N_2O$	203.0	297.1	3.4	1.7E-08~1.7E-05 (6.7E-09~6.7E-06)	4.41	16.48	15.81
Imazamox		C15H19N3O4	120.2	306.1	2.5	1.6E-08~1.6E-05 (6.5E-09~6.5E-06)	3.04	15.34	15.35
Imazaquin		$C_{17}H_{17}N_3O_3$	122.5	312.1	3.5	1.6E-08~1.6E-05 (6.4E-09~6.4E-06)	3.01	15.32	15.46
Imidacloprid		$C_9H_{10}ClN_5O_2$	148.6	256.1	2.6	2.0E-08~2.0E-05 (7.8E-09~7.8E-06)	2.23	14.67	14.68

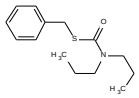
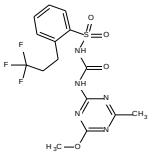
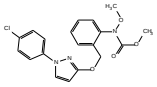
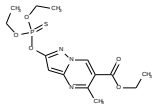
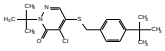
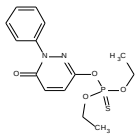
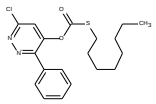
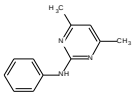
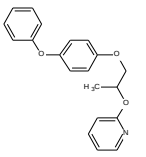
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Indoxacarb		$C_{22}H_{17}ClF_3N_3O_7$	171.1	528.1	6.1	9.5E-09~9.5E-06 (3.8E-09~3.8E-06)	2.37	14.79	15.23
Iprovalicarb		$C_{18}H_{28}N_2O_3$	123.5	321.2	4.8	1.6E-08~1.6E-05 (6.2E-09~6.2E-06)	3.03	15.33	15.42
Isoproturon		C12H18N2O	115.1	207.1	4.1	2.4E-08~2.4E-05 (9.7E-09~9.7E-06)	3.25	15.52	15.85
Isopyrazam		$C_{20}H_{23}F_2N_3O$	126.0	360.2	5.9	1.4E-08~1.4E-05 (5.6E-09~5.6E-06)	2.91	15.23	15.57
Linuron		$C_9H_{10}Cl_2N_2O_2$	193.3	249.0	4.7	2.0E-08~2.0E-05 (8.0E-09~8.0E-06)	2.34	14.76	14.81
Malaoxon		$C_{10}H_{19}O_7PS$	119.4	315.1	3.7	1.6E-08~1.6E-05 (6.3E-09~6.3E-06)	2.59	14.97	15.28
Malathion		$C_{10}H_{19}O_6PS_2$	125.0	331.0	5.3	1.5E-08~1.5E-05 (6.0E-09~6.0E-06)	2.67	15.03	14.53
Mecarbam		$C_{10}H_{20}NO_5PS_2$	125.2	330.1	5.5	1.5E-08~1.5E-05 (6.1E-09~6.1E-06)	2.55	14.94	14.98
Mepanipyrim		$C_{14}H_{13}N_3$	117.6	224.1	5.1	2.2E-08~2.2E-05 (8.9E-09~8.9E-06)	3.69	15.88	16.16

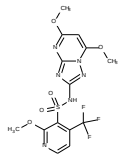
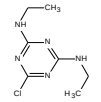
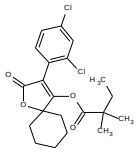
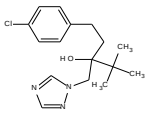
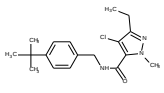
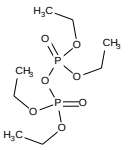
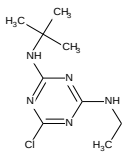
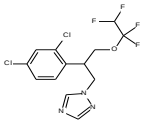
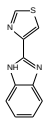
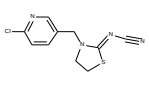
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Metaflumizone		$C_{24}H_{16}F_6N_4O_2$	132.2	507.1	6.3	9.9E-09~9.9E-06 (3.9E-09~3.9E-06)	2.3	14.73	15.14
Metalaxyl		$C_{15}H_{21}NO_4$	119.3	280.2	4.1	1.8E-08~1.8E-05 (7.1E-09~7.1E-06)	2.95	15.27	15.50
Metazachlor		$C_{14}H_{16}ClN_3O$	155.3	278.1	4.4	1.8E-08~1.8E-05 (7.2E-09~7.2E-06)	3.22	15.49	15.29
Metconazole		$C_{17}H_{22}ClN_3O$	160.3	320.2	5.3	1.6E-08~1.6E-05 (6.2E-09~6.2E-06)	3.48	15.71	15.42
methamidophos		$C_2H_8NO_2PS$	108.4	142.1	0.9	3.5E-08~3.5E-05 (1.4E-08~1.4E-06)	1.91	14.40	14.63
Methiocarb sulfoxide		$C_{11}H_{15}NO_3S$	119.9	242.1	2.4	2.1E-08~2.1E-05 (8.3E-09~8.3E-06)	2.43	14.84	15.31
Metosulam		$C_{14}H_{13}Cl_2N_5O4S$	219.5	418.0	4.1	1.2E-08~1.2E-05 (4.8E-09~4.8E-06)	2.77	15.12	15.10
Metrafenone		$C_{19}H_{21}BrO_5$	244.6	409.1	6.1	1.2E-08~1.2E-05 (4.9E-09~4.9E-06)	3.34	15.59	15.37
Metribuzin		$C_8H_{14}N_4OS$	116.8	215.1	3.6	2.3E-08~2.3E-05 (9.3E-09~9.3E-06)	2.9	15.23	15.58
Metsulfuron-methyl		$C_{14}H_{15}N_5O_6S$	126.5	382.1	3.7	1.3E-08~1.3E-05 (5.2E-09~5.2E-06)	2.52	14.91	15.22

Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Monolinuron		$C_9H_{11}ClN_2O_2$	147.4	215.1	4.1	2.3E-08~2.3E-05 (9.3E-09~9.3E-06)	2.33	14.75	15.14
Myclobutanil		$C_{15}H_{17}ClN_4$	157.7	289.1	4.9	1.7E-08~1.7E-05 (6.9E-09~6.9E-06)	3.1	15.39	15.32
Napropamide		$C_{17}H_{21}NO_2$	121.2	272.2	5.1	1.8E-08~1.8E-05 (7.3E-09~7.3E-06)	3.82	15.99	15.77
Nuarimol		$C_{17}H_{12}ClFN_2O$	160.2	315.1	4.5	1.6E-08~1.6E-05 (6.3E-09~6.3E-06)	3.03	15.33	15.11
Ofurace		$C_{14}H_{16}ClNO_3$	155.4	282.1	4.1	1.8E-08~1.8E-05 (7.1E-09~7.1E-06)	2.79	15.13	15.18
Omethoate		$C_5H_{12}NO_4PS$	112.7	214.0	1.3	2.3E-08~2.3E-05 (9.3E-09~9.3E-06)	2.3	14.73	15.36
Oxadiargyl		$C_{15}H_{14}Cl_2N_2O_3$	206.0	341.0	5.9	1.5E-08~1.5E-05 (5.9E-09~5.9E-06)	2.86	15.19	14.23
Oxadixyl		$C_{14}H_{18}N_2O_4$	118.5	279.1	3.4	1.8E-08~1.8E-05 (7.2E-09~7.2E-06)	2.76	15.11	14.88
Oxycarboxin		$C_{12}H_{13}NO_4S$	121.5	268.1	3.1	1.9E-08~1.9E-05 (7.5E-09~7.5E-06)	2.66	15.03	14.74

Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Oxydemeton-methyl		$C_6H_{15}O_4PS_2$	119.6	247.0	1.7	2.0E-08~2.0E-05 (8.1E-09~8.1E-06)	2.3	14.73	15.32
Oxyfluorfen		$C_{15}H_{11}ClF_3NO_4$	157.3	362.0	6.4	1.4E-08~1.4E-05 (5.5E-09~5.5E-06)	2.61	14.99	14.95
Paclobutrazol		$C_{15}H_{20}ClN_3O$	157.5	294.1	4.6	1.7E-08~1.7E-05 (6.8E-09~6.8E-06)	3.7	15.89	15.54
Paraoxon-methyl		$C_8H_{10}NO_6P$	111.0	248.0	3.4	2.0E-08~2.0E-05 (8.1E-09~8.1E-06)	1.88	14.38	14.76
Penconazole		$C_{13}H_{15}Cl_2N_3$	201.2	284.1	5.3	1.8E-08~1.8E-05 (7.0E-09~7.0E-06)	3.33	15.58	15.41
Pencycuron		$C_{19}H_{21}ClN_2O$	163.8	329.1	5.9	1.5E-08~1.5E-05 (6.1E-09~6.1E-06)	3.83	16.00	15.58
Penflufen		$C_{18}H_{24}FN_3O$	144.0	318.2	5.3	1.6E-08~1.6E-05 (6.3E-09~6.3E-06)	2.97	15.28	15.74
Penthiopyrad		$C_{16}H_{20}F_3N_3OS$	126.9	360.1	5.9	1.4E-08~1.4E-05 (5.6E-09~5.6E-06)	2.57	14.95	15.32
Pethoxamide		$C_{16}H_{22}ClNO_2$	158.5	296.1	5.3	1.7E-08~1.7E-05 (6.8E-09~6.8E-06)	3.55	15.77	15.31
Phosalone		$C_{12}H_{15}ClNO_4PS_2$	168.8	368.0	6.0	1.4E-08~1.4E-05 (5.4E-09~5.4E-06)	2.63	15.00	14.26

Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Phosphamidon		$C_{10}H_{19}ClNO_5P$	149.6	300.1	3.2	1.7E-08~1.7E-05 (6.7E-09~6.7E-06)	2.76	15.11	15.67
Phoxim		$C_{12}H_{15}N_2O_3PS$	121.7	299.1	6.0	1.7E-08~1.7E-05 (6.7E-09~6.7E-06)	2.79	15.13	14.56
Pirimicarb		$C_{11}H_{18}N_4O_2$	114.9	239.2	2.2	2.1E-08~2.1E-05 (8.4E-09~8.4E-06)	2.51	14.90	16.09
Pirimicarb-desmethyl		$C_{10}H_{16}N_4O_2$	113.7	225.1	1.7	2.2E-08~2.2E-05 (8.9E-09~8.9E-06)	2.55	14.94	15.97
Pirimiphos-methyl		$C_{11}H_{20}N_3O_3PS$	120.9	306.1	5.8	1.6E-08~1.6E-05 (6.5E-09~6.5E-06)	3.4	15.64	16.31
Prochloraz		$C_{15}H_{16}Cl_3N_3O_2$	274.4	376.0	4.8	1.3E-08~1.3E-05 (5.3E-09~5.3E-06)	3.26	15.53	15.08
Profenofos		$C_{11}H_{15}BrClO_3PS$	230.2	372.9	6.2	1.3E-08~1.3E-05 (5.3E-09~5.3E-06)	2.36	14.78	14.75
Propamocarb		$C_9H_{20}N_2O_2$	111.7	189.2	1.2	2.7E-08~2.7E-05 (1.1E-08~1.1E-05)	2.82	15.16	15.84
Propiconazole		$C_{15}H_{17}Cl_2N_3O_2$	208.4	342.1	5.4	1.5E-08~1.5E-05 (5.8E-09~5.8E-06)	3.57	15.78	15.39
Propyzamide		$C_{12}H_{11}Cl_2NO$	198.2	256.0	5.0	2.0E-08~2.0E-05 (7.8E-09~7.8E-06)	2.49	14.89	14.64

Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Prosulfocarb		$C_{14}H_{21}NOS$	123.3	252.1	6.2	2.0E-08~2.0E-05 (7.9E-09~7.9E-06)	3.48	15.71	15.61
Prosulfuron		$C_{15}H_{16}F_3N_5O_4S$	127.3	420.1	4.7	1.2E-08~1.2E-05 (4.8E-09~4.8E-06)	2.25	14.69	14.93
Pyraclostrobin		$C_{19}H_{18}ClN_3O_4$	165.5	388.1	5.9	1.3E-08~1.3E-05 (5.2E-09~5.2E-06)	3.22	15.49	15.59
Pyrazophos		$C_{14}H_{20}N_3O_5PS$	125.4	374.1	5.8	1.3E-08~1.3E-05 (5.3E-09~5.3E-06)	2.88	15.21	15.47
Pyridaben		$C_{19}H_{25}ClN_2OS$	172.6	365.1	7.1	1.4E-08~1.4E-05 (5.5E-09~5.5E-06)	4.36	16.44	15.39
Pyridaphenthion		$C_{14}H_{17}N_2O_4PS$	124.6	341.1	5.0	1.5E-08~1.5E-05 (5.9E-09~5.9E-06)	2.92	15.24	15.30
Pyridate		$C_{19}H_{23}ClN_2O_2S$	173.0	379.1	7.5	1.3E-08~1.3E-05 (5.3E-09~5.3E-06)	3.5	15.73	14.83
Pyrimethanil		$C_{12}H_{13}N_3$	115.2	200.1	3.6	2.5E-08~2.5E-05 (1.0E-08~1.0E-05)	3.53	15.75	16.09
Pyriproxyfen		$C_{20}H_{19}NO_3$	125.6	322.1	6.5	1.6E-08~1.6E-05 (6.2E-09~6.2E-06)	3.99	16.13	15.92

Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Pyroxsulam		$C_{14}H_{13}F_3N_6O_5S$	126.7	435.1	3.8	1.2E-08~1.2E-05 (4.6E-09~4.6E-06)	2.43	14.84	15.48
Simazine		$C_7H_{12}ClN_5$	145.0	202.1	3.4	2.5E-08~2.5E-05 (9.9E-09~9.9E-06)	3.34	15.59	15.81
Spirodiclofen		$C_{21}H_{24}Cl_2O_4$	191.1	411.1	7.2	1.2E-08~1.2E-05 (4.9E-09~4.9E-06)	3.25	15.52	14.73
Tebuconazole		$C_{16}H_{22}ClN_3O$	159.3	308.2	5.1	1.6E-08~1.6E-05 (6.5E-09~6.5E-06)	3.58	15.79	15.43
Tebufenpyrad		$C_{18}H_{24}ClN_3O$	162.7	334.2	6.2	1.5E-08~1.5E-05 (6.0E-09~6.0E-06)	3.63	15.83	15.44
TEPP		$C_8H_{20}O_7P_2$	111.0	291.1	3.1	1.7E-08~1.7E-05 (6.9E-09~6.9E-06)	2.94	15.26	15.18
Terbuthylazine		$C_9H_{16}ClN_5$	148.4	230.1	4.7	2.2E-08~2.2E-05 (8.7E-09~8.7E-06)	3.34	15.59	15.95
Tetraconazole		$C_{13}H_{11}Cl_2F_4N_3O$	203.2	372.0	5.0	1.3E-08~1.3E-05 (5.4E-09~5.4E-06)	3.14	15.43	15.21
Thiabendazole		$C_{10}H_7N_3S$	118.6	202.0	1.7	2.5E-08~2.5E-05 (9.9E-09~9.9E-06)	2.81	15.15	15.84
Thiacloprid		$C_{10}H_9ClN_4S$	152.0	253.0	3.2	2.0E-08~2.0E-05 (7.9E-09~7.9E-06)	2.57	14.95	14.97

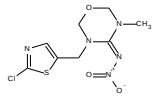
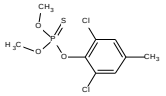
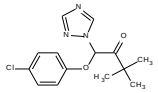
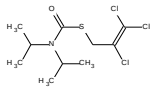
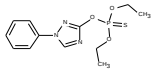
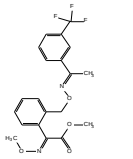
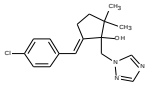
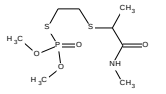
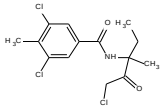
Name	Structure	Formula	ID	m/z	RT	C_{UT} (C_{DTU}) (mol/L)	log*IE*_pred	log*RF*_pred	log*RF*_Measured
Thiamethoxam		$C_8H_{10}ClN_5O_3S$	155.6	292.0	4.7	1.7E-08~1.7E-05 (6.8E-09~6.8E-06)	2.06	14.53	15.38
Tolclofos-methyl		$C_9H_{11}Cl_2O_3PS$	201.6	301.0	6.0	1.7E-08~1.7E-05 (6.6E-09~6.6E-06)	2.41	14.82	14.09
Triadimefon		$C_{14}H_{16}ClN_3O_2$	156.1	294.1	5.0	1.7E-08~1.7E-05 (6.8E-09~6.8E-06)	3.37	15.62	15.30
Triallate		$C_{10}H_{16}Cl_3NOS$	267.2	304.0	6.8	1.7E-08~1.7E-05 (6.6E-09~6.6E-06)	2.39	14.80	14.38
Triazophos		$C_{12}H_{16}N_3O_3PS$	122.1	314.1	5.3	1.6E-08~1.6E-05 (6.4E-09v6.4E-06)	2.92	15.24	15.77
Trifloxystrobin		$C_{20}H_{19}F_3N_2O_4$	126.4	409.1	6.1	1.2E-08~1.2E-05 (4.9E-09~4.9E-06)	2.86	15.19	15.86
Triticonazole		$C_{17}H_{20}ClN_3O$	160.9	318.1	4.7	1.6E-08~1.6E-05 (6.3E-09~6.3E-06)	3.43	15.67	15.26
Vamidothion		$C_8H_{18}NO_4PS_2$	122.6	288.0	2.3	1.7E-08~1.7E-05 (6.9E-09~6.9E-06)	3.25	15.52	15.35
Zoxamide		$C_{14}H_{16}Cl_3NO_2$	269.5	336.0	3.6	1.5E-08~1.5E-05 (6.0E-09~6.0E-06)	2.72	15.08	15.25

Table S2 Comparison of the concentration prediction errors between spiking the calibration set into pure solvent and into corresponding matrices on QQQ

compound	mean_error_solvent	mean_error_matrix	log*IE_pred*
Fluazifop-*p*-buthyl	32.0	33.7	2.32
Pyridate	28.7	30.1	3.68
Trifloxystrobin	15.5	16.0	2.62
Thiamethoxam	14.3	15.6	1.90
Indoxacarb	9.8	10.4	2.26
Phosphamidon	9.5	9.9	2.43
Pirimicarb-desmethyl	8.5	8.8	2.72
Pirimicarb	7.6	7.7	2.87
Metaflumizone	7.3	7.8	2.09
Prosulfuron	7.3	7.9	1.88
Pyraclostrobin	6.7	6.7	3.24
Triazophos	6.5	6.7	2.91
EPN	6.1	5.9	2.50
Chlorfenvinphos	6.0	6.3	2.32
Pirimiphos-methyl	5.9	5.8	3.31
Pyridaben	5.6	6.2	4.39
Pyrazophos	5.5	5.6	2.88
Zoxamide	5.5	5.7	2.66
Fluoxastrobin	5.3	5.3	3.29
Oxydemeton-methyl	5.3	5.6	2.32
Thiabendazole	5.2	5.3	2.87
Diazinon	4.9	4.9	3.29
Aldicarb-sulfone	4.4	4.9	1.36
Omethoate	4.4	4.7	2.19
Azoxystrobin	4.3	4.2	3.15
Isoproturon	4.1	4.1	3.10
Methiocarb sulfoxide	4.0	4.2	2.47
Ethoprophos	4.0	4.0	2.81
Heptenophos	4.0	4.2	2.22
Spirodiclofen	3.9	4.0	3.48
Clofentezine	3.9	3.8	2.87
Fenbuconazole	3.9	4.0	3.81
Cyprodinil	3.8	3.7	3.55
Tolclofos-methyl	3.7	3.6	2.57
Malaoxon	3.6	3.8	2.46

compound	mean_error_solvent	mean_error_matrix	log*IE_pred*
Monolinuron	3.6	3.7	2.35
Vamidothion	3.5	3.7	2.55
Fosthiazate	3.5	3.7	2.52
Propamocarb	3.5	3.6	3.00
Metsulfuron-methyl	3.5	3.6	2.52
Ofurace	3.5	3.6	2.53
Mepanipyrim	3.3	3.2	3.65
Dimethomorph	3.2	3.3	3.67
Triticonazole	3.1	3.2	3.69
Paraoxon-methyl	3.1	3.3	1.84
Flusilazole	3.0	3.0	3.05
Metconazole	2.9	3.0	3.82
Phosalone	2.9	2.8	2.58
Mecarbam	2.8	2.9	2.51
Tebuconazole	2.8	3.0	3.83
Fenthion	2.8	2.9	2.73
Fenthion Sulfone	2.7	2.6	2.16
Fenthion -sulfoxid	2.7	2.8	2.62
Fenpropimorph	2.7	2.9	4.40
Buprofezine	2.6	2.9	4.49
Cadusafos	2.4	2.5	3.70
Profenofos	2.4	2.4	2.46
Clethodim	2.3	2.4	3.82
Benfuracarb	2.3	2.3	3.49
Hexythiazox	2.3	2.3	2.99
TEPP	2.2	2.3	2.46
Carboxin	2.2	2.2	2.80
Fenoxycarb	2.2	2.2	3.11
Metalaxyl	2.1	2.1	3.1
Ethion	2.1	2.2	3.36
Tebufenpyrad	2.1	2.2	3.89
Pyridaphenthion	2.1	2.1	2.95
Phoxim	2.1	2.0	2.55
Clomazone	2.1	2.1	2.93
Metribuzin	2.0	2.0	2.96
Fenamiphos	2.0	2.0	3.09
Paclobutrazol	2.0	2.1	3.72

compound	mean_error_solvent	mean_error_matrix	log*IE_pred*
Flutriafol	2.0	2.0	2.82
Tetraconazole	1.9	1.9	3.02
Linuron	1.9	2.0	2.53
Carbendazim	1.9	1.9	3.11
Fenamiphos-sulfoxide	1.9	1.9	2.82
Bupirimate	1.9	1.8	4.05
Pyrimethanil	1.8	1.8	3.61
Malathion	1.8	1.8	2.45
Acetamiprid	1.8	1.8	2.86
Iprovalicarb	1.8	1.8	3.12
Atrazine	1.7	1.7	3.54
Fenamiphos-sulphone	1.7	1.7	2.58
Nuarimol	1.7	1.7	2.75
methamidophos	1.7	1.8	1.70
Difenoconazole	1.6	1.6	3.49
Fenpropidin	1.6	1.5	4.32
Simazine	1.6	1.6	3.43
Imidacloprid	1.6	1.7	2.23
Pencycuron	1.5	1.6	4.01
Fenazaquin	1.5	1.6	4.47
Boscalid	1.5	1.5	2.83
Prochloraz	1.5	1.5	2.95
Thiacloprid	1.5	1.5	2.76
imazalil	1.4	1.4	3.71
Penconazole	1.4	1.4	3.22
Prosulfocarb	1.4	1.5	3.78
Oxadixyl	1.4	1.5	2.64
Propyzamide	1.4	1.4	2.43
Pyriproxyfen	1.4	1.4	4.00
Oxycarboxin	1.3	1.3	2.43
Propiconazole	1.3	1.3	3.48
Ethoxyquin	1.3	1.3	3.8
Myclobutanil	1.3	1.3	3.17
Triallate	1.3	1.2	2.5
Fenarimol	1.3	1.3	3.07
Triadimefon	1.3	1.3	3.03
Dichlorvos	1.2	1.3	1.78

Table S3 Predicted log*IE* Values of All Compounds and Their Mean Concentration Prediction Error in Solvent and Matrices (UT)

			Concentration Prediction Error							
Name	**logIE_{pred}**	**logRF_{pred}**	**Solvent**	**Matrices**	**Barley**	**Maize**	**Oat**	**Rice**	**Rye**	**Wheat**
1-Naphtylacetamide	3.06	15.36	2.82	2.63	2.44	3.05	2.52	2.58	2.54	2.65
Acetamiprid	2.5	14.89	2.5	3.24	3.88	3.13	2.82	2.95	3.88	2.75
Acetochlor	2.82	15.16	3.96	4.19	3.86	4.27	4.14	4.38	4.36	4.11
Aldicarb-sulfone	1.31	13.90	6.21	6.04	3.77	6.17	4.33	7.69	7.02	7.26
Ametoctradin	3.49	15.72	1.92	2.68	2.37	2.51	3.6	2.69	2.54	2.36
Atrazine	3.52	15.74	1.6	1.40	1.16	1.4	1.42	1.53	1.41	1.47
Azoxystrobin	3.17	15.45	2.13	3.91	4.17	5.02	3.73	3.32	3.99	3.24
Benfuracarb	3.09	15.38	2.63	2.46	2.78	1.38	5.87	2	1.17	1.58
Boscalid	3.38	15.63	4.39	3.77	3.64	3.02	4.71	3.6	3.8	3.87
Bupirimate	4.11	16.23	1.49	1.67	1.28	1.84	1.82	1.69	2.04	1.36
Buprofezine	3.53	15.75	2.66	2.17	2.05	2.43	1.65	2.29	2.33	2.24
Cadusafos	3.16	15.44	1.24	1.45	1.31	1.47	1.51	1.42	1.69	1.31
Carbendazim	2.87	15.20	3.09	2.54	2.65	2.47	2.38	2.91	2.55	2.27
Carbosulfan	3.92	16.07	3.92	64.63	42.19	55.48	173.77	54.51	39.68	22.14
Carboxin	2.9	15.23	1.42	2.09	2.78	2.07	1.99	2	1.98	1.69
Chlorfenvinphos	2.46	14.86	2.94	4.45	4.41	3.94	6.96	4.21	3.96	3.21
Chloridazon	2.46	14.86	3.39	4.33	6.81	2.96	4.12	4.5	3.55	4.04
Chlorsulfuron	2.46	14.86	1.26	1.46	1.73	2.19	1.3	1.2	1.24	1.12
Clethodim	3.22	15.49	1.51	1.28	1.31	1.19	1.37	1.29	1.23	1.29
Clofentezine	2.71	15.07	3.17	3.69	3.44	3.67	5.03	3.31	3.09	3.57
Clomazone	3.13	15.42	1.87	1.86	2.24	1.81	1.79	1.82	1.78	1.69
Cyprodinil	3.55	15.77	3.57	4.07	4.37	4.2	3.62	4.11	4.11	4.01
Desmedipham	2.94	15.26	5.01	5.15	7.21	4.55	4.09	3.73	6.5	4.82
Diazinon	3.14	15.43	6.87	7.18	7.83	8.23	5.04	6.73	7.56	7.67
Dichlorvos	1.58	14.13	2.76	3.08	3.68	3.01	2.57	2.97	3.53	2.72
Difenoconazole	3.92	16.07	1.65	1.43	1.42	1.41	1.49	1.37	1.38	1.5
Dimethachlor	2.89	15.22	1.99	2.40	3.33	2.27	2.18	2.39	2.18	2.02
Dimethomorph	3.6	15.81	4.39	4.01	4.91	3.14	3.74	3.92	4.1	4.27
Dodemorph	4.48	16.54	3.04	2.98	2.69	3	3.17	2.98	3.18	2.88
EPN	2.8	15.14	11.75	11.28	8.6	12.19	12.37	12.3	11.32	10.9
Ethion	3.78	15.96	7.25	7.41	7.84	6.69	7.41	7.22	7.37	7.94
Ethoprophos	2.82	15.16	2.62	2.90	2.83	3.05	2.72	3.31	3.14	2.33
Ethoxyquin	3.66	15.86	1.27	1.82	1.34	1.22	1.24	1.32	3.5	2.31
Fenamiphos	3.15	15.43	1.22	1.44	1.3	1.42	1.67	1.53	1.51	1.18
Fenamiphos-sulfoxide	3.14	15.43	1.29	1.26	1.41	1.19	1.35	1.11	1.25	1.24
Fenamiphos-sulphone	2.99	15.30	3.77	3.31	3.54	2.87	4.04	3.01	3.48	2.93
Fenarimol	3.24	15.51	1.83	1.75	1.58	1.67	2.15	1.42	1.73	1.92
Fenazaquin	3.92	16.07	1.86	1.78	1.42	1.97	1.5	2.19	1.91	1.69
Fenbuconazole	3.6	15.81	2.49	4.88	8.54	2.26	2.36	9.54	4.03	2.54
Fenoxycarb	3.01	15.32	2.7	2.37	2.39	1.65	3.28	2.15	2.56	2.21
Fenpropidin	4.28	16.37	1.73	2.20	3.64	1.98	1.9	1.95	1.84	1.87
Fenpropimorph	4.32	16.41	2.19	2.15	2.39	2.04	2.02	2.15	2.16	2.12
Fenthion	2.87	15.20	1.73	2.07	2.73	2.08	1.71	2.27	2.1	1.55
Fenthion Sulfone	2.48	14.88	1.43	1.53	1.37	1.52	1.94	1.52	1.53	1.29
Fenthion -sulfoxid	2.72	15.08	5.99	6.69	6.78	9.28	5.59	5.5	6.82	6.16
Fluazifop-*p*-buthyl	2.93	15.25	4.19	6.24	6.02	6.61	6.25	6.82	6	5.71
Fluometuron	2.02	14.49	13.69	15.85	17.9	14.62	16.33	16.57	14.95	14.73
Fluopicolide	2.22	14.66	4.91	5.22	5.27	5.05	4.49	4.5	5.07	6.96
Fluopyram	2.67	15.03	3.94	5.12	5.7	4.37	7.45	4.85	4.64	3.73

Fluoxastrobin	3.18	15.46	2.06	3.76	4.36	3.73	4.43	3.86	3.57	2.61
Flurochloridone	2.03	14.50	1.3	1.49	1.52	1.43	2.01	1.29	1.52	1.17
Flurprimidole	2.66	15.03	2.05	2.65	4.66	2.02	2.32	2.23	2.47	2.2
Flusilazole	3.48	15.71	1.3	1.34	1.36	1.37	1.3	1.32	1.28	1.43
Flutriafol	3.15	15.43	1.15	1.31	1.38	1.36	1.22	1.32	1.28	1.29
Fluxapyroxad	2.81	15.15	1.22	1.57	1.52	1.29	1.24	2.82	1.2	1.34
Fosthiazate	2.78	15.13	1.49	1.63	1.73	1.62	1.51	1.84	1.71	1.35
Halosulfuron methyl	2.47	14.87	3.01	3.42	3.83	3.68	3.49	3.02	3.01	3.49
Heptenophos	2.5	14.89	1.72	1.77	1.56	1.74	2.07	1.76	1.68	1.82
Hexythiazox	3.01	15.32	2.89	3.01	3.71	2.91	2.84	2.52	2.82	3.25
imazalil	4.41	16.48	3.02	2.97	2.81	2.98	3.04	3.03	3.09	2.88
Imazamox	3.04	15.34	1.05	1.56	1.75	1.36	1.63	1.84	1.4	1.4
Imazaquin	3.01	15.32	1.5	2.01	2.84	2.11	1.83	1.87	1.7	1.71
Imidacloprid	2.23	14.67	1.29	2.09	3.58	1.88	2.06	1.74	1.62	1.68
Indoxacarb	2.37	14.79	3.87	4.98	5.53	5.31	4.63	5.28	5.11	4.04
Iprovalicarb	3.03	15.33	1.45	2.32	1.58	2.61	1.95	3.99	2.12	1.68
Isoproturon	3.25	15.52	2.35	2.91	2.18	3.12	3.12	3.05	3.01	2.95
Isopyrazam	2.91	15.23	2.33	3.44	4.08	3.08	3.44	3.67	3.52	2.86
Linuron	2.34	14.76	1.85	1.95	1.37	1.15	3.25	2.13	1.74	2.06
Malaoxon	2.59	14.97	2.16	4.17	2.84	4.45	3.16	4.66	5.38	4.53
Malathion	2.67	15.03	3.19	2.29	2.21	2.17	1.96	3.17	2.2	2
Mecarbam	2.55	14.94	1.38	1.79	1.85	1.89	1.5	2.3	1.79	1.41
Mepanipyrim	3.69	15.88	2.13	2.30	1.99	2.26	1.93	2.4	3.01	2.18
Metaflumizone	2.3	14.73	2.96	2.95	2.59	2.98	2.55	2.9	3.03	3.63
Metalaxyl	2.95	15.27	1.81	2.80	2.86	2.5	2.06	2.35	5.26	1.79
Metazachlor	3.22	15.49	1.4	1.48	1.53	1.64	1.44	1.44	1.39	1.43
Metconazole	3.48	15.71	1.48	1.38	1.31	1.35	1.48	1.38	1.34	1.39
methamidophos	1.91	14.40	1.65	1.36	1.14	2	1.24	1.26	1.2	1.29
Methiocarb sulfoxide	2.43	14.84	3.21	3.78	5.19	3.47	2.96	3.82	3.45	3.76
Metosulam	2.77	15.12	1.66	2.33	3.06	2.51	2.41	1.99	2.19	1.82
Metrafenone	3.34	15.59	1.13	1.35	1.28	1.29	1.36	1.22	1.67	1.27
Metribuzin	2.9	15.23	2.32	2.66	2.5	2.69	2.57	2.78	2.74	2.66
Metsulfuron-methyl	2.52	14.91	2.19	3.48	4.03	2.72	3.03	3.5	4.7	2.87
Monolinuron	2.33	14.75	3.15	3.77	4.05	3.56	3.92	3.58	3.91	3.58
Myclobutanil	3.1	15.39	1.15	1.21	1.22	1.23	1.15	1.24	1.2	1.2
Napropamide	3.82	15.99	1.65	1.59	1.78	1.52	1.55	1.51	1.51	1.66
Nuarimol	3.03	15.33	1.29	1.35	1.27	1.35	1.68	1.31	1.27	1.23
Ofurace	2.79	15.13	1.45	1.66	2.22	1.34	1.79	1.91	1.35	1.36
Omethoate	2.3	14.73	4.51	5.64	3.44	5.96	6.73	5.74	6.91	5.03
Oxadiargyl	2.86	15.19	5.23	5.08	4.81	5.24	6.53	4.92	3.32	5.67
Oxadixyl	2.76	15.11	1.67	1.46	1.51	1.35	1.63	1.37	1.43	1.49
Oxycarboxin	2.66	15.03	1.91	1.72	1.91	1.83	1.75	1.61	1.49	1.71
Oxydemeton-methyl	2.3	14.73	4.71	5.10	3.02	6.07	5.73	6.15	3.63	5.97
Oxyfluorfen	2.61	14.99	1.31	1.52	1.84	1.35	1.8	1.49	1.1	1.54
Paclobutrazol	3.7	15.89	1.8	1.52	1.37	1.43	1.78	1.53	1.58	1.45
Paraoxon-methyl	1.88	14.38	2.4	3.31	3.46	2.26	2.78	4.55	3.02	3.79
Penconazole	3.33	15.58	1.23	1.47	1.64	1.29	1.3	1.4	1.63	1.53
Pencycuron	3.83	16.00	1.97	1.61	1.51	1.45	1.85	1.44	1.56	1.86
Penflufen	2.97	15.28	3.73	3.76	4.03	3.21	4.07	4.27	3.61	3.36
Penthiopyrad	2.57	14.95	2.59	2.91	3.36	2.7	2.91	3.04	2.78	2.66
Pethoxamide	3.55	15.77	2.26	2.22	2.06	2.27	2.22	2.25	2.22	2.32
Phosalone	2.63	15.00	3.92	3.82	3.45	3.99	3.55	3.49	3.71	4.74
Phosphamidon	2.76	15.11	4.65	5.15	4.85	5.28	5.28	5.35	5.22	4.91
Phoxim	2.79	15.13	3.92	4.13	3.63	4.14	4.89	4.49	3.78	3.85

Pirimicarb	2.51	14.90	15.99	15.70	14.49	14.22	17.11	15.78	16.4	16.18
Pirimicarb-desmethyl	2.55	14.94	11.79	12.47	9.13	12.52	15.92	12.32	12.78	12.13
Pirimiphos-methyl	3.4	15.64	5.22	5.02	5.95	5.65	3.3	4.82	5.08	5.31
Prochloraz	3.26	15.53	1.36	1.32	1.37	1.25	1.4	1.28	1.26	1.36
Profenofos	2.36	14.78	1.76	1.52	1.44	1.46	1.25	1.72	1.56	1.69
Propamocarb	2.82	15.16	4.92	5.09	4.48	5.54	5.12	5.28	5.73	4.4
Propiconazole	3.57	15.78	1.49	1.30	1.29	1.24	1.28	1.49	1.22	1.28
Propyzamide	2.49	14.89	1.29	1.40	1.35	1.35	1.66	1.36	1.34	1.32
Prosulfocarb	3.48	15.71	1.22	1.31	1.3	1.28	1.38	1.35	1.24	1.29
Prosulfuron	2.25	14.69	1.94	2.21	2.16	2.32	2.82	1.7	2.24	2.03
Pyraclostrobin	3.22	15.49	1.74	4.27	3.97	3.98	7	3.88	3.74	3.03
Pyrazophos	2.88	15.21	2	4.45	4.59	3.87	6.14	5.36	3.77	2.96
Pyridaben	4.36	16.44	9.02	15.41	17.77	14.04	19.49	12.42	13.94	14.81
Pyridaphenthion	2.92	15.24	1.27	1.36	1.55	1.46	1.17	1.29	1.36	1.34
Pyridate	3.5	15.73	5.22	14.11	11.36	20.3	24.08	13.52	9.24	6.17
Pyrimethanil	3.53	15.75	2.44	2.16	3.16	2.18	1.44	2.22	1.94	2
Pyriproxyfen	3.99	16.13	1.59	1.61	1.5	1.47	2.01	1.38	1.74	1.54
Pyroxsulam	2.43	14.84	4.81	8.33	7.23	9.03	10.13	8.29	8.26	7.05
Simazine	3.34	15.59	2.1	1.84	2.17	1.71	1.59	1.95	1.78	1.82
Spirodiclofen	3.25	15.52	4.1	10.28	11.38	10.86	12.59	7.86	9.32	9.66
Tebuconazole	3.58	15.79	1.76	1.57	1.54	1.6	1.52	1.58	1.47	1.69
Tebufenpyrad	3.63	15.83	1.89	2.16	2.43	2.36	2.14	1.76	2.17	2.09
TEPP	2.94	15.26	1.26	1.40	1.56	1.53	1.46	1.27	1.35	1.25
Terbuthylazine	3.34	15.59	3.31	2.89	2.25	3	2.95	3.15	2.85	3.15
Tetraconazole	3.14	15.43	1.1	1.20	1.12	1.12	1.18	1.35	1.25	1.16
Thiabendazole	2.81	15.15	5.95	4.74	4.14	4.96	4.35	6.03	4.31	4.63
Thiacloprid	2.57	14.95	1.52	1.78	3.21	1.35	1.25	1.63	1.73	1.48
Thiamethoxam	2.06	14.53	9.27	11.95	11.61	14.94	11.74	11.75	11.69	9.96
Tolclofos-methyl	2.41	14.82	3.21	3.99	3.61	4.06	5.31	3.51	3.95	3.49
Triadimefon	3.37	15.62	1.62	1.60	1.41	1.64	1.91	1.61	1.49	1.54
Triallate	2.39	14.80	1.14	1.57	1.4	1.28	2.86	1.32	1.23	1.3
Triazophos	2.92	15.24	3.62	4.81	4.94	4.71	5.34	4.79	5	4.1
Trifloxystrobin	2.86	15.19	5	6.63	6.27	6.77	7.18	6.47	6.7	6.38
Triticonazole	3.43	15.67	1.96	2.32	4.63	1.74	1.72	2.53	1.76	1.56
Vamidothion	3.25	15.52	1.52	1.78	2.6	1.58	1.65	1.73	1.53	1.61
Zoxamide	2.72	15.08	3.05	4.60	5.34	4.24	4.43	5.33	4.56	3.67

Table S4 Predicted log IE Values of All Compounds and Their Mean concentration Prediction Error in Solvent and Matrices (DTU)

			Concentration Prediction Error							
Name	logIE_{pred}	logRF_{pred}	Solvent	Matrices	Barley	Maize	Oat	Rice	Rye	Wheat
1-Naphtylacetamide	2.75	13.07	11.57	18.45	18.27	18.51	14.51	23.97	16.85	18.61
Acetamiprid	2.45	12.92	1.23	1.69	2.25	1.64	1.41	1.96	1.41	1.49
Acetochlor	2.92	13.15	6.06	7.78	14.28	7.61	6.17	5.69	6.23	6.67
Aldicarb-sulfone	1.58	12.49	5.13	9.83	11.04	13.36	9.04	7.06	8.96	9.51
Ametoctradin	3.53	13.45	2.08	1.91	1.8	1.85	2.11	2.08	1.97	1.63
Atrazine	3.32	13.35	1.51	1.26	1.16	1.19	1.43	1.36	1.14	1.27
Azoxystrobin	3.24	13.31	1.69	1.45	1.59	1.22	1.6	1.6	1.2	1.48
Boscalid	3.35	13.36	2.28	3.13	3.49	3.78	2.79	2.58	2.94	3.18
Bupirimate	4.07	13.71	2.36	1.86	1.92	1.7	2.03	2.12	1.64	1.74
Buprofezine	3.36	13.37	2.91	2.12	1.81	1.89	2.51	2.55	1.99	1.97
Cadusafos	3.13	13.25	1.55	2.85	2.24	2.28	3.26	3.19	2.94	3.19
Carbendazim	2.59	12.99	1.24	1.85	2.09	2.14	2.14	1.33	1.54	1.86
Carboxin	2.79	13.09	1.45	2.35	2.17	3.43	1.88	1.61	1.96	3.07
Chlorfenvinphos	2.49	12.94	2.37	1.97	1.87	1.73	2.26	2.43	1.79	1.76
Chloridazon	2.5	12.95	1.83	1.61	1.78	1.71	1.59	1.45	1.63	1.47
Chlorsulfuron	2.4	12.90	1.12	1.23	1.21	1.2	1.21	1.32	1.27	1.16
Clethodim	3.1	13.24	1.23	1.55	1.45	1.38	1.95	1.5	1.18	1.86
Clofentezine	2.74	13.06	4.46	8.39	9.25	9.45	9.34	5.77	9.53	6.99
Clomazone	3.17	13.27	2.04	2.42	2.9	2.97	1.79	1.44	3.32	2.07
Cyprodinil	3.54	13.45	1.55	1.35	1.2	1.31	1.49	1.58	1.3	1.23
Desmedipham	3.03	13.21	1.75	3.08	3.44	3.32	2.41	2.7	3.18	3.44
Diazinon	3.21	13.29	5.4	3.75	3.3	3.34	3.81	4.52	3.82	3.71
Dichlorvos	1.56	12.49	1.55	1.28	1.16	1.18	1.25	1.48	1.29	1.33
Difenoconazole	3.77	13.57	1.97	1.48	1.22	1.41	1.69	1.65	1.36	1.56
Dimethachlor	2.82	13.10	1.57	1.50	1.78	1.42	1.63	1.25	1.48	1.41
Dimethomorph	3.68	13.52	3.38	4.66	5.24	4.74	3.83	4.92	4.42	4.78
Dodemorph	4.49	13.92	1.93	2.61	2.99	2.93	2.35	2.08	2.66	2.66
EPN	2.75	13.07	11.74	10.84	9.48	8.96	10.95	11.08	11.13	13.46
Ethion	3.71	13.54	8.49	16.03	23.7	13.74	22.8	9.46	13.03	13.45
Ethoprophos	2.8	13.09	1.71	1.31	1.17	1.24	1.54	1.49	1.27	1.13
Ethoxyquin	3.69	13.53	2.35	4.93	7.52	2.65	7.11	3.65	2.27	6.35
Fenamiphos	3.13	13.25	1.2	1.51	1.39	1.47	1.31	1.27	1.87	1.74
Fenamiphos-sulfoxide	3.15	13.26	1.33	1.89	2.29	2.64	1.64	1.51	1.56	1.67
Fenamiphos-sulphone	3	13.19	1.6	1.60	1.57	1.91	1.33	1.61	1.7	1.5
Fenarimol	3.3	13.34	1.32	1.38	1.25	1.2	1.93	1.29	1.35	1.25
Fenazaquin	3.69	13.53	3.8	2.91	2.57	2.62	3.1	3.4	3.04	2.75
Fenbuconazole	3.64	13.50	1.37	1.43	1.36	1.4	1.23	1.27	1.42	1.89
Fenoxycarb	2.88	13.13	2.05	2.92	2.66	2.69	3.32	2.6	2.59	3.66
Fenpropidin	4.24	13.80	1.65	1.41	1.51	1.29	1.55	1.46	1.35	1.27
Fenpropimorph	4.26	13.81	1.21	1.83	2.13	2.02	1.6	1.32	1.87	2.01
Fenthion	3.03	13.21	4.99	9.16	9.04	16.16	7.39	6.46	8.82	7.1
Fenthion Sulfone	2.81	13.10	1.57	1.51	1.28	1.56	1.3	1.21	1.26	2.45
Fenthion -sulfoxid	2.59	12.99	2.23	2.11	2.53	2.29	2.12	2.28	2.27	1.19
Fluazifop-*p*-buthyl	2.93	13.16	4.04	3.33	2.62	2.86	4.52	3.88	3.33	2.76
Fluometuron	1.95	12.68	3.86	2.74	2.52	2.39	3.51	3.24	2.63	2.17
Fluopicolide	2.24	12.82	3.03	2.28	2.07	1.99	2.46	2.88	2.13	2.15
Fluopyram	2.66	13.02	3.36	2.62	2.27	2.24	2.5	3.47	2.82	2.4
Fluoxastrobin	3.11	13.24	1.73	1.43	1.33	1.42	1.58	1.36	1.3	1.58
Flurochloridone	1.96	12.68	1.25	1.83	1.8	1.58	1.75	1.41	1.59	2.82

Flurprimidole	2.62	13.00	2.99	2.87	2.66	2.56	3.13	3.49	2.49	2.88
Flusilazole	3.38	13.38	2.33	2.30	2.18	2.09	2.35	2.51	2.45	2.19
Flutriafol	3.08	13.23	2.09	1.78	1.57	1.8	1.8	1.99	1.56	1.94
Fluxapyroxad	2.98	13.18	1.53	2.13	1.96	2.57	1.82	2.33	2.03	2.08
Fosthiazate	2.79	13.09	1.88	2.27	3.24	2.18	1.67	2.02	2.19	2.34
Halosulfuron methyl	2.49	12.94	1.38	1.27	1.41	1.31	1.4	1.18	1.14	1.16
Heptenophos	2.52	12.96	1.57	1.95	2.49	1.84	1.5	1.87	1.67	2.34
Hexythiazox	2.91	13.15	1.76	2.43	2.71	2.47	2.26	2.21	2.47	2.44
imazalil	4.22	13.79	1.34	2.10	2.44	2.08	1.76	1.75	2.08	2.46
Imazamox	2.98	13.18	1.7	1.71	1.66	1.91	1.8	1.75	1.67	1.46
Imazaquin	3.03	13.21	2.37	1.97	1.73	1.64	2.15	2.33	1.99	1.96
Imidacloprid	2.19	12.79	1.54	1.80	1.87	1.88	1.6	1.75	1.7	2
Indoxacarb	2.37	12.88	1.76	1.35	1.2	1.24	1.55	1.69	1.25	1.19
Iprovalicarb	3.07	13.22	1.15	1.55	1.42	2.08	1.58	1.12	1.21	1.86
Isoproturon	3.27	13.32	1.26	1.36	1.4	1.4	1.35	1.25	1.37	1.36
Isopyrazam	3.01	13.20	1.86	1.60	1.43	1.89	1.48	1.96	1.53	1.28
Linuron	2.32	12.86	1.71	1.94	2.1	1.99	1.68	1.59	2.26	2.03
Malaoxon	2.55	12.97	1.44	1.38	1.5	1.27	1.51	1.32	1.38	1.27
Malathion	2.64	13.01	2.94	4.33	5.44	4.43	4	3.44	4.77	3.87
Mecarbam	2.55	12.97	5.9	6.17	9.8	6.67	5.57	4.92	4.77	5.26
Mepanipyrim	3.64	13.50	1.83	1.41	1.29	1.3	1.41	1.6	1.42	1.43
Metaflumizone	2.35	12.87	1.98	1.58	1.28	1.65	1.54	1.84	1.46	1.7
Metalaxyl	3.05	13.21	1.43	1.30	1.29	1.17	1.23	1.28	1.28	1.57
Metazachlor	3.31	13.34	2.37	3.20	2.88	3.76	2.86	3.09	3.45	3.13
Metconazole	3.44	13.41	1.78	1.65	1.35	1.52	1.95	1.85	1.72	1.53
Methiocarb sulfoxide	2.26	12.83	1.28	1.40	1.27	1.57	1.4	1.37	1.44	1.37
Metosulam	2.77	13.08	1.15	1.21	1.18	1.22	1.26	1.2	1.16	1.23
Metrafenone	3.29	13.33	1.33	1.69	1.71	2.12	1.83	1.3	1.46	1.69
Metribuzin	2.71	13.05	1.84	1.77	1.91	1.55	1.93	1.8	1.75	1.7
Metsulfuron-methyl	2.47	12.93	1.97	1.65	1.74	1.52	1.74	1.73	1.61	1.58
Monolinuron	2.34	12.87	1.49	2.04	2.46	2.11	1.92	1.83	1.83	2.11
Myclobutanil	3.27	13.32	1.5	1.27	1.33	1.2	1.34	1.36	1.22	1.14
Napropamide	3.89	13.63	1.2	1.53	1.78	1.64	1.35	1.24	1.39	1.77
Nuarimol	3	13.19	1.41	1.36	1.31	1.44	1.35	1.6	1.21	1.23
Ofurace	2.86	13.12	1.41	1.53	1.72	1.54	1.56	1.28	1.51	1.55
Omethoate	2.16	12.78	2.11	3.51	4.84	4.2	2	1.92	5.47	2.63
Oxadiargyl	2.78	13.08	2.02	4.08	3.74	3.9	3.88	2.68	6.5	3.8
Oxadixyl	2.78	13.08	2.73	3.66	3.87	4.31	3.03	3.02	3.76	3.99
Oxycarboxin	2.61	13.00	1.8	3.16	2.36	5.67	2.29	3.95	2.13	2.55
Oxydemeton-methyl	2.19	12.79	1.2	1.66	1.86	1.85	1.38	1.73	1.49	1.67
Oxyfluorfen	2.57	12.98	3.26	5.90	7.59	4.51	6.86	4.51	4.31	7.61
Paclobutrazol	3.67	13.52	1.7	1.39	1.39	1.2	1.54	1.44	1.32	1.44
Paraoxon-methyl	1.85	12.63	1.15	1.36	1.26	1.34	1.49	1.18	1.58	1.32
Penconazole	3.34	13.36	1.91	1.77	1.64	1.52	1.93	1.95	1.86	1.74
Pencycuron	3.65	13.51	1.33	1.70	2.16	1.86	1.37	1.36	1.31	2.11
Penflufen	3.07	13.22	2.9	1.85	1.61	1.81	2.03	2.17	1.89	1.58
Penthiopyrad	2.57	12.98	2.17	1.97	1.73	1.73	2.66	2.1	1.91	1.71
Pethoxamide	3.58	13.47	1.31	2.18	2.32	2.16	2.14	2.04	2.44	1.98
Phosalone	2.58	12.98	1.84	3.67	5.67	3.57	3.03	3.14	3.34	3.28
Phosphamidon	2.71	13.05	3.18	2.23	2.19	1.61	2.7	2.5	2.42	1.93
Phoxim	2.86	13.12	1.28	1.27	1.12	1.27	1.2	1.3	1.54	1.19
Pirimicarb	2.61	13.00	3.21	2.45	1.9	1.92	2.75	3.2	2.6	2.34
Pirimicarb-desmethyl	2.59	12.99	2.48	1.83	1.31	1.69	1.96	2.41	2	1.6
Pirimiphos-methyl	3.42	13.40	4.41	3.22	2.7	2.67	3.26	4.72	3	2.96

Prochloraz	3.31	13.34	1.75	1.44	1.55	1.72	1.18	1.19	1.27	1.74
Profenofos	2.47	12.93	2	1.61	1.53	1.45	1.67	1.84	1.65	1.51
Propamocarb	2.71	13.05	2.01	3.84	3.79	3.91	3.85	3.16	3.67	4.64
Propiconazole	3.59	13.48	2.2	1.57	1.35	1.41	1.82	1.8	1.66	1.37
Propyzamide	2.56	12.97	1.93	2.26	2.68	2.36	2	2.05	2.19	2.25
Prosulfocarb	3.39	13.38	2.43	2.99	3.75	2.79	2.22	2.58	3	3.6
Prosulfuron	2.24	12.82	1.47	1.67	1.73	1.66	1.52	1.36	1.59	2.16
Pyraclostrobin	3.46	13.42	1.53	1.50	1.63	1.32	1.63	1.66	1.39	1.37
Pyrazophos	2.94	13.16	1.87	1.47	1.35	1.3	1.22	2.12	1.34	1.47
Pyridaben	3.99	13.68	1.6	2.50	2.96	2.97	1.96	1.88	2.16	3.04
Pyridaphenthion	2.9	13.14	1.31	1.52	2.07	1.95	1.47	1.26	1.16	1.2
Pyridate	3.5	13.44	2.08	2.78	2.82	3.07	2.08	2.97	3.09	2.65
Pyrimethanil	3.47	13.42	1.19	1.42	1.57	1.78	1.28	1.11	1.38	1.42
Pyriproxyfen	3.86	13.61	2.69	2.02	1.97	1.83	2.06	2.44	1.95	1.89
Pyroxsulam	2.32	12.86	2.43	2.18	2.04	2	2.46	2.58	2.1	1.92
Simazine	3.24	13.31	1.45	1.37	1.34	2.09	1.15	1.19	1.26	1.2
Spirodiclofen	3.18	13.28	1.89	1.82	2.51	1.87	1.85	1.34	1.59	1.73
Tebuconazole	3.58	13.47	1.34	1.32	1.29	1.32	1.3	1.43	1.26	1.33
Tebufenpyrad	3.51	13.44	2.22	2.02	2.21	2.25	1.76	1.62	1.94	2.34
TEPP	2.99	13.19	2	2.87	2.84	3.73	3	2.13	2.57	2.97
Terbuthylazine	3.24	13.31	4.02	2.90	2.37	2.72	3.22	3.53	2.83	2.74
Tetraconazole	3.05	13.21	2.29	1.97	1.82	1.76	1.92	2.21	1.95	2.15
Thiabendazole	2.4	12.90	1.79	1.47	1.63	1.47	1.45	1.51	1.26	1.48
Thiacloprid	2.44	12.92	1.21	1.66	2.06	2.03	1.22	1.26	1.43	1.97
Thiamethoxam	2.16	12.78	2.28	3.67	5.39	4.13	3.34	2.83	2.72	3.61
Tolclofos-methyl	2.35	12.87	3.01	5.17	4.77	4.75	4.13	4.87	6.37	6.14
Triadimefon	3.38	13.38	1.32	1.32	1.34	1.22	1.22	1.26	1.46	1.39
Triallate	2.38	12.89	1.28	2.50	2.19	2.43	3.62	3.26	1.79	1.73
Triazophos	3.03	13.21	1.66	1.27	1.14	1.12	1.49	1.3	1.28	1.27
Trifloxystrobin	2.98	13.18	2.79	1.94	1.49	1.64	2.4	2.43	1.92	1.77
Triticonazole	3.45	13.41	1.76	1.41	1.32	1.24	1.59	1.52	1.42	1.34
Vamidothion	3.05	13.21	1.4	2.45	3.01	3.62	1.76	1.66	2.17	2.5
Zoxamide	2.68	13.03	1.58	1.36	1.15	1.44	1.36	1.44	1.5	1.26

Table S5 Mean Concentration Prediction Error by Concentration Level in UT.

conc_level	solvent	matrices	barley	maize	oat	rice	rye	wheat
1	3.2	4.4	5.0	4.2	4.9	4.6	4.1	3.7
2	3.3	3.9	4.1	3.9	4.0	4.0	4.0	3.5
3	3.1	3.9	3.9	3.8	4.4	3.9	3.8	3.5
4	2.8	3.5	3.3	3.4	4.4	3.3	3.2	3.1
5	2.6	3.4	3.2	3.3	4.3	3.2	3.1	3.0
6	2.8	4.0	3.8	4.1	5.3	3.8	3.8	3.3

Table S6 Mean Concentration Prediction Error by Concentration Level in DTU.

conc_level	solvent	matrices	barley	maize	oat	rice	rye	wheat
1	2.3	2.4	2.2	2.5	2.3	2.6	2.3	2.4
2	2.2	2.5	2.6	2.6	2.4	2.5	2.3	2.4
3	2.6	3.0	3.4	2.8	3.2	2.9	2.9	2.7
4	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.2
5	2.0	2.3	2.4	2.4	2.2	2.1	2.2	2.3
6	2.1	2.5	2.7	2.7	2.4	2.2	2.5	2.6

Table S7 Matrix effect of all the test pesticides in matrices on QQQ (UT).

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
1-Naphtylacetamide	97.42	95.46	92.47	103.79	102.21	103.35
Benfuracarb	42.2	52.12	42.09	46.65	52.27	48.29
Acetamiprid	113.07	113.76	104.89	107.4	110	120.16
Acetochlor	86.37	84.83	91.05	81.87	82.28	88.26
Aldicarb-sulfone	87.85	104.21	95.06	85.92	88.43	103.21
Ametoctradin	98.36	97.59	94.22	102.48	95.65	101.57
Atrazine	105.37	100.95	101.54	104.87	101.87	100.83
Azoxystrobin	86.4	84.12	92.07	91.42	97.09	86.87
Boscalid	84.5	90.23	82.18	91.89	90.12	89.39
Bupirimate	99.74	95.69	95.31	98.03	97.34	98.06
Buprofezine	98.05	110.2	90.84	104.73	111.29	111.16
Cadusafos	94.96	107.19	76.11	92.25	97	103.39
Carbendazim	100.46	99.77	93.8	97.85	101.12	108.98
Carbosulfan	6.7	4.74	2.32	5.74	11.95	5.2
Carboxin	92.11	96.1	92.08	88.39	94.68	93.85
Chlorfenvinphos	98.72	97.66	101.15	106.53	108.63	104.85
Chloridazon	103.93	94.41	100.95	98.22	104.16	96.45
Chlorsulfuron	95.4	91.4	93.85	95.53	105.36	91.84
Clethodim	86.75	99.48	91.6	101.58	92.74	102.75
Clofentezine	109.11	92.31	86.33	104.47	87.64	107.88
Clomazone	97.11	101.99	89.56	102.89	105.68	104.04

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Cyprodinil	100.62	103.1	96.94	95.84	99.33	100.27
Desmedipham	99.81	90.21	91.02	110.83	98.63	82.27
Diazinon	92.54	101.18	85.77	99.52	102.98	93.72
Dichlorvos	92.37	95.22	94.07	99.51	88.55	96.82
Difenoconazole	105.49	98.36	96.77	100.24	97.92	102.9
Dimethachlor	99.91	98.78	101.22	104	103.59	100.53
Dimethomorph	97.32	107.7	100.14	108.1	109.08	101.72
Dodemorph	95.21	99.89	104.03	93.78	108.42	104.33
EPN	95.19	80.69	76.61	77.5	90.48	91.56
Ethion	74.75	82.47	74.9	81.47	92	86.83
Ethoprophos	99.55	95.05	87.02	110.26	103.73	95.87
Ethoxyquin	102.8	104.54	97.07	99.12	101.11	105.42
Fenamiphos	93.59	115.61	95.19	107.43	107.35	111
Fenamiphos-sulfoxide	103.41	113.45	107.72	109.11	114.07	118.02
Fenamiphos-sulphone	70.65	112.08	87.14	99.43	119.24	104.21
Fenarimol	97.07	97.84	84.45	94.2	95.3	104.62
Fenazaquin	85.9	82.14	69.57	91.77	82.82	91.56
Fenbuconazole	89.58	89.03	85.91	93.82	89.96	98.8
Fenoxycarb	59.26	275.99	91.87	108.92	112.1	115.78
Fenpropidin	97.75	92.73	93.43	98.05	98.96	99.52
Fenpropimorph	94.4	98.57	102.27	100.22	97.25	96.5
Fenthion	106.39	114.66	98.29	109.63	103.16	116.37

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Fenthion Sulfone	98.25	62.48	91.51	100.25	94.24	106.3
Fenthion -sulfoxid	92.42	108.66	103.31	103.6	99.2	115.27
Fluazifop-p-buthyl	88.2	93.26	90.3	98.88	96.72	92.47
Fluometuron	99.65	98.8	104.43	94.67	106.57	101.84
Fluopicolide	102.49	102.43	96.32	103.75	100.36	93.8
Fluopyram	101.45	101.64	89.4	95.08	100.52	100.93
Fluoxastrobin	94.59	102.37	95.68	99.96	105.44	90.7
Flurochloridone	94.79	92.46	86.9	100.35	146.21	96.55
Flurprimidole	102.44	114.72	93.24	100.93	107.38	100.29
Flusilazole	105.17	102.14	97.91	104.51	95.1	97
Flutriafol	95	114.22	109.3	93.92	95.47	101.06
Fluxapyroxad	142.91	88.37	82.45	92.06	97.41	83.66
Fosthiazate	104.14	124.8	106.77	114.68	98.98	122.99
Halosulfuron methyl	95.27	103.74	92.7	97.44	100.03	93.26
Heptenophos	100.14	101.03	98.13	102.4	115.13	109.8
Hexythiazox	83.73	91.39	51.94	93.33	90.23	94.91
imazalil	104.5	108.61	104.31	98.28	105.7	97.08
Imazamox	90.51	98.94	91.18	98.57	100.44	95.01
Imazaquin	113.23	113.53	108.22	116.61	115.65	110.44
Imidacloprid	105.12	103.13	94.15	97.78	104.95	101.25
Indoxacarb	92.68	102.68	91.58	101.22	99.69	99.61
Iprovalicarb	89.45	95	90.26	102.64	99.65	99.16

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Isoproturon	91.72	103.42	98.89	96.46	104.68	98.34
Isopyrazam	92.7	102.69	102.9	105.92	101.82	102.42
Linuron	98.68	100.13	94.55	96.71	100.7	99.82
Malaoxon	85.33	93.43	90.89	92.09	90.15	92.7
Malathion	86.04	82.31	105.34	88.99	105.66	92.78
Mecarbam	90.08	83.68	89.64	103.29	98.99	83.62
Mepanipyrim	100.57	91.23	92.62	97.3	94.79	100.4
Metaflumizone	97.44	97.96	83.97	97.88	99.17	110.57
Metalaxyl	103.41	115.13	99.81	108.77	103.25	107.57
Metazachlor	99.19	93.77	95.5	93.27	92.8	94.71
Metconazole	100.06	95.4	96.2	100.53	94.97	95.23
methamidophos	78.99	88.7	83.3	87.32	80.22	75.03
Methiocarb sulfoxide	107.76	108.74	95.36	112.48	114.66	113.13
Metosulam	103.61	109.69	103.14	104.35	108.11	104.56
Metrafenone	102.89	93.04	93.1	98.53	111.04	99.15
Metribuzin	101.67	103.35	98.19	99.98	99	101.42
Metsulfuron-methyl	104.86	102.99	97.95	89.47	105.56	97.98
Monolinuron	109.14	111.27	97.79	102.49	96.88	103.47
Myclobutanil	103.5	97.94	95.49	107.05	105.69	107.17
Napropamide	92.66	97.23	94.17	93.58	93.91	90.49
Nuarimol	93.8	93.49	81.19	98.15	99.45	97.83
Ofurace	107.2	90.46	92.31	96.04	107.36	97.42

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Omethoate	98.44	98.35	109.29	104.5	103.03	100.55
Oxadiargyl	97.28	84.09	96.43	87.13	85.13	75.87
Oxadixyl	105.79	119.04	95.57	98.12	92.8	91.39
Oxycarboxin	99.64	92.95	96.1	108.72	103.75	95.89
Oxydemeton-methyl	98.55	97.81	103.46	94.8	108.43	98.28
Oxyfluorfen	94.71	110.19	89.92	109.86	99.79	108.33
Paclobutrazol	94.41	99.48	84.33	98.05	103.91	96.72
Paraoxon-methyl	97.69	90.7	90.72	108.03	96.26	119.64
Penconazole	96.88	97.02	89.07	99.15	94.98	87.6
Pencycuron	104.14	99.53	85.2	100.63	98.13	105.59
Penflufen	93.91	93.21	103.39	105.56	105.04	91.05
Penthiopyrad	115.47	109.6	97.42	98.31	111.04	95.86
Pethoxamide	98.1	90.87	97.89	96.81	110.77	92.99
Phosalone	87.2	94.26	91.09	87.72	84.63	81.41
Phosphamidon	100.61	103.39	107.74	99.43	110.99	104.66
Phoxim	87.79	82.58	75.89	91.56	87.67	82.27
Pirimicarb	96.64	86.36	96.75	98.17	95.19	92.25
Pirimicarb-desmethyl	98.72	102.96	104.03	102.62	103.31	93.79
Pirimiphos-methyl	94.03	102.57	93.78	100.97	102.37	95.71
Prochloraz	110.92	103.2	100.2	100.29	97.24	105.41
Profenofos	90.54	88.11	83.13	103.51	94.31	105.74
Propamocarb	105.67	103.45	105.37	105.25	101.26	102.67

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Propiconazole	101.02	100.12	87.4	97.73	103.13	100.44
Propyzamide	103.79	99.02	93.22	104.53	99.08	86.11
Prosulfocarb	85.08	81.65	81.46	94.32	94.99	85.23
Prosulfuron	86.79	104.19	98.57	85.9	100.9	98.83
Pyraclostrobin	94.7	87.35	91.96	91.59	88.57	90.54
Pyrazophos	100.6	94.72	97.25	112.45	97.35	97.89
Pyridaben	50.31	52.09	46.59	60.91	51.61	57.2
Pyridaphenthion	99.27	105.5	100.19	100.45	107.43	102.6
Pyridate	38.07	26.8	13.42	55.42	66.23	46.07
Pyrimethanil	103.83	103.03	92.2	98.09	102.38	102
Pyriproxyfen	110.86	102.28	92.21	117.61	104.71	115.35
Pyroxsulam	95.24	105.64	108.94	97.91	109.97	98.7
Simazine	97.59	100.38	103.5	98.42	97.69	98.2
Spirodiclofen	39.38	42.36	46.84	53.93	52.69	49.77
Tebuconazole	86.58	93.33	91.77	101.22	93.38	89.11
Tebufenpyrad	107.77	114.34	91.82	105.81	111.47	118.31
TEPP	86.66	87.77	88.26	88.99	95.23	89.27
Terbuthylazine	95.1	94.24	94.26	97.22	92.69	88.06
Tetraconazole	113.19	98.06	91.17	117.06	97.92	113.44
Thiabendazole	89.68	93.05	84.98	93.95	93.39	93.81
Thiacloprid	92.58	106.91	95.99	94.57	102.49	97.32
Thiamethoxam	113.04	113.72	103.88	106.78	103.2	113.8

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Tolclofos-methyl	87.85	79.69	74.12	113.55	100.51	103.2
Triadimefon	84.06	80.12	74.89	81.17	89.83	90.14
Triallate	105.05	98.79	76.95	103.52	100.88	90.79
Triazophos	98.26	93.71	92.76	105.63	92.85	93.15
Trifloxystrobin	96.48	110.25	106.64	105.61	104.44	102.53
Triticonazole	83.46	72.34	85.71	62.25	89.89	98.6
Vamidothion	87.94	96.76	91.9	97.15	98.58	94.96
Zoxamide	101.47	108.43	100.31	110.97	104.51	107.99

Table S8 Matrix effect of all the test pesticides in matrices on QTOF (DTU).

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
1-Naphtylacetamide	66.4	67.29	82	74.87	68.43	88.23
Acetamiprid	64.55	61.11	73.43	66.65	64.58	82.12
Acetochlor	58.44	64.64	74.69	73.43	67.83	89.84
Aldicarb sulfone	44.09	39.55	60.53	53.75	44.86	72.11
Ametoctradin	63.33	66.09	75.31	71.17	69.22	85.68
Atrazine	68.72	64.86	69.36	74.36	68.44	87.47
Azoxystrobin	68.02	66.72	77.07	75.88	71.68	89.52
Boscalid	69.36	77.43	83.71	79.02	76.27	89.31
Bupirimate	33.19	64.01	74.64	70.51	68.03	86.44
Buprofezine	60.52	63.8	72.68	67.37	66.21	82.57
Cadusafos	68.25	64.89	77.04	70.55	66	85.44
Carbendazim	60.75	61.2	74.41	67.37	63.95	82.15
Carboxin	73.79	68.29	78.01	72.02	68.6	90.44
Chlorfenvinphos	63.31	68.37	77.2	75.3	70.15	91.37
Chloridazon	61.41	53.91	70.5	71.53	65	83.04
Chlorsulfuron	81.77	77.57	87.42	85.78	82.35	104.37
Clethodim	73	74.33	79.17	75.37	75.47	89.37
Clofentezine	70.72	71.08	72.97	74.38	69.6	89.95
Clomazone	61.4	69.87	76.2	70.63	70.27	87.3
Cyprodinil	64.3	65.34	73.38	69.83	67.86	89.87
Desmedipham	50.32	58.63	70.02	63.47	57.46	81.85
Diazinon	55.39	63.01	69.09	67.32	67.46	83.82
Dichlorvos	57.79	59.3	70.61	64.39	61.1	83.42

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Difenoconazole	70.96	73.84	77.91	74.25	72.97	86.13
Dimethachlor	58.85	64.34	75.78	69.52	68.89	85.88
Dimethomorph	71.99	72.46	83.05	74.4	71.65	87
Dodemorph	59.35	61.73	70.08	66.06	64.64	82.4
EPN	69.69	69.92	72.02	73.34	73.27	86.83
Ethion	62.43	63.64	58.24	74.39	70.57	90.02
Ethoprophos	69.31	65.27	76.41	73.69	68.06	91.96
Ethoxyquin	74.51	69.61	81.11	77.53	73.69	101.5
Fenamiphos	62.44	65.73	72.33	66.16	64.15	81.03
Fenamiphos-sulfoxide	62.29	62.43	74.1	70.62	68.98	85.55
Fenamiphos-sulphone	70.4	63.59	74.17	74.86	69.49	88
Fenarimol	74.14	80.95	82.36	83.12	83.12	96.41
Fenazaquin	67.16	67.45	80.47	75.07	71.74	91.25
Fenbuconazole	72.89	75.38	80.62	79.25	77.37	90.8
Fenoxycarb	77.67	74.18	83.89	82.87	75.43	95.88
Fenpropidin	68.6	68.39	78.5	76.86	73.26	94.47
Fenpropimorph	62.77	63.43	72.11	69.17	68.76	87.68
Fenthion	67.76	61.06	71.16	70.72	71.76	86.86
Fenthion -sulfone	62.66	67.98	77.25	70.22	68.45	85.97
Fenthion -sulfoxid	59.63	62.57	72.03	64.26	65.02	81.78
Fluazifop-p-buthyl	66.51	68	77.25	73.54	70.88	87.24
Fluometuron	71.34	69.28	70.92	77.38	71.07	91.09
Fluopicolide	75.91	73.67	81.49	82.26	79.74	91.92
Fluopyram	73.34	68.95	77.09	76.14	74.61	88.18
Fluoxastrobin	70.94	66.92	81.42	77.05	70.72	89.34
Flurochloridone	66.89	68.16	76.56	69.07	69.76	85.38
Flurprimidole	70.32	73.52	80.17	72.43	73.39	86.52
Flusilazole	80.26	67	79.33	82.83	84.34	95.73
Flutriafol	68.4	68.35	63.87	73.09	70.18	87.27
Fluxapyroxad	65.34	67.71	75.44	72.41	70.88	86.07
Fosthiazate	57.01	62.57	82.21	65.91	67.41	83.5
Halosulfuron methyl	61.48	65.03	76.86	67.75	66.85	84.52
Heptenophos	58.13	65.68	75.7	70.1	68.04	86.79
Hexythiazox	68.34	73.98	83.3	79.02	76.88	91.08
Imazamox	79.07	74.58	89.37	83.27	79.66	95.3
Imazaquin	67.57	67.49	78.01	73.37	72.32	91.07
Imidacloprid	77.86	86.63	99.06	87.25	91.44	97.29
Indoxacarb	72.44	72.73	78.5	76.66	71.13	84.61
Iprovalicarb	68.87	65.05	75.55	73.86	69.47	88.1
Isoproturon	64.32	60.48	71.08	68.55	64.51	82.12
Isopyrazam	70.61	64.66	77.52	69.43	68.04	85.74
Linuron	69.99	68.1	78.76	74.02	69.26	88.56
Malaoxon	69.56	62.12	73.52	71.3	68.3	85.5
Malathion	55.19	58.5	75.35	66.43	62.87	85.83
Mecarbam	52.25	62.41	79.11	70.27	68.5	84.03

Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Mepanipyrim	64.53	64.36	74.95	69.07	65.17	85.3
Metaflumizone	72.78	78.15	84.3	80.94	79.47	92.17
Metalaxyl	68.08	66.12	81.42	73.37	70.36	92.84
Metazachlor	81.6	67.31	77.73	80.17	79.96	89.81
Metconazole	66.94	71.65	79.58	76.21	74.19	88.38
Methiocarb sulfoxide	65.27	54.22	75.54	75.03	69.58	89.63
Metosulam	76.56	70.34	76.67	76.2	75.83	90.89
Metrafenone	71.3	64.73	77.1	74.6	69.86	85.16
Metribuzin	77.58	73.03	86.58	82.56	79.03	100.19
Metsulfuron-methyl	57.82	65.92	74.95	72.32	69.71	89.3
Monolinuron	79.93	72.36	75.47	84.17	78.95	95.55
Myclobutanil	81.69	76.82	82.64	86.02	81.76	96.8
Napropamide	68.98	66.52	77.66	74.82	70.41	90.89
Nuarimol	82.34	80.21	83.62	86.06	84.91	98.59
Ofurace	64.8	65.67	82.77	69.93	68.04	85.67
Omethoate	58.59	59.87	75.85	67.81	66.48	85.13
Oxadiargyl	66.58	65.5	71.52	68.08	66.99	82.54
Oxadixyl	63.85	61.28	70.63	73.03	62.78	88.92
Oxycarboxin	67.11	62.38	75.72	76.08	67.73	92.85
Oxydemeton-methyl	59.24	56.45	74.91	73.38	63.26	87.8
Oxyfluorfen	71.2	75.78	78.7	82.1	75.73	97.37
Paclobutrazol	48.03	52.58	77.12	73.09	71.28	85.61
Paraoxon-methyl	74.27	69.07	78.8	75.63	73.4	91.39
Penconazole	70.35	71.31	79.03	77.03	72.68	90.37
Pencycuron	65.41	60.39	72.27	70.07	65.91	82.78
Penflufen	67.69	65.88	77.17	73.88	70.6	89.43
Penthiopyrad	61.12	67.1	74.95	71.99	69.15	83.05
Pethoxamide	68.66	65.53	75.43	74.06	70.14	90.61
Phosalone	61.3	56.87	65.25	65.72	63.74	75.62
Phosphamidon	66.15	64.31	73.42	74.02	67.53	90.23
Phoxim	65.27	57.51	71.29	69.48	65.86	84.16
Pirimicarb	68.65	60.36	77.92	73.71	70.66	91.01
Pirimicarb-desmethyl	57.43	61.8	73.05	63.64	64.55	83.92
Pirimiphos-methyl	63.31	65.45	72.01	68.81	68	87.19
Prochloraz	69.96	73.59	80.25	73.42	73.93	85.58
Profenofos	60.67	69.22	77.29	73.89	71.54	88.27
Propamocarb	74.98	67.41	81.27	77.99	73.27	93.73
Propiconazole	71	68.77	77.77	73.85	72.47	86.32
Propyzamide	66.5	72.52	80.79	81.28	77.26	90.77
Prosulfocarb	61.98	67.27	76.7	74.16	72.39	88.37
Prosulfuron	71.86	68.55	81.79	78.02	75.9	91.62
Pyraclostrobin	68.63	67.21	77.93	71.56	68.76	87.14
Pyrazophos	62.32	74.29	75.94	77.15	75.11	85.67
Pyridaben	64.14	69.71	80.13	72.9	69.55	85.41
Pyridaphenthion	71.98	66.39	83.72	77.34	69.11	92.44

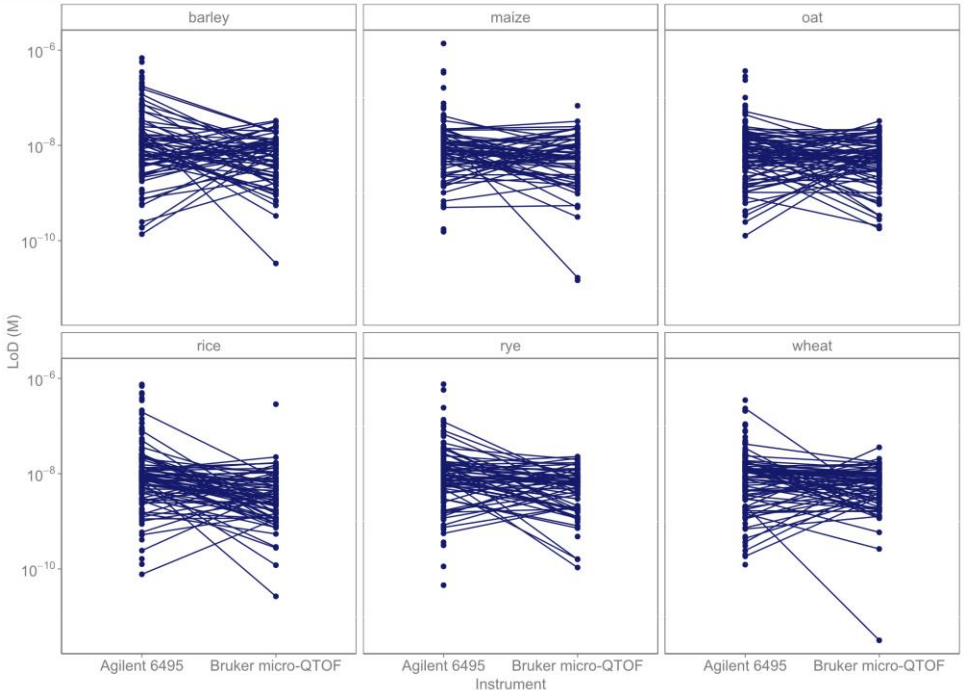
Compounds	barley%	maize%	oat%	rye%	wheat%	rice%
Pyridate	63.71	71.23	80.97	73.16	69.51	90.36
Pyrimethanil	68.2	67.68	78.1	75.66	71.55	95.97
Pyriproxyfen	44.22	66.36	77.82	72.49	69.11	89.69
Pyroxsulam	63.67	71.41	77.15	76.13	71.21	92.15
Simazine	57.87	61.32	72.05	65.61	65.93	85.43
Spirodiclofen	70.21	65.88	79.31	76.64	69.54	91.05
Tebuconazole	70.67	71.94	81.87	77.25	75.24	92.53
Tebufenpyrad	68.18	76.17	81.72	80.82	77.66	90.54
TEPP	55.37	59.88	69.11	64.1	60.99	81.55
Terbuthylazine	65.19	66.56	79.21	70.53	69.29	88.77
Tetraconazole	74.54	75.23	81.76	81.25	77.84	93.44
Thiabendazole	63.05	59.98	73.44	69.19	63.76	81.88
Thiacloprid	58.64	61.65	71.77	69.79	67.93	86.1
Thiamethoxam	60.73	60.81	82.65	80.86	73.38	87.91
Tolclofos-methyl	65.11	68.25	78.16	69.53	77.41	87.41
Triadimefon	81.86	78.01	84.48	86.06	81.5	98.31
Triallate	62.05	62.92	74.58	77.3	71.3	86.92
Triazophos	66.42	67.63	75.81	70.61	68.77	85.12
Trifloxystrobin	61.18	66.03	74.24	66.34	67.23	84.48
Triticonazole	69.89	67.79	75.82	73.6	70.13	86.04
Vamidothion	52.86	53.83	69.4	63.71	61.36	81.45
Zoxamide	74.01	65.98	81.93	76.46	72.48	86.85

Table S9 LoD and LoQ values for studied pesticides. All valueas are in mol/L.

Available at:

https://www.sciencedirect.com/science/article/pii/S0308814620303228?via%3Dihub

Figure S1 The comparison of LoD values for UT and DTU.



Appendix IV

IV. Other supplementary information relevant to 3.scientific research (240)

Table S1. Compounds information used for ion source optimization and eluent composition optimization. The compounds detected in each method were marked as "√" (Eluent optimization, ESI and APCI ion source optimization in flow injection mode)

Compounds	Eluent Optimiz	ESI infusion	APCI infusion	log*P*	MV	Formula	M+H	RT
Pesticides								
1-Naphtylacetamide	√	√	√	2.24	157.16	$C_{12}H_{11}NO$	186.0913	8.4
Acetamiprid	√	√	√	1.06	189.65	$C_{10}H_{11}ClN_4$	223.0745	7.4
Acetochlor	√			2.99	240.9	$C_{14}H_{20}ClNO_2$	270.1255	14.2
Aldicarb	√			1.36	174.8	$C_7H_{14}N_2O_2S$	191.0849	
Ametoctradin	√			4.56	237.05	$C_{15}H_{25}N_5$	276.2183	13.3
Atrazine	√	√	√	2.66	169.86	$C_8H_{14}ClN_5$	216.1010	10.6
Azinphos-ethyl	√	√		3.58	241.2	$C_{12}H_{16}N_3O_3PS_2$	346.0443	14.5
Benfuracarb	√	√	√	4.29	350	$C_{20}H_{30}N_2O_5S$	411.1948	17.3
Benfluralin (Benefin)		√	√	5.19	250.65	$C_{13}H_{16}F_3N_3O_4$	336.1166	
Bitertanol	√	√		3.92	295.7	$C_{20}H_{23}N_3O_2$	338.1863	14.0
Bromuconazole	√	√	√	3.32	219.65	$C_{13}H_{12}BrCl_2N_3O$	375.9614	12.9
Bromophos-ethyl			√	5.73	247.78	$C_{10}H_{12}BrCl_2O_3PS$	392.8878	
Bupirimate	√	√		2.64	262.74	$C_{13}H_{24}N_4O_3S$	317.1642	13.8
Buprofezine	√	√		3.85	272.28	$C_{16}H_{23}N_3OS$	306.1635	17.1
Carvone	√	√	√	2.8	159.76	$C_{10}H_{14}O$	151.1117	
Cadusafos	√	√		3.83	252.8	$C_{10}H_{23}O_2PS_2$	271.0950	15.5
Carbendazim	√	√	√	1.62	134.54	$C_9H_9N_3O_2$	192.0767	4.6
Carbosulfan	√	√	√	5.2	346.14	$C_{20}H_{32}N_2O_3S$	381.2206	20.8
Chlorfenapyr		√	√	4.76	266.21	$C_{15}H_{11}BrClF_3N_2O$	406.9768	
Chloridazon	√	√	√	1.11	177.32	$C_{10}H_8ClN_3O$	222.0429	6.8
Chlorsulfuron	√	√	√	1.92	232.89	$C_{12}H_{12}ClN_5O_4S$	358.0371	10.3
Clethodim	√	√	√	3.76	312.47	$C_{17}H_{26}ClNO_3S$	360.1395	17.0
Clofentezine	√			3.56	216.8	$C_{14}H_8Cl_2N_4$	303.0199	15.8
Clomazone	√	√	√	2.36	195.99	$C_{12}H_{14}ClNO_2$	240.0786	11.9
Cymoxanil	√			0.73	155.53	$C_7H_{10}N4O_3$	199.0826	
Cyazofamid**	√			1.33	234.98	$C_{13}H_{13}ClN_4O_2S$	325.0520	15.0
Cyflumetofen	√	√	√	5.72	368.48	$C_{24}H_{24}F_3NO_4$	448.1730	13.1

DMPF	√	√		2.65	171.76	$C_{10}H_{14}N_2$	163.1230	4.9
Dazomet	√	√	√	0.62	125.34	$C_5H_{10}N_2S_2$	163.0358	11.4
Desmedipham	√			3.45	230.35	$C_{16}H_{16}N_2O_4$	301.1183	12.6
Dichlofluanid	√			3.53	214.45	$C_9H_{11}Cl_2FN_2O_2S_2$	332.9696	15.0
Dimethachlor	√	√		2.63	224.36	$C_{13}H_{18}ClNO_2$	256.1099	11.7
Dimethomorph	√	√	√	3.31	315.07	$C_{21}H_{22}ClNO_4$	388.1310	12.3
Dinoterb		√		3.52	178.23	$C_{10}H_{12}N_2O_5$	241.0819	
Diphenylamine		√	√	3.12	155.47	$C_{12}H_{11}N$	170.0964	
Ditalimphos	√			3.26	216.49	$C_{12}H_{14}NO_4PS$	300.0454	14.2
Dodemorph	√	√		2.63	224.36	$C_{18}H_{35}NO$	282.2791	9.7
EPN	√			4.06	241.13	$C_{14}H_{14}NO_4PS$	324.0454	16.6
Ethiofencarb/Methiocarb	√	√	√	2.39	199.17	$C_{11}H_{15}NO_2S$	226.0896	12.4
Ethoprophos	√	√		3.26	219.04	$C_8H_{19}O_2PS_2$	243.0637	13.3
Ethoxyquin	√	√		3.5	222.5	$C_{14}H_{19}NO$	218.1539	6.0
Fenamiphos	√			3.22	264.72	$C_{13}H_{22}NO_3PS$	304.1131	13.0
Fenamiphos-sulfoxide	√			1.68	261.05	$C_{13}H_{22}NO_4PS$	320.1080	8.8
Fenamiphos-sulphone	√			1.67	277.22	$C_{13}H_{22}NO_5PS$	336.1029	9.9
Fenarimol	√	√	√	3	240.64	$C_{17}H_{12}Cl_2N_2O$	331.0399	12.9
Fenazaquin	√			5.29	277.9	$C_{20}H_{22}N_2O$	307.1805	18.0
Fenoxycarb	√			3.91	262.35	$C_{17}H_{19}NO_4$	302.1387	14.3
Fenpropathrin	√			5.82	309.61	$C_{22}H_{23}NO_3$	350.1751	19.1
Fenthion	√			3.8	221.87	$C_{10}H_{15}O_3PS_2$	279.0273	15.4
Fenthion -sulfoxid	√			2.21	218.27	$C_{10}H_{15}O_4PS_2$	295.0222	10.1
Fenthion oxon-sulfon	√			0.26	224.87	$C_{10}H_{15}O_6PS$	295.0400	7.9
Fenthion oxon-sulfoxide	√	√		0.48	210.13	$C_{10}H_{15}O_5PS$	279.0451	6.9
Fluazifop-p-buthyl	√	√	√	4.14	313.66	$C_{19}H_{20}F_3NO_4$	384.1417	17.7
Fluazinam	√	√	√	5.16	263.3	$C_{13}H_4Cl_2F_6N_4O_4$	464.9586	17.5
Flufenoxuron	√	√	√	5.88	319.42	$C_{21}H_{11}ClF_6N_2O_3$	489.0435	17.9
Fluometuron	√	√	√	2.36	179.79	$C_{10}H_{11}F_3N_2O$	233.0896	10.5
Fluopicolide	√	√	√	4.38	253.07	$C_{14}H_8Cl_3F_3N_2O$	382.9727	13.8
Fluopyram	√	√	√	4.23	279.18	$C_{16}H_{11}ClF_6N_2O$	397.0537	13.8
Fluoxastrobin	√	√	√	4.28	318.94	$C_{21}H_{16}ClFN_4O_5$	459.0866	14.5
Flupyrsulfuron-methyl-Natrium	√	√		1.93	300.07	$C_{15}H_{13}F_3N_5NaO_7S$	488.0458	12.6
Flurochloridone	√			3.07	213.04	$C_{12}H_{10}Cl_2F_3NO$	312.0164	14.3
Flurprimidole	√			2.58	243.94	$C_{15}H_{15}F_3N_2O_2$	313.1158	12.4
Fluxapyroxad	√	√	√	3.5	267.86	$C_{18}H_{12}F_5N_3O$	382.0973	13.4

Fosthiazate	√	√		1.38	223.39	$C_9H_{18}NO_3PS_2$	284.0538	10.6
Halosulfuron methyl	√			1.2	259.29	$C_{13}H_{15}ClN_6O_7S$	435.0484	13.1
Heptenophos	√	√		2.42	183.96	$C_9H_{12}ClO_4P$	251.0234	11.3
Hexythiazox	√	√	√	3.71	267.58	$C_{17}H_{21}ClN_2O_2S$	353.1085	17.9
Imazamox	√			0.61	231.44	$C_{15}H_{19}N_3O_4$	306.1448	7.0
Imazaquin	√			1.9	229.45	$C_{17}H_{17}N_3O_3$	312.1343	9.2
Indoxacarb	√	√	√	3.7	343.31	$C_{22}H_{17}ClF_3N_3O_7$	528.0780	16.6
Iodosulfuron-methyl-Natrium	√			2.4	280.91	$C_{14}H_{13}IN_5NaO_6S$	529.9602	11.9
Iprovalicarb	√			3.29	307.2	$C_{18}H_{28}N_2O_3$	321.2173	13.1
Isofenphos-methyl	√			3.84	281.81	$C_{14}H_{22}NO_4PS$	332.1080	15.8
Isopyrazam	√			4.44	261.09	$C_{20}H_{23}F_2N_3O$	360.1882	15.9
Jodfenphos			√	5.07	220.66	$C_8H_8Cl_2IO_3PS$	412.8426	
Lufenuron			√	5.99	313.24	$C_{17}H_8Cl_2F_8N_2O_3$	510.9857	17.3
Mecarbam	√			2.84	260.17	$C_{10}H_{20}NO_5PS_2$	330.0593	14.8
Mepanipyrim	√	√	√	3.37	191.89	$C_{14}H_{13}N_3$	224.1182	13.7
Metaflumizone	√	√	√	5.72	375.71	$C_{24}H_{16}F_6N_4O_2$	507.1250	17.4
Metalaxyl	√	√		1.88	249.85	$C_{15}H_{21}NO_4$	280.1543	10.9
Metazachlor	√			2.28	232.15	$C_{14}H_{16}ClN_3O$	278.1055	11.6
Methidathion	√			1.83	188.56	$C_{14}H_{16}ClN_3O$	278.1055	12.5
Methiocarb sulfone	√			1.38	210.21	$C_{11}H_{15}NO_4S$	258.0794	8.0
Methiocarb sulfoxide	√	√		1.29	192.52	$C_{11}H_{15}NO_3S$	242.0845	6.6
Methoxyfenozide	√	√		3.62	335.35	$C_{22}H_{28}N_2O_3$	369.2173	14.0
Metosulam	√	√	√	3.03	251.33	$C_{14}H_{13}Cl_2N_5O_4S$	418.0138	10.9
Metrafenone	√	√	√	4.04	312.29	$C_{19}H_{21}BrO_5$	409.0645	16.5
Metsulfuron-methyl	√			1.52	258.84	$C_{14}H_{15}N_5O_6S$	382.0816	9.9
Mevinphos	√			0.43	184.22	$C_7H_{13}O_6P$	225.0522	7.7
Monocrotophos	√			0.38	186.67	$C_7H_{14}NO_5P$	224.0682	5.4
Monolinuron	√	√	√	2.16	164.59	$C_9H_{11}ClN_2O_2$	215.0582	10.8
Myclobutanil	√	√		3.19	247.89	$C_{15}H_{17}ClN_4$	289.1214	13.1
Napropamide	√	√		3.44	251.34	$C_{17}H_{21}NO_2$	272.1645	13.7
Oryzalin		√	√	3.44	251.34	$C_{12}H_{18}N_4O_6S$	347.1020	
Nuarimol	√			2.5	232.9	$C_{17}H_{12}ClFN2O$	315.0695	11.8
Ofurace	√			1.29	217.29	$C_{14}H_{16}ClNO_3$	282.0891	10.8
Oxadiargyl	√			3.99	263.36	$C_{15}H_{14}Cl_2N_2O_3$	341.0454	16.0
Oxadixyl	√	√		0.98	219.84	$C_{14}H_{18}N_2O_4$	279.1339	9.1
Oxycarboxin	√			0.71	191.93	$C_{12}H_{13}NO_4S$	268.0638	8.2
Oxyfluorfen	√			4.48	257.83	$C_{15}H_{11}ClF_3NO_4$	362.0401	17.5

Paraoxon-methyl	√	√		1.26	178.66	$C_8H_{10}NO_6P$	248.0318	9.1
Parathion-methyl	√			2.94	186.52	$C_8H_{10}NO_5PS$	264.0090	13.5
Pencycuron	√	√	√	4.34	268.65	$C_{19}H_{21}ClN_2O$	329.1415	16.0
Penflufen	√			4.03	281.38	$C_{18}H_{24}FN_3O$	318.1976	14.3
Penthiopyrad	√			3.64	275.74	$C_{16}H_{20}F_3N_3OS$	360.1352	15.2
Pethoxamide	√			3.4	271.24	$C_{16}H_{22}ClNO_2$	296.1412	14.0
Phenthoate	√			3.99	251.22	$C_{12}H_{17}O_4PS_2$	321.0379	15.5
Phosalone	√		√	4.43	254.88	$C_{12}H_{15}ClNO_4PS_2$	367.9941	16.2
Phosmet	√			4.43	254.88	$C_{11}H_{12}NO_4PS_2$	318.0018	13.0
Phoxim	√			3.98	245.7	$C_{12}H_{15}N_2O_3PS$	299.0614	16.2
Profenofos	√	√	√	4.4	252.08	$C_{11}H_{15}BrClO_3PS$	372.9424	16.5
Propamocarb	√	√	√	1.15	196.65	$C_9H_{20}N_2O_2$	189.1597	2.8
Propyzamide	√	√		3.42	203.52	$C_{12}H_{11}Cl_2NO$	256.0290	13.4
Prosulfocarb	√	√		4.1	239.63	$C_{14}H_{21}NOS$	252.1417	16.8
Prosulfuron	√	√		2.79	286.66	$C_{15}H_{16}F_3N_5O_4S$	420.0948	12.7
Prothiofos	√		√	5.46	255.71	$C_{11}H_{15}Cl_2O_2PS_2$	344.9701	
Pyrazophos			√	3.6	270.08	$C_{14}H_{20}N_3O_5PS$	374.0934	15.6
Pyridaben	√	√	√	5.44	323.53	$C_{19}H_{25}ClN_2OS$	365.1449	19.2
Pyridaphenthion	√	√	√	2.92	260.64	$C_{14}H_{17}N_2O_4PS$	341.0719	13.5
Pyridate	√	√	√	6.76	320.26	$C_{19}H_{23}ClN_2O_2S$	379.1241	20.5
Pyriproxyfen	√	√		4.48	279.03	$C_{20}H_{19}NO_3$	322.1438	17.7
Pyroxsulam	√	√	√	2	257.9	$C_{14}H_{13}F_3N_6O_5S$	435.0693	10.1
Simazine	√	√	√	2.29	152.98	$C_7H_{12}ClN_5$	202.0854	9.0
Spinosad A	√	√		5.01	627.29	$C_{41}H_{65}NO_{10}$	732.4681	11.9
Spinosad D	√			5.01	627.29	$C_{42}H_{67}NO_{10}$	746.4838	12.4
Spirodiclofen	√			5.83	319.61	$C_{21}H_24Cl_2O_4$	411.1124	19.5
Spiromesifen	√	√		5.25	328.88	$C_{23}H_{30}O_4$	371.2217	19.2
Tau-fluvalinate		√	√	6.58	383.05	$C_{26}H_{22}ClF_3N_2O_3$	503.1344	
Sulcotrione			√	1.19	217.68	$C_{14}H_{13}ClO_5S$	329.0245	
TEPP	√			0.73	240.7	$C_8H_{20}O_7P_2$	291.0757	8.7
Tebufenpyrad	√	√	√	4.41	293.16	$C_{18}H_{24}ClN_3O$	334.1681	16.7
Teflubenzuron	√	√	√	4.41	293.16	$C_{14}H_6Cl_2F_4N_2O_2$	380.9815	16.1
Terbuthylazine	√	√	√	2.99	186.04	$C_9H_{16}ClN_5$	230.1167	12.4
Tetraconazole	√	√	√	3.71	247.19	$C_{13}H_{11}Cl_2F_4N_3O$	372.0288	13.6
Thiabendazole	√			2.39	143.08	$C_{10}H_7N_3S$	202.0433	5.2
Thiophanate-methyl		√	√	1.26	228.95	$C_{12}H_{14}N_4O_4S_2$	343.0529	
Thiometon	√			3.2	200.25	$C_6H_{15}O_2PS_3$	247.0045	15.5

Tolclofos-methyl	√			4.4	214.86	$C_{12}H_{14}N_4O_4S_2$	343.0529	16.1
Tolylfluanid	√			3.69	230.73	$C_{10}H_{13}Cl_2FN_2O_2S_2$	346.9852	15.9
Triallate	√			4.93	240.07	$C_{10}H_{16}Cl_3NOS$	304.0091	18.5
Trichlorfon	√	√		0.73	163.52	$C_4H_8Cl_3O_4P$	256.9298	6.3
Triclopyr	√			2.68	153.64	$C_7H_4Cl_3NO_3$	255.9329	11.1
Triflumuron	√	√	√	4.52	243.09	$C_{15}H_{10}ClF_3N_2O_3$	359.0405	15.2
Tritosulfuron	√	√	√	3.11	271.69	$C_{13}H_9F_6N_5O_4S$	446.0352	13.1
Vamidothion	√	√		0.84	231.71	$C_8H_{18}NO_4PS_2$	288.0488	6.6
Zoxamide	√			4.21	261.04	$C_{14}H_{16}Cl_3NO_2$	336.0319	15.5
Mycotoxins	√							
Beauvericin	√			5.69	695.82	$C_{45}H_{57}N_3O_9$	784.4168	19.3
15-acetyldeoxynivalenol	√			-0.17	237.49	$C_{17}H_{22}O_7$	339.1438	7.0
Alternariol monomethyl ether	√			3.82	190.75	$C_{15}H_{12}O_5$	273.0757	12.6
Zearalenone	√			3.96	272.35	$C_{18}H_{22}O_5$	319.1540	12.7
Fumonisins B1	√			1.59	575.19	$C_{34}H_{59}NO_{15}$	722.3957	9.1
Fumonisins B2	√			3.03	577.31	$C_{34}H_{59}NO_{14}$	706.4008	10.3
Diacetoxyscirpenol (DAS)	√			1.4	278.57	$C_{19}H_{26}O_7$	367.1751	9.3
Ochratoxin A	√			4.42	283.24	$C_{20}H_{18}ClNO_6$	404.0895	12.7
Aflatoxin B1	√			1.08	199.58	$C_{17}H_{12}O_6$	313.0707	8.9
Aflatoxin B2	√			0.92	206.26	$C_{17}H_{14}O_6$	315.0863	9.4
Aflatoxin G1	√			0.73	205.66	$C_{17}H_{12}O_7$	329.0656	8.4
Aflatoxin G2	√			0.66	212.33	$C_{17}H_{14}O_7$	331.0812	8.9
Enniatin A	√			4.67	666.77	$C_{36}H_{63}N_3O_9$	682.4637	20.5
Enniatin A1	√			4.21	650.26	$C_{35}H_{61}N_3O_9$	668.4481	19.8
Enniatin B	√			3.42	617.25	$C_{33}H_{57}N_3O_9$	640.4168	18.4
Enniatin B1	√			3.81	633.75	$C_{34}H_{59}N_3O_9$	654.4324	19.2
Sterigmatocysin	√			2.64	214.75	$C_{18}H_{12}O_6$	325.0707	13.1

Note. MV is the molecular volume. Eluent Optimiz means the compounds detected in eluents optimization part in 3.1.1.1. ESI infusion means the compounds detected in flow injection model in 3.1.1.1, APCI infusion means the compounds detected in flow injection model in 3.1.1.1.