



HAL
open science

Développement d'une approche non-ciblée par empreinte pour caractériser la qualité sanitaire chimique de matrices agro-alimentaires complexes

Grégoire Delaporte

► **To cite this version:**

Grégoire Delaporte. Développement d'une approche non-ciblée par empreinte pour caractériser la qualité sanitaire chimique de matrices agro-alimentaires complexes. Chimie analytique. Université Paris Saclay (COMUE), 2018. Français. NNT : 2018SACLA034 . tel-02304147

HAL Id: tel-02304147

<https://theses.hal.science/tel-02304147>

Submitted on 3 Oct 2019

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

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

Développement d'une approche non-ciblée par empreinte pour caractériser la qualité sanitaire chimique de matrices agro-alimentaires complexes

Thèse de doctorat de l'Université Paris-Saclay préparée à AgroParisTech
(Institut des sciences et industries du vivant et de l'environnement)

École doctorale n°581
Agriculture, alimentation, biologie, environnement et santé (ABIES)
Spécialité de doctorat : Chimie analytique

Thèse présentée et soutenue à Paris, le 18/12/2018, par

Grégoire Delaporte

Composition du Jury :

Carlos Afonso Professeur des Universités, Université de Rouen (UMR COBRA)	Président du Jury
Frédérique Courant Maître de Conférences HDR, Université de Montpellier (UMR HydroSciences)	Rapporteur
Laurent Debrauwer Ingénieur de Recherches HDR, INRA (Toxalim)	Rapporteur
Christophe Junot Directeur de recherche, CEA Saclay (Institut des sciences du vivant Frédéric Joliot)	Examineur
Valérie Camel Professeur, AgroParisTech (UMR GENIAL)	Directrice de thèse
Mathieu Cladière Maître de Conférences, AgroParisTech (UMR GENIAL)	Invité – Co-encadrant

REMERCIEMENTS

J'aimerais tout d'abord remercier ma directrice de thèse la Pr Valérie Camel et mon co-encadrant le Dr Mathieu Cladière pour leur encadrement et leur aide tout au long de cette thèse, depuis le concours de l'école doctorale jusqu'à la soutenance. Votre gentillesse, les échanges constructifs et francs que j'ai pu avoir avec vous, ainsi que la liberté que vous m'avez laissée dans la réalisation de mes travaux et mon organisation ont été essentiels pour moi lors de ce doctorat.

J'aimerais remercier toutes les personnes qui ont contribué à ces travaux. Merci en particulier à Dr Delphine Jouan-Rimbaud Bouveresse (grâce à toi j'ai pu dompter Matlab en un temps record !), Even Le Roux (tes bons conseils, ta disponibilité et ton implication à mes côtés m'ont été précieuses lors de l'utilisation de la spectrométrie de masse), et enfin Brigitte Polet et Pascale Lieben pour leur aide et leur disponibilité lors de l'utilisation de l'Orbitrap. Un grand merci également à Stéphane Besançon pour son aide au quotidien au laboratoire. Je voudrais remercier d'une manière générale l'ensemble des personnels de l'ex-équipe IAQA pour leur gentillesse et leur accueil, et ainsi que ceux des équipes CALIPRO et GMPA pour leur accueil lorsque que je venais faire des analyses.

Je voudrais exprimer ma reconnaissance envers les membres de mon comité de thèse, Yann Guitton et Christophe Junot pour les discussions, toujours extrêmement constructives et bienveillantes que j'ai pu avoir avec eux ainsi que pour leurs précieux conseils.

Merci aux membres du jury d'avoir accepté d'évaluer mon travail, et notamment les rapporteurs de cette thèse, le Dr Laurent Debrauwer et la Dr Frédérique Courant, ainsi que les examinateurs, le Pr Carlos Afonso et le Dr Christophe Junot.

Merci aux membres de l'école doctorale ABIES, Dr. Irina Vassilleva, Dr. Alexandre Pery et Sylvie Ponsonnet et Corinne Fiers pour leur gentillesse, leur disponibilité tout au long de cette thèse.

Un grand merci aux doctorants et stagiaires du laboratoire de chimie analytique et de l'équipe CALIPRO pour avoir fait de ce doctorat un moment riche en raclettes, en fondues (dédicace Dr. Miguel), en bulgogi (merci Jeehuyn !), en amitiés (Aurélia, Thibault, Pierre) planches de charcuterie et tapas (merci au café La belle mine, notre fournisseur officiel), et plus généralement en bonne humeur.

Enfin, j'aimerais remercier ma famille pour le soutien et les encouragements qu'ils m'ont apporté tout au long de mes études d'ingénieur puis lors de mon doctorat, ainsi que Constance qui m'a soutenu tout au long de cette thèse.

VALORISATION DES TRAVAUX DE THÈSE

Publications dans des revues à comité de lecture

Mathieu Cladière, **Grégoire Delaporte**, Even Le Roux, Valérie Camel (2018). Multi-class analysis for simultaneous determination of pesticides, mycotoxins, process-induced toxicants and packaging contaminants in tea. *Food Chemistry*, 242:113–121. doi: 10.1016/j.foodchem.2017.08.108

Grégoire Delaporte, Mathieu Cladière, Delphine Jouan-Rimbaud Bouveresse, Valérie Camel (2019). Untargeted food contaminant detection using UHPLC-HRMS combined with multivariate analysis: feasibility study on tea. *Food Chemistry*, 277:54-62. doi:10.1016/j.foodchem.2018.10.089.

Grégoire Delaporte, Mathieu Cladière, Valérie Camel (2019). Untargeted food chemical safety assessment: a proof-of-concept on two analytical platforms and contamination scenarios of tea. *Food Control*, 98:510-519. doi:10.1016/j.foodcont.2018.12.004

Grégoire Delaporte, Mathieu Cladière, Valérie Camel (2019). Missing value imputation and data cleaning in untargeted food chemical safety assessment by LC-HRMS. *Chemometrics and Intelligent Laboratory Systems* (in press), doi:10.1016/j.chemolab.2019.03.005

Communication en congrès internationaux

Grégoire Delaporte, Mathieu Cladière, Delphine Jouan-Rimbaud Bouveresse, Valérie Camel (2017). Towards a foodomics approach for food chemical safety assessment. *8th International Symposium on Recent Advances in Food Analysis (RAFA)*, Prague, République Tchèque. **Communication orale** ayant bénéficié d'une bourse du congrès.

Grégoire Delaporte, Mathieu Cladière, Valérie Camel (2018). Towards untargeted chemical food safety assessment. *XXII International Mass Spectrometry Conference (IMSC)*, Florence, Italie. **Communication par affiche** ayant obtenu une bourse du Réseau Français de Métabolomique et Fluxomique (RFMF).

Communication en congrès nationaux

Grégoire Delaporte, Mathieu Cladière, Valérie Camel (2016). Simultaneous quantification of multi-class contaminants (pesticides, mycotoxins, neoformed and migrants from packaging) in tea by UHPLC-HRMS, *33^{ème} Journées Françaises de Spectrométrie de Masse (JFSM)*, Bordeaux. **Communication par affiche**.

Grégoire Delaporte, Mathieu Cladière, Delphine Jouan-Rimbaud Bouveresse, Valérie Camel (2017). Towards a foodomics approach for food chemical safety assessment, *Spectrométrie de Masse, Métabolomique et Analyse Protéomique (SMMAP) 2017*, Paris. **Communication par affiche.**

Grégoire Delaporte, Mathieu Cladière, Delphine Jouan-Rimbaud Bouveresse, Valérie Camel (2018). Analytical and data treatment strategies from metabolomics for untargeted contaminants detection in complex food products, *Chimiométrie 2018*, Paris. **Communication orale.**

Jeux de données

Valérie Camel, Mathieu Cladière, **Grégoire Delaporte** (2018). Untargeted food contaminant detection using UHPLC-HRMS combined with multivariate analysis: Feasibility study on tea (Dataset 1 - Development), *Metabolights repository* (MTBLS752),
<https://www.ebi.ac.uk/metabolights/MTBLS752>

Valérie Camel, Mathieu Cladière, **Grégoire Delaporte** (2018). Untargeted food contaminant detection using UHPLC-HRMS combined with multivariate analysis: Feasibility study on tea (Dataset 2 - Validation), *Metabolights repository* (MTBLS754),
<https://www.ebi.ac.uk/metabolights/MTBLS754>

Valérie Camel, Mathieu Cladière, **Grégoire Delaporte** (2019). Untargeted food chemical safety assessment: A proof-of-concept on two analytical platforms and contamination scenarios of tea (Green tea study), *Metabolights repository* (MTBLS771),
<https://www.ebi.ac.uk/metabolights/MTBLS771>

Valérie Camel, Mathieu Cladière, **Grégoire Delaporte** (2019). Untargeted food chemical safety assessment: A proof-of-concept on two analytical platforms and contamination scenarios of tea (Black tea study), *Metabolights repository* (MTBLS772),
<https://www.ebi.ac.uk/metabolights/MTBLS772>

Divers

Organisation de l'édition 2017 du Forum Doc' Avenir de l'école doctorale ABIES

1^{er} prix du concours "Mon Projet Professionnel en 180 secondes" 2017 de l'école doctorale ABIES

TABLE DES MATIÈRES

Remerciements	i
Valorisation des travaux de thèse	iii
Table des matières	v
Liste des figures.....	ix
Liste des tableaux	x
Liste des abréviations	xiii
Introduction générale.....	1
Chapitre 1 Contexte, état de l’art et méthodologie.....	3
1. La qualité sanitaire chimique d’un aliment	5
1.1. Définitions et enjeux.....	5
1.2. Contexte réglementaire et outils de gestion du risque européens.....	6
1.3. Les différentes classes de contaminants	8
1.4. La qualité sanitaire d’un produit : l’exemple du thé.....	17
2. L’évaluation de la qualité sanitaire chimique des aliments.....	23
2.1. Aperçu des composés d’intérêt.....	24
2.2. Approches ciblées.....	26
2.3. L’analyse non ciblée des contaminants alimentaires.....	46
3. Conclusion générale de l’étude bibliographique	58
4. Méthodologie des travaux de thèse	60
5. Références	62
Chapitre 2 Développement de l’approche ciblée multi-classes.....	73
1. Introduction et résumé de l’article.....	75
2. Article.....	76
2.1. Introduction	77
2.2. Materials and methods.....	78

2.3.	Results and discussion.....	84
2.4.	Conclusion.....	91
2.5.	Funding.....	91
2.6.	References	91
2.7.	Supplementary materials	94
3.	Conclusion.....	96
Chapitre 3 Développement de l'approche non-ciblée		97
1.	Introduction et résumé de l'article.....	99
2.	Article.....	100
2.1.	Introduction	101
2.2.	Material and methods	103
2.3.	Results and discussion.....	109
2.4.	Conclusion.....	115
2.5.	Funding.....	116
2.6.	Acknowledgements	116
2.7.	References	117
2.8.	Supplementary materials	119
3.	Conclusion.....	123
Chapitre 4 Étude méthodologique du prétraitement des données		125
1.	Introduction et résumé de l'article.....	127
2.	Article.....	128
2.1.	Introduction	128
2.2.	Material and methods	131
2.3.	Results and discussion.....	137
2.4.	Conclusion.....	145
2.5.	Funding.....	146

2.6.	References	146
2.7.	Supplementary materials	148
3.	Conclusion.....	155
Chapitre 5	Validation et application de la preuve de concept.....	157
1.	Introduction et résumé de l'article.....	159
2.	Article.....	160
2.1.	Introduction	160
2.2.	Material & methods.....	162
2.3.	Results & discussion.....	168
2.4.	Conclusion.....	178
2.5.	Funding.....	179
2.6.	Acknowledgements	179
2.7.	References	179
2.8.	Supplementary materials	181
3.	Conclusion.....	186
Chapitre 6	Discussion générale.....	187
	Références	196
	Conclusion générale et perspectives.....	199

LISTE DES FIGURES

Figure 1.1 Schéma simplifié du procédé de fabrication des thés verts et noirs (adapté de Harbowy et al., 1997).....	18
Figure 1.2 Composition de l'extrait sec d'une infusion de thé (% massiques, solide/solide, données pour le thé noir et le thé vert) (Harbowy et al., 1997)	20
Figure 1.3 Résumé des principales informations fournies par le RASFF sur le thé (2014-2017).....	22
figure 1.4 Principales étapes d'une analyse par couplage chromatographie/spectrométrie de masse....	29
Figure 1.5 Description des modes de balayage en MS/MS (Rondeau, 2017)	39
Figure 1.6 Processus générique d'acquisition et de traitement des données pour la détection non-ciblée des contaminants de l'aliment.....	48
Figure 2.1.1 Base Peak Ion chromatogram (BPI) of tea leave extracts along with signal intensity and retention of targeted contaminants in ESI + and ESI-.....	85
Figure 2.1.2 Percentage of 32 target contaminants in recovery range for spiked tea brew.....	87
Figure 2.1.3 Matrix effects observed upon analysis of tea leave extracts	88
Figure 3.1 Data processing steps	105
Figure 3.2 ICA score plots for filtered and unfiltered data matrix for both ionization modes on the development data set	110
Figure 3.3 Score plots of discriminating ICs (both ionization modes) for the validation data set	115
Figure 3.4 PCA score plots for filtered and unfiltered data matrix for both ionization modes on the development data.....	123
Figure 4.1 Workflow implemented	133
Figure 4.2 Data matrix layout and decision tree for MV classification	134
Figure 4.3 Venn diagrams for number of ions* selected by each pretreatment combination on both data sets for positive and negative ionization modes	140
Figure 4.4 Representation of MV rate against m/z , RT and median peak areas of ions.....	153
Figure 4.5 Score plots of ICA output for data set #1 acquired in positive ionization mode	154
Figure 4.6 Score plots of ICA output for data set #1 acquired in negative ionization mode.....	154
Figure 5.1 Generic workflow used for untargeted food safety assessment	161

Figure 5.2 Untargeted data treatment workflow.....	166
Figure 5.3 Score plot of ICA output for green tea samples analyzed on ToF platform	169
Figure 5.4 Score plots of ica output on black tea samples analyzed on orbitrap platform.....	172
Figure 5.5 PCA of raw data before and after drift correction (Green tea, ToF data acquired in positive ionization mode).....	185
Figure 5.6 Score plots of ICA output on green tea samples analyzed on Orbitrap platform.....	185
Figure 5.7 Score plots of ica output on black tea samples analyzed on ToF platform.....	186
Figure 6.1 Récapitulatif des expériences non-ciblées réalisées.....	191
Figure 6.2 Représentation des pertes en « traceurs » au fil des étapes de la preuve de concept pour trois jeux de données	193

LISTE DES TABLEAUX

Tableau 1.1 Principaux textes législatifs européens sur la sécurité sanitaire chimique des aliments.....	8
Tableau 1.2 Principales sources de contamination chimique de l'aliment	9
Tableau 1.3 Contaminants et composés interférents représentatifs lors d'une analyse LC-MS du thé. 25	
Tableau 1.4 Aperçu des travaux existants sur les approches ciblées d'analyse de contaminants.....	27
Tableau 1.5 Nombre d'occurrences trouvées sur la plateforme "Web of Science" pour les trois grands types de chromatographie (recherche au 18/07/2018, période 1956-2018)	31
Tableau 1.6 Nombre d'occurrences relevées sur la plateforme "web of science" pour l'ESI et l'APCI (recherche au 18/07/2018, période 1956-2018).....	35
Tableau 1.7 Principales caractéristiques des analyseurs de masse couramment couplés à des méthodes chromatographiques (Christian et al., 2014; Rondeau, 2017)	36
Tableau 1.8 Détails des publications traitant de la détection non-ciblée des contaminants de l'aliment	48
Tableau 1.9 Caractéristiques des outils d'intégration de pics en LC-MS	52
Table 2.1 Targeted compounds along with their instrumental optimized parameters and method performances.	80
Table 2.2 Recovery, RSD, method limit of quantification (MLOQ) and maximum residue limit (MRL) for the 32 targeted compounds in tea leaves quantified with matrix-matched calibration.	90
Table 2.3 Ionization parameters for ESI+ and ESI-.	94

Table 2.4 Recoveries, RSDs and estimated MLOQs in tea brew for compounds spiked between 10 and 100 µg/kg extracted by LLE or D&S method	95
Table 3.1 Discriminating features for positive ionization mode and respective putative annotation..	112
Table 3.2 Discriminating features for negative ionization mode and respective putative annotation.	113
Table 3.3 Information on “tracer” contaminants used in this study	120
Table 3.4 Full parameters and their corresponding values for peak extraction using XCMS	122
Table 4.1 Main characteristics of studied data sets (Delaporte et al., 2019)	132
Table 4.2 Combination of pretreatment methods tested.....	136
Table 4.3 Property summary of data sets (for both ionization modes).....	137
Table 4.4 Performances of the whole workflow depending on the MV imputation / filtration methods combination and data sets.....	141
Table 4.5 Main features of MV imputation methods	144
Table 4.6 Main features of filtration methods	144
Table 4.7 Information on spiking contaminants used in this study	150
Table 4.8 Full parameters and their corresponding values for peak extraction using XCMS.....	152
Table 5.1 Contamination scenarios considered	164
Table 5.2 Data sets available and their respective metrics	168
Table 5.3 Annotated contaminants in green tea samples.....	170
Table 5.4 Detection rates for contamination case n°1	171
Table 5.5 Annotated contaminants in black tea samples.....	175
Table 5.6 Detection rates for contamination case n°2	176
Table 5.7 Main characteristics of studied compounds	182
Table 5.8 Source parameters for ToF platform	183
Table 5.9 Source parameters for Orbitrap platform	183
Table 5.10 Parameters for peak extraction by XCMS.....	184

LISTE DES ABRÉVIATIONS

Abréviation utilisée	Signification
AA	Acides Aminés
ACN	Acétonitrile
AMM	Autorisation de Mise sur le Marché
ANSES	Agence Nationale de Sécurité Sanitaire de l'alimentation, de l'environnement et du travail
APCI	Ionisation Chimique à Pression Atmosphérique
ASE	Extraction par Solvant Accélérée (=PLE)
BPA	Bisphénol A
CV	Coefficient de Variation
DAD	Détecteur à barrette de diodes
DCM	Dichlorométhane
dSPE	Extraction en phase solide dispersive
CE	Communauté Européenne
CEE	Communauté Économique Européenne
EFSA	Autorité Européenne de sécurité des Aliments
AEM	Agence Européenne du Médicament
ESI	Ionisation par électrospray
UE	Union Européenne
FA	Acide formique
FAO	Organisation des Nations Unies pour l'Alimentation et l'Agriculture
FC	Facteur multiplicatif (traduction de l'anglais "Fold Change")
FC _{min}	Facteur multiplicatif minimum pertinent
FWHM	Largeur maximale à mi-hauteur
GC	Chromatographie en phase Gazeuse
GCB	Carbone graphite
HILIC	Hydrophilic Interaction Liquid Chromatography
HRMS	Spectrométrie de masse à Haute Résolution
ICA	Analyse en Composantes Indépendantes
KNN	Méthode des k plus proches voisins
LC ou HPLC	Chromatographie Liquide (à Hautes Performances)
LLE	Extraction Liquide/Liquide
LOD	Limite de Détection
LOESS	Régression locale (en anglais "Locally weighted Scatterplot Smoother")
LOQ	Limite de Quantification
MAR	Manquant de façon aléatoire
MCAR	Manquant de façon complètement aléatoire
MNAR	Manquant de façon non-aléatoire
LMR	Limite Maximale de Résidu
MRM	Surveillance de réactions multiples (traduction de l'anglais "Multiple Reaction Monitoring")

LMT	Limite Maximale Tolérable
NaCit	Citrate de Sodium
NaN	Pas un nombre (traduction de l'anglais "Not a Number")
PA	Alcaloïdes Pyrrolizidiniques
HAP	Hydrocarbure Aromatique Polycyclique
PCA	Analyse en Composantes Principales
PCB	PolyChloroBiphényle
PFC	PerFluoroCarbone
PFP	PentaFluoroPhényle
PLE	Extraction par Solvant Pressurisé (=ASE)
PLS (-R ou - DA)	Régression (ou Analyse de Données) des moindres carrés partiels
POP	Polluant Organique Persistant
PPDB	Pesticides Properties DataBase
PQN	Normalisation par Quotient Probabilistique
PSA	Amine Secondaire et Primaire
PVPP	PolyVinylPolyPyrrolidone
Q	Quadripôle
QC	Contrôle qualité
RASFF	Rapid Alert System for Food and Feed
RF (méthode)	Forêt d'arbres décisionnels
RT	Temps de Rétention
SFC	Chromatographie en Phase Supercritique
SIM	Surveillance d'ions sélectionnés
SLE	Extraction Solide/Liquide
LMS	Limite de Migration Spécifique
SPE	Extraction sur Phase Solide
SRM	Surveillance de réaction simple
SVD	Décomposition en Valeurs Singulières
TBBPA	Tétrabromobisphénol A
ToF	Temps de Vol
UHPLC	Chromatographie Liquide à Ultra-Hautes Performances
USE	Extraction assistée par ultrasons
UV	Ultraviolet

INTRODUCTION GÉNÉRALE

La contamination des aliments par des composés chimiques toxiques est une importante problématique de santé publique aujourd'hui. Ces composés ont une grande diversité de sources, tant naturelles qu'anthropiques (ex : moisissures, matériaux au contact, résidus de produits phytopharmaceutiques utilisés lors de la production), et présentent une variété importante de structures chimiques. Une analyse exhaustive des contaminants de l'aliment nécessite donc la multiplication des méthodes d'analyse. Les contaminants sont généralement présents dans les aliments à l'état de traces (entre le $\mu\text{g}/\text{kg}$ et le mg/kg), voire d'ultra-traces (en dessous du $\mu\text{g}/\text{kg}$), ce qui rend leur détection et leur quantification d'autant plus difficile que les produits agro-alimentaires sont chimiquement complexes et très divers.

L'évaluation de la qualité sanitaire chimique des aliments repose actuellement sur l'implémentation de multiples méthodes analytiques dites « ciblées » pour obtenir un aperçu exhaustif des contaminants présents dans un échantillon. Cela entraîne un coût important, nécessite beaucoup de temps, et l'exhaustivité n'est pas garantie car tout composé non recherché ne sera pas détecté. On peut donner comme exemple le scandale du lait en poudre contaminé à la mélamine en Chine en 2008, ou plus proche de nous, la présence d'un insecticide interdit (le fipronil) dans des œufs en Europe en 2017. Il est donc nécessaire de développer de nouvelles approches analytiques « non ciblées », à même de détecter des composés non attendus (voire inconnus) dans les produits agro-alimentaires de façon rapide et fiable.

En effet, les deux décennies passées ont vu l'élargissement du spectre d'application des méthodes analytiques ciblées, grâce aux progrès réalisés en matière d'instrumentation, notamment dans le domaine de la spectrométrie de masse à haute résolution. Dans le même temps, des approches analytiques non-ciblées sont apparues dans le domaine des sciences de la vie avec le développement des sciences « omiques », et notamment de la métabolomique. Ces développements ont été rendus possibles grâce aux progrès de la spectrométrie de masse à haute résolution et de l'informatique observés à ce moment. Au début des années 2010, ces approches ont été identifiées comme très prometteuses pour l'évaluation non-ciblée de la qualité sanitaire chimique des aliments, et une première preuve de concept a été proposée dès 2013 par une équipe Suédoise.

Depuis, quelques autres méthodes ont été proposées, mais ce champ d'études en est toujours à ses balbutiements. Les principales limites de ces approches que nous avons identifiées, et auxquelles cette thèse va essayer d'apporter des éléments de réponse sont :

- Des performances restant à prouver (en termes de sensibilité et d'exhaustivité notamment) sur des situations complexes susceptibles d'être rencontrées lors de l'étude de contaminations « réelles » ;

- Un manque d'études méthodologiques sur les différentes étapes, notamment de traitement de données, et un manque de compréhension de l'influence de chaque étape et outil sur le résultat final ;
- Un manque de recul sur la robustesse de ces approches, par exemple lors de leur implémentation sur des plateformes instrumentales différentes, ou dans des situations différentes de celle de développement.

Le thé (sous forme de feuilles, mais aussi d'infusion) a été choisi comme produit d'étude pour la réalisation de ces travaux. Notre choix s'est porté sur ce produit en raison de sa complexité de composition chimique, de sa très large consommation ainsi que de son intérêt du point de vue de la contamination chimique, avec des signalements fréquents sur les réseaux d'alerte sanitaire européens.

Le chapitre 1 présente de façon détaillée la problématique de la sécurité sanitaire chimique des aliments, et notamment concernant notre produit d'étude. Un état de l'art des approches analytiques pour l'évaluation de la qualité sanitaire chimique du thé (et des produits agro-alimentaire d'une façon générale) sera également réalisé. Une attention particulière sera portée aux derniers développements en matière de méthodes à spectre large et non-ciblées.

Le second chapitre présente le développement de la méthode d'analyse d'une sélection représentative de contaminants par LC-HRMS dans le thé sous forme de boisson et de feuilles (produit fini). En effet, l'évaluation des performances de l'approche non-ciblée qui sera développée nécessite de disposer d'une méthode d'analyse ciblée performante pour les contaminants investigués dans ce travail.

Le chapitre 3 expose la preuve de concept d'approche non-ciblée développée dans ce travail, et notamment un processus de traitement de données automatisé pour la détection des contaminants de l'aliment. Cette approche a été appliquée à deux situations contrastées de contamination du thé vert.

Le chapitre suivant traite de l'étude méthodologique de deux étapes importantes de prétraitement des données : la gestion des valeurs manquantes et la filtration de la matrice de données. Les données et résultats sont discutés en regard des deux plateformes analytiques UHLC-HRMS utilisées dans ce travail.

Le chapitre 5 détaille l'approche non-ciblée finalement mise au point, et son application à la détection en aveugle de contaminants dans le thé dans plusieurs situations complexes. Les résultats sont à nouveau discutés en fonction de la plateforme UHPLC-HRMS utilisée. Enfin, le dernier chapitre se veut une discussion des différents résultats, et une mise en perspective des faits marquants de ce travail.

CHAPITRE 1
CONTEXTE, ÉTAT DE L'ART ET MÉTHODOLOGIE

1. LA QUALITÉ SANITAIRE CHIMIQUE D’UN ALIMENT

Cette partie vise à poser le contexte général de ces travaux de thèse. On commencera par quelques définitions ainsi qu’un rappel des enjeux liés à la qualité sanitaire chimique. Ensuite, on présentera les contaminants de l’aliment ainsi que le contexte réglementaire en vigueur à l’heure actuelle en Europe. Enfin, le produit d’étude de cette thèse (le thé) sera présenté.

1.1. DÉFINITIONS ET ENJEUX

Les termes « contaminant (ou contamination) de l’aliment » et « sécurité sanitaire de l’aliment » peuvent désigner une problématique liée aux composés chimiques, mais peuvent aussi renvoyer à un problème microbiologique. Dans la mesure où seule la partie chimique est traitée ici, les mots, « contaminant », « contamination » et « sécurité sanitaire » renvoient implicitement par la suite à la contamination chimique uniquement.

Selon l’Autorité Européenne de Sécurité des Aliments (EFSA) : « Les contaminants sont des substances chimiques qui n’ont pas été ajoutées intentionnellement aux aliments destinés à l’alimentation humaine ou animale » (EFSA, 2018).

On peut également se référer à la définition donnée par la Commission Européenne dans le règlement CEE n°315/93 (Commission européenne, 1993) : « On entend par ‘contaminant’ toute substance qui n’est pas intentionnellement ajoutée à la denrée alimentaire, mais qui est cependant présente dans celle-ci comme un résidu de la production (y compris les traitements appliqués aux cultures et au bétail et dans la pratique de la médecine vétérinaire), de la fabrication, de la transformation, de la préparation, du traitement, du conditionnement, de l’emballage, du transport ou du stockage de ladite denrée, ou à la suite de la contamination par l’environnement ». Cette définition a le mérite de circonscrire l’étude de la contamination à un périmètre suffisamment restreint pour être traité dans le cadre d’une thèse, bien qu’encore très large, tout en étant exhaustif dans les potentielles origines de la contamination.

Par ailleurs, on peut distinguer deux grands types d’enjeux liés à la sécurité sanitaire des aliments selon les acteurs concernés.

1.1.1. ENJEUX DE SANTÉ PUBLIQUE

Les acteurs s’intéressant à ces enjeux sont en grande majorité des autorités publiques (ministères par exemple), des agences publiques (ANSES ou Santé Publique France dans l’hexagone, EFSA au niveau européen par exemple) ainsi que des Organisations Non-Gouvernementales (ONG).

Le premier enjeu pour les autorités publiques est la gestion du risque sanitaire et la protection des consommateurs au jour le jour par le biais de contrôles de routine pour la détection et la prévention des crises sanitaires. Les méthodes analytiques utilisées pour cela sont quasi-exclusivement des méthodes

opérées en routine dans des laboratoires de référence accrédités. D'autres méthodes peuvent parfois être mises en œuvre dans des situations de crise, quand survient une question spécifique à laquelle les méthodes officielles ne peuvent pas répondre.

Évaluer de façon fiable et la plus exhaustive possible la qualité sanitaire des aliments est crucial pour évaluer les risques et prévenir les crises sanitaires par la législation appropriée. En effet, cette information permet d'évaluer l'exposition des populations et ainsi de fixer les niveaux de contaminants acceptables pour leur offrir une protection satisfaisante. Elle sert ainsi à la fois à l'évaluation (par exemple en répondant à la question « dans les conditions actuelles quelle est l'exposition des populations pour tel contaminant ? »), puis indirectement à la gestion du risque (« quelles sont les actions à entreprendre pour réduire cette exposition ? »). Les méthodes analytiques utilisées pour cette évaluation sont à la fois des méthodes officialisées opérées en routine, ainsi que des méthodes de recherche spécifiques, offrant de meilleures performances.

Par ailleurs, il y a de la part du grand public une toujours plus grande sensibilisation aux thématiques de sécurité sanitaire, qui s'accompagne également d'une demande d'informations. Cela amène les autorités nationales et européennes à accorder une place importante à la communication auprès du grand public sur les données de sécurité sanitaire.

1.1.2. ENJEUX ÉCONOMIQUES

L'évaluation de la qualité sanitaire des aliments constitue également un enjeu économique clé pour les acteurs de l'industrie agro-alimentaire. En effet, un bon contrôle de la qualité sanitaire d'un produit permet de détecter de façon précoce de potentielles crises sanitaires. Ces crises peuvent entraîner des pertes financières importantes pour les compagnies (rappel de produit, dégradation de l'image de marque par exemple), d'où la nécessité d'évaluer de façon fiable la qualité sanitaire. Les méthodes analytiques utilisées dans ce cas sont des méthodes de référence, mais des méthodes « maison » plus performantes sont souvent également implémentées en interne, surtout au sein des grandes multinationales.

L'évaluation de la qualité sanitaire des aliments constitue donc un enjeu majeur à la fois pour les autorités de sûreté ainsi que l'industrie. Cependant, cette évaluation soulève plusieurs challenges en termes de chimie analytique. En effet, les matrices agro-alimentaires sont très complexes chimiquement, et les contaminants retrouvés la plupart du temps à l'état de traces (de l'ordre du $\mu\text{g}/\text{kg}$) (Castro-Puyana, Pérez-Míguez, Montero, & Herrero, 2017). La détection et la quantification des contaminants peuvent donc devenir un réel défi, puisque, comme nous allons le voir dans la prochaine partie, ces derniers peuvent venir de sources très variées, entraînant donc une très grande diversité de structures chimiques.

1.2. CONTEXTE RÉGLEMENTAIRE ET OUTILS DE GESTION DU RISQUE EUROPÉENS

On placera le point de vue réglementaire au niveau européen, puisque les règlements européens sont directement et entièrement obligatoires dans tous les États membres de l'Union Européenne. Des

spécificités peuvent exister au sein des États, mais la plupart du temps dans le sens d’une plus grande protection. Dans certaines situations spécifiques, des pays membres peuvent bénéficier d’une dérogation ; ce cas sera abordé plus tard.

Les contaminants de l’aliment sont règlementés en Europe par le règlement (CEE n°315/93), dont est issue la définition des contaminants citée précédemment. Ce règlement a été amendé à deux reprises, en 2003 et 2009. Ce texte a pour objectif d’harmoniser le cadre législatif sur les contaminants pour permettre la libre circulation des marchandises au sein de l’Union Européenne (UE), et protéger la santé des consommateurs.

Il vient s’ajouter un règlement (CE n°1881/2006) fixant la limite maximale tolérable (LMT) de certains contaminants dans de nombreux groupes d’aliments (Commission Européenne, 2006). Ce règlement est régulièrement amendé pour tenir à jour les teneurs maximales au vu de l’évolution des connaissances.

À ce cadre général viennent s’ajouter des règlements spécifiques à certaines catégories de molécules.

Les résidus de produits chimiques utilisés lors de la production (par exemple les pesticides et les produits vétérinaires) font l’objet de « Limites Maximales de Résidus » (LMR) régies par le règlement CE n°396/2005 (Commission européenne, 2005), faisant lui aussi l’objet d’amendements réguliers pour les maintenir à jour. Quant aux matériaux au contact des denrées alimentaires, ils sont règlementés par des « Limites de Migration Spécifiques » (LMS) régies par le règlement CE n°1935/2004 (Commission européenne, 2004a), avec un règlement spécifique (EU n°10/2011) pour les matière plastiques (Commission européenne, 2011a).

Au-delà de ces règlements, les États membres sont libres d’appliquer des LMR et des réglementations autres que celles de l’UE du moment que celles-ci sont plus protectrices que les LMR européennes (on peut penser notamment à l’interdiction du bisphénol A en France (ANSES, 2018)). Des dérogations vers une situation moins protectrice peuvent parfois être observées lorsqu’elles sont justifiées par un contexte socio-économique particulier (exemple : le cas des pays bordant la mer Baltique pour les taux de certains polluants organiques dans des poissons (Commission européenne, 2011b)).

Les analyses et les contrôles dans l’agro-alimentaire sont régis par le règlement européen CE n°882/2004 (Commission européenne, 2004c). La Commission Européenne a également émis un document guide sur le développement et la validation de méthodes de dosage de résidus dans les produits agro-alimentaires (European Commission - DG Health and Food Safety, 2015). En fonction des contaminants analysés et des produits testés, des consignes spécifiques peuvent s’appliquer. Celles-ci sont détaillées dans les règlements relatifs aux contaminants spécifiques (exemple : CE n°396/2005 pour les pesticides).

L’évolution de ce cadre réglementaire dans les années 2000 s’est faite dans un contexte de mondialisation des circuits de production, et également de façon concomitante à un certain nombre de crises sanitaires liées aux aliments. Une des plus médiatique a été celle du lait adultéré à la mélamine

(Organisation Mondiale de la Santé, 2008) qui a provoqué plusieurs décès en Chine en 2008. Plus proche de nous, la présence sur le marché européen d'œufs contaminés par un insecticide interdit (le fipronil) pendant plusieurs mois montre que des améliorations des système d'alerte et d'évaluation de la qualité sanitaire des produits agro-alimentaires sont toujours possibles et nécessaires (Ministère de l'Agriculture et de l'Alimentation, 2017).

Afin de faciliter la circulation de l'information relative à d'éventuels problèmes sanitaires (un produit présentant un risque pour le consommateur par exemple) au sein des pays membres, un système d'alerte rapide a été mis en place (RASFF : « Rapid Alert System for Food and Feed ») pour les pays membres ainsi que pour certains pays tiers (Norvège, Liechtenstein, Islande et Suisse) (Commission européenne, 2002). Ce système permet un partage rapide de l'information sur un échantillon ne respectant pas les règles communautaires, et ainsi de prévenir de potentielles crises sanitaires. Un résumé des règlements Européens portant sur les contaminants de l'aliment se trouve en Tableau 1.1.

TABLEAU 1.1 PRINCIPAUX TEXTES LÉGISLATIFS EUROPÉENS SUR LA SÉCURITÉ SANITAIRE CHIMIQUE DES ALIMENTS

Règlement N°	Objet
CEE n°315/93	Règlementation générale sur la contamination des aliments par des produits chimiques
CE n°1881/2006	Limites maximales d'un certain nombre de contaminants
CE n°853/2004	Règles d'hygiène applicables aux produits alimentaires d'origine animale (notamment vis-à-vis des biotoxines marines)
UE n° 2016/239	Teneurs maximales en alcaloïdes tropanoïques dans les aliments pour bébé
UE n°2015/1940	Teneurs maximales en ergot du seigle
CE n°1107/2009	Mise sur le marché des produits phytosanitaires
CE n°396/2005	Définition des LMR des produits phytosanitaires dans les aliments
CE n°1935/2004	Matériaux au contact des aliments (général)
CE n°10/2011	Matières plastiques au contact des aliments
UE n°2017/2158	Teneurs en acrylamide dans les aliments et mesures de réduction

1.3.LES DIFFÉRENTES CLASSES DE CONTAMINANTS

Les contaminants de l'aliment peuvent être des composés organiques (structure chimique basée sur le carbone) ou bien inorganiques (voire organométalliques). Dans le cadre de ce travail de thèse, seuls les contaminants organiques seront abordés et traités.

Comme indiqué précédemment, les contaminants peuvent être aussi bien d’origine naturelle qu’anthropique. Le corollaire de cette diversité de sources est une diversité chimique des contaminants potentiels très importante. Ces principales sources sont répertoriées dans le Tableau 1.2. Certains contaminants peuvent être classés dans plusieurs catégories en fonction du scénario de contamination envisagé. Par exemple, les hydrocarbures aromatiques polycycliques (HAP) peuvent être à la fois considérés comme des contaminants de l’environnement (ex : gaz d’échappement des moteurs thermiques), ou comme des contaminants issus de procédés (ex : cuisson au feu de bois). De même, les problématiques liées à des contaminants d’origine naturelle (comme les mycotoxines) peuvent être aggravées ou modifiées par des facteurs anthropiques comme le réchauffement climatique.

TABLEAU 1.2 PRINCIPALES SOURCES DE CONTAMINATION CHIMIQUE DE L’ALIMENT

Sources naturelles	Sources anthropiques
Végétaux <i>Ex : contamination d’infusions par des alcaloïdes pyrrolizidiniques</i>	Résidus de produits chimiques (naturels ou synthétiques) utilisés lors de la production <i>Ex : résidus de pesticides dans des fruits</i>
Champignons (macroscopiques ou microscopiques) <i>Ex : contamination du seigle par les alcaloïdes produits par l’ergot du seigle</i>	Résidus de produits chimiques synthétiques présents dans l’environnement <i>Ex : polluants organiques persistants dans les poissons gras</i>
	Migrants de matériaux au contact <i>Ex : migration du bisphénol A dans les boîtes de conserve</i>
	Contaminants issus de procédés <i>Ex : acrylamide dans le café</i>
Contaminants inorganiques (<i>origine naturelle ou anthropique</i>) <i>Ex : métaux lourds accumulés dans des végétaux</i>	

La classification par source permet de catégoriser relativement aisément les potentiels contaminants, la variabilité chimique observée au sein de chacune des sources restant très importante. Il n’est pas pertinent d’établir une liste exhaustive des potentiels contaminants de l’aliment ici, car cela dépasse grandement le sujet de cette thèse. On résumera donc les principales sources de contamination de l’aliment connues. On commencera par définir chacune des sources, avec éventuellement les scénarios de contamination typiques et les problématiques sanitaires associées, on donnera ensuite le contexte réglementaire européen (et français en cas de différences notables). Puis les structures chimiques seront rapidement évoquées (présence ou non d’un noyau commun, diversité, polarité etc.) ainsi que les niveaux de concentration attendus dans l’aliment. On se concentrera sur les molécules organiques uniquement (c.à.d. basées sur le carbone), les problématiques de chimie analytique posées par l’analyse des inorganiques étant différentes de celles des molécules organiques.

1.3.1. CONTAMINANTS DE L'ALIMENT D'ORIGINE NATURELLE

1.3.1.1. CONTAMINANTS ISSUS DE VÉGÉTAUX

Les toxines végétales sont des molécules toxiques synthétisées par des organismes vivants du règne végétal, tant microscopiques (ex : micro-algues) que macroscopiques (ex : plantes). Ces contaminants se retrouvent dans les aliments par plusieurs voies :

- (i) Certaines micro-algues sont connues pour synthétiser des toxines qui peuvent, par accumulation dans la chaîne alimentaire, se retrouver dans la chair des poissons consommés par l'Homme. Un des cas de ce type de contamination est la présence de ciguatoxines dans des poissons tropicaux, entraînant des épidémies de ciguatera ;
- (ii) Certaines plantes comestibles en temps normal peuvent, suite à de mauvaises conditions de stockage, ou bien à un usage inapproprié (par exemple une consommation à maturation insuffisante), synthétiser des composés toxiques. C'est le cas par exemple des pommes de terre, qui, mal stockées, peuvent synthétiser des molécules de la classe des glyco-alcaloïdes (Santé Canada, 2011) ;
- (iii) Certaines plantes, non comestibles et produisant des toxines, peuvent se retrouver mélangées à une étape de la production à des plantes destinées à la consommation, et ainsi contaminer ces dernières, notamment par des alcaloïdes pyrrolizidiniques (Mulder, Sánchez, These, Preiss-Weigert, & Castellari, 2015). Par ailleurs, cette contamination peut également se faire sans qu'il y ait de contact entre les espèces non-comestibles et les autres (Nowak et al., 2016). Les mécanismes de ces transferts sont cependant encore très mal connus.

Les toxines synthétisées par les plantes présentent une diversité de structure très importante, même si on retrouve un certain nombre de noyaux structuraux (par exemple les alcaloïdes).

Certaines de ces toxines font l'objet d'une réglementation spécifique, c'est le cas pour plusieurs biotoxines marines [règlement CE n°853/2004 (Commission européenne, 2004b)], ou encore les alcaloïdes tropaniques dans les aliments pour bébé [règlement UE n°2016/239 (Commission européenne, 2016)].

De plus, des cas réguliers (une trentaine par an en moyenne) sont rapportés *via* le RASFF (catégories « biotoxines » et « biocontaminants » suivant les cas), montrant donc que ces toxines constituent un sujet de préoccupation très actuel.

Les problématiques sanitaires soulevées par ces contaminants portent à la fois sur une toxicité aiguë et chronique. Un des exemples le plus emblématique de toxicité aiguë pour ces contaminants concerne la famille des ciguatoxines (Institut Louis Maladré, 2014), qui sont la première cause des intoxications liées à la consommation de poisson dans le monde (Hossen et al., 2013). Les molécules de cette famille,

synthétisées par des micro-algues des récifs tropicaux (algues de la famille *Gambierdiscus*), ont pour caractéristiques d’être bioaccumulables (elles s’accumulent dans les organismes vivants, notamment dans les tissus gras), bioamplifiables (leurs concentrations augmentent le long de la chaîne alimentaire) et biotransformables (leur toxicité augmente au fur et à mesure de leur métabolisation par les différents organismes de la chaîne alimentaire). Ainsi, la concentration et la toxicité de ces toxines augmentent depuis les poissons brouteurs qui consomment les algues, jusqu’à leurs prédateurs, tous ces poissons étant consommés par l’Homme. Ce sont des neurotoxines, qui sont classées parmi les toxines les plus puissantes, puisque une exposition à 2 ng/kg de poids corporel (soit une dose unique de 140 ng pour un individu de 70 kg) suffit pour provoquer des symptômes chez 50 % des consommateurs (Institut Louis Maladré, 2014).

En ce qui concerne les problématiques de toxicité chronique (qui peuvent de plus se combiner avec des problématiques de toxicité aiguë), la thématique émergente de la présence des alcaloïdes pyrrolizidiniques dans les aliments est un bon exemple (Knutsen et al., 2017; Mulder et al., 2015). Ces toxines, synthétisées par les végétaux de nombreuses familles différentes, peuvent se retrouver dans divers produits alimentaires (infusions, miels, produits animaux, etc.). Ces molécules se retrouvent à l’état de traces par contact entre des plantes saines et des plantes contaminées, ou par des mécanismes de transfert encore mal connus (Nowak et al., 2016)). Ces contaminants sont très présents dans des infusions ou des thés commerciaux, et ce à des niveaux de l’ordre de 0,1-1 mg/kg (Mulder et al., 2015). Cependant, au regard du manque de données relatives à l’occurrence de ces composés dans les aliments, des travaux sont en cours pour l’évaluation du risque relatif à ces molécules (Knutsen et al., 2017).

1.3.1.2. TOXINES FONGIQUES

Parmi les toxines d’origine fongique (champignons macroscopiques et moisissures), on fera une distinction entre les toxines synthétisées par des champignons susceptibles d’être consommés par l’Homme et les toxines synthétisées par des moisissures se retrouvant de façon accidentelle à grande échelle dans l’aliment. La raison d’être de cette distinction est que ces situations ne recoupent pas les mêmes problématiques sanitaires. La première soulève uniquement des problématiques de toxicité aiguë, portant sur des cas ponctuels et saisonniers, alors que les problématiques soulevées par la deuxième concernent à la fois des cas de toxicité chronique et aiguë et portent sur des aliments consommés quotidiennement par l’ensemble de la population (Brimer, 2011). On se focalisera donc essentiellement sur la deuxième situation.

On peut toutefois rapidement résumer la problématique des toxines synthétisées par des champignons directement consommés par l’Homme. La ressemblance entre les champignons réputés comestibles et ceux toxiques est parfois telle que des personnes non averties ramassant par eux-mêmes des champignons peuvent être amenées à consommer des champignons toxiques. En ce cas, la toxicité est en général exclusivement aiguë, et peut parfois être mortelle (environ 1 000 intoxications par an, dont une vingtaine graves et trois mortelles en moyenne en France ((Agence nationale de sécurité sanitaire

de l'alimentation de l'environnement et du travail, 2015)). Ce problème de toxicité survient en cas de consommation d'un champignon toxique, ou bien de contamination d'une récolte de champignons comestibles par un champignon toxique. Les toxines fongiques responsables de cette toxicité peuvent être de nature variées, mais les plus communes sont des familles de toxines peptidiques telles que les amatoxines (Brimer, 2011). Cette problématique concerne surtout des récoltes destinées à une consommation privée, au contraire de la deuxième situation, celle des mycotoxines.

En effet, celles-ci concernent potentiellement l'ensemble de la population ainsi que des aliments consommés au quotidien et en grande quantité (ex : céréales, fruits), les enjeux sanitaires sont donc ici cruciaux. Ces toxines sont des métabolites secondaires synthétisés par certaines souches pathogènes de différentes familles de moisissures, principalement les familles *Fusarium*, *Penicillium* et *Aspergillus* (Fremy, 2009). Une des caractéristiques du risque posé par les mycotoxines est qu'il est influencé par de nombreux facteurs. En effet, toutes les souches des trois familles de moisissure citées ci-dessus ne produisent pas des mycotoxines. Par ailleurs, le potentiel pathogène d'une souche varie énormément en fonction des conditions environnementales (température, humidité, lumière par exemple).

Actuellement, il existe plusieurs grandes familles structurales identifiées de mycotoxines, les plus connues étant les plus problématiques d'un point de vue sanitaire et sont donc réglementées (CEE n°315/93 pour la réglementation générale). Ce sont les aflatoxines, les ochratoxines, les trichothécènes (déoxynivalénol et toxines T-2 et HT-2), la patuline, les fumonisines, la zéaralénone (règlement CE n°1881/2006) et les alcaloïdes de l'ergot (réglementation touchant la présence du champignon lui-même et non les molécules [règlement UE n°2015/1940 (Commission européenne, 2015a)]. Comme mentionné ci-dessus, les mycotoxines sont issues du métabolisme secondaire des moisissures. Or, l'état actuel des connaissances suggère qu'environ 80 % de ce métabolisme demeure inconnu (Hautbergue et al., 2017), ce qui peut laisser raisonnablement penser qu'il existe encore d'autres mycotoxines potentiellement dangereuses pour l'Homme à découvrir.

Les mycotoxines touchent une très large variété d'aliment et sont la plupart du temps des molécules thermostables, c'est-à-dire ne se dégradant pas avec la température ; on les retrouve donc même après cuisson. Elles ne sont pas non plus dégradées lors des procédés de fermentation. La totalité des aliments d'origine végétale peuvent être potentiellement contaminés par des mycotoxines, de même que certains produits d'origine animale tels que les produits issus du lait, du sang ou des abats.

Les mycotoxines ont tout d'abord été connues de l'Homme pour leur toxicité aiguë. Le premier cas documenté d'intoxication par des mycotoxines (appelée « mycotoxicose aiguë ») remonte à plus d'un millénaire. Il s'agit d'une mycotoxicose causée par les alcaloïdes de l'ergot du seigle, appelée ergotisme ou « mal des ardents » en référence aux symptômes développés par les personnes intoxiquées (gangrènes, délires, abcès notamment). Ce type de crise a un caractère exceptionnel en Europe et concerne surtout à l'heure actuelle l'alimentation animale.

Les mycotoxines peuvent avoir également une toxicité chronique. Les principaux organes cibles sont le foie, les reins et le système nerveux. Certaines possèdent également un caractère mutagène et/ou reprotoxique (Fremy, 2009). Ces effets, de même que l’omniprésence des mycotoxines, ont amené les autorités européennes à fortement réglementer plusieurs mycotoxines (la plus basse LMT reportée pour des mycotoxines est de 0,025 µg/kg pour l’aflatoxine M₁ dans les aliments pour nourrissons - à titre de comparaison, c’est 200 fois moins que la plus basse LMR reportée pour les pesticides qui est de 5 µg/kg pour le difenoconazole dans le lait (Commission européenne, 2018)).

1.3.2. CONTAMINANTS DE L’ALIMENT D’ORIGINE ANTHROPIQUE

1.3.2.1. RÉSIDUS DE SUBSTANCES ACTIVES UTILISÉES LORS DE LA PRODUCTION

Durant le procédé de fabrication d’un aliment, des produits chimiques sont susceptibles d’être utilisés lors de nombreuses étapes et dans des buts divers (par exemple pour la protection des cultures, le stockage ou le traitement de maladies). Les mêmes normes et règles s’appliquent à ces composés quelle que soit leur origine, synthétique (pour une utilisation en agriculture « conventionnelle ») ou bien naturelle (pour une utilisation en agriculture biologique par exemple). Des traces de ces molécules, ainsi que de leurs métabolites et sous-produits, peuvent donc se retrouver dans le produit final sous l’appellation de « résidus ». Ces composés présentent souvent une certaine toxicité et peuvent donc constituer un risque sanitaire potentiel pour l’être humain qu’il convient d’évaluer et de maîtriser.

L’utilisation de substances chimiques pour lutter contre des ravageurs des cultures ou des adventices (ou dans le cas des produits animaux contre des parasites et maladies) est connue depuis la découverte de l’agriculture (IUPAC, 2010). Jusqu’au milieu du XX^{ème} siècle, ces produits étaient très majoritairement d’origine naturelle (organiques ou inorganiques), mais l’essor de la chimie de synthèse au sortir de la Seconde Guerre Mondiale a conduit à l’utilisation massive de nouveaux produits synthétiques. Il existe maintenant plusieurs dizaines de familles de pesticides de synthèse. Si les premières molécules mises sur le marché avaient un spectre d’action extrêmement large, conduisant à d’importants problèmes de destruction des milieux naturels ainsi que de santé des personnes en contact, les générations de produits suivantes ont été conçues pour avoir des actions de plus en plus ciblées. Ceci étant, la présence de résidus de pesticides (molécule « mère » + métabolites et produits de dégradation éventuels) dans les aliments soulève de nombreux problèmes suspectés avec plus ou moins de certitudes d’effet sur la santé des consommateurs, y compris pour les molécules actuellement utilisées. La présence de ces molécules dans l’environnement n’est pas sans soulever d’importants problèmes, de même que l’exposition des travailleurs les manipulant, mais ces aspects débordent l’objet de notre étude ici.

Il peut être très délicat d’estimer le risque lié à la présence d’un pesticide dans l’alimentation puisque, si la toxicité aiguë d’une substance est aisément connue, quantifiable et contrôlable, la toxicité chronique est bien plus délicate à évaluer. Toute substance active (pesticide ou produit vétérinaire) mise sur le marché en Union Européenne (UE) passe par un processus d’Autorisation de Mise sur le Marché

(AMM) qui évalue le risque lié non seulement au(x) substance(s) active(s), mais aussi à la formulation (c.à.d. la substance active, le co-formulant et tous les excipients) et les bénéfices que cette combinaison apporte. Pour l'obtention de cette AMM (qui se décide maintenant par zone au niveau européen, pour une utilisation bien définie, règlement CE n°1107/2009 (Commission européenne, 2009)), le fabricant doit déposer un dossier dans un des pays membre de la zone où il veut commercialiser la formulation, pour laquelle il fournit des données toxicologiques et environnementales (par exemple sur leur potentiel d'accumulation ou de transfert) qui permettront ensuite aux autorités de sûreté sanitaire de formuler un avis scientifique puis, en mettant en balance les risques face aux bénéfices apportés par le produit, de décider ou non de l'AMM. Le fabricant demande cette AMM pour une liste d'usage bien définis, et l'autorité peut : (i) approuver le dossier pour tous les usages demandés (ii) approuver pour certains des usages uniquement (iii) refuser le dossier. Elle est également libre de demander des études complémentaires sur certains points qui lui semblent fragiles. L'autorisation est accordée pour 10 ans, au bout desquels l'évaluation doit de nouveau être conduite.

En Europe, pour gérer le risque lié à la présence de résidus de pesticides dans l'alimentation, la Commission Européenne (CE) fixe des LMR pour chaque molécule mise sur le marché, et pour chaque denrée. Les LMR pour les pesticides sont définies par le règlement spécifique CE n° 396/2005 (Commission européenne, 2005). Pour la définition des LMR, la Commission s'appuie sur un avis scientifique rendu par l'EFSA (ou l'Agence Européenne du Médicament –EMA– dans le cas des substances vétérinaires) basé sur les données toxicologiques, mais aussi sur l'état de l'art des méthodes analytiques (cela ne fait pas de sens de définir une LMR non quantifiable analytiquement). Il est cependant plus fréquent que les méthodes analytiques s'adaptent aux nouvelles LMR que l'inverse. Une LMR est fixée pour chaque combinaison denrée alimentaire – molécule active en fonction de la Dose Journalière Admissible (DJA), basée sur des études toxicologiques et épidémiologiques disponibles, ainsi que sur la consommation de la denrée concernée, de façon à ne pas dépasser cette dose par les consommateurs. Les LMR usuelles dans les aliments s'échelonnent entre 5 µg/kg et des niveaux de l'ordre du mg/kg. Les produits pour lesquels les données sont insuffisantes pour fixer une LMR fiable se voient attribuer une LMR protectrice par défaut de 10 µg/kg.

Il peut exister des effets de synergie entre les différentes molécules (appelés « effet cocktail ») qui peuvent modifier les seuils de risque acceptables définis par les autorités. Ces effets cocktails sont encore mal connus et font l'objet de nombreuses recherches à l'heure actuelle (Kortenkamp et al., 2012).

1.3.2.2. RÉSIDUS DE PRODUITS CHIMIQUES PRÉSENTS DANS L'ENVIRONNEMENT

Il peut y avoir un transfert de contaminants entre l'environnement et l'aliment. Ce transfert peut arriver à toutes les étapes de la vie du produit, que ce soit lors la première étape de la production (élevage ou culture), le transport, la transformation ou le stockage car tous les compartiments de l'environnement (eau, sol et air) sont susceptibles d'être contaminés par une grande diversité de molécules organiques synthétiques ou naturelles pouvant avoir des propriétés toxiques. Une autre voie d'entrée des

contaminants de l'environnement dans l'aliment est l'alimentation du bétail pour les produits d'origine animale (EFSA (European Food Safety Authority), 2008). Ces problématiques concernent surtout des problématiques de toxicité chronique. Leurs LMT sont fixées par le règlement CE n°1881/2006.

On peut distinguer deux grandes catégories de contaminants environnementaux se retrouvant dans l'aliment en fonction de leur persistance dans l'environnement.

Tout d'abord, on trouve les contaminants persistants dans l'environnement et se diffusant dans tous les compartiments. On peut parler de contaminant « en héritage » (traduction de l'anglais « legacy contaminants »). Ils ont la propriété d'être extrêmement stables dans l'environnement, avec des temps de demi-vie de plusieurs années voire décennies ou siècles. Des traités internationaux ont été mis en place pour tenter de réduire la présence de ce type de composés dans l'environnement avec par exemple la Convention de Stockholm qui a défini une liste de « Polluants Organiques Persistants » (POP) (WHO/UNEP, 2008b). Les composés visés par cette convention ont, en plus d'être stables, les propriétés de se diffuser largement dans tous les compartiments de l'environnement, d'être bioaccumulables ainsi que d'être toxiques à la fois pour l'être humain, la faune et la flore à des faibles niveaux d'exposition. On y retrouve des molécules telles que les polychlorobiphényles (PCB), certains retardateurs de flammes bromés, ou encore des pesticides organochlorés tels que le DDT et ses métabolites. D'autres molécules ne sont pas classées comme POP, mais possèdent des propriétés relativement similaires en termes de stabilité et de diffusion dans l'environnement. On peut citer par exemple certains pesticides comme l'atrazine ou encore les composés perfluorés (PFC (Santé Canada, 2009)). Certains de ces composés (c'est notamment le cas pour ceux inscrits à l'annexe A de la Convention de Stockholm (WHO/UNEP, 2008a)) sont interdits dans l'UE, de même que certains pesticides problématiques tels que l'atrazine. Pour ces composés (qui sont parfois formés de façon accidentelle comme les dioxines), d'importants efforts ont été mis en œuvre dans l'UE pour réduire la contamination des milieux naturels et donc l'exposition des populations (Commission européenne, 2010). Cependant, certains de ces composés sont toujours utilisés dans d'autres régions du globe, une crise sanitaire par le biais de l'importation est donc toujours possible (Commission européenne, 2015b), ainsi que par le biais de rejets accidentels locaux (EFSA (European Food Safety Authority), 2008).

Les contaminants de l'environnement peuvent être également des composés non-persistants ayant comme source des activités humaines à proximité du lieu de production de l'aliment. Très souvent, il s'agit d'une contamination ponctuelle dans le temps et/ou dans l'espace (par exemple, taux élevé de HAP pour des légumes cultivés à proximité d'un axe routier). On peut rencontrer dans cette catégorie une très large gamme de sources de contamination, installation industrielle, champs cultivés, axe routier, etc. Les contaminants retrouvés ont également une grande diversité de structure, depuis les hydrocarbures aromatiques polycycliques (HAP) aux hydrocarbures en passant par toutes les familles de pesticides.

Pour beaucoup de contaminants environnementaux (pesticides exceptés), les LMT (Règlement CE n°1881/2006) sont à des niveaux situés entre le ng/kg (pour les PCB, dioxines et assimilés) et le µg/kg (pour les HAP par exemple), soit des niveaux plus faibles que ceux tolérés pour les pesticides par exemple. Tout comme d'autres contaminants, l'alimentation est un des principaux facteurs d'exposition de la population pour les contaminants environnementaux (Fréry et al., 2017).

1.3.2.3. MIGRANTS DE MATÉRIAUX AU CONTACT

Tout au long de leur cycle de vie, les aliments se trouvent au contact de matériaux, que ce soit lors de leur production et transformation, mais aussi de leur transport, leur stockage ou encore la préparation par le consommateur. Tous ces matériaux sont susceptibles de transférer des composés dans l'aliment, pouvant potentiellement soulever des problèmes sanitaires. Les matériaux pouvant se trouver au contact des aliments sont réglementés au niveau européen par le règlement CE n°1935/2004 modifié par le règlement CE n°596/2009 (Commission européenne, 2004a). Dans ce règlement, les matériaux autorisés sont divisés en 17 groupes différents (exemple : groupe 1 « matériaux et objets actifs et intelligents » ; groupe 2 « colles » ; groupe 3 « céramiques », etc.). Chaque groupe peut être « soumis à des mesures spécifiques » pour en gérer le risque associé. Ces mesures spécifiques sont décrites ensuite dans des règlements en fonction des groupes de matériaux (CE n°10/2011 pour les matières plastiques, CE n°42/2007 pour la cellulose régénérée, n°84/500/EEC pour la céramique et n°93/11EEC pour les caoutchoucs à titre d'exemples). Les matériaux pour lesquels il n'existe pas de réglementation spécifique au niveau européen sont gérés par les différents États membres. La réglementation européenne impose aux matériaux destinés au contact alimentaire de ne pas transférer « aux denrées alimentaires des constituants en une quantité susceptible de présenter un danger pour la santé humaine, d'entraîner une modification inacceptable de la composition des denrées, ou d'entraîner une altération des caractères organoleptiques de celles-ci ». Par ailleurs, les produits autorisés pour la synthèse et la constitution des emballages sont définis grâce au principe d'une liste positive. Pour l'utilisation d'un matériau non expressément autorisé, le fabricant doit demander une autorisation.

Ceci étant posé, il est nécessaire de rappeler qu'il n'existe aucune matière connue comme étant 100 % inerte. Le transfert de molécules entre deux matières (ex : l'aliment et son emballage) va être conditionné par de nombreux facteurs tels que la température, le temps de contact ou encore la nature même des deux produits.

Cette problématique est extrêmement complexe pour plusieurs raisons :

- i) La composition exacte d'un produit d'emballage n'est pas nécessairement connue avec précision, même si les matières premières le sont. C'est notamment le cas pour des matériaux issus de procédés de polymérisation, où peuvent cohabiter des résidus de monomère, des oligomères (polymères à courte chaîne) et le polymère constituant

le matériau proprement-dit. C’est également vrai pour les matériaux issus du recyclage.

- ii) Les interactions entre produits alimentaires et matériaux ne sont pas toujours bien appréhendées.

On peut citer comme exemples de ce type de contaminants les composés de type bisphénols (le plus connu étant le bisphénol A, interdit en France mais autorisé en Europe), les phtalates ou encore les colorants de type benzophénone.

Pour les matières plastiques il existe des Limites de Migration Spécifiques (LMS) pour une liste de composés, plus une limite de migration globale qui est fixée à 10 mg/dm² de matière en contact. Ces limites s’appliquent à tous les composés entrant dans la composition de l’emballage (polymères et additifs éventuels compris). Les LMS s’échelonnent, en fonction des composés, entre environ 50 µg/kg jusqu’à des niveaux de l’ordre de 100 mg/kg. Tout composé non autorisé est limité par défaut à 10 µg/kg. Pour certains enfin, le risque associé est tel qu’aucun résidu n’est autorisé dans l’aliment : c’est le cas par exemple du bisphénol F.

1.3.2.4. CONTAMINANTS ISSUS DE PROCÉDÉS

Les contaminants issus des procédés constituent la dernière classe des contaminants organiques présentés ici. Les aliments sont constitués de plusieurs centaines de composés chimiques différents. Au cours de leur transformation (séchage, cuisson ou encore raffinage par exemple), ces composés peuvent réagir entre eux et ainsi former de nouvelles molécules, dont certaines peuvent poser des problèmes de sécurité sanitaire.

Le cadre réglementaire de ces composés est le règlement CEE n°315/93, complété par le règlement CE n°1881/2006 pour les LMT de certains contaminants (quatre types de molécules identifiées comme contaminants issus de procédés sont règlementées). Parmi ces molécules, on retrouve des composés comme les HAP, qui sont notamment générés par les procédés de fumage, mais également par la cuisson au barbecue et dont certains sont règlementés.

Certaines molécules n’ont pas des taux règlementés au niveau européen mais font l’objet de valeurs de référence (les industriels concernés par un dépassement doivent mettre en place des mesures de réduction). C’est le cas par exemple de l’acrylamide avec le règlement UE n°2017/2158 (Commission européenne, 2017).

1.4.LA QUALITÉ SANITAIRE D’UN PRODUIT : L’EXEMPLE DU THÉ

1.4.1. GÉNÉRALITÉS SUR LE THÉ

Le thé est la deuxième boisson consommée au monde derrière l’eau (Chang, 2015). Cette boisson est réalisée à partir de feuilles de théier (*Camellia Sinensis*), un arbuste des zones tropicales et subtropicales. Découverte il y a 5 000 ans en Chine, cette boisson s’est ensuite diffusée dans toute l’Asie, avant de

gagner l'Europe à l'époque des grandes découvertes vers les années 1600 (Harbowy, Balentine, Davies, & Cai, 1997). À l'heure actuelle, le thé est cultivé dans toutes les zones tropicales et subtropicales dans le monde (incluant l'Amérique du sud et Centrale, l'Afrique, l'Océanie et une production anecdotique en Europe).

En 2013, les principaux pays producteurs étaient la Chine avec 40 % de la production mondiale en tonnage et l'Inde avec 25 % de la production mondiale en tonnage (Chang, 2015). Cependant, ces pays ayant également une très grande consommation intérieure (30 et 20 % de la consommation mondiale respectivement), le plus gros exportateur est le Kenya avec 23 % des exportations mondiales, devant la Chine (18 %), le Sri Lanka (17 %) et l'Inde (12 %). À titre d'information, en 2013, pour une production mondiale de 5 millions de tonnes, l'UE a importé 242 000 tonnes de thé (dont 15 000 pour la France) selon la *Food and Agriculture Organization* (FAO) (Chang, 2015). Le premier pays consommateur de thé en Europe est le Royaume-Uni avec plus de 116 000 tonnes consommées en 2013.

Il existe environ 1 500 types de thés selon la *UK Tea & Infusions Association* (UK Tea & Infusions Association, 2018), classés selon leurs procédés de fabrication, leurs terroirs ou leurs cultivars. Cette association distingue six grands types de thés : noirs, verts, oolongs, blancs, en briques et aromatisés. On se concentrera sur les deux types de thé les plus consommés : les thés verts et les thés noirs. Ces thés diffèrent dans leurs procédés de production, le thé noir subissant une étape d'oxydation enzymatique (appelée « fermentation » mais ne faisant pas intervenir de micro-organismes) comme montré sur la Figure 1.1 (Harbowy et al., 1997).

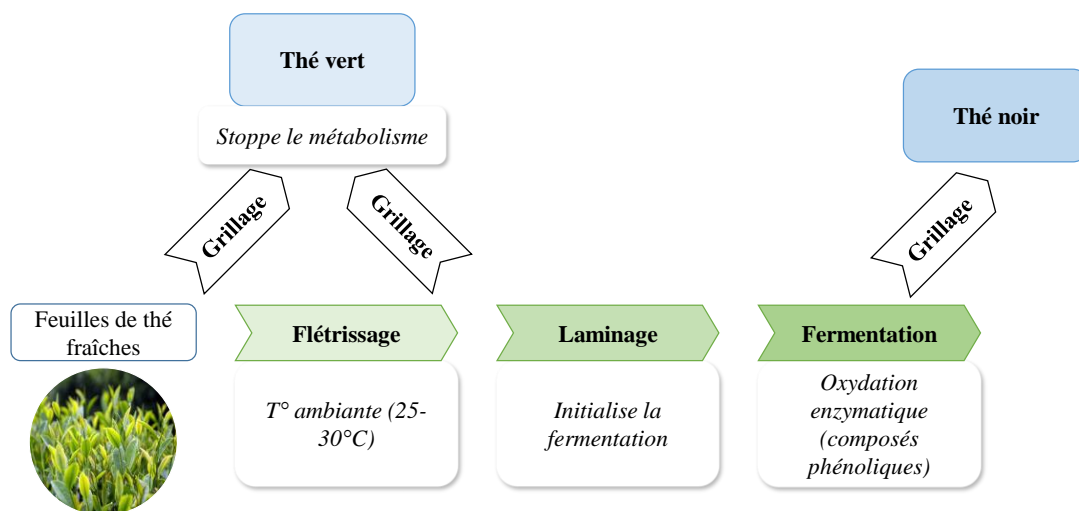


FIGURE 1.1 SCHÉMA SIMPLIFIÉ DU PROCÉDÉ DE FABRICATION DES THÉS VERTS ET NOIRS (ADAPTÉ DE HARBOWY ET. AL., 1997)

Pour le thé vert, les feuilles peuvent être flétries à température ambiante (on parle alors de thé « jaune »), puis elles sont immédiatement chauffées à haute température pour dénaturer les enzymes présentes et ainsi bloquer toute fermentation. Ce chauffage se fait à la vapeur (procédé employé au Japon et en Corée)

ou par un grillage à sec (procédé traditionnel employé en Chine). Les feuilles sont ensuite séchées jusqu’à une teneur en eau de l’ordre de 3 % puis peuvent éventuellement être mises en forme (par exemple laminées ou tressées) pour donner le thé vert sous sa forme finie. Pour le thé noir, après flétrissage à température ambiante, les feuilles sont laminées, cela a pour effet de casser les cellules des feuilles afin de mettre les polyphénols au contact de l’air. Elles seront ensuite laissées à fermenter à température ambiante (25-30°C) pendant plusieurs heures le temps que l’oxydation enzymatique des phénols se produise. Enfin, comme pour le thé vert, les feuilles sont passées à haute température puis mises en forme. Le principal facteur affectant la qualité du thé noir est le temps de fermentation. Au-delà des grands types de thés (verts ou noirs), il en existe beaucoup d’autres qui diffèrent par leur degré d’oxydation, ainsi que par les procédés employés pour stopper la fermentation, la maturité des feuilles, le cultivar utilisé, le terroir, etc.

1.4.2. COMPOSITION DU THE

Parce que le produit final tel que consommé est une infusion des feuilles de thé transformées, la plupart des études de la composition du thé se concentrent sur l’infusion ou son extrait sec. En ce qui concerne le produit fini sous forme de feuilles, il existe une littérature abondante sur certaines classes de composés (notamment les composés phénoliques) et l’influence de divers paramètres (procédés ou origine géographique par exemple) sur ces classes spécifiques (Fraser et al., 2013, 2014). Cependant, aucune étude générale n’a été trouvée sur la composition du produit fini sous forme de feuilles, ou alors ce sont des données purement nutritionnelles et non de composition chimique fine (DTU, 2018). La grande base de données de composition des aliments Foodb (<http://foodb.ca/>) ne semble contenir que des informations sur les feuilles de thé fraîches, ce qui sera tout de même utile pour recouper certaines informations.

La principale source de données dont nous disposons pour le produit fini est l’ouvrage de Harbowy et al. (Harbowy et al., 1997). Dans cet ouvrage, les auteurs donnent un aperçu de la composition « moyenne » des extraits secs d’infusions de thés verts et noirs. Ces compositions sont résumées en Figure 1.2. Bien que très informatives et intéressantes pour nous, ces données doivent être prises avec un certain recul vis-à-vis de nos travaux. En effet :

- i) Les auteurs donnent une composition « générique » de l’extrait sec pour les thés verts et noirs, mais ceci sans indiquer ni les sources bibliographiques, ni les méthodes employées. Néanmoins, cet ouvrage est resté notre seule source d’information sur ce point, puisqu’il est cité comme référence par toutes les sources bibliographiques ultérieurement trouvées ;
- ii) On parle ici de la composition de l’extrait sec d’une infusion à l’eau chaude. Nul doute que l’extraction des composés est incomplète et que l’utilisation d’un autre

solvant pourra révéler des différences qualitatives (nouvelles molécules extraites) ou quantitatives.

Ceci étant dit, au vu des teneurs des composés mesurés, et en croisant avec les données existantes dans la base de données foodb (<http://foodb.ca/foods/38>), on peut raisonnablement penser que l'analyse de l'extrait sec de l'infusion indique de façon relativement fiable les principaux composants des feuilles (composés phénoliques dont flavonols, xanthines et dérivés, peptides et acides aminés (AA), et sucres) susceptibles d'être extraits avec un solvant organique.

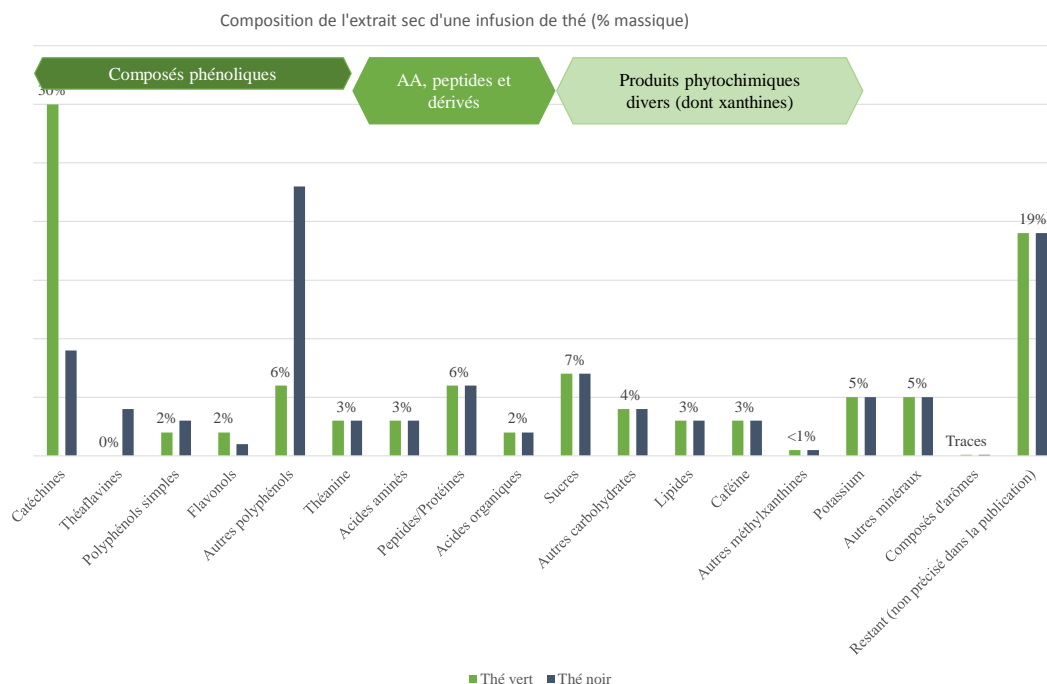


FIGURE 1.2 COMPOSITION DE L'EXTRAIT SEC D'UNE INFUSION DE THÉ (% MASSIQUES, SOLIDE/SOLIDE, DONNÉES POUR LE THÉ NOIR ET LE THÉ VERT) (HARBOWY ET AL., 1997)

À partir de ces informations de composition (Figure 1.2), croisées avec les données disponibles sur Foodb pour les feuilles fraîches, on peut dégager deux principales familles de molécules rencontrées dans le produit fini sous forme de feuilles. Tout d'abord, la classe des composés phénoliques, composée de plusieurs centaines de molécules différentes regroupées autour de quelques grands noyaux structurels. On en trouve une grande quantité dans le thé (de l'ordre de plusieurs dizaines de pourcent), que ce soit dans les feuilles fraîches ou dans l'infusion (et donc dans les feuilles transformées). Cette richesse en composés phénoliques, réputés pour leur impact bénéfique sur la santé, est d'ailleurs un des grands arguments marketing du thé en occident à l'heure actuelle (Harbowy et al., 1997). L'autre grande classe qu'on retrouve dans le thé est la classe des AA et de leurs dérivés (peptides et protéines par exemple). On les retrouve sous leur forme libre (6 % de la masse de l'extrait sec, dont la moitié de théanine), et sous forme de peptide ou de protéine (6 % de la masse de l'extrait sec). On retrouve ensuite une très large variété de molécules issues également du métabolisme de la plante (lipides, xanthines dont

caféine, caroténoïdes, chlorophylle et dérivés, composés d’arôme). Ces composés sont présents en plus faible quantité comparés aux deux grandes familles majoritaires, mais ils peuvent, selon les techniques d’analyse employées, soulever des problèmes spécifiques (exemples : phénomènes de suppression d’ions en spectrométrie de masse).

De fait de sa complexité chimique, l’analyse des molécules traces dans le thé peut rapidement devenir un réel challenge en matière de chimie analytique. C’est pour cela que ce produit est classé comme « complexe or unique commodity » dans le document guide pour la validation de méthodes d’analyse des pesticides dans l’alimentation de la Commission Européenne (European Commission - DG Health and Food Safety, 2015).

1.4.3. PROBLÉMATIQUES SANITAIRES LIÉES AU THÉ

Le thé étant une boisson de consommation quotidienne pour un grand nombre de personnes (15 % des hommes et 27 % des femmes en sont consommateurs réguliers en France (Hubert, Malvy, Preziosi, Galan, & Herberg, 2000)), il est particulièrement important de contrôler la qualité sanitaire de ce produit. C’est d’autant plus vrai que l’EFSA, dans son rapport sur les produits chimiques dans les aliments pour l’année 2016 (EFSA (European Food Safety Authority), 2016), indique que le thé est le troisième produit alimentaire le plus sujet à des dépassements de LMR pour les pesticides (22,9 % des échantillons analysés). Par exemple, l’imidaclopride (LMR de 50 µg/kg) a été détecté à 180 µg/kg en 2018, 130 mg/kg en 2016 et 230 mg/kg en 2015.

Une brève recherche sur le RASFF (période 2014-2017) renvoie quelques chiffres et données intéressants sur les problématiques sanitaires liées à la contamination chimique du thé. Ces informations sont résumées en Figure 1.3.

Les signalements dus aux contaminants chimiques ont entraîné en majorité des rejets à la frontière (104), mais également 7 alertes (émises lorsqu’un produit présente un risque sévère et qu’une action peut être requise dans un autre pays membre que celui ayant émis l’alerte) ainsi que des informations pour suivi (16, quand un risque a été identifié pour un produit placé sur le marché, mais que celui-ci ne requière pas une action rapide), des informations pour attention (4, quand le produit n’a pas été placé sur le marché en dehors du pays ayant émis l’alerte ou bien qu’il n’est plus sur le marché).

Tout d’abord, nous voyons que l’immense majorité des alertes du RASFF sur le thé concernent la présence d’une contamination chimique (131 cas sur 138). En pratique, un produit est signalé sur le RASFF dès qu’il y a un dépassement constaté de la teneur maximale de la substance retrouvée dans l’aliment concerné, ou la présence d’une molécule interdite. Parmi ces signalements, plusieurs molécules reviennent de façon récurrente. On peut citer pour exemple l’anthraquinone qui est utilisée comme adjuvant de produits phytosanitaires, et des pesticides tels que l’acétamipride, l’imidaclopride, le tolfenpyrad, la propargite et le diméthoate entre autres (teneurs moyennes de dépassement mesurées sur l’année 2017 : 200, 170, 320, 360 et 170 µg/kg respectivement). Cependant, ce sont des niveaux

« anormaux » (c.à.d. reportés en cas de dépassement des LMR) ne concernant qu'une petite minorité d'échantillons, les niveaux usuels ont donc une grande probabilité d'être plus faibles. On voit également apparaître sporadiquement des contaminations ponctuelles plus surprenantes par des molécules telles que la cocaïne ou le sildénafil (principe actif pharmaceutique), signe que l'évaluation de la qualité sanitaire d'un produit peut toujours révéler des surprises.

Quelques chiffres sur le Rapid Alert System for Food and Feed (RASFF) et le thé

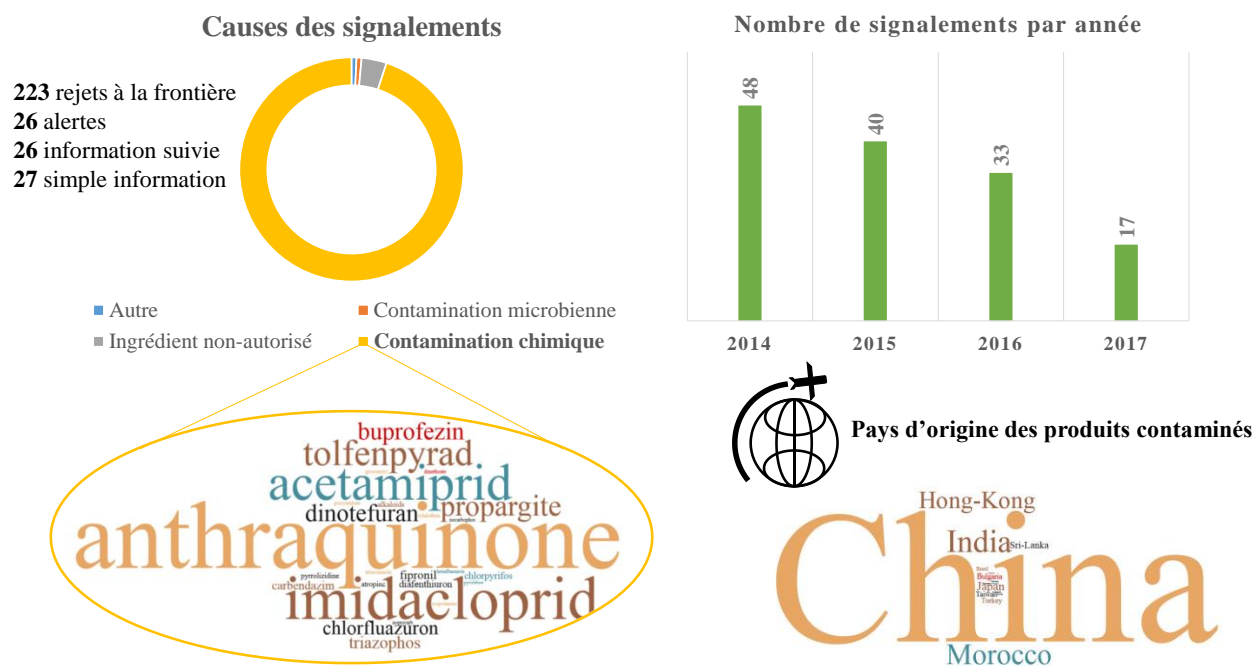


FIGURE 1.3 RÉSUMÉ DES PRINCIPALES INFORMATIONS FOURNIES PAR LE RASFF SUR LE THÉ (2014-2017)

De façon intéressante, le nombre de signalements pour le thé diminue de façon régulière depuis plusieurs années. On pourrait expliquer cela par une meilleure adaptation des entreprises exportant vers l'Europe au contexte réglementaire européen.

On peut également visualiser les pays d'origine des produits ayant fait l'objet d'un signalement. Sans surprise, les deux pays les plus concernés sont la Chine et l'Inde, ces pays étant les deux premiers producteurs mondiaux de thé. La présence d'Hong Kong et du Maroc dans les pays de provenance de denrées signalées s'explique par l'importance du premier dans le commerce mondial, et par le fait que le second exporte beaucoup de produits transformés vers l'Europe.

L'étude de la littérature nous apporte des renseignements sur les niveaux de contaminations dues à d'autres classes de composés. En ce qui concerne les mycotoxines, elles semblent présentes à de faibles niveaux dans les thés puisque Monbaliu et. al. (Monbaliu, Wu, Zhang, Van Peteghem, & De Saeger,

2010) n'ont quantifié qu'une seule mycotoxine (fumonisine B1 à 76 µg/kg dans un échantillon), le signal étant sous la limite de détection pour toutes les autres (11 µg/kg pour le déoxynivalénol, 25 µg/kg pour la fumonisine B2, 2,6 µg/kg pour l'ochratoxine A à titre d'exemple). Aucune donnée n'a été trouvée pour les migrants des matériaux au contact, et une publication peut être relevée concernant un contaminant issu de procédés, l'acrylamide (Liu, Zhao, Yuan, Chen, & Hu, 2008), pour lequel des niveaux compris entre 10 et 100 µg/kg sont rapportés.

2. L'ÉVALUATION DE LA QUALITÉ SANITAIRE CHIMIQUE DES ALIMENTS

Dans cette partie, nous nous intéresserons aux différentes approches analytiques permettant l'évaluation de la qualité sanitaire des aliments.

Les contaminants chimiques de l'aliment ont pour particularité de se trouver à l'état de traces, c'est-à-dire à des niveaux de l'ordre du µg/kg au mg/kg comme on l'a vu dans la partie 1 de ce chapitre. Par ailleurs, les aliments, et particulièrement notre produit d'étude (le thé), sont composés de plusieurs centaines de molécules différentes dans des quantités potentiellement importantes, de l'ordre du % ou de la dizaine de %, soit un rapport de plus de 1 000 par rapport aux potentiels contaminants.

L'analyse de composés traces (tels que les contaminants) dans les aliments a été rendue possible grâce aux développements des instruments analytiques modernes dans la deuxième moitié du 20^{ème} siècle (McGorin, 2009). Les techniques de choix pour ces analyses sont celles basées sur les sciences séparatives (chromatographie en phase liquide –LC– et chromatographie en phase gazeuse –GC– principalement) et leurs perfectionnements (chromatographie en phase liquide à ultra-haute performance, UHPLC par exemple). Ces techniques sont couplées à divers détecteurs qui peuvent être de type spectrométriques (détecteur UV-DAD par exemple), conductimétriques (détecteur à décharge corona par exemple), ou encore à la spectrométrie de masse (MS). Cette dernière technique est de nos jours la plus utilisée pour l'analyse de traces organiques, grâce à ses performances en termes de sélectivité, de sensibilité et de sa large gamme de molécules analysables (Frenich, Romero-González, & del Mar Aguilera-Luiz, 2014; Hird, Lau, Schuhmacher, & Krska, 2014). Par ailleurs, elle a l'avantage, comparée à ses concurrentes, de pouvoir analyser plusieurs centaines de composés simultanément et ce même dans des produits complexes et avec des coélutions chromatographiques. De plus, elle a la capacité de fournir des informations sur la structure des composés analysés, permettant la confirmation de l'identité de la molécule analysée, et limitant grandement de ce fait le risque de faux-positif. Enfin, les spectromètres de masse récents offrent une répétabilité et une robustesse comparable avec celles des autres détecteurs (Masiá, Suarez-Varela, Llopis-Gonzalez, & Picó, 2016).

Pour l'analyse des contaminants de l'aliment, des approches dites « ciblées » ont tout d'abord été développées. Comme leur nom l'indique, ces approches « ciblent » un certain nombre de composés

connus, pour lesquels on recherche souvent une information quantitative. La complexité de l'évaluation de la qualité sanitaire chimique des aliments (tous les contaminants n'étant pas forcément connus ou attendus), ainsi que l'amélioration des performances des instruments et l'essor de l'informatique dans les années 2000 a permis l'apparition de nouvelles approches dites « non-ciblées » permettant la détection de composés qui ne sont pas forcément attendus *a priori* (Antignac et al., 2011; Castro-Puyana et al., 2017; Hu & Xu, 2013).

2.1. APERÇU DES COMPOSÉS D'INTÉRÊT

Avant toutes choses, il est nécessaire d'avoir un aperçu des composés impliqués, tant au niveau des contaminants que des potentiels composés interférents. Le Tableau 1.3 présente une sélection de molécules (à la fois des contaminants et des composés interférents) susceptibles d'être rencontrées lors d'une analyse des contaminants du thé par LC-MS. De la composition du thé rapportée dans la littérature et présentée en section 1.4.2 de ce chapitre (croisée avec ce qui reporté dans la base de données Foodb), on a retenu trois molécules comme étant représentatives en termes de structure et de famille chimique des composés du thé susceptibles d'interférer avec nos composés d'intérêt. Ces composés représentent trois familles de molécules présentes en grande quantité dans le thé, à savoir les composés phénoliques (représentés ici par la catéchine), les acides aminés et dérivés (représentés par la théanine), et enfin les xanthines (représentées par la caféine).

Pour les contaminants, le choix d'une sélection la plus représentative possible a été plus délicate car la diversité de famille et de structure chimique est très grande comme on a pu le voir dans la partie 1. Le point de départ de cette sélection a été d'avoir des représentants du plus grand nombre de sources de contamination possible, ainsi que d'être le plus pertinent possible par rapport aux contaminants identifiés comme problématiques, notamment *via* les alertes rapportées sur le RASFF (partie 1.4.3 de ce chapitre). Par ailleurs, on ne s'intéresse qu'aux composés analysables avec la technologie LC-MS, qui est la technique prédominante pour l'analyse des contaminants de l'aliment (Frenich et al., 2014), bien que certaines familles de contaminants nécessitent l'emploi d'autres techniques de séparation ou de détection. On peut citer par exemple la famille des HAP non analysable en LC-MS.

On remarque que les contaminants et les composés de la matrice possèdent des poids moléculaires similaires, on ne pourra donc pas mettre en place une stratégie basée sur cette caractéristique pour séparer les deux. Par ailleurs, même si les deux catégories de composés semblent avoir des gammes de log Kow (log du coefficient de partage octanol/eau, représentatif du caractère lipophile d'un composé) relativement proches, on observe que certains contaminants (propargite, tolfenpyrad et ochratoxin A en particulier) présentent un caractère lipophile plutôt supérieur à celui observé pour les constituants du thé.

TABLEAU 1.3 CONTAMINANTS ET COMPOSÉS INTERFÉRENTS REPRÉSENTATIFS LORS D’UNE ANALYSE LC-MS DU THÉ

Nom	Abrév.	Famille chimique (<i>source</i>)	log K _{ow}	pKa	Formule brute	Masse monoisotopique	Niveaux usuels dans le thé (µg/kg)**
(a) Contaminants représentatifs (parmi ceux analysables en LC-MS)							
Acrylamide	AA	Amide (<i>néoformé</i>)	-0,67	N/A*	C ₃ H ₅ NO	71,0371	10-100 (Liu, 2008)
Hydroxyméthylfurfural	HMF	Furane (<i>néoformé</i>)	-0,09	N/A	C ₆ H ₆ O ₃	126,0317	N/A
Imidaclopride	IMID	Néonicotinoïde (<i>pesticide</i>)	0,57	N/A	C ₉ H ₁₀ ClN ₅ O ₂	255,0523	170 (RASFF)
Diméthoate	DIME	Organophosphate (<i>pesticide</i>)	0,7	N/A	C ₅ H ₁₂ NO ₃ PS ₂	228,9996	170 (RASFF)
Fumonisine B1	FB1	Fumonisine (<i>mycotoxine</i>)	-0,5	3,16	C ₃₄ H ₅₉ NO ₁₅	721,3885	76 (Monbaliu et al., 2010)
Aldicarbe	ALD	Carbamate (<i>pesticide</i>)	1,15	N/A	C ₇ H ₁₄ N ₂ O ₂ S	190,0776	N/A
Amétryne	AME	Triazine (<i>pesticide</i>)	2,63	4,1	C ₉ H ₁₇ N ₅ S	227,1205	<10 (H. Chen, Yin, Wang, Jiang, & Liu, 2014)
Dichlorvos	DIVO	Organochloré (<i>pesticide</i>)	1,9	N/A	C ₄ H ₇ Cl ₂ O ₄ P	219,9459	N/A
Diuron	DION	Phenylurée (<i>pesticide</i>)	2,87	N/A	C ₉ H ₁₀ Cl ₂ N ₂ O	232,0170	N/A
Ochratoxine A	OTA	Ochratoxine (<i>mycotoxine</i>)	4,74	7,1	C ₂₀ H ₁₈ ClNO ₆	403,0823	< 2,6 (Monbaliu, 2010)
Acide 2,4-Dichlorophenoxyacétique	2,4-D	Herbicide auxinique (<i>pesticide</i>)	-0,82	2,73	C ₈ H ₆ Cl ₂ O ₃	219,9694	N/A
Tolfenpyrad	TOLF	Pyrazole (<i>pesticide</i>)	5,61	4,1	C ₂₁ H ₂₂ ClN ₃ O ₂	383,1401	320 (RASFF)
Propargite	PROP	Organosulfite (<i>pesticide</i>)	5,7	N/A	C ₁₉ H ₂₆ O ₄ S	350,1552	360 (RASFF)
Fenoprop	2,4,5-TP	Herbicide auxinique (<i>pesticide</i>)	2,84	2,84	C ₉ H ₇ Cl ₃ O ₃	267,9461	N/A
Bisphénol A	BPA	Bisphénol (<i>migrant de matériaux au contact</i>)	3,3	10,3	C ₁₅ H ₁₆ O ₂	228,1150	50 (Gao et al., 2017)
(b) Composés interférents représentatifs							
Catéchine	CAT	Flavonoïdes (<i>composé phénolique</i>)	1,8	9	C ₁₅ H ₁₄ O ₆	290,0790	2,6 g/kg (Foodb)
Théanine	THE	Glutamines et dérivés (<i>acide aminé</i>)	-2,8	2,16	C ₇ H ₁₄ N ₂ O ₃	174,1004	3,9 g/kg (Foodb)
Caféine	CAF	Xanthines	-0,255	-0,92	C ₈ H ₁₀ N ₄ O ₂	194,0804	80 mg/kg (Foodb)

* N/A = Non applicable

** Pour les contaminants, niveau rapporté sur le RASFF ; si non rapporté, niveau trouvé dans la littérature

Nous allons donc maintenant voir les méthodes et stratégies analytiques permettant l’analyse de ces contaminants. On verra d’abord les approches ciblées, puis les approches non-ciblées seront abordées.

2.2. APPROCHES CIBLÉES

Les approches ciblées peuvent se découper en deux grandes phases :

- i) L'échantillon subit tout d'abord une étape de traitement, à la suite de laquelle la ou les molécules d'intérêt (appelées « analytes »), ainsi qu'éventuellement un certain nombre de composés interférents se trouveront sous une forme analysable (le plus souvent une solution). Ce traitement peut également permettre de concentrer ou de purifier les molécules d'intérêt en fonction de la technique utilisée par la suite, de l'échantillon à analyser ou des niveaux recherchés.
- ii) L'extrait final est ensuite analysé grâce à une méthode instrumentale (souvent un couplage entre une séparation chromatographique et un détecteur).

Un tableau récapitulatif de publications notables portant sur le développement de méthodes d'analyse ciblées de contaminants dans le thé a été réalisé (Tableau 1.4). Plusieurs catégories de travaux ont été sélectionnées pour faire ce récapitulatif. Dans chacune de ces catégories, 2 à 3 publications considérées comme représentative de l'état de l'art ou bien d'un intérêt particulier ont été sélectionnées. Seules les publications utilisant la technique de la LC-MS ont été considérées.

La première catégorie rassemble le développement de méthodes d'analyse mono-résidu de composés spécifiques (par exemple ici l'acrylamide ou le bisphénol A) dans le thé. La deuxième catégorie regroupe des publications étudiant le développement de méthodes pour analyser d'une famille de composés chimiques, aussi bien des molécules de synthèse (par exemple une classe de pesticide dans le thé sous forme d'infusion ou de feuilles) que des contaminants d'origine naturelle (par exemple les mycotoxines). La troisième est composée de publications proposant une méthode de quantification de plusieurs familles de pesticides dans le thé sous ses deux formes. La quatrième rassemble des publications ne traitant pas spécifiquement du thé (dans lesquelles il peut être étudié), mais proposant un point de vue critique sur un des aspects du processus analytique (traitement de l'échantillon ou instrumentation) avec plusieurs classes de contaminants (au moins 3) sur plusieurs matrices différentes.

Les niveaux analysés pour les pesticides se situent aux alentours des plus basses LMR existantes (c.à.d. entre 1 et 10 µg/kg de produit). Pour les mycotoxines, les niveaux analysés sont du même ordre de grandeur que ceux des pesticides (de l'ordre du µg/kg jusqu'à la centaine du µg/kg). Pour les contaminants naturels autres que les mycotoxines (qui ne sont pas réglementés spécifiquement pour le thé), les niveaux usuels dans le thé sont très mal connus (Mulder et al., 2015), les niveaux analysés sont donc très bas [jusqu'à 0,05 µg/kg (Chung & Lam, 2018)].

Le choix de la méthode de préparation d'échantillon dépendant en partie de l'instrument utilisé. On commencera donc par présenter les principaux couplages chromatographie-spectrométrie de masse utilisés pour l'analyse des contaminants de l'aliment (dont le couplage LC-MS en particulier), puis les méthodes de préparation de l'échantillon seront abordées.

TABLEAU 1.4 APERÇU DES TRAVAUX EXISTANTS SUR LES APPROCHES CIBLÉES D'ANALYSE DE CONTAMINANTS

Catégorie	Analytes	Matrice	Extraction	Purification	Chromatographie	Spectrométrie de masse	Méthode de calibration	Niveaux analysés (µg/kg)	Référence
Molécule unique dans le thé	BPA ou TBBPA	Thé (feuilles)	LLE (H ₂ O/ACN + 0,7% acide acétique) assisté de MgSO ₄ et NaCl	dSPE (C18, nanotubes de carbone, et échange d'ions)	Phase stationnaire : C18 (UHPLC) Phase mobile : H ₂ O / MeOH à 0,25 mL/min	Source : ESI (-) Analyseurs : QqQ opéré en MRM	En matrice	BPA 8/20 TBBPA 0,4/10	Gao et. al., 2017
	Acrylamide	Thé (feuilles)	LLE (H ₂ O/ACN avec 4g MgSO ₄ et NaCl)	Concentration par évaporation puis SPE (mode d'échange d'ions mixte), test de l'élution avec H ₂ O, MeOH, ACN, acétone et isopropanol	Phase stationnaire : C18 (HPLC) Phase mobile : ACN et H ₂ O avec 0,1% FA à 0,4 mL/min	Source : ESI (+) Analyseur : QqQ opéré en MRM	En solvant et en matrice	1	Liu, J et. al., 2008
Une classe de contaminants (contaminants de source naturelle)	28 alcaloïdes pyrrolizidiniques et dérivés	Thé (infusion)	Dilution (H ₂ SO ₄ , 0,05 M dans H ₂ O) ajustement du pH puis SPE (-C18), lavage H ₂ O puis ACN, élution séparée avec MeOH puis MeOH avec 2,5% NH ₄ OH, éluats tamponnés puis évaporés, combinés et filtrés		Phase stationnaire : C18 (UHPLC) Phase mobile : ACN et H ₂ O avec 0,5 mM/L de formate d'ammonium et 2 mM/L de FA à 0,5 mL/min	Source : ESI (+) Analyseur : Q-Trap en mode MRM	En solvant	0,05	Chung, S, 2018
	28 mycotoxines	Thé et tisanes (feuilles et infusion)	Feuilles: Extraction par solvant (acétate d'éthyle acidifié avec 1% de FA) Infusion: 2g de feuilles pendant 2 min dans 200 mL d'H ₂ O, prélèvement de 5mL extraits par SPE (C18), élution avec MeOH, évaporation de l'extrait et reconstitution dans H ₂ O/MeOH (60/40) acidifié avec 0,3% FA	Concentration par évaporation puis séparation de 2 extraits: Extrait 1: SPE (NH ₂), récupération directe de l'éluat, redilution dans H ₂ O/ACN (16/84) Extrait 2: SPE (C18) conditionnée avec ACN/H ₂ O (84/16), récupération directe, passage de fraction 1 sur la même colonne et récupération du tout et évaporation. Reconstitution dans H ₂ O/MeOH (60/40) acidifié avec 0,3% FA	Phase stationnaire : C18 (UHPLC) Phase mobile : H ₂ O et MeOH (0,3% de FA pour chaque) à 0,55 mL/min	Source : ESI (+) Analyseur : QqQ opéré en MRM	En matrice	4-200 (feuilles) 1-100 (infusion)	Monbaliu et. al., 2010
Une classe de contaminants (contaminants de source synthétique)	7 benzylurées	Thé oolong (feuilles + infusions)	Feuilles : USE (ACN, 50°C) Infusion : 5g de feuilles dans 110 mL d'eau bouillante pendant 3 min, puis 50 mL prélevés et LLE avec 15 mL de DCM (x3), évaporation à 1 mL	Feuilles: SPE (silicate de magnésium et charbon actif), élution avec ACN/toluène (25/75), évaporation/reconstitution dans MeOH Infusion : même SPE/concentration que pour les feuilles	Phase stationnaire : -C18 (UHPLC) Phase mobile : MeOH et H ₂ O (0,002 mol/L acétate d'ammonium) à 0,3 mL/min	Source : ESI (-) Analyseur : QqQ opéré en mode "full MS2"	En solvant	Plus bas niveaux testés 12,5 (feuilles) et 2,13 (infusions)	Chen, L et. al., 2014
	8 néonicotinoïdes	Thés (6 types, dont vert, noir et blanc, feuilles)	LLE : H ₂ O puis ACN en présence de NaSO ₄	dSPE avec PVPP, GCB et MgSO ₄ évaporé à 40°C sous N ₂ et reconstitution (ACN/H ₂ O 15/85)	Phase stationnaire : C18 (UHPLC) Phase mobile : H ₂ O (5mM formate d'ammonium)/ACN à 0,2 mL/min	Source : ESI (+) Analyseur : QqQ opéré en MRM	En solvant couplé avec une dilution de l'échantillon	10	Jiao, W et al., 2016

G.Delaporte - Développement d'une approche non-ciblée par empreinte pour caractériser la qualité sanitaire chimique de matrices agro-alimentaires complexes

Pesticides multi-familles sur le thé (infusion et/ou feuilles)	146 pesticides (multi-classes)	Thé vert (feuilles)	LLE (ACN et H ₂ O + NaCl et tampon phosphate à pH 7)	SPE (C18), conditionnement et élution ACN, évaporation de l'éluat et dissolution dans ACN/Toluène (3:1), deuxième SPE GCB/PSA (conditionnement et élution ACN/Toluène 3:1), évaporation et reconstitution dans MeOH	Phase stationnaire : C18 (UHPLC) Phase mobile : H ₂ O / MeOH (5 mM d'acétate d'ammonium dans chaque) à 0,3 mL/min	Source : ESI (+) Analyseurs : Q-ToF et Q-Orbitrap opérés en balayage complet	En matrice	10	Saito-Shida et. al., 2018
	172 pesticides (multi-classes)	Thés variés (vert, noir, oolong, tisane, feuilles et infusions)	Feuilles: LLE (H ₂ O et ACN acidifié avec 1% acide acétique, assistés par NaOAc et MgSO ₄) Infusion: 5g dans 150 mL d'H ₂ O bouillante, LLE avec ACN acidifié avec 1% d'acide acétique assistés par NaOAc et MgSO ₄	Feuilles: dSPE (MgSO ₄ , GCB, PSA), évaporation et reconstitution H ₂ O (0,1 mM acétate d'ammonium)/MeOH Infusion: dSPE (MgSO ₄ , C18, PSA), évaporation et reconstitution H ₂ O (0,1 mM acétate d'ammonium)/MeOH	Phase stationnaire : C18 (UHPLC, core-shell) Phase mobile : H ₂ O (10 mM acétate d'ammonium) / ACN à 0,3/0,5 mL/min (gradient de débit)	Source : ESI (+) Analyseurs : QqQ opéré en MRM	En matrice	4	Wang et. al., 2014
Multi-classes (au moins 3) dans plusieurs matrices	323 pesticides, 55 mycotoxines, 11 toxines de plantes	Thé (feuilles), poireau, blé	LLE : H ₂ O puis ACN assistés par MgSO ₄ et NaCl	dSPE (C18)	Phase stationnaire : C18 (UHPLC) Phase mobile : ESI+ : H ₂ O + 5mM formate d'ammonium et MeOH (chacun acidifié avec 0,1% FA) ESI-: H ₂ O et MeOH chacun avec 5mM d'acétate d'ammonium Débit de 0,4 mL/min	Source : ESI (+/-) Analyseur : Q-Orbitrap en mode balayage complet et dd-MS/MS	En matrice	<10	Dzuman et. al., 2015
	625 contaminants (426 pesticides, 117 produits pharmaceutiques ou vétérinaires, 42 migrants de l'emballage, 10 composés perfluorés, 21 mycotoxines, 9 nitrosamines)	Pots pour bébés (végétal et animal, dopage de l'extrait)	Extraction par solvant directe (ACN + 0,1% FA, assistée par NaCO ₃ H et MgSO ₄)	dSPE (PSA et MgSO ₄), évaporation et reconstitution dans MeOH/H ₂ O (20/80)	Phase stationnaire : C18 (UHPLC) Phase mobile : H ₂ O et ACN (0,1% de FA pour chaque) à 0,5 mL/min	Source : ESI (+/-) Analyseur : Q-ToF opéré en balayage complet	Qualitatif	10	Pérez-Ortega et. al., 2016
	Développement sur 172 contaminants (pesticides, mycotoxines, toxines végétales) Application à 258 contaminants (136 pesticides, 36 toxines naturelles et 86 produits vétérinaires)	6 matrices solides et liquides (aliment pour chevaux, maïs, viande, lait, œuf, miel)	Extraction directe par ACN (+0,1% FA) sans purification 7 protocoles comparés (dont QuEChERS et plusieurs méthodes similaires)		Phase stationnaire : C18 (UHPLC) Phase mobile : H ₂ O /MeOH (1mM formate d'ammonium et 20 µL/L de FA dans chaque) à 0,4 mL/min	Source : ESI (+/-) Analyseurs : QqQ opéré en MRM et Q-Orbitrap opéré en balayage complet	En matrice	10-100	Mol et. al. 2008

TBBPA : Tétrabromobisphénol A ; LLE : Extraction Liquide/Liquide ; ACN : Acétonitrile ; MRM : Multiple Reaction Monitoring ; FA : Acide Formique ; DCM : Dichlorométhane ; USE : Extraction assistée par ultrasons ; PVPP : Polyvinylpyrrolidone ; GCB : carbone graphite ; PSA : Amines secondaires et primaires

2.2.1. PRINCIPE GÉNÉRAL DES COUPLAGES CHROMATOGRAPHIE-SPECTROMÉTRIE DE MASSE

Les principales étapes d’une analyse par un couplage entre une technique chromatographique et la spectrométrie de masse sont présentées en figure 1.4.

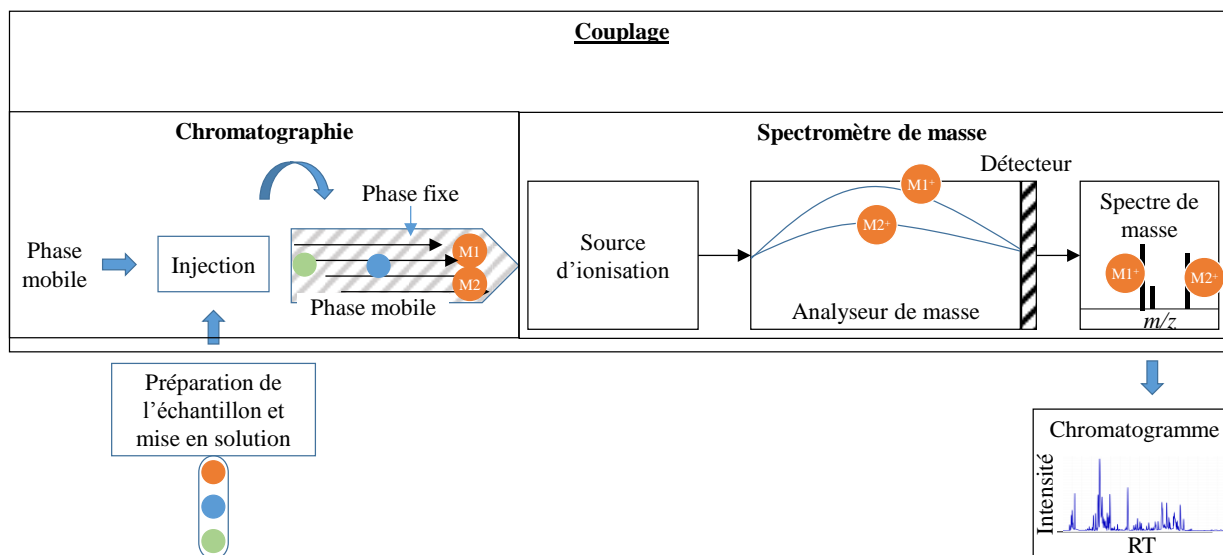


FIGURE 1.4 PRINCIPALES ÉTAPES D'UNE ANALYSE PAR COUPLAGE CHROMATOGRAPHIE/SPECTROMÉTRIE DE MASSE

Après la préparation et la mise en solution de l'échantillon à analyser, une certaine quantité de l'extrait obtenu (pouvant varier selon les instruments, souvent de l'ordre de la dizaine de μL sur les instruments récents) est injectée dans le système. Les différents constituants du mélange sont ensuite séparés par chromatographie puis envoyés vers le spectromètre de masse. Pour pouvoir être analysés à l'aide de ce dernier, les composés chimiques présents doivent se trouver sous forme d'ions, le spectromètre de masse séparant les molécules selon leurs rapports masse sur charge (m/z). Les analytes sont donc ionisés dans une source d'ionisation, dont la nature dépend à la fois de la technique chromatographique utilisée et des propriétés physico-chimiques des analytes. Les ions formés sont ensuite séparés suivant leurs rapports m/z dans un analyseur de masse, et les signaux correspondants récupérés grâce au détecteur. Un fichier de données en trois dimensions (temps de rétention –RT–, rapport masse/charge – m/z – et intensité), duquel sont extraits les spectres de masse et des chromatogrammes sont enfin générés par un logiciel en vue de leur exploitation.

2.2.1.1. TECHNIQUES CHROMATOGRAPHIQUES

La chromatographie est une technique permettant de séparer les composants d'un mélange suivant leurs affinités avec deux phases non-miscibles appelées phase mobile et phase stationnaire (la phase mobile passant à travers la phase stationnaire). Le principe général d'une séparation chromatographique qui se

trouve en figure 1.4 est le suivant : les différents composés (appelés « analytes ») se trouvant dans l'échantillon sont mélangés à la phase mobile, constituée d'un fluide, et cette dernière passe à travers une phase stationnaire non-miscible avec la phase mobile. Certains composés vont se trouver plus ou moins retenus par la phase stationnaire en fonction de leurs propriétés physico-chimiques et de leur structure chimique, et ils vont ainsi être séparés. D'une manière générale, la rétention d'un composé dépend de ses interactions respectives avec la phase stationnaire et la phase mobile, mais aussi éventuellement des interactions des phases stationnaires et mobiles entre elles (Caude & Jardey, 1996).

On peut tout d'abord classer les techniques chromatographiques selon la nature des interactions provoquant la séparation des analytes. On distingue ainsi la chromatographie d'adsorption (interactions entre les analytes et la surface de la phase stationnaire), la chromatographie de partage (basée sur la distribution des analytes entre la phase mobile et la phase stationnaire), la chromatographie ionique (basée sur des échanges d'ions entre les analytes, la phase stationnaire et la phase mobile) et enfin la chromatographie d'exclusion stérique (basée sur la différence de taille entre les analytes lors de leur passage dans une phase stationnaire à la taille de pores contrôlée) (Christian, Dasgupta, & Schug, 2014). On peut aussi faire une distinction par la nature des phases mobiles utilisées. Il existe ainsi la chromatographie en phase liquide (LC), la chromatographie en phase gazeuse (GC) et la chromatographie en phase supercritique (SFC) (Caude & Jardey, 1996).

Le choix du type de chromatographie se fait par rapport à la nature des analytes et à leurs concentrations attendues. Pour les thématiques de sécurité sanitaire chimique des aliments (analytes de petite taille, moyennement polaires à apolaires), les chromatographies de partage sont les techniques les plus utilisées (Christian et al., 2014).

Parmi celles-ci, une rapide recherche sur la plateforme de référencement d'articles scientifiques « Web of Science » montre que les techniques en phase gazeuse (GC) et en phase liquide (LC) sont très largement majoritaires par rapport à la chromatographie en phase supercritiques (SFC) comme montré dans le Tableau 1.5. On se concentrera donc pour le moment sur les deux premières techniques (LC et GC).

La chromatographie de partage en phase gazeuse utilise un gaz comme phase mobile (typiquement He, N₂ ou H₂), et une phase stationnaire le plus souvent composé d'un capillaire de silice fondue dont la surface externe est recouverte d'une couche de polymère de protection et la surface interne est recouverte d'un film d'un liquide non-volatile ou d'un gel fonctionnalisé (Christian et al., 2014). La nature de cette phase stationnaire est ce qui va le plus influencer la séparation, la principale variable d'ajustement étant sa polarité. Les colonnes utilisées sont très fines (de l'ordre de 0,1-0,5 mm de diamètre interne) et très longues (entre 10 et 100 m). Les interactions avec la phase mobile sont négligeables, cette dernière étant un gaz inerte (Caude & Jardey, 1996) et servant uniquement de vecteur. Les analytes doivent se trouver en phase gazeuse dans la colonne et donc soit introduits directement sous forme gazeuse, soit sous forme

liquide. Dans ce dernier cas, il est nécessaire de les vaporiser pour faire passer les analytes en phase gazeuse. Cela limite l’application de cette technique à des molécules volatiles ou au moins non-thermolabiles (c.à.d. stables à haute température), les températures rencontrées lors de la vaporisation de l’échantillon pouvant atteindre plusieurs centaines de degrés (Tranchant, 1996)

TABLEAU 1.5 NOMBRE D’OCCURRENCES TROUVÉES SUR LA PLATEFORME "WEB OF SCIENCE" POUR LES TROIS GRANDS TYPES DE CHROMATOGRAPHIE (RECHERCHE AU 18/07/2018, PÉRIODE 1956-2018)

Mots clés et opérateurs logiques (plateforme Web of Science)	Nombre de résultats
(LC OU liquid chromatography OU HPLC OU UHPLC OU UPLC) ET (safety OU contaminants) ET food	3 361
(GC OU gas chromatography) ET (safety OU contaminants) ET food	1 573
(SFC OU supercritical fluid chromatography) ET (safety OU contaminants) ET food	66

Dans les applications de sécurité sanitaire chimique des aliments, les composés types analysés grâce à cette technique sont les pesticides apolaires (Li et al., 2013), les polluants organiques persistants tels que les PCB (Bayen, Koroleva, Lee, & Obbard, 2005) ou encore les HAP (Grover, Singh, & Pal, 2013). Certains analytes thermolabiles ou non-volatils peuvent être rendus analysables en GC grâce à une dérivation chimique consistant à opérer une réaction chimique sur l’analyte pour modifier ses propriétés dans le sens voulu. Par exemple, une silylation permet d’augmenter la volatilité d’une molécule (Christian et al., 2014). Un des exemples les plus connus de cette approche en analyse agro-alimentaire est l’analyse des acides gras par GC grâce à une dérivation (Seppänen-Laakso, Laakso, & Hiltunen, 2002). Faire une dérivation suppose donc qu’il existe une fonctionnalité commune à tous les analytes ciblés pour pouvoir effectuer la réaction de façon sélective sur ces dernières. Cette condition est donc souvent incompatible avec l’analyse multi-classes des contaminants de l’aliment qui possèdent une grande diversité de structures chimiques comme on a pu le voir dans la partie 1.4.3 de ce chapitre.

Pour la chromatographie de partage en phase liquide la phase mobile est un solvant (ou un mélange de solvant), et la phase stationnaire un gel (le plus souvent de silice) pouvant être fonctionnalisé. On distingue deux grands « modes » de chromatographie de partage en phase liquide, le mode dit « normal » (phase stationnaire polaire, phase mobile apolaire à modérément polaire) ou le mode dit « inverse » (phase stationnaire apolaire, phase mobile polaire à modérément apolaire). Actuellement, le mode « inverse » est largement prédominant par rapport au mode « normal » (Christian et al., 2014). Pour les applications à l’analyse des contaminants de l’aliment (donc des petites molécules apolaires ou modérément polaires), c’est la technique de choix comme on peut le voir dans le Tableau 1.4.

En LC de « phase inverse », la séparation s'effectue sur une colonne remplie de particules (de silice – SiOH– le plus souvent, plus rarement de carbone graphite ou d'alumine) greffées avec des fonctions chimiques (le greffon le plus commun étant une chaîne octadécyle avec 18 carbones) et à la granulométrie contrôlée (des granulométries allant de 1,7 à 5 µm sont utilisées actuellement pour les applications analytiques). Les colonnes de silice greffées avec des chaînes C18 sont devenues la norme en chromatographie de « phase inverse », bien qu'il existe d'autres types de greffons comme les C8 (moins hydrophobes que les C18) ou bien les phényles (apportant des interactions π - π) entre autres, ces greffons pouvant être combinés pour ajuster la rétention des colonnes.

La chromatographie liquide peut s'opérer soit à composition de solvant constante (mode « isocratique »), ou faire varier cette dernière au cours du temps (mode « gradient »), ce dernier mode étant le plus utilisé pour la séparation de mélanges complexes car il permet la séparation de composés possédant une large plage de polarité (Christian et al., 2014). Dans un gradient chromatographique, la force éluante (=la capacité d'un mélange de solvant à éluier un composé piégé dans la colonne) doit être croissante au cours du temps pour permettre la séparation des constituants d'un mélange. En chromatographie liquide de « phase inverse », un mélange d'une solution aqueuse (appelée « phase aqueuse ») et d'une solution organique (appelée « phase organique », typiquement MeOH ou ACN) est couramment utilisé. Les gradients de chromatographie de « phase inverse » commencent systématiquement par une certaine proportion de phase aqueuse (celle-ci étant moins éluante que la phase organique) qui diminue au cours du temps (augmentant donc la force éluante du mélange passant dans la colonne). Des modifiants peuvent être ajoutés dans les deux phases pour contrôler le pH dans la colonne ou bien encore assister la détection des analytes (Christian et al., 2014).

Il existe un troisième mode de chromatographie de partage qui est la chromatographie d'interactions hydrophiles (HILIC), qui utilise une phase stationnaire et une phase mobile polaires. C'est une technique utilisée pour l'analyse de composés très polaires, mais elle n'a pas été reportée pour des analyses de contaminants à spectres larges. Elle ne sera donc pas discutée ici.

Historiquement, la principale technique des séparations utilisée dans l'analyse des contaminants de l'aliment par couplage est la GC, car c'est la première à avoir fait l'objet d'un couplage à la spectrométrie de masse (Masiá et al., 2016). Cependant, l'essor des techniques de LC-MS observé dans les années 1990 a rendu cette dernière technique de plus en plus populaire, car elle permet à la fois d'analyser des composés non-analysables en GC (comme les composés les plus polaires et ceux thermolabiles), mais aussi une grande partie de ceux pour lesquels la GC est traditionnellement utilisée (Masiá et al., 2016). Par ailleurs, les récents développements en LC, avec notamment l'apparition d'appareils dits à « Ultra Hautes Performances » (UHPLC) ont permis de raccourcir les temps d'analyses (un cycle d'analyse complet en UHPLC peut prendre moins de 10 minutes en routine) tout en gardant la même capacité de rétention (Christian et al., 2014), alors que pour les couplages GC, la durée d'un cycle d'analyse est souvent compris entre 30 minutes et 1 heure. Tout cela, couplé à l'important développement des

techniques LC-MS a fait que cette technologie est devenue la technologie centrale dans l'analyse des contaminants de l'aliment à l'heure actuelle (Masiá et al., 2016). Par la suite, nous nous concentrerons donc uniquement sur les applications utilisant la LC-MS, les autres technologies comme la GC-MS n'étant mentionnées qu'à titre informatif.

2.2.1.2. DÉTECTION PAR SPECTROMÉTRIE DE MASSE

2.2.1.2.i. Interfaces et sources d'ions

Le rôle de la source d'ions est de transformer les analytes (initialement sous forme neutre la plupart du temps) en ions. Son rôle est également de faire l'interface entre un milieu à une pression proche de celle de l'atmosphère ou supérieure (la sortie de la séparation chromatographique) et un milieu sous vide poussé (le spectromètre de masse). Le choix de la source est dépendant de la technique de séparation utilisée avant, ainsi que de la nature des analytes.

Dans le cas d'une chromatographie en phase gazeuse, il existe deux principales sources, les sources par ionisation électronique (EI), et les sources par ionisation chimique (CI) (Christian et al., 2014). La première, qui est la plus répandue de par sa simplicité et son coût modéré par rapport aux autres analyseurs de masse, est connue pour donner lieu à des phénomènes de fragmentation des analytes importants. Cette fragmentation est reproductible et permet la construction de bases de données fiables (Christian et al., 2014). La seconde donne lieu à plus ou moins de fragmentation suivant la nature du réactif utilisé, les réactifs préférés étant souvent ceux ne générant que peu de fragmentation et permettant donc l'obtention d'ions pseudo-moléculaires (Christian et al., 2014). Ces deux sources se trouvent sous pression réduite.

Dans le cas de la LC, les analytes sont des composés souvent non-volatils et potentiellement thermolabiles, et il est de plus nécessaire de se débarrasser du solvant de la phase mobile. Les sources d'ions utilisées après une LC sont à pression atmosphérique (Christian et al., 2014). Il en existe deux principales : l'ionisation par électrobulbion (couramment désignée par le terme anglais d'ionisation par « electrospray », abrégé en « ESI ») et l'ionisation chimique à pression atmosphérique (abrégée en « APCI »). Les deux sont des sources d'ions dites « douces », formant des ions pseudo-moléculaires (appelés « adduits ») et ne provoquant pas de forts phénomènes de fragmentation (Christian et al., 2014).

L'ionisation de composés en solution par ESI est obtenue en appliquant une importante différence de potentiel (plusieurs kV) entre un capillaire où passe l'échantillon à faible débit (typiquement de l'ordre du $\mu\text{L}/\text{min}$ à la centaine de $\mu\text{L}/\text{min}$) et l'entrée du spectromètre de masse (le plus souvent un cône métallique percé). Le sens de cette différence de potentiel détermine la polarité d'ionisation (mode positif ou mode négatif). L'ionisation se fait en solution.

Cette différence de potentiel va permettre la formation d'un « cône de Taylor » et la formation de gouttelettes chargées de plus en plus petites et *in fine* la formation et le passage d'ions en phase gazeuse. Le mécanisme de formation de ces ions est toujours discuté, avec plusieurs modèles théoriques proposés (Christian et al., 2014; Kebarle & Peschke, 2000). Cette désolvatation est assistée par un flux de gaz

neutre (le plus souvent N_2) ainsi que par une température relativement élevée (moins cependant que dans une source destinée à la GC), de l'ordre de la centaine de degrés Celsius en source (certaines zones pouvant être plus chaudes). Les spectres produits par une source ESI présentent très peu de fragmentation et beaucoup d'adduits (les plus courants sont $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$ en mode positif et $[M-H]^-$ et $[M+Cl]^-$ en mode négatif).

L'ESI est un procédé d'ionisation compétitif, c.à.d. que les différentes espèces présentes dans la solution sont en concurrence pour être ionisées. C'est pourquoi de forts effets de suppression (ou plus rarement d'augmentation) de signal sont souvent observés dans ce mode lors d'un changement de milieu (typiquement entre une injection en solvant et une injection d'un échantillon), particulièrement pour les milieux complexes (Christian et al., 2014). L'ESI est également très sensible aux contaminants interférents couramment rencontrés tels que les tensio-actifs (Christian et al., 2014) et nécessite donc des précautions particulières d'utilisation pour s'en prévenir, en particulier sur les consommables utilisés. Pour obtenir des résultats quantitatifs malgré ce phénomène de compétition, il est donc nécessaire *a minima* d'estimer l'ampleur de ces phénomènes sur les analytes, ainsi qu'éventuellement de les compenser soit grâce à l'utilisation de standards internes d'analytes marqués avec des isotopes stables (^{13}C ou deutérium classiquement), soit par l'utilisation d'un milieu proche des échantillons réels pour effectuer la courbe d'étalonnage (technique dite de la calibration en matrice). En ESI, l'ionisation des analytes peut être assistée par des additifs présents dans la phase mobile ou bien co-introduits en source avec celle-ci. Il s'agit souvent de sels favorisant la formation d'adduits (par exemple $CH_3CO_2-NH_4$, acétate d'ammonium favorisant la formation d'adduit $[M+NH_4]^+$). Contrôler le pH de la phase mobile pour favoriser la dissociation en solution (et donc l'ionisation) des analytes est un autre moyen d'assister l'ionisation (Christian et al., 2014). Enfin, l'ESI peut donner lieu à des composés multichargés (par exemple des adduits $[M+2H]^{2+}$), mais cela n'est pas discuté ici car cela concerne surtout les grandes molécules polaires (typiquement les protéines ou bien certains polymères) et très peu les petites molécules tels que les contaminants de l'aliment (Christian et al., 2014).

L'autre source d'ions couramment utilisée est l'APCI, qui est une source dans laquelle les ions sont formés en phase gazeuse, contrairement à l'ESI dans laquelle les ions sont formés en solution. La source APCI est une source ESI qui a été modifiée par l'ajout d'une aiguille à décharge Corona au niveau du spray. La différence de potentiel est appliquée, non entre le capillaire et l'entrée du spectromètre comme pour l'ESI, mais entre l'aiguille et l'entrée du spectromètre. C'est une technique souvent utilisée pour les composés trop apolaires pour être ionisés par ESI (Carroll, Dzidic, Horning, & Stillwell, 2006).

Pour les applications de sécurité sanitaire chimique des aliments l'utilisation de sources ESI est très largement majoritaire comme on peut le voir dans le Tableau 1.6. Par ailleurs, comme indiqué dans le Tableau 1.4, l'ESI a déjà été éprouvée pour l'analyse simultanée de plusieurs classes de contaminants de l'aliment à des niveaux pertinents.

TABLEAU 1.6 NOMBRE D’OCCURRENCES RELEVÉES SUR LA PLATEFORME "WEB OF SCIENCE" POUR L’ESI ET L’APCI (RECHERCHE AU 18/07/2018, PÉRIODE 1956-2018)

Mots clés (plateforme Web of Science)	Nombre de résultats
(ESI OU electrospray) ET (safety OU contaminants) ET food	366*
(APCI OU atmospheric pressure chemical ionisation) ET (safety OU contaminants) ET food	44

* Nombre très certainement sous-estimé car il s’agit de la technique « par défaut » et donc parfois non précisée dans les mots-clés de recherche

2.2.1.2.ii. Analyseurs de masses

Une fois les ions formés, ils sont séparés selon leurs rapports m/z grâce à l’analyseur de masse. Deux principales grandeurs servent à caractériser les performances d’un analyseur de masse : le pouvoir de résolution et l’erreur de mesure de masse.

Le pouvoir de résolution d’un analyseur de masse est sa capacité à distinguer deux masses, et est calculé comme le rapport entre la masse mesurée et la différence minimale de masse mesurable ($R = \frac{m}{\Delta m}$).

Δm est souvent mesurée à mi-hauteur du pic (FWHM « full-width half maximum » en anglais). Par exemple, une résolution de 20 000 à m/z 200 va permettre de distinguer la masse 200,00 de la masse 200,01.

L’erreur de mesure de masse est l’erreur relative, exprimée en partie par millions (ppm), entre la masse mesurée et la masse exacte théorique. Elle est calculée par l’équation suivante :

$$\text{erreur de masse} = \frac{m_{\text{mesurée}} - m_{\text{théorique}}}{m_{\text{théorique}}} \times 10^6$$

Le choix de l’analyseur de masse est fonction des performances attendues (en termes de résolution par exemple), ainsi que des caractéristiques de la séparation (largeur des pics chromatographiques, nombre de points d’acquisition nécessaires entre autres).

Les caractéristiques des principaux analyseurs de masse couramment couplés à la chromatographie sont montrées dans le Tableau 1.7.

Le plus courant des analyseurs de masse est le quadripôle (Q). Cet analyseur basse résolution ($R \sim 10^3$) est constitué de quatre électrodes cylindriques ou hyperboliques disposées parallèlement et auxquelles sont appliquées des combinaisons de tensions continues et oscillantes. La composante continue varie au cours du scan et sélectionne successivement les différents ions (Rondeau, 2017). En faisant varier la tension au cours du temps, on peut ainsi acquérir un spectre de masse. Ces analyseurs sont relativement compacts, peu coûteux et possèdent des fréquences d’acquisition compatibles avec les conditions chromatographiques (voir Tableau 1.7). Le temps passé sur chaque ion détermine la sensibilité d’un quadripôle. Plus le nombre d’ions à mesurer est faible, plus il sera sensible à fréquence d’acquisition

constante. Par ailleurs, il possède une bonne gamme dynamique (« rapport entre le plus grand et le plus petit signal détectable pour un ion » (Rondeau, 2017)), ce qui en fait un analyseur prisé pour les applications quantitatives. Cet analyseur est également très utilisé pour guider les faisceaux d'ions au sein des instruments : il est alors placé en mode radiofréquence, la composante de tension continue est supprimée, ce qui a pour effet de focaliser le faisceau d'ions au centre des 4 électrodes (Christian et al., 2014). Il est également utilisé en complément d'autres analyseurs de masse pour réaliser des expériences de spectrométrie de masse en tandem.

TABLEAU 1.7 PRINCIPALES CARACTÉRISTIQUES DES ANALYSEURS DE MASSE COURAMMENT COUPLÉS À DES MÉTHODES CHROMATOGRAPHIQUES (CHRISTIAN ET AL., 2014; RONDEAU, 2017)

Analyseur	Gamme de m/z	Résolution (FWHM)	Erreur (ppm)	Gamme dynamique	Fréquence d'acquisition (Hz)	Vide nécessaire (mbar)	Coût
Quadripôle	10^3	10^3	100	10^7	1-20	10^{-6}	€
Piège à ions	10^3	10^3 - 10^4	100	10^2 - 10^5	1-30	10^{-3}	€
Temps-de-vol (ToF)	10^4 - 10^5	10^4 - 10^5	1-10	10^2 - 10^6	1 - 10^6	10^{-6} - 10^{-7}	€€
Orbitrap	10^3	$>10^5$	1-5	10^3 - 10^4	1-100	10^{-10}	€€€

Les pièges à ions (abrégiés Trap) fonctionnent sur le même principe que les quadripôles, à la différence près que les électrodes ne sont pas des barreaux cylindriques mais délimitent une zone en trois dimensions (on parle de trappe 3D, il existe également des trappes de conception linéaire) dans laquelle les ions vont être confinés, puis expulsés et détectés l'un après l'autre suivant leur m/z grâce à l'application de tensions adéquates pour acquérir un spectre de masse. Ces analyseurs présentent des caractéristiques similaires aux quadripôles en termes de coût, de résolution et d'erreur de masse (Christian et al., 2014). Les pièges 3D souffrent en revanche, contrairement aux quadripôles, d'effets dits de « charges d'espace » limitant leur gamme dynamique (ces effets sont moins prononcés lors de l'utilisation de trappes linéaires (Rondeau, 2017)). De ce fait, les pièges à ions sont moins utilisés de nos jours que les quadripôles pour les analyses quantitatives, mais elles possèdent des propriétés intéressantes pour l'analyse structurale puisqu'elles sont les seuls analyseurs avec lesquels on peut réaliser des expériences de spectromètre de masse à plusieurs niveaux (MS^n) avec un seul analyseur. Enfin tout comme les quadripôles, elles sont fréquemment combinées à d'autres analyseurs de masse pour réaliser des expériences de MS/MS ou de MS^n .

Les analyseurs temps de vol (Time-of-Flight ou ToF) sont des analyseurs à haute résolution qui possèdent le principe de base le plus simple de tous les analyseurs de masse : tous les ions sont accélérés avec un même potentiel (c.à.d. la même énergie) puis dérivent dans une zone libre de champ sous vide poussé, appelé « tube de vol ». Les ions recevant tous la même énergie cinétique initiale, leurs vitesses

respectives seront fonction de leurs masses et de leurs charges ($E_c = \frac{1}{2} \times m \times v^2 = q \times E \times S$, où m la masse, v la vitesse, q la charge, E la différence de potentiel appliquée et S la distance d'accélération (Rondeau, 2017)). Grâce à la mesure précise du temps de dérive on peut donc déterminer le m/z de chaque ion. Pour atteindre une haute résolution (supérieure à 10 000), il est nécessaire à la fois que les distributions de l'énergie cinétique et de la position des ions au départ soient les plus étroites possibles, et que la distance de vol soit la plus longue possible. Les spectromètres récents atteignent des résolutions autour de 50 000 grâce à l'introduction de technologies telles que l'extraction retardée des ions et les sources d'ions orthogonales, qui permettent une bonne homogénéité des conditions de départ des ions, ainsi que du réflectron, un jeu de lentilles ioniques agissant comme un miroir et permettant à la fois d'allonger le trajet de champ libre sans allonger le tube de vol ainsi que de réduire la dispersion des énergies cinétiques des ions (Rondeau, 2017). Les principales caractéristiques des analyseurs ToF sont leur fréquence d'acquisition, qui peut atteindre plusieurs kHz sans perte de résolution, ainsi que leur gamme de m/z , allant des petites molécules (l'analyse de composés avec de faibles m/z peut parfois être problématique car la mesure de leur temps de dérive est entachée d'une erreur relative plus grande (Christian et al., 2014)) à des macromolécules. Les analyseurs ToF sont souvent de grande taille (le tube de vol fait souvent au minimum 1 m), ce qui implique d'être capable de maintenir un vide poussé ($\sim 10^{-7}$ mbar) dans une enceinte de relativement grande dimension. Cet analyseur est fréquemment combiné avec un quadripôle pour réaliser des expériences de MS/MS.

Le dernier type d'analyseur décrit ici est l'Orbitrap. C'est le plus récent de tous les analyseurs de masse puisque le premier Orbitrap commercial a été mis sur le marché en 2006 (Scigelova & Makarov, 2013). L'Orbitrap est un piège à ions constitué de deux électrodes, une externe en forme d'anneau entourant une électrode interne en forme de fuseau. Les ions sont piégés entre les deux électrodes et l'application de champs adéquats va permettre de les faire orbiter autour de l'électrode centrale avec une fréquence fonction du m/z de l'ion. L'électrode externe va acquérir le « courant image » ainsi généré, et une transformée de Fourier permet enfin de passer de l'espace des fréquences à l'espace des m/z et d'ainsi générer le spectre de masse (Rondeau, 2017). Les résolutions atteintes par ces analyseurs ont comme ordre de grandeur 10^4 - 10^5 , la résolution étant inversement liée à la fréquence d'acquisition (plus les ions orbitent longtemps dans l'analyseur, plus la qualité du signal enregistré est bonne) (Scigelova & Makarov, 2013). Cela en fait les analyseurs de masse couplables à la chromatographie possédant la plus haute résolution. Comme les pièges à ions, l'Orbitrap peut être sujet à des effets charges d'espace diminuant ses performances (notamment en terme de qualité de la mesure de masse, résolution et précision) (Scigelova & Makarov, 2013). Cet analyseur fonctionne sous vide extrêmement poussé (de l'ordre de 10^{-10} mbar), mais présente l'avantage, comparé aux analyseurs temps de vol par exemple, d'être très compact (une cellule Orbitrap ne mesure que quelques centimètres) et est également très sensible (quelques ions suffisent à générer un signal fiable sous réserve que le temps de scan soit

suffisamment long (Scigelova & Makarov, 2013). Il est souvent couplé avec d'autres analyseurs à basse résolution (piège à ions ou quadripôle).

2.2.1.2.iii. *Stratégies d'analyse par spectrométrie de masse*

On distingue généralement deux grandes familles de stratégies d'analyse par spectrométrie de masse, celles à un seul niveau (MS simple), et celles à niveaux multiples (spectrométries de masse en tandem ou MS/MS et MSⁿ).

Les stratégies de MS simples ne requièrent l'utilisation que d'un seul analyseur. Elles peuvent consister à acquérir tous les ions présents dans une certaine gamme de m/z (mode de balayage complet) ou bien à sélectionner uniquement certains ions de m/z donnés (mode « Selected Ions Monitoring », abrégé en SIM). Cette dernière ne peut s'appliquer qu'à des analyseurs pouvant faire fonction de filtre (typiquement les quadripôles et les pièges à ions). Le mode de balayage complet permet d'acquérir un spectre de masse complet de l'échantillon. Il est à la fois utilisé dans les analyses quantitatives ciblées comme on peut le voir dans le Tableau 1.4, mais aussi dans les analyses qualitatives ou non-ciblées. Le mode SIM présente l'avantage d'être plus sensible, mais est réservé à l'analyse ciblée de par sa sélectivité.

Les stratégies à niveaux multiples consistent à effectuer plusieurs analyses de masse consécutives au sein d'une même acquisition. Ces analyses peuvent être consécutives dans l'espace (utilisation de plusieurs analyseurs placés en série entre lesquels les ions peuvent être fragmentés, on parle alors de MS/MS dans l'espace) ou dans le temps (un analyseur effectue plusieurs balayages successifs sur un même paquet d'ions, entrecoupés d'étapes de fragmentation, on parle de MS/MS dans le temps), ce type d'analyse est effectuée uniquement par les analyseurs de type « piège à ions » (Christian et al., 2014). Lors d'une analyse en spectrométrie de masse en tandem dans le temps, il est possible de réaliser plusieurs cycles de fragmentation/sélection, on parle alors de MSⁿ.

L'utilisation de stratégies MSⁿ et MS/MS a pour objectif soit l'élucidation structurale de molécules inconnues (MS/MS et MSⁿ), soit l'augmentation de la sélectivité et de la sensibilité d'une méthode (MS/MS surtout) (Christian et al., 2014). La MSⁿ avec de nombreux niveaux étant surtout utilisée pour l'élucidation structurale de molécules complexes inconnues, elle ne sera donc pas discutée ici. Pour présenter les principaux modes de balayages existant en MS/MS, on va prendre l'exemple de l'instrument le plus commun : le triple quadripôle (QqQ). Cet instrument est constitué de trois quadripôles en série, celui du milieu étant en mode radiofréquence et servant de guide d'ions et de cellule de collision, le premier et le dernier sont placés dans différents modes de balayage suivant l'objectif visé. Quatre grands modes de balayages existent en MS/MS (Figure 1.5).

Le balayage d'ions produits (Figure 1.5.a) permet de construire le spectre de fragmentation d'une molécule, et ainsi d'élucider la structure d'un composé inconnu ou de construire une base de données de spectres de fragmentation. Le balayage d'ions précurseur (b) est utile dans le cas où plusieurs analytes

possèdent un fragment commun. Le balayage de pertes de neutre (c) permet de repérer les molécules perdant un même fragment neutre (par exemple une molécule d’eau) lors de l’étape d’activation, et enfin, le mode de SRM ou MRM (d) permet de sélectionner spécifiquement un ion à fragmenter puis de suivre un de ses fragments. Ce dernier mode est très utilisé pour l’analyse des contaminants de l’aliment, de par sa sélectivité et sa sensibilité (Hird et al., 2014).

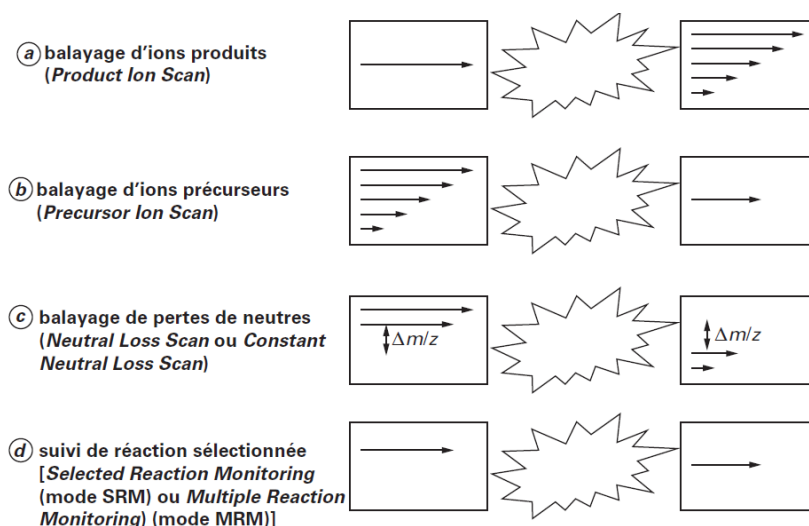


FIGURE 1.5 DESCRIPTION DES MODES DE BALAYAGE EN MS/MS (RONDEAU, 2017)

Enfin, il existe des modes de balayage lors desquels une première acquisition est effectuée dans un mode (typiquement en balayage complet), et la suivante dans un autre (souvent en MS/MS), avec la possibilité d’adapter la deuxième en fonction de la première (par exemple faire du MRM avec les n ions les plus intenses de la première). Ces modes sont spécifiques de chaque constructeur, ils ne seront donc mentionnés que quand leur utilisation sera rapportée.

2.2.1.3. APPLICATIONS EXISTANTES DES COUPLAGES LC-MS À L’ANALYSE DES CONTAMINANTS DU THÉ

Un aperçu d’une sélection de publications portant sur l’analyse des contaminants du thé et de matrices associées (tisanes et herbes notamment) se trouve en Tableau 1.4. Les conditions chromatographiques utilisées sont relativement homogènes entre les publications. Une colonne de type C18 est ainsi systématiquement utilisée (la granulométrie dépendant de la génération de l’instrument utilisé, soit de type HPLC soit UHPLC et donc pouvant supporter des pressions plus élevées), et l’élution réalisée avec un mélange d’eau et d’un solvant organique (MeOH ou ACN suivant les cas, MeOH semblant prédominant pour les analyses en mode négatif, et ACN surtout utilisé en mode positif). Des additifs sont quasiment systématiquement utilisés pour tamponner le pH de la phase mobile (à l’exception des méthodes pour l’analyse du BPA et de l’acrylamide). Les additifs les plus fréquemment rapportés sont l’acétate d’ammonium, le formate d’ammonium et l’acide formique. Le débit de phase mobile utilisé varie en fonction des publications mais les valeurs rapportées sont de l’ordre de 0,1-0,5 mL/min. La

source d'ionisation utilisée est l'ESI dans tous les cas (les travaux utilisant d'autres sources que l'ESI comme l'APCI sont rares comme on a pu le voir dans le Tableau 1.6). Quand la source est opérée dans les deux modes d'ionisation, on reporte à la fois l'utilisation de phases mobiles identiques pour chaque mode (2 publications), mais aussi d'un mélange de solvants par mode (une publication).

En ce qui concerne la partie MS, on distingue les approches utilisant des acquisitions à basse résolution (Q et Trap, 8 occurrences sur 11 publications) et les approches utilisant la haute résolution (ToF et Orbitrap, souvent couplés avec un analyseur basse résolution en amont, 5 occurrences sur 11 publications). Une étude propose la comparaison d'analyseurs à haute et basse résolution (Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008), et une autre la comparaison de deux analyseurs à haute résolution (Saito-Shida, Hamasaka, Nemoto, & Akiyama, 2018). Les approches utilisant des analyseurs basse résolution utilisent systématiquement des stratégies de type MS/MS (le plus souvent avec un analyseur de type QqQ, une utilisation de Q-Trap rapportée), avec le mode MRM (à l'exception d'une publication utilisant le mode balayage d'ion produit). Les méthodes utilisant des analyseurs à haute résolution utilisent souvent des instruments hybrides (Q-ToF et Q-Orbitrap), mais ils sont presque toujours opérés en mode balayage complet (sans sélection d'ion préalable). On relève l'utilisation d'un mode appelé « data-dependent MS/MS » (dd-MS/MS, spécifique des instruments Thermo Orbitrap consistant à effectuer un scan complet puis de la MRM sur les n ions les plus intenses) dans une publication (Dzuman, Zachariasova, Veprikova, Godula, & Hajslova, 2015) en complément du balayage complet pour confirmer l'identité des analytes. Par ailleurs, on relève deux publications majeures supplémentaires sur l'étude de la faisabilité de l'utilisation d'instruments à haute résolution en remplacement des instruments hybrides (Kellmann, Muenster, Zomer, & Mol, 2009; Lehotay et al., 2008). La première (Kellmann et al., 2009) présente l'étude des exigences en termes de résolution et de précision en masse pour l'analyse qualitative et quantitative des contaminants de l'aliment ; elle présente notamment l'apport d'instruments à très haute résolution de type Orbitrap pour ce type d'analyse. La seconde (Lehotay et al., 2008) est une revue discutant des performances de MS nécessaires à l'identification de résidus de contaminants dans des aliments, des apports des différentes techniques et des écueils les plus courants rencontrés. Les différentes études utilisant la MS à haute résolution présentées (voir en Tableau 1.4) montrent que les performances de ces instruments sont désormais comparables à ceux utilisant la MS/MS pour la quantification de composés traces, aussi bien en termes de sensibilité que de sélectivité.

Enfin, l'ESI étant un mode d'ionisation compétitif, et le thé une matrice très complexe (European Commission - DG Health and Food Safety, 2015), la correction des effets de matrice est souvent nécessaire pour obtenir des résultats satisfaisants (Saito-Shida et al., 2018). La méthode la plus efficace et la plus employée pour compenser ces phénomènes semble l'utilisation d'une gamme de calibration en matrice (7 occurrences dans 10 publications dans le Tableau 1.4, une publication portant sur des résultats qualitatifs (Pérez-Ortega et al., 2016)). On peut voir que la calibration par un étalon interne marqué, bien que théoriquement la plus efficace, n'est pas utilisée dans les publications sélectionnées. Cela peut

s'expliquer par le coût de ces standards qui font que leur implémentation sur une méthode multi-résidus à spectre large n'est pas envisageable. Dans ce cas, il serait nécessaire qu'une même molécule marquée soit utilisée pour corriger la mesure de plusieurs analytes différents, diminuant donc la qualité de la correction.

Une étude assez complète, bien que sur une autre matrice que le thé (aliments pour bébé), des effets de matrice et de la faisabilité de l'implémentation de la HRMS a été publiée en 2016 et est présentée dans le Tableau 1.4 (Pérez-Ortega et al., 2016). Cette étude aborde notamment les éventuels problèmes de coélution chromatographique et de résolution du spectromètre de masse pouvant survenir lors d'une analyse à spectre large. Les auteurs mettent notamment en avant la difficulté d'inclure des analytes de familles et structures chimiques différents dans une même méthode LC-MS, notamment à cause de l'aspect compétitif de l'ionisation en ESI.

2.2.2. TRAITEMENTS DE L'ÉCHANTILLON POUR LES ANALYSES CIBLÉES ET LEURS APPLICATION AU THÉ

Dans une analyse ciblée, l'objectif du traitement de l'échantillon est de récupérer le ou les analytes de façon quantitative (c.à.d. les analytes présents dans l'échantillon analysé sont récupérés intégralement) et la plus sélective possible (la co-extraction des composés interférents est minimisée). L'extraction d'un analyte depuis un échantillon repose sur sa distribution entre l'échantillon et une deuxième phase non-miscible avec l'échantillon appelée extractant. Ce dernier est choisi de manière à ce que les analytes aient plus d'affinité pour lui que pour l'échantillon, et que les potentiels composés interférents en aient le moins possible (on cherche à maximiser la récupération des analytes et à minimiser celle des composés interférents) (Camel, 2014). L'action de l'agent d'extraction peut être assistée par des conditions physiques (température ou pression par exemple (Mendiola, Herrero, Cifuentes, & Ibañez, 2007)) et/ou chimiques (pH de la solution par exemple (Mol, Plaza-Bolaños, Zomer, Rijk, et al., 2008)). Par ailleurs, plusieurs méthodes d'extraction peuvent être combinées pour s'assurer de la récupération de la totalité des analytes ou pour enlever des composés interférents.

L'affinité entre un composé et une phase (dans notre cas le solvant d'extraction par exemple) est régie par quatre grandes interactions (Camel, 2014), avec à chaque fois une même idée, « qui se ressemble s'assemble » (un analyte sera d'autant mieux extrait qu'il aura des interactions en commun avec son agent d'extraction) :

- i) les interactions dipôle-dipôle (appelées forces de Van de Waals), faisant intervenir les moments dipolaires des analytes et des agents d'extraction. Ces forces sont gouvernées par la présence d'hétéroatomes (Cl, F, O, N, P par exemple) dans la structure chimique de l'analyte, des composés de la matrice et de l'agent d'extraction

- ii) les liaisons hydrogènes, impliquant la présence d'atomes donneurs de proton (souvent O ou N) et accepteurs de proton (ayant une charge locale négative due par exemple à une forte électronégativité) dans le mélange
- iii) les interactions électrostatiques de type « empilement- π » (aussi appelées interactions π - π), impliquant le recouvrement des orbitales π de deux cycles aromatiques
- iv) les interactions hydrophobes qui veulent que, pour minimiser l'entropie, les composés hydrophobes placés dans un milieu hydrophile auront tendance à se lier entre eux de façon non-covalente. On observe ces interactions particulièrement en présence de longues chaînes carbonées et en milieu aqueux.

Nos échantillons sont du thé sous forme de feuilles (échantillon solide) ou sous forme d'infusion (échantillon liquide). Une méthode de traitement de l'échantillon sera développée pour chacune de ces formes. Dans le cas des échantillons solides, les analytes peuvent se trouver aussi bien sur la surface de l'échantillon qu'en son sein (on peut citer comme exemple les pesticides dits « systémiques » qui diffusent dans l'intégralité de la plante). Le développement de la méthode se faisant sur des échantillons dopés artificiellement, cette question ne se pose pas pour l'instant mais elle pourra se poser à l'avenir lors de l'analyse d'échantillons réels. Pour les échantillons liquides, les analytes sont solubilisés dans l'échantillon.

Le thé est un produit ayant une composition chimique complexe comme on a pu le voir en 1.1.4 de ce chapitre, de nombreuses molécules interférentes sont donc susceptibles d'être co-extraites avec les analytes comme présenté dans le. Le produit sous sa forme solide (feuilles) risque d'apporter plus d'interférents que la forme liquide, du fait de la première sélection que l'infusion réalise ainsi que de la dilution qu'elle représente par rapport au produit solide. Les principaux interférents à attendre sont les composés phénoliques, présents en grande quantité dans le thé, mais également les composés de la classe des xanthines qui rassemble notamment la caféine et la théobromine. Ces interférents sont des composés polaires et possédant pour beaucoup des structures aromatiques ainsi que la possibilité de former des liaisons H par le biais de groupements -OH. Dans notre cas, les analytes (qui sont les contaminants potentiels du thé) ont une très grande diversité de structure comme on l'a vu en partie 1.1.3 de ce chapitre et dans le

. Un contaminant peut donc avoir une structure aromatique ou non, être très polaire ou très apolaire avec tous les intermédiaires existants. Dans notre étude, nous nous restreindrons aux composés analysables par la technologie LC-MS (avec une source d'ions de type ESI), celle-ci étant devenue une technique de choix pour l'analyse des contaminants de l'aliment grâce à son champ d'application, sa polyvalence et son choix d'analyseurs de masse (haute et basse résolution) comme on a pu le voir, ainsi qu'à cause de la tendance à utiliser des pesticides de plus en plus polaires (Frenich et al., 2014; Hird et al., 2014). La diversité de structure potentielle est encore très large puisque cela inclut aussi bien des composés

modérément apolaires que des petites molécules relativement polaires ainsi que des molécules de taille plus importante.

En fonction de l’état de l’échantillon (solide ou liquide), on relève trois grandes techniques de préparation d’échantillon pour l’analyse multi-résidus de contaminants traces dans les aliments, toutes trois représentées dans le Tableau 1.4.

La première technique d’extraction ne s’applique qu’aux échantillons sous forme liquide (ou pour purifier les extraits liquides d’échantillons solides) ; c’est l’extraction en phase solide (SPE). Elle consiste à faire passer l’échantillon liquide à travers un support solide. Ce support est souvent une cartouche de gel de silice éventuellement fonctionnalisé (à l’image des phases stationnaires de LC en « phase inverse » vues dans la partie 2.2.1.1), un gel d’alumine, du carbone graphite ou bien un support polymérique (Camel, 2014; Christian et al., 2014). L’objectif de cette technique est de séparer les composés interférents des analytes en se basant sur les différences d’affinité existantes entre ces composés et le support solide. Elle peut comporter plusieurs étapes suivant les analytes et les composés interférents visés. La première étape est toujours de mettre en milieu le support avec un solvant ou un mélange de solvant. L’échantillon est ensuite déposé sur le support et différents mélanges de solvant sont utilisés pour éluer le plus sélectivement possible les composés interférents ou les analytes. Toute la mise au point des méthodes utilisant cette technique repose sur le choix de la phase stationnaire, des mélanges de solvant, ainsi que des volumes déposés pour obtenir un extrait le plus purifié possible (Andrade-Eiroa, Canle, Leroy-Cancellieri, & Cerdà, 2016). Il existe également des techniques dérivées de la SPE. On peut citer comme exemple la SPE dispersive (dSPE), qui consiste à mélanger l’échantillon ou l’extrait liquide avec la phase de SPE puis à agiter et récupérer le surnageant. Cette technique est notamment utilisée dans le protocole dit « QuEChERS » (« Quick, Efficient, Cheap, Rugged and Safe ») consistant à faire une extraction liquide/liquide (LLE) suivie d’une dSPE (Anastassiades, Lehota, & Schenck, 2003), très utilisé dans l’analyse de traces comme on peut le voir dans le Tableau 1.4 (6 occurrences sur les 11 publications sélectionnées utilisent cette combinaison).

La deuxième, s’appliquant aussi bien aux échantillons liquides que solides, est l’extraction par solvant. Elle consiste à extraire les analytes par un solvant non-miscible avec l’échantillon. L’extraction peut éventuellement être assistée par un apport d’énergie extérieure, par exemple sous forme de micro-ondes, de chaleur ou de pression, ou d’une agitation (Christian et al., 2014). Dans le cas d’un échantillon solide, on parle d’extraction solide/liquide (SLE), dans le cas d’un échantillon liquide d’extraction liquide/liquide (LLE). Dans le cas d’un échantillon liquide aqueux, il est possible d’utiliser un solvant d’extraction miscible avec l’échantillon ; dans ce cas, l’ajout de sels (typiquement $MgSO_4$ et le citrate de sodium) permettra de provoquer le partage de phases voulu en augmentant la polarité de la phase aqueuse, et sa saturation par ces derniers va favoriser le transfert des analytes dans la phase organique. Cette stratégie est également utilisée pour l’extraction d’analytes depuis des échantillons solides avec un mélange de solvants organiques et aqueux, séparés ensuite grâce à l’ajout de sel (Mol, Plaza-Bolaños,

Zomer, De Rijk, et al., 2008). Parmi les travaux rapportés dans le Tableau 1.4 utilisant cette technique, les principaux solvants utilisés sont des solvants modérément polaires tels que l'acétonitrile (ACN), le méthanol (MeOH) ou encore l'acétate d'éthyle (AcEth), parfois mélangés à une phase aqueuse (Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008). Les extraits obtenus par cette technique sont soit analysés directement, soit subissent des étapes de purification, communément par SPE ou dSPE (Lacina et al., 2012; Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008; Samsidar, Siddiquee, & Shaarani, 2018). Ils peuvent également subir une étape d'évaporation/concentration pour changer de solvant et/ou concentrer l'extrait (Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008).

Enfin, la dernière technique, qui ne concerne que les échantillons liquides, consiste à injecter directement l'échantillon dans l'instrument, éventuellement après une dilution par une solution aqueuse ou un solvant organique pour se placer dans des conditions compatibles avec la méthode chromatographique, ainsi que souvent une filtration pour ne pas endommager l'appareil. Cette technique, appelée en anglais « dilute & shoot » (D&S) présente l'avantage de nécessiter très peu de manipulation et d'être très facile à mettre en œuvre. Elle est très utilisée dans le domaine de la métabolomique, couplée à des instruments à haute résolution (Gika, Theodoridis, Plumb, & Wilson, 2014). Elle a été utilisée avec succès par Mol (Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008) pour l'analyse d'un grand nombre de contaminants dans une matrice liquide aqueuse (lait) à des niveaux pertinents (jusqu'à 10 µg/kg).

Les échantillons sous forme liquide sont traités dans 5 des 11 publications sélectionnées (4 publications sur des matrices de thé ou assimilées, 1 publication sur une matrice laitière). Les méthodes traitant de l'analyse d'échantillons de thé (L. Chen et al., 2014; Chung & Lam, 2018; Monbaliu et al., 2010; Wang, Cheung, & Leung, 2014) utilisent pour deux d'entre elles une première extraction par LLE, suivie d'une purification par SPE (L. Chen et al., 2014) ou par dSPE (Wang et al., 2014), une concentration par évaporation/reconstitution terminant le traitement de l'échantillon. Ces publications étudient l'analyse de pesticides d'une et plusieurs classes respectivement, avec des limites de quantification (LOQ) de l'ordre de 1-10 µg/kg. Deux autres publications étudient l'utilisation de la technique de la SPE pour l'analyse d'infusions de thés (Chung & Lam, 2018; Monbaliu et al., 2010). Ces SPE sont suivies d'une étape d'évaporation/reconstitution pour se placer dans le solvant d'injection et pour concentrer les analytes. Les molécules analysées sont respectivement des alcaloïdes pyrrolizidiniques et des mycotoxines. Les LOQ rapportées sont de l'ordre du dixième de µg/kg pour la méthode d'analyse des alcaloïdes et entre 1 et 100 µg/kg pour les mycotoxines. Toutes ces publications utilisent des spectromètres de masse basse résolution hybrides (QqQ ou Q-Trap) opérés en mode MRM comme décrit en 2.2.1.2. L'approche « dilute & shoot » n'a quant à elle pas été encore appliquée à une infusion de thé à notre connaissance, mais on en retrouve une application intéressante, entre autre sur une matrice de type « lait » par Mol et son équipe (Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008). L'intérêt de cette étude pour notre travail est que 4 familles de molécules sont analysées (pesticides, mycotoxines, toxines

végétales et médicaments vétérinaires) et 7 méthodes de traitement de l'échantillon et 2 instruments (LC-MS/MS et LC-HRMS) comparés. La méthode « dilute & shoot » montre de bonnes performances sur la matrice liquide y compris dès 10 µg/kg, et ce même sans purification. Par ailleurs, cette méthode semble montrer une meilleure répétabilité que la LLE.

Une très grande majorité des publications sélectionnées concernent une matrice solide (10 sur 11, parmi lesquelles 8 sur du thé ou assimilé, 2 sur d'autres matrices). Parmi celles étudiant le thé, la majorité utilisent des méthodes basées sur la LLE couplée à une purification par SPE ou dSPE. Trois publications portent sur l'analyse d'une ou plusieurs classes de pesticides avec cette technique (Jiao et al., 2016; Saito-Shida et al., 2018; Wang et al., 2014) avec des LOQ de l'ordre de 10 µg/kg, avec une utilisation de la HRMS opérée en mode de balayage complet (Saito-Shida et al., 2018), cette publication présentant l'intérêt de comparer deux analyseurs de masse à haute résolution (Orbitrap et ToF). Les deux autres publications utilisent des analyseurs QqQ opérés en MRM. Deux publications portent sur l'analyse de classes de composés spécifiques (Gao et al., 2017; Liu et al., 2008) (BPA et dérivés respectivement, et acrylamide), avec des LOQ de l'ordre du µg/kg avec des analyseurs de type QqQ et une extraction par LLE/SPE. Cette approche de traitement de l'échantillon est par ailleurs éprouvée par Dzuman et al. (Dzuman et al., 2015) pour l'analyse de 3 classes de contaminants (pesticides, mycotoxines et toxines végétales) dans plusieurs matrices, dont du thé. Les LOQ atteintes sont de l'ordre de 10 µg/kg avec l'utilisation d'un instrument à haute résolution opéré en mode de balayage complet. Enfin, on retrouve l'extraction par solvant directe, suivie d'une SPE pour deux publications (L. Chen et al., 2014; Monbaliu et al., 2010) portant respectivement sur l'analyse d'une classe de pesticides et de mycotoxines. La première publication (L. Chen et al., 2014) utilise une extraction assistée par ultrasons. Les deux publications utilisent un analyseur QqQ opéré en MRM et les LOQ sont de l'ordre de la dizaine de µg/kg.

Enfin, deux publications traitent de l'analyse de nombreuses classes de contaminants (au moins 4) dans plusieurs matrices (Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008; Pérez-Ortega et al., 2016). La première (Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008) traite surtout de l'étape d'extraction et propose une comparaison de 7 méthodes de traitement de l'échantillon dans 5 matrices solides en plus de la matrice liquide vue précédemment, et ce sur plusieurs centaines de composés. La méthode de choix retenue par les auteurs est une extraction à l'acétonitrile aqueux et il est montré que l'étape de purification n'est pas nécessaire pour obtenir des recouvrements satisfaisants, y compris à des bas niveaux (10 µg/kg) sur une grande diversité de molécules. Ces résultats ont été confirmés en 2012 par une autre étude (Lacina et al., 2012). La deuxième (Pérez-Ortega et al., 2016) traite plus de la partie analytique (chromatographie/spectrométrie de masse), et comment cette dernière est affectée par des composés interférents co-extraits de la matrice.

2.2.3. DISCUSSION GÉNÉRALE DES MÉTHODES ANALYTIQUES POUR L'ANALYSE MULTI-CLASSE DES CONTAMINANTS DANS LE THÉ

Toutes les méthodes analytiques sélectionnées sur le thé (feuilles ou boisson) ne s'intéressant qu'à une classe de contaminants (Tableau 1.4) ont en commun d'utiliser une méthode de traitement de l'échantillon incluant une purification de l'extrait par SPE ou dSPE pour maîtriser les effets de matrice (Saito-Shida et al., 2018). Dans leur majorité, ces travaux utilisent des spectromètres de masse en tandem et basse résolution (le plus souvent QqQ, un usage de Q-Trap relevé) opéré en MRM, à l'exception d'un utilisant un instrument ToF et un Orbitrap (Saito-Shida et al., 2018) opérés en mode de balayage complet.

Les méthodes multi-classes sont très peu reportées sur le thé, la seule (analyse de pesticides, mycotoxines et toxines naturelles) utilise un protocole de type QuEChERS pour la préparation d'échantillon suivi d'une analyse LC-HRMS (Orbitrap) opérée en mode de balayage complet (Dzuman et al., 2015). En étendant la recherche à d'autres matrices, on trouve plusieurs travaux suggérant que l'étape de purification n'est pas nécessaire lors de l'utilisation de LC-HRMS (Lacina et al., 2012; Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008). L'utilisation d'une telle méthode pour l'analyse multi-classes semble donc réalisable, et elle aurait de plus l'avantage d'être très générique. Ces travaux présentent un intérêt particulier puisque, comme on a pu le voir, il n'existe pas à l'heure actuelle de méthodes multi-classes sur des matrices de thé ou infusion. Enfin, cette méthode aurait un bon potentiel d'application à une approche non-ciblée, du fait de sa faible sélectivité et de sa simplicité de mise en œuvre.

Nous allons maintenant nous intéresser à l'analyse non-ciblée des contaminants de l'aliment, et à l'apport des méthodes ciblées multi-classes à cette problématique.

2.3.L'ANALYSE NON CIBLÉE DES CONTAMINANTS ALIMENTAIRES

2.3.1. PROBLÉMATIQUE ET DÉFIS ANALYTIQUES DE L'ANALYSE NON-CIBLÉE DES CONTAMINANTS DE L'ALIMENT

Une méthode destinée à l'analyse non-ciblée de contaminants de l'aliment comporte deux volets distincts : la méthode analytique à proprement parler (traitement de l'échantillon et analyse instrumentale), et la méthode de traitement de données visant à détecter, dans le jeu de données généré, les potentiels contaminants et à donner des hypothèses quant à leur identité (Antignac et al., 2011; Knolhoff & Croley, 2016). Ces méthodes doivent de plus être implémentées dans le cadre d'une stratégie d'analyse (des échantillons et des données) visant à maximiser la détection des contaminants en dépit des facteurs de variation externes (dérive instrumentale, répétabilité de la méthode analytique entre autres).

Les contaminants de l'aliment se trouvent à des niveaux de traces (de l'ordre de la dizaine de $\mu\text{g}/\text{kg}$), et les matrices agro-alimentaires sont très complexes, particulièrement les matrices végétales comme le

thé (Camel, 2014; Pongsuwan et al., 2008). Par ailleurs, la diversité de structures chimiques des contaminants est très grande, comme on a pu le voir dans la partie 1.3 de ce chapitre. La méthode analytique devra donc être capable d’être à la fois assez sensible et avec un spectre le plus large possible. Ce sera également un défi pour le processus de traitement de données, car il sera nécessaire d’extraire des signaux potentiellement d’intensité beaucoup plus faible que les signaux de la matrice pour être à même de détecter les contaminants (Knolhoff & Croley, 2016). La capacité à détecter une contamination inconnue exclut *de facto* certaines approches dites de criblage, pouvant être qualifiées de « semi-ciblées », consistant à identifier l’ensemble des pics détectés *via* une recherche au sein d’une base de données préexistante (Knolhoff & Croley, 2016; Lehotay, Sapozhnikova, & Mol, 2015; Pérez-Ortega et al., 2016). On se concentrera donc sur les approches réellement non-ciblées, à même de détecter et de caractériser une contamination inconnue. Cependant, les considérations des études de criblage étant tout de même pertinentes pour notre étude (notamment vis-à-vis des performances nécessaires des instruments utilisés), il y sera fait référence dans cette partie.

2.3.2. APERÇU DES TRAVAUX EXISTANTS

Les principales publications traitant de la détection non-ciblée des contaminants sont détaillées dans le Tableau 1.8, et listées par ordre chronologique de parution. Il est à noter que, contrairement à ce qui se fait dans le domaine de la métabolomique (Hardy & Taylor, 2007), aucune des publications relevées ici ne propose un accès libre aux données brutes.

Cinq publications majeures sont recensées. Les matrices étudiées sont relativement variées, certains produits revenant de façon récurrente (par exemple le jus d’orange et les formules infantiles). Les niveaux de contamination considérés vont de niveaux réalistes en cas de dépassement de LMR (autour de la dizaine de $\mu\text{g}/\text{kg}$) à des niveaux très importants (de l’ordre du mg/kg) qui relèvent probablement davantage de problématiques de toxicité aiguë que de toxicité chronique.

Des informations données par ces publications, et en se basant sur les revues de la littérature sur le sujet (Antignac et al., 2011; Knolhoff & Croley, 2016), on peut établir un schéma global du processus d’analyse et de traitement de données pour la détection non-ciblée des contaminants (voir Figure 1.6). Nous commencerons tout d’abord par présenter les stratégies globales guidant le développement et l’implémentation des méthodes non-ciblées, puis nous présenterons brièvement les méthodes analytiques utilisées en se basant sur le Tableau 1.8 et ce qui a été vu auparavant à leur sujet, et enfin, les outils et méthodes de traitement de données seront présentés et discutés.

TABLEAU 1.8 DÉTAILS DES PUBLICATIONS TRAITANT DE LA DÉTECTION NON-CIBLÉE DES CONTAMINANTS DE L'ALIMENT

Produit d'étude <i>État</i>	Préparation d'échantillon <i>Instrument</i>	Niveaux	Contaminants d'études (nombre)	Mode de sélection des pics	Logiciel	Publication
Jus d'orange Liquide	D&S <i>UHPLC-ToF (C18, ESI⁺)</i>	~ mg/kg	Pesticides, mycotoxines, médicaments (26)	Comparaison aux contrôles	TracMass 2 <i>libre</i>	(Tengstrand, Rosén, Hellenäs, & Åberg, 2013)
Miel Solide	LLE <i>UHPLC-Q-Orbitrap (C18, ESI⁺)</i>	~ 10 µg/kg	Pesticides, antibiotiques (83)	Détection de massif isotopique	XCMS et R <i>libre</i>	(Cotton et al., 2014)
Formules infantiles Liquide	D&S <i>UHPLC-ToF (HILIC, ESI⁺)</i>	~ mg/kg	Mélatamine (1)	Analyse multivariée et univariée	MarkerLynx™ <i>propriétaire</i>	(Inoue et al., 2015)
Jus d'orange et formules infantiles Liquide et solide	D&S et extraction par solvant <i>UHPLC-Q-ToF (C18, ESI⁺)</i>	~ 10 µg/kg	Alcaloïdes (4)	Analyse univariée	MassHunter <i>propriétaire</i>	(Knolhoff, Zweigenbaum, & Croley, 2016)
Lait Liquide	D&S <i>UHPLC-ToF (C18, ESI⁺)</i>	5-400 µg/kg	Pesticides (19)	Comparaison aux contrôles	TracMass 2 <i>libre</i>	(Kunzelmann, Winter, Åberg, Hellenäs, & Rosén, 2018)

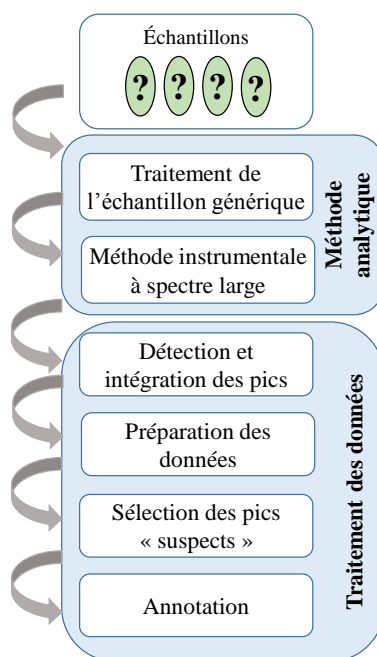


FIGURE 1.6 PROCESSUS GÉNÉRIQUE D'ACQUISITION ET DE TRAITEMENT DES DONNÉES POUR LA DÉTECTION NON-CIBLÉE DES CONTAMINANTS DE L'ALIMENT

2.3.3. STRATÉGIES D’ANALYSE

La méthodologie analytique utilisée pour la détection non-ciblée de contaminants vient de celle utilisée en métabolomique (Antignac et al., 2011; Castro-Puyana et al., 2017; Hu & Xu, 2013; Knolhoff & Croley, 2016). Elle a initialement pour objectif de détecter un changement de composition chimique (qualitatif et/ou quantitatif) dû à des molécules qui ne sont pas forcément connues entre plusieurs groupes d’échantillons ayant, par exemple, subi des traitements différents.

Ces approches ont initialement été développées sur des matrices biologiques (sang, tissus par exemple), et rapidement étendues à d’autres applications, notamment dans l’analyse de produits agro-alimentaires sous l’appellation de « foodomics » (Cifuentes, 2009; Herrero, Simó, García-Cañas, Ibáñez, & Cifuentes, 2012). Leur potentiel pour détecter de façon non-ciblée les contaminants de l’aliment a vite été mis en avant dans des revues bibliographiques par de nombreux auteurs (Antignac et al., 2011; Castro-Puyana et al., 2017; Herrero et al., 2012; Knolhoff & Croley, 2016).

Les études métabolomiques ayant pour but de détecter des biomarqueurs d’un phénomène consistent à comparer la composition chimique d’un ou plusieurs groupes d’échantillons d’intérêt avec un échantillon de référence grâce à une méthode d’analyse la plus exhaustive possible, dans le but d’annoter (c.à.d. émettre une hypothèse quant à la structure chimique) ou d’identifier (c.à.d. assigner avec certitude la structure) les composés qui diffèrent qualitativement ou quantitativement entre les groupes d’échantillons (Dunn et al., 2011; Gika et al., 2014). De grands jeux de données sont utilisés, soit plusieurs dizaines d’échantillons et plusieurs milliers de composés détectés dans le cas des approches basées sur la MS (Dunn et al., 2011; Gika et al., 2014). Leur exploration est rendue possible par l’utilisation d’outils de traitement de données tels que les analyses multivariées, ou encore la fouille de données (« data mining »). Une attention particulière doit être portée aux possibles biais, qu’ils soient analytiques (dérive instrumentale, présence d’analyses aberrantes par exemple), ou bien liés aux échantillons en eux-mêmes (hétérogénéité de la prise d’échantillon par exemple). Généralement, le biais analytique est géré grâce à l’injection des échantillons dans un ordre aléatoire, l’utilisation d’échantillon de contrôle qualité (QC : souvent un « échantillon moyen » formé par le mélange d’une fraction de chaque échantillon ou de leur extrait) qui est injecté régulièrement au cours de la séquence analytique (pour détecter une éventuelle dérive instrumentale), et l’injection régulière de blancs (c.à.d. de phase mobile) pour vérifier l’absence de phénomène d’effet mémoire (De Livera et al., 2015; Gika et al., 2014; Wehrens et al., 2016). Le biais lié aux échantillons en eux-mêmes doit quant à lui être géré en amont par un échantillonnage robuste et des bonnes pratiques de laboratoire.

Les informations recherchées dans les études métabolomiques ne sont pas forcément les mêmes que dans celles sur la qualité sanitaire. En effet, on recherche un changement du métabolome dû à un facteur donné (c.à.d. les changements de concentration ou l’apparition d’un grand nombre de molécules, le tout avec des facteurs interférents très forts), tandis que pour étudier la qualité sanitaire on recherche un nombre restreint de composés traces qui peuvent disposer d’une signature chimique particulière

(exemple : massif isotopique, fragmentation). Par ailleurs, si la métabolomique vise à la découverte de biomarqueurs fiables, elle ne vise pas à les découvrir tous, alors qu'il est nécessaire d'avoir une vision la plus exhaustive possible des contaminants lors de l'évaluation de la qualité sanitaire chimique d'un produit. On retrouve aussi en sécurité sanitaire des aliments, de même qu'en analyse environnementale, plusieurs approches originales dédiées à la recherche de signatures chimiques spécifiques, typiquement la présence de composés halogénés (Andersen, Hansen, Jørgensen, & Nielsen, 2016; Cariou, Omer, Léon, Dervilly-Pinel, & Le Bizec, 2016; Cotton et al., 2014; Roullier et al., 2016).

2.3.4. MÉTHODES ANALYTIQUES

Comme mentionné précédemment, les méthodes analytiques utilisées pour les approches de type métabolomique (traitement de l'échantillon et méthode instrumentale) doivent avoir le champ d'application le plus large possible dans le but de disposer d'une vision la plus exhaustive possible de la composition chimique des échantillons. Cela se vérifie dans les approches proposées pour l'analyse des contaminants de l'aliment. En effet, toutes les méthodes de traitement de l'échantillon employées (Tableau 1.8) ont en commun de ne pas comporter d'étapes de purification et ainsi d'être les plus génériques possibles. On retrouve les grandes familles de méthodes décrites en 2.2.2, à savoir une extraction par solvant suivie d'une LLE pour une matrice solide (Cotton et al., 2014), mais également des méthodes de type « dilute & shoot » (Inoue et al., 2015; Knolhoff et al., 2016; Kunzelmann et al., 2018; Tengstrand et al., 2013) pour des matrices liquides, et un exemple d'extraction directe par solvant pour une matrice solide (Knolhoff et al., 2016).

La technique d'analyse semble faire consensus car toutes les publications reportées utilisent une méthode générique d'UHPLC-HRMS avec une colonne de silice greffée C18 dans 4 cas sur 5, et une méthode HILIC dans un des cas (mais pour l'analyse d'une molécule particulière, en l'occurrence la mélamine). Les deux types d'analyseurs à haute résolution sont utilisés (ToF et Orbitrap), et le mode de balayage exclusif est le balayage complet. On trouve également des expériences de MS/MS pour confirmer l'identité des contaminants détectés (Cotton et al., 2014). Dans les publications, l'ionisation se fait exclusivement en mode positif, et le mode négatif n'est jamais testé, ni la comparaison du comportement du processus de traitement des données sur les deux modes. Cela semble un manque majeur dans les études actuelles qu'il sera intéressant d'investiguer dans ce travail.

2.3.5. DÉVELOPPEMENTS DANS LES TRAITEMENTS ET L'INTERPRÉTATION DES DONNÉES ISSUES D'ANALYSES NON-CIBLÉES

Comme illustré sur la Figure 1.6 et souligné dans plusieurs revues de la littérature (Antignac et al., 2011; Hu & Xu, 2013; Knolhoff & Croley, 2016), le processus de traitement de données peut se diviser en quatre grandes étapes : (i) l'extraction des pics depuis les fichiers de données, (ii) le prétraitement et la préparation des données, (iii) la détection des composés et des échantillons « suspects », (iv)

l’annotation des composés suspects. Pour chaque étape, on présentera l’état de l’art pour la détection des contaminants, en nous intéressant également à ce qui existe pour les études métabolomiques.

2.3.5.1. DÉTECTION ET INTÉGRATION DES PICS

L’analyse d’un échantillon par LC-HRMS génère plusieurs milliers de pics par fichier, et ce d’autant que les produits agro-alimentaires sont très complexes (Knolhoff & Croley, 2016). L’extraction des pics est donc une étape cruciale, pour laquelle de nombreux outils, aussi bien libres et ouverts que propriétaires, existent.

Les outils ouverts présentent la caractéristique d’être génériques, c.à.d. de ne pas être conçu pour traiter uniquement des données provenant d’un instrument précis. Cela est un avantage lorsqu’il s’agit de développer une approche commune à plusieurs instruments, mais cela les rend potentiellement moins performants que les outils dédiés développés par les constructeurs. À notre connaissance, il n’existe pas dans la littérature de comparaison entre outils libres et outils propriétaires pour l’intégration des pics en LC-MS. Une comparaison entre les principales caractéristiques des outils libres et propriétaires, ainsi qu’une liste des principaux outils existants sont proposées dans le Tableau 1.9.

Dans notre cas, nous désirons développer une méthode qui soit la plus générique possible. Par ailleurs, à l’image de ce qui se fait dans le domaine de la métabolomique (Hardy & Taylor, 2007), une ouverture des données et des processus et outils de traitement de données permettraient de développer de meilleurs outils et de tester ceux existants sur de multiples jeux de données, ce qui n’existe pas pour les applications portant sur la qualité sanitaire. On retrouve cette volonté de partage des données également dans le domaine des analyses d’authenticité des aliments (Esslinger, Riedl, & Fauth-Hassek, 2014; Riedl, Esslinger, & Fauth-Hassek, 2015). Par ailleurs, il est prévu dans notre projet de comparer les données acquises sur différentes plateformes LC-HRMS ; l’utilisation d’un procédé de traitement de données « neutre » nous permettrait d’effectuer de réelles comparaisons pertinentes, alors que l’utilisation de deux procédés différents rendrait cette comparaison hasardeuse. Dans cette optique, l’utilisation d’outils libres et ouverts semble donc préférable à l’utilisation d’outils propriétaires.

Parmi les outils libres rapportés (Tableau 1.9), le plus fréquent est XCMS (Smith, Want, O’Maille, Abagyan, & Siuzdak, 2006) (1 631 citations de l’article fondateur relevées au 26/07/2018 sur la plateforme Web of Science), suivi de MZmine 2 (Pluskal, Castillo, Villar-Briones, & Orešič, 2010) (647 citations relevées au 26/07/2018), MetAlign (Lommen, 2009) (342 citations relevées au 26/07/2018), et enfin TracMass 2 (Tengstrand, Lindberg, & Åberg, 2014) avec 14 citations, peu utilisé mais qui mérite d’être cité ici pour son utilisation dans deux publications traitant de la sécurité sanitaire chimique (Kunzelmann et al., 2018; Tengstrand et al., 2013). XCMS, une boîte à outils de modules pour l’extraction et l’alignement des pics depuis des fichiers LC (ou GC)-MS, est donc de loin l’outil le plus utilisé, aidé en cela par une importante communauté scientifique qui s’est constituée autour, donnant naissance à des améliorations de certains modules (Tautenhahn, Bottcher, & Neumann, 2008), et

permettant son accessibilité au plus grand nombre grâce au développement d'interfaces graphiques et de plateformes de calcul en ligne tels que Workflow4Metabolomics (Giacomini et al., 2015) ou encore XCMS-Online (Tautenhahn, Patti, Rinehart, & Siuzdak, 2012).

TABLEAU 1.9 CARACTÉRISTIQUES DES OUTILS D'INTÉGRATION DE PICS EN LC-MS

	Outils libres	Outils propriétaires
Avantages	Génériques, gratuits, modulaires, communauté scientifique importante, autorisent le partage des données, possibilité d'implémenter des méthodes « maison »	Simple à utiliser, adaptés à chaque format de données, possibilité d'analyse des données en ligne avec les instruments dans des processus automatisés, support technique du fournisseur
Inconvénients	Plus complexes à implémenter, mises à jour dépendantes de la communauté, beaucoup de paramètres à régler	Très chers (~10 k€), partage avec la communauté scientifique compliqué, impossibilité d'analyser les données d'autres instruments, peu modulaires, souvent compliqués à interfacer avec d'autres outils de traitement de données
Principaux logiciels	MZmine, XCMS, TracMass, MetAlign, Matlab, R	MarkerLynx (Waters), MassHunter (Agilent), TraceFinder (Thermo)

De façon similaire, le deuxième outil le plus utilisé (MZmine 2) est un logiciel proposant une série de modules pour l'intégration des pics de fichiers LC et GC-MS. Il propose également des outils de traitement de données et de visualisation, ainsi que la mise en place de procédés automatisés. MetAlign propose quant à lui une interface graphique complète pour l'intégration des pics issus de fichiers de spectrométrie de masse. Notre tentative de le faire fonctionner s'est soldée par un échec qui n'a pas pu être résolu (problème de compatibilité), ce logiciel n'a donc pas été retenu. Une publication intéressante compare MZmine 2, XCMS et MetAlign pour l'intégration de pics dans des fichiers LC-MS (comparaison intégration automatique vs. intégration manuelle d'un fichier LC-MS d'un standard analytique connu) (Coble & Fraga, 2014). Il apparaît que, si MetAlign est le meilleur des trois pour le taux de détection de pics (79,5% vs. 45,1% pour XCMS et 35,2% pour MZmine), MZmine et XCMS sont bien meilleurs en termes de justesse de la mesure de l'aire du pic (biais absolu de 3,1 et 4,5% pour XCMS et MZmine, 22% pour MetAlign, avec des écarts types relatifs standards respectivement de 3,8 ; 4,8 et 21%). Enfin, comme on l'a vu, XCMS est implémenté sur de nombreuses plateformes en ligne à la différence de MZmine. La puissance de calcul et la capacité de mémoire vive nécessaire pouvant être rédhibitoire lors de l'analyse de fichiers bruts LC-HRMS (un fichier peut peser plusieurs centaines de Mo, et un jeu de données en contient plusieurs dizaines), c'est un avantage net que de pouvoir externaliser le calcul.

Comme évoqué dans la publication mentionnée plus haut (Coble & Fraga, 2014), l'intégration des pics depuis un fichier LC-MS n'est pas complète. Il semble néanmoins possible d'améliorer ces performances de XCMS par l'application d'algorithmes d'optimisation de ses paramètres. Les travaux

les plus aboutis en ce sens ont conduit à la publication du package IPO (Libiseller et al., 2015), qui permet l’optimisation de plusieurs paramètres de XCMS. Le fonctionnement de cet algorithme repose sur un critère de qualité de pic qui est la détection, à côté du pic monoisotopique, du pic correspondant au ^{13}C .

À la fin de l’étape d’extraction des pics, on récupère un tableau de données (aussi appelé matrice de données) contenant les intensités des pics pour chaque échantillon et chaque pic (c.à.d. pour chaque combinaison m/z / RT pour laquelle un pic aura été trouvé). Ce tableau peut éventuellement être complété par des tableaux annexes contenant par exemple des informations sur les pics (m/z , RT, motifs isotopiques détectés, etc.) ou sur les échantillons (ordre d’injection, groupe, etc.).

2.3.5.2. PRÉTRAITEMENTS

Les prétraitements ont pour objectifs de préparer les données à leur analyse et de maximiser la chance de détection des composés d’intérêt. On distingue trois grandes étapes : (i) la complétion du tableau de données, (ii) la correction de dérive analytique, (iii) la filtration et le nettoyage des données, (iv) la mise à l’échelle et la normalisation (Antignac et al., 2011; Di Guida et al., 2016; Knolhoff & Croley, 2016). Certains auteurs placent la filtration dans l’étape d’analyse des données (Godzien, Ciborowski, Angulo, & Barbas, 2013).

Tout d’abord, la première étape consiste à s’assurer que le tableau de données qui sera analysé est complet, c.à.d. qu’il n’y a pas de valeurs manquantes ou aberrantes (valeurs infinies, ou dans certains cas zéros ou valeurs négatives). La présence de ces valeurs peut, au mieux, dégrader la qualité du résultat (moins de données disponibles), au pire empêcher totalement l’analyse des données (par exemple s’il y a utilisation d’une méthode ne les tolérant pas, comme un passage en log avec des zéros par exemple). Cet aspect du prétraitement n’est pas abordé dans les études dédiées à la qualité sanitaire chimique des aliments, il faut donc se référer aux travaux existants en métabolomique. On relève trois grands types d’approches pour gérer ces valeurs. La première consiste à les imputer avec une valeur fixe (souvent la moyenne de toutes les mesures, ou la plus petite mesure) (Wehrens et al., 2016) : cette approche est connue pour donner des résultats de médiocre qualité (Di Guida et al., 2016; Wei et al., 2018). Il existe ensuite des approches consistant à se servir de méthodes de traitement de données pour tenter de prédire les valeurs manquantes ou aberrantes ; on trouve notamment plusieurs méthodes basées sur des méthodes multivariées (on peut citer comme exemple la décomposition en valeurs singulières –SVD–, ou encore le remplacement par analyse en composante principale Bayésienne), ainsi que sur des méthodes de classification (méthode des k plus proches voisins -KNN-) ou encore d’apprentissage (méthode des forêts aléatoires -RF-) (Di Guida et al., 2016; Hrydziuszkowski & Viant, 2012; Wei et al., 2018). Plusieurs comparaisons de ces méthodes ont été réalisées sur des jeux de données LC-MS pour la métabolomique (Di Guida et al., 2016; Wei et al., 2018). Les résultats sont difficiles à croiser car les méthodologies employées ne sont pas les mêmes (évaluation des outils par rapport à des valeurs de référence pour Wei et al. ; évaluation du procédé global dans le cas de Di Guida et al.). Il apparaît

toutefois que les méthodes RF et SVD semblent être celles donnant les meilleures performances. La dernière approche pour la gestion des valeurs manquantes et aberrantes est de revenir aux données brutes pour intégrer les pics manquants. C'est ce que propose l'outil « `xcms.fillPeaks` » inclus dans la suite XCMS (Smith et al., 2006). Cette approche garantit d'imputer les valeurs manquantes par des valeurs les plus justes possibles au regard des données réelles. Cependant, cette approche est susceptible de générer un certain nombre de zéros si les signaux possèdent un bruit très bas (typiquement à cause de l'application d'un filtre au moment de leur génération).

Pour résumer, la première approche (imputer une valeur unique) est très simple à implémenter, mais est connue pour donner de mauvais résultats. La deuxième (utilisation de méthodes de traitement de données pour imputer les valeurs manquantes) semble donner de bons résultats, et est très utilisée mais plus compliquée d'implémentation. En effet, il existe de nombreuses méthodes différentes et le choix de la méthode semble critique sur la qualité des résultats (Di Guida et al., 2016; Wei et al., 2018). Enfin, la dernière approche (ré-analyse des données brutes) est censée donner les valeurs les plus justes, et son implémentation peut se faire en ligne avec l'extraction des pics par XCMS (ce package étant lui-même implémenté dans plusieurs interfaces graphiques et plateformes de calcul comme on l'a vu auparavant). Cependant, c'est une méthode « clés-en-main », et sur laquelle l'utilisateur a peu de possibilité d'action. Dans ce travail, cette approche sera choisie dans un premier temps, et éventuellement complétée par une approche basée sur le traitement de données si des manquements apparaissent.

Les études de type métabolomique utilisent souvent des séquences analytiques de grande taille (plusieurs centaines d'injections), et sont donc souvent l'objet de phénomènes de dérive analytique (Wehrens et al., 2016). Cette étape n'a pas fait l'objet d'études particulières pour la sécurité sanitaire chimique des aliments. La présence d'échantillon QC (voir 2.3.3) permet de visualiser si la séquence présente une dérive, et de la corriger si nécessaire. Pour chaque pic, on peut modéliser la dérive par divers modèles, les plus communs étant les modèles linéaires et la méthode LOESS (« Locally Estimated Scatterplot Smoothing », (Dunn, Wilson, Nicholls, & Broadhurst, 2012). Grâce à la modélisation de la dérive instrumentale, on peut ensuite corriger le signal pour chaque échantillon en fonction de son ordre d'injection. Ces outils sont notamment implémentés dans la plateforme de calcul en ligne W4M (Giacomoni et al., 2015) basée sur XCMS.

Pour pouvoir détecter les contaminants traces, il est nécessaire de réduire la taille de la matrice de données (Knolhoff et al., 2016). Dans ce but, l'utilisation de tests statistiques univariés (test de Student, t-test), couplés avec le calcul d'un « fold change » (FC, signal moyen du groupe d'échantillons d'étude / signal moyen du groupe d'échantillons de contrôle, que l'on peut traduire comme « facteur multiplicatif ») a été implémentée (Knolhoff et al., 2016). Toutefois, comme les auteurs l'expliquent dans cette étude, l'utilisation de seuils de filtration fixes (traditionnellement p-values < 0,05 pour les t-tests et FC > 2) est limitante (chaque pic ayant son propre comportement et qualité de mesure). Aussi, les auteurs encouragent-ils les discussions sur cette étape. Par ailleurs, une méthodologie alternative

pour cette filtration a été proposée en provenance du champ de la métabolomique, par le calcul d’un FC minimum pertinent par le biais d’un calcul de propagation d’erreur (Ortmayr, Charwat, Kasper, Hann, & Koellensperger, 2017). À ce jour, cette approche n’a pas été testée sur des thématiques d’analyse agro-alimentaire.

Les méthodes de normalisation et de mise à l’échelle ont fait l’objet d’études pour la détection non-ciblée des contaminants de l’aliment. Les transformations log sont ainsi couramment implémentées dans ce but (Antignac et al., 2011; Godzien et al., 2013), les données de type LC-MS ayant la particularité de présenter une grande gamme dynamique. Ces transformations sont couramment couplées, comme cela se fait également en métabolomique, à des méthodes de mises à l’échelle dont les plus courantes sont « Autoscaling » (1/écart type du pic) et « Pareto scaling » (1/racine carré de l’écart type du pic) (Yi et al., 2016), la deuxième étant souvent préférée car elle distord moins les données (Antignac et al., 2011; Hu & Xu, 2013). Une autre méthode de normalisation, PQN (Probabilistic Quotient Normalization) a également été identifiée comme prometteuse dans les études métabolomiques (Di Guida et al., 2016; Gorrochategui, Jaumot, Lacorte, & Tauler, 2016; Habchi et al., 2017) et consiste à utiliser les QC pour compenser un potentiel effet dû à la préparation de l’échantillon (notamment de dilution) (Dieterle, Ross, Schlotterbeck, & Senn, 2006).

2.3.5.3. DÉTECTION DES ÉCHANTILLONS ET DES PICS SUSPECTS PAR ANALYSE MULTIVARIÉE

Les étapes de détection des pics, puis de prétraitement des données permettent d’obtenir un tableau de données prêt pour l’analyse. Pour détecter les échantillons suspects et les composés associés, les méthodes statistiques de choix sont les méthodes multivariées (Antignac et al., 2011; Castro-Puyana et al., 2017; Hu & Xu, 2013). On distingue deux grandes familles de méthodes multivariées : les méthodes supervisées, et les méthodes non-supervisées (Antignac et al., 2011; Yi et al., 2016).

Dans une méthode supervisée, la méthode utilise à la fois les données, et des informations quantitatives ou qualitatives sur les échantillons pour être à même de prédire l’état du système à partir des mesures (Brereton, 2007). Cela peut être de prédire une grandeur quantitative (on parle alors de régression), ou qualitative (on parle de classification). La plus connue de ces méthodes est la méthode des moindres carrés partiels (Partial Least Square ou PLS), opérant en régression (PLS-R) ou en classification (PLS-DA) (Gromski et al., 2015). En analyse agro-alimentaire par MS, ces méthodes sont très utilisées pour les problématiques d’authenticité et de qualité des produits (Callao & Ruisánchez, 2018; Riedl et al., 2015), qui sont des problématiques de classification. Le but est davantage d’établir un modèle pouvant prédire si un échantillon est authentique ou adultéré, ce qui diffère des problématiques de sécurité sanitaire où on cherche à détecter des molécules potentiellement inconnues. Les approches de régression sont également utilisées pour quantifier des phénomènes, par exemple des adultérations (Tsopelas, Konstantopoulos, & Kakoulidou, 2018). Une des principales caractéristiques de ces méthodes est de nécessiter des étapes de validation pour s’assurer que le modèle obtenu est bien porté par l’information

présente dans les échantillons et non par un artefact statistique (Gromski et al., 2015). Cette validation peut être interne (le jeu de données est séparé en un sous-jeu de construction du modèle, et un sous-jeu de validation) ou externe (utilisation d'un jeu de données complètement indépendant du modèle pour en tester la validité). Ces méthodes peuvent également être appliquées à des travaux exploratoires mais cela nécessite l'implémentation de nombreuses procédures de validation (Gorochategui et al., 2016). Elles présentent un potentiel pour les applications de sécurité sanitaire chimiques des aliments (Antignac et al., 2011; Knolhoff & Croley, 2016), mais aucune application n'a été reportée à ce jour.

Les méthodes non-supervisées sont des méthodes qui n'utilisent que les données, sans informations sur les échantillons (elles ne nécessitent donc pas de procédures de validation comme les méthodes supervisées) (Brereton, 2007). Ce sont des méthodes exploratoires qui permettent de réduire la dimensionnalité des données, et ainsi de visualiser les structures pouvant exister au sein des échantillons (Berrueta, Alonso-Salces, & Héberger, 2007). Ces méthodes sont utilisées pour des problématiques d'authenticité et d'adultération (Callao & Ruisánchez, 2018). Elles ont également été appliquées avec succès sur des thématiques de sécurité sanitaire chimique des aliments (Inoue et al., 2015; Knolhoff et al., 2016; Tengstrand et al., 2013). Le grand avantage de ces méthodes est qu'avec une exploration à l'aveugle du jeu de données, toute nouvelle contamination même non-attendue ou inconnue peut en théorie être détectée, contrairement à des méthodes utilisant des modèles comme les méthodes supervisées, pour lesquelles les même variables (dans notre cas, des pics chromatographiques) doivent être mesurées dans le jeu de données de construction du modèle et dans les échantillons que l'on souhaite étudier. La plus connue des méthodes non-supervisées, et la plus utilisée pour l'exploration de données en chimie analytique est l'analyse en composantes principales (PCA) (Lavine, 2006). Le but de la PCA est de réduire la dimensionnalité des données, tout en conservant l'information présente (Lavine, 2006). Cette réduction est atteinte en combinant les variables (dans notre cas, les pics chromatographiques) existantes pour former de nouvelles variables appelées « composantes ». Cela permet de visualiser les données, et ainsi de détecter si un motif (par exemple, une séparation entre groupes d'échantillons) est visible, ainsi que de détecter les variables qui en sont responsables (Latrille, 2016). Avec la PCA, les composantes sont construites de façon à en maximiser la variance des individus (=des échantillons) pour chaque composante principale (PC). Il existe des méthodes alternatives à la PCA, on peut notamment citer l'analyse en composantes indépendantes (ICA) (Jouan-Rimbaud Bouveresse & Rutledge, 2016), développée initialement en traitement du signal pour décomposer des signaux complexes en signaux « sources » (c.à.d. ne représentant qu'un seul phénomène) indépendants. Son application pour les problématiques de sécurité sanitaire chimique de l'aliment n'a pas encore été évaluée, mais cette méthode présente des résultats prometteurs sur des études métabolomiques par MS (Habchi et al., 2017; Scholz, Gatzek, Sterling, Fiehn, & Selbig, 2004), ainsi que pour l'évaluation des emballages alimentaires (Habchi et al., 2018).

2.3.5.4. ANNOTATION

Une fois les signaux suspectés comme étant responsables d'une contamination isolés, il est nécessaire d'émettre des hypothèses quant à leur identité. Quatre niveaux de confiance d'identification ont été définis en métabolomique pour les expériences de découvertes de biomarqueurs (Dunn et al., 2013; Sumner et al., 2007). Au niveau 1, un composé est qualifié « d'identifié », c'est-à-dire qu'on a pu comparer ses propriétés (deux propriétés indépendantes ou plus) avec celles d'un standard analysé dans les mêmes conditions (par exemple temps de rétention et spectre de masse). Au niveau 2, un composé est qualifié d'« annoté putativement » (en anglais « putatively annotated »), c.à.d. que la proposition d'identité se fait en se basant sur une ou plusieurs informations contenues dans des bases de données publiques (par exemple, masse exacte correspondante et massif isotopique). Au niveau 3, on parle d'« annotation putative de la classe du composé ». Cela se base sur une observation permettant de déduire l'appartenance du composé à une classe, sans pour autant pouvoir donner de proposition d'identité précise (par exemple, lors de la présence d'un massif isotopique caractéristique d'un composé chloré, mais sans possibilité de générer de formule brute). Enfin, le niveau 4 recoupe tous les composés inconnus, pour lesquels aucune hypothèse ne peut être donnée quant à leur identité.

La plupart des publications reportées se contentent du niveau 2 (annotation putative) (Inoue et al., 2015; Knolhoff et al., 2016; Kunzelmann et al., 2018), une seule publication allant jusqu'à l'identification complète pour certains composés (Cotton et al., 2014).

La méthodologie généralement employée consiste à utiliser des bases de données en ligne à spectre large (HMDB, ou bien Metlin par exemple), éventuellement couplées à des générateurs de formules brutes (Antignac et al., 2011; Gorrochategui et al., 2016; Knolhoff & Croley, 2016). Il existe beaucoup de bases de données de spectres de masse de composés chimiques. Certaines sont généralistes (Metlin, MassBank, HMDB, KEGG parmi les principales (Milman & Zhurkovich, 2016)), d'autres sont spécifiques aux composés toxiques, comme T3DB (Wishart et al., 2015). Il existe aussi des bases de données ne présentant pas de données spectrales, mais pouvant receler des informations utiles pour l'annotation (formule brute, propriétés physico-chimiques), par exemple la Pesticides Properties DataBase (PPDB) (Lewis, Tzilivakis, Warner, & Green, 2016), ainsi que la base de données OpenFoodTox éditée par l'EFSA (<https://www.efsa.europa.eu/en/microstrategy/openfoodtox>) (Dorne et al., 2017). Suivant la stratégie adoptée et les bases de données utilisées, il est possible de les interroger directement en ligne, ou bien de télécharger tout ou partie de certaines bases pour des utilisations en local.

Les contaminants de l'aliment, et de manière générale les composés exogènes au vivant, peuvent présenter des motifs chimiques récurrents. Un de ces motifs est la présence d'atomes halogénés dans la formule brute (Cariou et al., 2016; Cotton et al., 2014; Roullier et al., 2016). Celui-ci est aisément détectable en spectrométrie de masse grâce à la différence de masse et le rapport caractéristique de son isotope M+2 par rapport au pic monoisotopique. Cette stratégie a été employée, à la fois dans des

matrices environnementales et alimentaires, par plusieurs études (Cariou et al., 2016; Cotton et al., 2014; Roullier et al., 2016). Une autre stratégie consiste à tirer parti du défaut de masse inhérent aux composés halogénés (Andersen et al., 2016; Cariou et al., 2016).

3. CONCLUSION GÉNÉRALE DE L'ÉTUDE BIBLIOGRAPHIQUE

Dans la première partie, nous avons vu la complexité de la problématique de l'évaluation de la qualité sanitaire chimique, notamment concernant notre produit d'étude, le thé. C'est en effet un produit complexe composé de plusieurs centaines de molécules qui peuvent être très concentrées (de l'ordre du %). Les principales classes de composés naturels du thé identifiées sont les composés phénoliques, les dérivés d'acides aminés et les xanthines. Ses contaminants quant à eux sont variés en termes de sources et de structures chimiques, et les niveaux autorisés, pour ceux règlementés, à l'état de traces (depuis le $\mu\text{g}/\text{kg}$ jusqu'au mg/kg). Les principales sources de contamination identifiées sont : la matière première (principales classes de contaminants : pesticides, mycotoxines, alcaloïdes et contaminants de l'environnement), les procédés et le stockage (principales classes de contaminants : mycotoxines, HAP, néoformés) et enfin les matériaux au contact (migrants notamment). Un tableau présentant des contaminants et des composés naturels représentatif a été réalisé (voir Tableau 1.3).

L'évaluation de la qualité sanitaire chimique d'un aliment se fait actuellement majoritairement par des méthodes dites « ciblées » visant à quantifier avec précision un certain nombre de contaminants connus. Ces méthodes sont composées de deux grandes étapes : la préparation de l'échantillon suivie d'une analyse instrumentale. En raison de sa polyvalence et de son champ d'application, la technique « LC-MS » a été choisie dans ce travail. Un aperçu d'une sélection de méthodes d'analyse de contaminants dans le thé et quelques autres matrices par LC-MS est présenté en Tableau 1.4. Plusieurs enseignements ont pu être tirés de ces méthodes ciblées. Premièrement, il est nécessaire d'utiliser une méthode de préparation de l'échantillon la plus générique possible en raison de la diversité de structures chimiques rencontrées. La méthode type relevée est constituée d'une extraction (par LLE, extraction directe par solvant ou dilution par solvant suivant la nature de l'échantillon), éventuellement complétée par une purification sommaire par SPE ou dSPE et évaporation/reconstitution pour concentrer l'extrait. Les méthodes instrumentales rapportées reposent sur l'utilisation d'une méthode générique de type « phase inverse » pour la partie LC. Le spectromètre de masse utilisé est soit un appareil à basse résolution (le plus souvent QqQ) pour effectuer de la MS/MS en mode MRM, soit un spectromètre de masse à haute résolution (Orbitrap ou ToF) opéré en mode de balayage complet. La source d'ionisation utilisée est systématiquement l'ESI, opérée en positif et/ou en négatif. Il est intéressant de noter que les méthodes basées sur la HRMS présentent des performances similaires en termes de sensibilité, de sélectivité et de recouvrements que les méthodes basées sur la MS/MS à basse résolution. De plus, certains travaux (Mol, Plaza-Bolaños, Zomer, De Rijk, et al., 2008), confirmés par la suite (Lacina et al., 2012), suggèrent qu'il

est possible d'obtenir de bons résultats quantitatifs même sans l'étape de purification lors de l'utilisation de la HRMS en mode de balayage complet. Il n'existe pas encore, à notre connaissance, de méthodes (traitement de l'échantillon + méthode instrumentale) proposant l'analyse de plus de trois classes de contaminants distincts dans le thé.

Le développement de méthodes analytiques au champ de plus en plus large et l'essor des études métabolomiques ont conduit, au début des années 2010, à identifier les approches de type « métabolomiques » comme prometteuses pour l'évaluation de la qualité sanitaire chimique des aliments (Antignac et al., 2011; Castro-Puyana et al., 2017). Plusieurs publications reportant des travaux préliminaires ont été relevées (voir Tableau 1.8), et un processus générique a pu être déterminé (voir Figure 1.6). Comme pour l'analyse ciblée multi-classes, les méthodes analytiques utilisées reposent sur une combinaison d'un traitement générique de l'échantillon et d'une méthode LC-HRMS à spectre large. Les outils de traitement des données utilisés varient suivant les auteurs, mais une méthodologie commune peut être dégagée (voir Figure 1.6). On distingue deux approches, une consistant à comparer les échantillons « suspects » avec de échantillons de contrôle pour détecter d'éventuels contaminants (Inoue et al., 2015; Knolhoff et al., 2016; Kunzelmann et al., 2018; Tengstrand et al., 2013), et l'autre consistant à détecter la présence de motifs chimiques caractéristiques de beaucoup de contaminants (en l'occurrence, la présence d'hétéroatomes halogénés dans une molécules) au sein des données (Cotton et al., 2014). La première a l'avantage d'être à même de détecter en théorie n'importe quelle contamination ; cependant, la nécessité d'avoir un échantillon témoin (c.à.d. non-contaminé) similaire à l'échantillon à caractériser présente un obstacle, et ce d'autant plus pour des produits complexes et sujets à de fortes variabilité inter-échantillons (Antignac et al., 2011; Knolhoff & Croley, 2016; Knolhoff et al., 2016). La deuxième approche quant à elle ne nécessite pas de disposer d'un échantillon témoin, mais elle est limitée en termes de diversité chimique détectable. Les niveaux de contaminants analysés par les deux approches sont pertinents par rapport aux réglementations européennes (aux alentours de la dizaine de $\mu\text{k}/\text{kg}$). Combiner les deux approches permettrait à la fois de pouvoir proposer une détection totalement non-ciblée des contaminants, et en même temps de faciliter le processus d'annotation, considéré par plusieurs auteurs (y compris dans le champ de la métabolomique) comme le plus chronophage et le plus délicat (Antignac et al., 2011; Gorrochategui et al., 2016).

Les études existantes sur l'analyse non-ciblée des contaminants (voir Tableau 1.8) n'utilisent les résultats que d'un seul mode d'ionisation (généralement le mode positif). Or, il semble raisonnable de penser que l'utilisation des deux modes d'ionisation permettrait d'élargir encore le champ d'application des méthodes actuelles (Antignac et al., 2011), comme cela se pratique par exemple en métabolomique (Gika et al., 2014) ou en analyse ciblée de contaminants (Mol, Plaza-Bolaños, Zomer, Rijk, et al., 2008). Les matrices utilisées sont relativement simples et présentent peu de variabilité (lait, formules infantiles et jus d'orange industriels). Il est donc nécessaire d'élargir le domaine de ces méthodes à des produits plus complexes comme le thé par exemple. Les processus de traitement de données proposés jusque-là

ne sont appliqués que dans le cas pour lequel ils ont été développés, à l'exception notable de l'approche proposée par Tengstrand en 2013 (Tengstrand et al., 2013) qui a été appliquée avec succès par Kunzelman (Kunzelmann et al., 2018) à un autre jeu de données (même instrument et même méthode, mais matrice différente et niveaux de contaminants plus faibles). Pour s'assurer de la robustesse des approches proposées, il serait nécessaire d'évaluer leurs performances sur de multiples cas de contamination (variation de contaminants, de matrices et de niveaux de contamination), mais également sur des jeux de données provenant d'instruments de technologies variées (par exemple analyseurs ToF Orbitrap). Pour ce faire, il est nécessaire de disposer d'un processus de traitement des données générique, et qui ne soit pas dépendant d'une plateforme instrumentale particulière. Cela implique de travailler avec des outils libres et ouverts, à l'image des travaux de Tengstrand et de Cotton (Cotton et al., 2014; Tengstrand et al., 2013). Le développement d'un processus de traitement de données le plus automatisé possible est envisagé, ce processus incluant notamment, en fonction des résultats préliminaires qui seront obtenus, une étape de gestion des valeurs manquantes et aberrantes ainsi qu'une étape de nettoyage des données pour les préparer à l'analyse multivariée. Pour l'analyse multivariée, deux méthodes ont retenu notre attention, la première est la PCA, qui est, on l'a vu, la méthode multivariée de référence pour l'exploration des données, et la deuxième est l'ICA, qui montre des propriétés intéressantes pour résoudre les mélanges de signaux complexes et qui sera également testée. La dernière étape (l'annotation des composés suspects) se fera en combinant les approches « classiques » par recherche dans les bases de données et génération de formules brutes avec la détection des massifs isotopiques caractéristiques liés aux composés halogénés.

Enfin, le développement de processus et de stratégies de traitement de données pourrait être accéléré grâce à l'ouverture des jeux de données utilisés comme cela se fait dans le domaine de la métabolomique (Hardy & Taylor, 2007; Salek, Haug, & Steinbeck, 2013). Les jeux de données utilisés durant cette thèse seront ainsi mis à disposition de la communauté scientifique *via* des plateformes d'échange de données.

4. MÉTHODOLOGIE DES TRAVAUX DE THÈSE

Le point de départ de ces travaux de thèse est le constat des limites que montrent les approches analytiques ciblées d'analyse des contaminants de l'aliment. La principale limite, mise en avant par plusieurs auteurs (Antignac et al., 2011; Castro-Puyana et al., 2017; Hu & Xu, 2013; Knolhoff & Croley, 2016) est leur incapacité à détecter un contaminant inconnu ou non-attendu. Les approches analytiques les plus prometteuses pour réussir cette détection sont celles venant des études métabolomiques selon les auteurs cités ci-dessus. Cependant, quelques différences opposent les études métabolomiques et l'évaluation de la qualité sanitaire chimique d'aliments. Tout d'abord, les matrices agro-alimentaires sont plus complexes et présentent une diversité plus grande que les matrices usuellement analysées en métabolomique (typiquement sang et urine) (Knolhoff & Croley, 2016). Les contaminants de l'aliment

présentent de plus une diversité chimique considérable. La méthode analytique utilisée devra donc avoir un spectre d’application suffisamment large pour être à même de détecter d’éventuels contaminants non-attendus.

La problématique scientifique n’est pas la même dans une étude métabolomique ou pour la détection non-ciblée des contaminants de l’aliment. Une étude métabolomique a souvent comme objectif d’élucider la réponse métabolique d’un organisme à un phénomène biologique (par exemple l’exposition à une molécule), par exemple en recherchant un biomarqueur fiable, qui sera validé par l’étude d’autres jeux d’échantillons. L’objectif final peut être de mieux comprendre un phénomène, ou bien d’utiliser le biomarqueur pour l’établissement d’un diagnostic. Dans chaque cas, le biomarqueur doit être identifié analytiquement (expérience de MS/MS avec un standard par exemple) et validé biologiquement (répétition des expériences) et sa voie métabolique élucidée (Gorrochategui et al., 2016). Dans notre cas, l’objectif est d’établir une approche qui permettrait de détecter et de caractériser de façon non-ciblée une contamination de l’aliment. La seule validation nécessaire est la validation analytique des contaminants détectés, avec les différents niveaux de confiance vus ci-dessus partie 2.3.5.4. L’approche en elle-même est validée par sa capacité à détecter en aveugle les contaminants présents dans un autre jeu d’échantillon que celui de développement. Ses performances sont jugées au regard des taux de détection obtenus sur des cas « tests », sa sensibilité par rapport aux limites réglementaires portant sur les contaminants, ainsi que leurs niveaux usuels relevés dans les aliments.

Enfin, les approches métabolomiques par HRMS sont étudiées depuis plus d’une dizaine d’année alors que la première preuve de concept non-ciblée de détection des contaminants date de 2013 (Tengstrand et al., 2013). Malgré les quelques travaux publiés sur le sujet par la suite (Cotton et al., 2016; Inoue et al., 2015; Knolhoff et al., 2016; Kunzelmann et al., 2018), il y a beaucoup moins de recul sur ces approches que sur les approches métabolomiques, et les méthodologies proposées manquent d’homogénéité et de tests de robustesse.

Cette thèse s’est déroulée en quatre grandes étapes :

- 1) Pour répondre aux deux premiers points soulevés ci-dessus (complexité des matrices agro-alimentaires et diversité des contaminants chimiques), une méthode analytique (traitement de l’échantillon + analyse instrumentale) générique ciblée a été développée pour la quantification d’un panel de 32 contaminants choisis pour leur diversité de structure et d’origine dans le produit d’étude (le thé). Cette méthode sera développée pour des niveaux de contamination pertinents au regard des réglementations existantes et des niveaux usuels reportés.
- 2) Grâce à cette méthode analytique, une preuve de concept de l’approche non-ciblée basée sur les approches métabolomiques a été développée. Appliquée tout d’abord sur une situation de contamination simple avec des contaminants connus, elle a été complétée avec une première validation sur un cas légèrement plus complexe.

- 3) Au regard des résultats obtenus lors de cette preuve de concept, un travail méthodologique a été mené sur le processus de traitement des données afin de l'améliorer et d'aboutir à un algorithme à même d'être implémenté de façon automatique sur tout nouveau jeu de données. Ces travaux ont également eu pour but de produire des connaissances sur le comportement d'outils usuels dans les études métabolomiques mais encore inédits pour l'évaluation de la contamination sanitaire chimique des aliments.
- 4) Enfin, le processus complet (procédé analytique et algorithme de traitement des données) a été appliqué à plusieurs situations de contamination en mobilisant plusieurs instruments. Cette validation a été menée en aveugle (les analyses, ainsi que le traitement des données ont été réalisés sans connaissance de la conception de l'étude).

Chaque étape a donné lieu à un article scientifique, présenté dans sa dernière version (parue, soumise ou pour soumission prochaine selon le cas).

5. RÉFÉRENCES

- Agence nationale de sécurité sanitaire de l'alimentation de l'environnement et du travail. (2015). Saisine n° 2015 - SA - 0180 Avis de l' Agence nationale de sécurité sanitaire de l' alimentation, de l' environnement et du travail relatif à «une demande d'avis lié à un projet d'arrêté relatif aux variétés comestibles de champignons de culture et sauva. Journal Officiel de l'Union Européenne, 38. Retrieved from www.anses.fr
- Anastassiades, M., Lehotay, S. J., & Schenck, F. J. (2003). Fast and Easy Multiresidue Method Employing Acetonitrile Extraction/Partitioning and " Dispersive Solid-Phase Extraction " for the Determination of Pesticide Residues in Produce. *Journal of AOAC International*, 86(2). Retrieved from <http://www.ingentaconnect.com/content/aoac/jaoac/2003/00000086/00000002/art00023>
- 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>
- Andrade-Eiroa, A., Canle, M., Leroy-Cancellieri, V., & Cerdà, V. (2016). Solid-phase extraction of organic compounds: A critical review (Part I). *TrAC - Trends in Analytical Chemistry*, 80, 641–654. <https://doi.org/10.1016/j.trac.2015.08.015>
- ANSES. (2018). Bisphénol A. Retrieved April 26, 2018, from <https://www.anses.fr/fr/content/bisphénol>
- 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>
- Bayen, S., Koroleva, E., Lee, H. K., & Obbard, J. P. (2005). Persistent organic pollutants and heavy metals in typical seafoods consumed in Singapore. *Journal of Toxicology and Environmental Health - Part A*, 68(3), 151–166. <https://doi.org/10.1080/15287390590890437>
- Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, 1158(1–2), 196–214. <https://doi.org/10.1016/j.chroma.2007.05.024>

- Brereton, R. G. (2007). *Applied Chemometrics for Scientists*. *Applied Chemometrics for Scientists*. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470057780>
- Brimer, L. (2011). *Chemical food safety*. *Chemical Food Safety*. <https://doi.org/10.1079/9781845936761.0000>
- Callao, M. P., & Ruisánchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control*, *86*, 283–293. <https://doi.org/10.1016/j.foodcont.2017.11.034>
- Camel, V. (2014). Extraction Methodologies: General Introduction. *Encyclopedia of Analytical Chemistry*, 1–24. <https://doi.org/10.1002/9780470027318.a9902>
- Cariou, R., Omer, E., Léon, A., Dervilly-Pinel, G., & Le Bizec, B. (2016). Screening halogenated environmental contaminants in biota based on isotopic pattern and mass defect provided by high resolution mass spectrometry profiling. *Analytica Chimica Acta*, *936*, 130–138. <https://doi.org/10.1016/j.aca.2016.06.053>
- Carroll, D. I., Dzidic, I., Horning, E. C., & Stillwell, R. N. (2006). Atmospheric Pressure Ionization Mass Spectrometry. In *Encyclopedia of Analytical Chemistry* (Vol. 17, pp. 337–406). <https://doi.org/10.1002/9780470027318.a6003>
- Castro-Puyana, M., Pérez-Míguez, R., Montero, L., & Herrero, M. (2017). Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability. *TrAC - Trends in Analytical Chemistry*, *93*, 102–118. <https://doi.org/10.1016/j.trac.2017.05.004>
- Caude, M., & Jardy, A. (1996). Méthodes chromatographiques. *Techniques de l'ingénieur*, *33*(0), 1–6.
- Chang, K. (2015). *World tea production and trade Current and future development*. *Food and Agriculture Organisation*.
- Chen, H., Yin, P., Wang, Q., Jiang, Y., & Liu, X. (2014). A Modified QuEChERS Sample Preparation Method for the Analysis of 70 Pesticide Residues in Tea Using Gas Chromatography-Tandem Mass Spectrometry. *Food Analytical Methods*, *7*(8), 1577–1587. <https://doi.org/10.1007/s12161-014-9791-0>
- Chen, L., Chen, J. F., Guo, Y., Li, J., Yang, Y., Xu, L., & Fu, F. (2014). Study on the simultaneous determination of seven benzoylurea pesticides in Oolong tea and their leaching characteristics during infusing process by HPLC-MS/MS. *Food Chemistry*, *143*, 405–410. <https://doi.org/10.1016/j.foodchem.2013.08.027>
- Christian, G., Dasgupta, P., & Schug, K. (2014). *Analytical Chemistry*.
- 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*(11), 3009–3018. <https://doi.org/10.1021/acs.jafc.7b06118>
- Cifuentes, A. (2009). Food analysis and foodomics. *Journal of Chromatography A*, *1216*(43), 7109. <https://doi.org/10.1016/j.chroma.2009.09.018>
- 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. <https://doi.org/10.1016/j.chroma.2014.06.100>
- Commission européenne. (1993). Règlement (CEE) n°315/93 du conseil du 8 février 1993 portant établissement des procédures communautaires relatives aux contaminants dans les denrées alimentaires. *Journal Officiel de l'Union Européenne*, 1–5. Retrieved from <http://data.europa.eu/eli/reg/1993/315/oj>
- Commission européenne. (2002). Règlement (CE) n°178/2002 du parlement européen et du conseil du 28 janvier 2002 établissant les principes généraux et les prescriptions générales de la législation alimentaire, instituant l'autorité européenne de sécurité des aliments et fixant des procédés. *Journal Officiel de l'Union Européenne*. Retrieved from <http://data.europa.eu/eli/reg/2002/178/oj>
- Commission européenne. (2004a). Règlement (CE) n° 1935/2004 du Parlement européen et du Conseil

- du 27 octobre 2004 concernant les matériaux et objets destinés à entrer en contact avec des denrées alimentaires et abrogeant les directives 80/590/CEE et 89/109/CEE. *Journal Officiel de l'Union Européenne*, 1–21. Retrieved from <http://data.europa.eu/eli/reg/2004/1935/oj>
- Commission européenne. (2004b). Règlement (CE) n° 853/2004 du parlement européen et du conseil u 29 avril 2004 fixant des règles spécifiques d'hygiène applicables aux denrées alimentaires d'origine animale. *Journal Officiel de l'Union Européenne*, 151. Retrieved from <http://data.europa.eu/eli/reg/2004/853/oj>
- Commission européenne. (2004c). Règlement (CE) no 882/2004 du parlement européen et du conseil du 29 avril 2004 relatif aux contrôles officiels effectués pour s'assurer de la conformité avec la législation sur les aliments pour animaux et les denrées alimentaires et avec les disposition. *Journal Officiel de l'Union Européenne*. Retrieved from <http://data.europa.eu/eli/reg/2004/882/oj>
- Commission européenne. (2005). Règlement (CE) no 396/2005 du parlement européen et du conseil du 23 février 2005 concernant les limites maximales applicables aux résidus de pesticides présents dans ou sur les denrées alimentaires et les aliments pour animaux d'origine végétale et anima. *Journal Officiel de l'Union Européenne*, 1–66. Retrieved from <http://data.europa.eu/eli/reg/2005/396/oj>
- Commission européenne. (2009). Règlement CE n°1107/2009 du parlement européen et du conseil du 21 octobre 2009 concernant la mise sur le marché des produits phytopharmaceutiques et abrogeants les directives 79/117/CEE et 91/414/CEE du Conseil. *Journal Officiel de l'Union Européenne*, 309(1), 1–50. Retrieved from <http://data.europa.eu/eli/reg/2009/1107/oj>
- Commission européenne. (2010). Communication de la commission au conseil, au parlement Européen et au comité économique et social européen Sur la mise en œuvre de la stratégie communautaire concernant les dioxines, les furannes Et les polychlorobiphényles (COM(2001)593) – troisième rapp. *Journal Officiel de l'Union Européenne*, 52–54. Retrieved from <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=LEGISSUM:121280>
- Commission européenne. (2011a). Règlement (UE) n°10/2011 de la commission du 14 janvier 2011 concernant les matériaux et objets en matière plastique destinés à entrer en contact avec des denrées alimentaires. *Journal Officiel de l'Union Européenne*. Retrieved from <http://data.europa.eu/eli/reg/2011/10/oj>
- Commission européenne. (2011b). Règlement (UE) n ° 1259/2011 de la Commission du 2 décembre 2011 modifiant le règlement (CE) n°1881/2006 en ce qui concerne les teneurs maximales en dioxines, en PCB de type dioxine et en PCB autres que ceux de type dioxine des denrées alimentaires. *Journal Officiel de l'Union Européenne*, 18–23. Retrieved from <http://data.europa.eu/eli/reg/2011/1259/oj>
- Commission européenne. (2015a). Règlement (UE) 2015/1940 de la Commission du 28 octobre 2015 modifiant le règlement (CE) n° 1881/2006 en ce qui concerne les teneurs maximales de certaines céréales brutes en sclérotés d'ergot et les dispositions relatives à la surveillance et aux rapport. *Journal Officiel de l'Union Européenne*, 4. Retrieved from <http://data.europa.eu/eli/reg/2015/1940/oj>
- Commission européenne. (2015b). Règlement d'exécution (UE) 2015/175 de la commission du 5 février 2015 fixant les conditions particulières applicables à l'importation de gomme de guar originaire ou en provenance de l'inde, en raison des risques de contamination par le pentachlorophénol. *Journal Officiel de l'Union Européenne*, 10–15. Retrieved from http://data.europa.eu/eli/reg_impl/2015/175/oj
- Commission européenne. (2016). Règlement (UE) 2016/239 de la Commission du 19 février 2016 modifiant le règlement (CE) n° 1881/2006 en ce qui concerne les teneurs maximales en alcaloïdes tropaniques dans certains aliments à base de céréales pour nourrissons et enfants en bas âge. *Journal Officiel de l'Union Européenne*, 3. Retrieved from <http://data.europa.eu/eli/reg/2016/239/oj>
- Commission européenne. (2017). Règlement (UE) 2017/2158 de la Commission du 20 novembre 2017

- établissant des mesures d’atténuation et des teneurs de référence pour la réduction de la présence d’acrylamide dans les denrées alimentaires. *Journal Officiel de l’Union Européenne*, 21. Retrieved from <http://data.europa.eu/eli/reg/2017/2158/oj>
- Commission européenne. (2018). Règlement (UE) 2018/70 de la commission du 16 janvier 2018 modifiant les annexes II, III et IV du règlement (CE) n° 396/2005. *Journal Officiel de l’Union Européenne*, (5), 24–52. Retrieved from <http://data.europa.eu/eli/reg/2018/70/oj>
- Commission Européenne. (2006). Règlement (UE) N° 1881/2006 de la commission du 19 décembre 2006 portant fixation de teneurs maximales pour certains contaminants dans les denrées alimentaires. *Journal Officiel de l’Union Européenne*, 20. Retrieved from <http://data.europa.eu/eli/reg/2006/1881/oj>
- Cotton, J., Leroux, F., Broudin, S., Marie, M., Corman, B., Tabet, J. C., ... Junot, C. (2014). High-resolution mass spectrometry associated with data mining tools for the detection of pollutants and chemical characterization of honey samples. *Journal of Agricultural and Food Chemistry*, 62(46), 11335–11345. <https://doi.org/10.1021/jf504400c>
- Cotton, J., Leroux, F., Broudin, S., Poirel, M., Corman, B., Junot, C., & Ducruix, C. (2016). Development and validation of a multiresidue method for the analysis of more than 500 pesticides and drugs in water based on on-line and liquid chromatography coupled to high resolution mass spectrometry. *Water Research*, 104, 20–27. <https://doi.org/10.1016/j.watres.2016.07.075>
- De Livera, A. M., Sysi-Aho, M., Jacob, L., Gagnon-Bartsch, J. A., Castillo, S., Simpson, J. A., & Speed, T. P. (2015). Statistical Methods for Handling Unwanted Variation in Metabolomics Data. *Analytical Chemistry*, 87(7), 3606–3615. <https://doi.org/10.1021/ac502439y>
- Di Guida, R., Engel, J., Allwood, J. W., Weber, R. J. M., Jones, M. R., Sommer, U., ... Dunn, W. B. (2016). Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics*, 12(5). <https://doi.org/10.1007/s11306-016-1030-9>
- Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics. *Analytical Chemistry*, 78(13), 4281–4290. <https://doi.org/10.1021/ac051632c>
- Dorne, J. Lou, Richardson, J., Kass, G., Georgiadis, N., Monguidi, M., Pasinato, L., ... Robinson, T. (2017). Editorial: OpenFoodTox: EFSA’s open source toxicological database on chemical hazards in food and feed. *EFSA Journal*, 15(1), 2016–2018. <https://doi.org/10.2903/j.efsa.2017.e15011>
- DTU. (2018). Fooddata - Teal leaves. Retrieved April 28, 2018, from <https://frida.fooddata.dk/ShowFood.php?foodid=537&2>
- Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-Mcintyre, S., Anderson, N., ... Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 6(7), 1060–1083. <https://doi.org/10.1038/nprot.2011.335>
- Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., ... Viant, M. R. (2013). Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, 9(SUPPL.1), 44–66. <https://doi.org/10.1007/s11306-012-0434-4>
- Dunn, W. B., Wilson, I. D., Nicholls, A. W., & Broadhurst, D. (2012). The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*, 4(18), 2249–2264. <https://doi.org/10.4155/bio.12.204>
- 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>
- EFSA. (2018). Contaminants dans l’alimentation humaine et animale. Retrieved December 8, 2017,

from <http://www.efsa.europa.eu/fr/topics/topic/contaminants-food-and-feed>

- EFSA (European Food Safety Authority). (2008). Statement of EFSA on the risks for public health due to the presence of dioxins in pork from Ireland. *EFSA Journal*, 6(12), 911. <https://doi.org/10.2903/j.efsa.2008.911>
- EFSA (European Food Safety Authority). (2016). *Chemicals in food 2016 : Overview of selected data collection*. <https://doi.org/10.2805/359743>
- Esslinger, S., Riedl, J., & Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Research International*, 60, 189–204. <https://doi.org/10.1016/j.foodres.2013.10.015>
- European Commission - DG Health and Food Safety. (2015). *Guidance document on analytical quality control and method validation procedures for pesticides residues analysis in food and feed* (Vol. 11945). Retrieved from https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_mrl_guidelines_wrkdoc_11945.pdf
- 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>
- Fraser, K., Lane, G. A., Otter, D. E., Hemar, Y., Quek, S. Y., Harrison, S. J., & Rasmussen, S. (2013). Analysis of metabolic markers of tea origin by UHPLC and high resolution mass spectrometry. *Food Research International*, 53(2), 827–835. <https://doi.org/10.1016/j.foodres.2012.10.015>
- Fremy, J.-M. (2009). Évaluation des risques liés à la présence de mycotoxines dans les chaînes alimentaires humaine et animale. *Afssa*, 339.
- Frenich, A. G., Romero-González, R., & del Mar Aguilera-Luiz, M. (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>
- Fréry, N., Fillol, C., Garnier, R., Falq, G., Bidondo, M.-L., Guldner, L., & Zeghnoun, A. (2017). Exposition de la population française aux substances chimiques de l'environnement – Étude ENNS 2006–2007. *Toxicologie Analytique et Clinique*, 29(4), 441–482. <https://doi.org/10.1016/j.toxac.2017.06.002>
- Gao, G., Chen, H., Zhu, L., Chai, Y., Ma, G., Wang, C., ... Lu, C. (2017). Simultaneous determination of bisphenol A and tetrabromobisphenol A in tea using a modified QuEChERS sample preparation method coupled with liquid chromatography-tandem mass spectrometry. *Analytical Methods*, 9(48), 6769–6776. <https://doi.org/10.1039/C7AY02145C>
- Giacomini, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., ... Caron, C. (2015). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics*, 31(9), 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>
- Gika, H. G., Theodoridis, G. A., Plumb, R. S., & Wilson, I. D. (2014). Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *Journal of Pharmaceutical and Biomedical Analysis*, 87, 12–25. <https://doi.org/10.1016/j.jpba.2013.06.032>
- Godzien, J., Ciborowski, M., Angulo, S., & Barbas, C. (2013). From numbers to a biological sense: How the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LC-MS. *Electrophoresis*, 34(19), 2812–2826. <https://doi.org/10.1002/elps.201300053>
- Gorrochategui, E., Jaumot, J., Lacorte, S., & Tauler, R. (2016). Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow. *TrAC - Trends in Analytical Chemistry*, 82, 425–442. <https://doi.org/10.1016/j.trac.2016.07.004>
- Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of

- convenience or a shotgun wedding. *Analytica Chimica Acta*, 879, 10–23. <https://doi.org/10.1016/j.aca.2015.02.012>
- Grover, I. S., Singh, S., & Pal, B. (2013). Priority PAHs in orthodox black tea during manufacturing process. *Environmental Monitoring and Assessment*, 185(8), 6291–6294. <https://doi.org/10.1007/s10661-012-3025-y>
- Habchi, B., Alves, S., Jouan-Rimbaud Bouveresse, D., Moslah, B., Paris, A., Lécluse, Y., ... Rathahao-Paris, E. (2017). An innovative chemometric method for processing direct introduction high resolution mass spectrometry metabolomic data: independent component–discriminant analysis (IC–DA). *Metabolomics*, 13(4), 0. <https://doi.org/10.1007/s11306-017-1179-x>
- Habchi, B., Kassouf, A., Padellec, Y., Rathahao-Paris, E., Alves, S., Rutledge, D. N., ... Ducruet, V. (2018). An untargeted evaluation of food contact materials by flow injection analysis-mass spectrometry (FIA-MS) combined with independent components analysis (ICA). *Analytica Chimica Acta*, 1022, 81–88. <https://doi.org/10.1016/j.aca.2018.03.042>
- Harbowy, M. E., Balentine, D. A., Davies, A. P., & Cai, Y. (1997). Tea Chemistry. *Critical Reviews in Plant Sciences*, 16(5), 415–480. <https://doi.org/10.1080/07352689709701956>
- Hardy, N. W., & Taylor, C. F. (2007). A roadmap for the establishment of standard data exchange structures for metabolomics. *Metabolomics*, 3(3), 243–248. <https://doi.org/10.1007/s11306-007-0071-5>
- Hautbergue, T., Puel, O., Tadrist, S., Meneghetti, L., Péan, M., Delaforge, M., ... Jamin, E. L. (2017). Evidencing 98 secondary metabolites of *Penicillium verrucosum* using substrate isotopic labeling and high-resolution mass spectrometry. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 1071, 29–43. <https://doi.org/10.1016/j.jchromb.2017.03.011>
- Herrero, M., Simó, C., García-Cañas, V., Ibáñez, E., & Cifuentes, A. (2012). Foodomics: MS-based strategies in modern food science and nutrition. *Mass Spectrometry Reviews*, 31(1), 49–69. <https://doi.org/10.1002/mas.20335>
- 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>
- Hossen, V., Velge, P., Turquet, J., Chinain, M., Laurent, D., & Krys, S. (2013). *La ciguatera : un état des lieux en France et dans l'Union européenne*. Retrieved from http://www.ciguatera-online.com/images/Telechargements/BEP-mg-BE56_cle4cd259.pdf
- Hrydziusko, O., & Viant, M. R. (2012). Missing values in mass spectrometry based metabolomics: An undervalued step in the data processing pipeline. *Metabolomics*, 8, 161–174. <https://doi.org/10.1007/s11306-011-0366-4>
- Hu, C., & Xu, G. (2013). Mass-spectrometry-based metabolomics analysis for foodomics. *TrAC - Trends in Analytical Chemistry*, 52, 36–46. <https://doi.org/10.1016/j.trac.2013.09.005>
- Hubert, A., Malvy, D., Preziosi, P., Galan, P., & Hercberg, S. (2000). La consommation de thé en France : phénomène culturel et social et contribution à l'équilibre nutritionnel et à l'état de santé. *Cahiers de Nutrition et de Diététique*, 35. <https://doi.org/CND-012-2000-35-S1-0007-9960-101019-ART10>
- Inoue, K., Tanada, C., Sakamoto, T., Tsutsui, H., Akiba, T., Min, J. Z., ... Toyo'Oka, T. (2015). Metabolomics approach of infant formula for the evaluation of contamination and degradation using hydrophilic interaction liquid chromatography coupled with mass spectrometry. *Food Chemistry*, 181, 318–324. <https://doi.org/10.1016/j.foodchem.2015.02.117>
- Institut Louis Maladré. (2014). Ciguatera Online. Retrieved February 8, 2018, from <http://www.ciguatera-online.com/index.php/fr/>
- IUPAC. (2010). History of Pesticide Use. Retrieved April 23, 2018, from https://agrochemicals.iupac.org/index.php?option=com_sobi2&sobi2Task=sobi2Details&catid=3

&sobi2Id=31

- Jiao, W., Xiao, Y., Qian, X., Tong, M., Hu, Y., Hou, R., & Hua, R. (2016). Optimized combination of dilution and refined QuEChERS to overcome matrix effects of six types of tea for determination eight neonicotinoid insecticides by ultra performance liquid chromatography-electrospray tandem mass spectrometry. *Food Chemistry*, 210, 26–34. <https://doi.org/10.1016/j.foodchem.2016.04.097>
- Jouan-Rimbaud Bouveresse, D., & Rutledge, D. N. N. (2016). *Independent Components Analysis: Theory and Applications. Data Handling in Science and Technology* (Vol. 30). <https://doi.org/10.1016/B978-0-444-63638-6.00007-3>
- Kebarle, P., & Peschke, M. (2000). On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions. *Analytica Chimica Acta*, 406(1), 11–35. [https://doi.org/10.1016/S0003-2670\(99\)00598-X](https://doi.org/10.1016/S0003-2670(99)00598-X)
- Kellmann, M., Muenster, H., Zomer, P., & Mol, H. (2009). Full Scan MS in Comprehensive Qualitative and Quantitative Residue Analysis in Food and Feed Matrices: How Much Resolving Power is Required? *Journal of the American Society for Mass Spectrometry*, 20(8), 1464–1476. <https://doi.org/10.1016/j.jasms.2009.05.010>
- 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>
- Knolhoff, A. M., Zweigenbaum, J. A., & Croley, T. R. (2016). Nontargeted Screening of Food Matrices: Development of a Chemometric Software Strategy to Identify Unknowns in Liquid Chromatography-Mass Spectrometry Data. *Analytical Chemistry*, 88(7), acs.analchem.5b04208. <https://doi.org/10.1021/acs.analchem.5b04208>
- Knutsen, H. K., Alexander, J., Barregård, L., Bignami, M., Brüschweiler, B., Ceccatelli, S., ... Binaglia, M. (2017). Risks for human health related to the presence of pyrrolizidine alkaloids in honey, tea, herbal infusions and food supplements. *EFSA Journal*, 15(7), 34. <https://doi.org/10.2903/j.efsa.2017.4908>
- Kortenkamp, A., Evans, R., Faust, M., Kalberlah, F., Scholze, M., & Schuhmacher-Wolz, U. (2012). Investigation of the state of the science on combined actions of chemicals in food through dissimilar modes of action and proposal for science-based approach for performing related cumulative risk assessment 1 Prepared. *EFSA Journal*, (January), 233. <https://doi.org/10.2903/SP.EFSA.2012.EN-232>
- 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*, 1–10. <https://doi.org/10.1007/s00216-018-1028-4>
- Lacina, O., Zachariasova, M., Urbanova, J., Vaclavikova, M., Cajka, T., & Hajslova, J. (2012). Critical assessment of extraction methods for the simultaneous determination of pesticide residues and mycotoxins in fruits, cereals, spices and oil seeds employing ultra-high performance liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A*, 1262, 8–18. <https://doi.org/10.1016/j.chroma.2012.08.097>
- Latrille, E. (2016). Grain 3 : ACP - 1. *CheMoocs*, 1–20.
- Lavine, B. K. (2006). Pattern recognition. *Critical Reviews in Analytical Chemistry*, 36(3–4), 153–161. <https://doi.org/10.1080/10408340600969411>
- Lehotay, S. J., Mastovska, K., Amirav, A., Fialkov, A. B., Martos, P. A., Kok, A. de, & Fernández-Alba, A. R. (2008). Identification and confirmation of chemical residues in food by chromatography-mass spectrometry and other techniques. *TrAC - Trends in Analytical Chemistry*, 27(11), 1070–1090. <https://doi.org/10.1016/j.trac.2008.10.004>
- Lehotay, S. J., Sapozhnikova, Y., & Mol, H. G. J. (2015). Current issues involving screening and identification of chemical contaminants in foods by mass spectrometry. *TrAC - Trends in*

- Analytical Chemistry*, 69, 62–75. <https://doi.org/10.1016/j.trac.2015.02.012>
- Lewis, K. A., Tzilivakis, J., Warner, D. J., & Green, A. (2016). An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment*, 22(4), 1050–1064. <https://doi.org/10.1080/10807039.2015.1133242>
- Li, X., Zhang, Z., Li, P., Zhang, Q., Zhang, W., & Ding, X. (2013). Determination for major chemical contaminants in tea (*Camellia sinensis*) matrices: A review. *Food Research International*, 53(2), 649–658. <https://doi.org/10.1016/j.foodres.2012.12.048>
- Libiseller, G., Dvorzak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., ... Magnes, C. (2015). IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics*, 16(1), 118. <https://doi.org/10.1186/s12859-015-0562-8>
- Liu, J., Zhao, G., Yuan, Y., Chen, F., & Hu, X. (2008). Quantitative analysis of acrylamide in tea by liquid chromatography coupled with electrospray ionization tandem mass spectrometry. *Food Chemistry*, 108(2), 760–767. <https://doi.org/10.1016/j.foodchem.2007.11.015>
- Lommen, A. (2009). Metalign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry*, 81(8), 3079–3086. <https://doi.org/10.1021/ac900036d>
- 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>
- McGorin, R. J. (2009). One Hundred Years of Progress in Food Analysis. *Journal of Agricultural and Food Chemistry*, 57(18), 8076–8088. <https://doi.org/10.1021/jf900189s>
- Mendiola, J. A., Herrero, M., Cifuentes, A., & Ibañez, E. (2007). Use of compressed fluids for sample preparation: Food applications. *Journal of Chromatography A*, 1152(1–2), 234–246. <https://doi.org/10.1016/j.chroma.2007.02.046>
- Milman, B. L., & Zhurkovich, I. K. (2016). Mass spectral libraries: A statistical review of the visible use. *TrAC - Trends in Analytical Chemistry*, 80, 636–640. <https://doi.org/10.1016/j.trac.2016.04.024>
- Ministère de l'Agriculture et de l'Alimentation. (2017). Fipronil dans les œufs : Questions - Réponses. Retrieved April 26, 2018, from <http://agriculture.gouv.fr/fipronil-dans-les-oeufs-questions-reponses>
- 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 matrices. *Analytical Chemistry*, 80(24), 9450–9459. <https://doi.org/10.1021/ac801557f>
- Mol, H. G. J., Plaza-Bolaños, P., Zomer, P., Rijk, T. C. De, 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 matrix. *Analytical Chemistry*, 80(24), 9450–9459. <https://doi.org/10.1021/ac801557f>
- Monbaliu, S., Wu, A., Zhang, D., Van Peteghem, C., & De Saeger, S. (2010). Multimycotoxin UPLC-MS/MS for tea, herbal infusions and the derived drinkable products. *Journal of Agricultural and Food Chemistry*, 58(24), 12664–12671. <https://doi.org/10.1021/jf1033043>
- Mulder, P. P. J., Sánchez, P. L., These, A., Preiss-Weigert, A., & Castellari, M. (2015). Occurrence of Pyrrolizidine Alkaloids in food. *EFSA Supporting Publications*, 12(8). <https://doi.org/10.2903/sp.efsa.2015.EN-859>
- Nowak, M., Wittke, C., Lederer, I., Klier, B., Kleinwächter, M., & Selmar, D. (2016). Interspecific transfer of pyrrolizidine alkaloids: An unconsidered source of contaminations of phytopharmaceuticals and plant derived commodities. *Food Chemistry*, 213, 163–168. <https://doi.org/10.1016/j.foodchem.2016.06.069>

- Organisation Mondiale de la Santé. (2008). Alerte et action au niveau mondial - Lait maternisé en poudre contaminé à la mélamine en Chine. Retrieved April 26, 2018, from http://www.who.int/csr/don/2008_09_19/fr/
- Ortmayr, K., Charwat, V., Kasper, C., Hann, S., & Koellensperger, G. (2017). Uncertainty budgeting in fold change determination and implications for non-targeted metabolomics studies in model systems. *The Analyst*, *142*(1), 80–90. <https://doi.org/10.1039/C6AN01342B>
- 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>
- Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, *11*, 395. <https://doi.org/10.1186/1471-2105-11-395>
- Pongsuwan, W., Bamba, T., Harada, K., Yonetani, T., Kobayashi, A., & Fukusaki, E. (2008). High-throughput technique for comprehensive analysis of Japanese green tea quality assessment using ultra-performance liquid chromatography with time-of-flight mass spectrometry (UPLC/TOF MS). *Journal of Agricultural and Food Chemistry*, *56*(22), 10705–10708. <https://doi.org/10.1021/jf8018003>
- Riedl, J., Esslinger, S., & Fahl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Analytica Chimica Acta*, *885*, 17–32. <https://doi.org/10.1016/j.aca.2015.06.003>
- Rondeau, D. (2017). *Spectrométrie de masse organique Analyseurs et méthodes en tandem ou MS n*.
- Roullier, C., Guitton, Y., Valery, M., Amand, S., Prado, S., Robiou Du Pont, T., ... Pouchus, Y. F. (2016). Automated Detection of Natural Halogenated Compounds from LC-MS Profiles-Application to the Isolation of Bioactive Chlorinated Compounds from Marine-Derived Fungi. *Analytical Chemistry*, *88*(18), 9143–9150. <https://doi.org/10.1021/acs.analchem.6b02128>
- Saito-Shida, S., Hamasaka, T., Nemoto, S., & Akiyama, H. (2018). Multiresidue determination of pesticides in tea by liquid chromatography-high-resolution mass spectrometry: Comparison between Orbitrap and time-of-flight mass analyzers. *Food Chemistry*, *256*(October 2017), 140–148. <https://doi.org/10.1016/j.foodchem.2018.02.123>
- Salek, R. M., Haug, K., & Steinbeck, C. (2013). Dissemination of metabolomics results: Role of MetaboLights and COSMOS. *GigaScience*, *2*(1), 2–4. <https://doi.org/10.1186/2047-217X-2-8>
- Samsidar, A., Siddiquee, S., & Shaarani, S. M. (2018). A review of extraction, analytical and advanced methods for determination of pesticides in environment and foodstuffs. *Trends in Food Science and Technology*, *71*(July 2017), 188–201. <https://doi.org/10.1016/j.tifs.2017.11.011>
- Santé Canada. (2009). Les composés perfluorés dans les aliments. Retrieved April 24, 2018, from <https://www.canada.ca/fr/sante-canada/services/aliments-nutrition/salubrite-aliments/contaminants-chimiques/contaminants-environnementaux/composes-perfluores-aliments.html>
- Santé Canada. (2011). Les glyco-alcaloïdes dans les aliments. Retrieved April 25, 2018, from <https://www.canada.ca/fr/sante-canada/services/aliments-nutrition/rapports-publications/salubrite-aliments/glyco-alcaloïdes-aliments.html>
- Scholz, M., Gatzek, S., Sterling, A., Fiehn, O., & Selbig, J. (2004). Metabolite fingerprinting: Detecting biological features by independent component analysis. *Bioinformatics*, *20*(15), 2447–2454. <https://doi.org/10.1093/bioinformatics/bth270>
- Scigelova, M., & Makarov, A. (2013). Fundamentals and Advances of Orbitrap Mass Spectrometry. In *Encyclopedia of Analytical Chemistry* (pp. 1–36). <https://doi.org/10.1002/9780470027318.a9309>
- Seppänen-Laakso, T., Laakso, I., & Hiltunen, R. (2002). Analysis of fatty acids by gas chromatography,

- and its relevance to research on health and nutrition. *Analytica Chimica Acta*, 465(1–2), 39–62. [https://doi.org/10.1016/S0003-2670\(02\)00397-5](https://doi.org/10.1016/S0003-2670(02)00397-5)
- Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), 779–787. <https://doi.org/10.1021/ac051437y>
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., ... Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3(3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- Tautenhahn, R., Bottcher, C., & Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, 9, 16. <https://doi.org/10.1186/1471-2105-9-504>
- Tautenhahn, R., Patti, G. J., Rinehart, D., & Siuzdak, G. (2012). XCMS online: A web-based platform to process untargeted metabolomic data. *Analytical Chemistry*, 84(11), 5035–5039. <https://doi.org/10.1021/ac300698c>
- 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>
- Tranchant, J. (1996). *Chromatographie en phase gazeuse*. [https://doi.org/10.1016/S0021-9673\(01\)98018-5](https://doi.org/10.1016/S0021-9673(01)98018-5)
- Tsopelas, F., Konstantopoulos, D., & Kakoulidou, A. T. (2018). Voltammetric fingerprinting of oils and its combination with chemometrics for the detection of extra virgin olive oil adulteration. *Analytica Chimica Acta*, 1015, 8–19. <https://doi.org/10.1016/j.aca.2018.02.042>
- UK Tea & Infusions Association. (2018). Types of tea. Retrieved April 27, 2018, from <https://www.tea.co.uk/types-of-tea>
- Wang, J., Cheung, W., & Leung, D. (2014). Determination of pesticide residue transfer rates (Percent) from dried tea leaves to brewed tea. *Journal of Agricultural and Food Chemistry*, 62(4), 966–983. <https://doi.org/10.1021/jf404123h>
- Wehrens, R., Hageman, J. A., van Eeuwijk, F., Kooke, R., Flood, P. J., Wijnker, E., ... de Vos, R. C. H. (2016). Improved batch correction in untargeted MS-based metabolomics. *Metabolomics*, 12(5), 88. <https://doi.org/10.1007/s11306-016-1015-8>
- Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., & Ni, Y. (2018). Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Scientific Reports*, 8(1), 663. <https://doi.org/10.1038/s41598-017-19120-0>
- WHO/UNEP. (2008a). All POPs listed in the Stockholm Convention. Retrieved April 24, 2018, from <http://www.pops.int/TheConvention/ThePOPs/AllPOPs/tabid/2509/Default.aspx>
- WHO/UNEP. (2008b). What are POPs? Retrieved April 24, 2018, from <http://www.pops.int/TheConvention/ThePOPs/tabid/673/Default.aspx>
- Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., ... Rappaport, S. M. (2015). T3DB: The toxic exposome database. *Nucleic Acids Research*, 43(D1), D928–D934. <https://doi.org/10.1093/nar/gku1004>
- Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., & Liang, Y. (2016). Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta*, 914, 17–34. <https://doi.org/10.1016/j.aca.2016.02.001>

CHAPITRE 2
DÉVELOPPEMENT DE L'APPROCHE CIBLÉE MULTI-
CLASSES

1. INTRODUCTION ET RÉSUMÉ DE L'ARTICLE

Pour effectuer des analyses non-ciblées fiables, il est nécessaire de disposer d'une méthode analytique robuste. Dans le cas des contaminants de l'aliment, pouvoir détecter de façon non-ciblé tout nouveau contaminant implique de disposer d'une méthode validée sur une large gamme de molécules dans un produit donné, et ce à des niveaux de contamination réalistes. Comme on l'a vu au chapitre 1, partie 2.2.3, il n'existe pas de méthode multi-classes permettant l'analyse simultanée de composés aussi divers que des contaminants de l'emballage, des pesticides et des contaminants issus de procédés dans le thé. Cependant, on a vu dans cette même partie que ce type de méthode a été développé avec succès sur d'autres matrices (aussi bien solide que liquide) que le thé.

Dans cet article, nous proposons une méthode (traitement de l'échantillon + méthode UHPLC-HRMS) générique pour l'analyse de 32 contaminants (appelés « traceurs ») dans le thé sous forme de boisson et de feuilles, choisis pour leur diversité de structures chimiques et d'origines. Ces 32 contaminants, ainsi que leurs propriétés physico-chimiques, sont présentés dans le Table 2.1 donné ci-dessous.

Des tests de recouvrement de ces 32 traceurs ont été menés dans des échantillons de thé (feuilles et boisson) à différents niveaux de dopage (10, 50 et 100 µg/kg – dans l'article qui suit, certains niveaux ne sont parfois pas présentés pour des raisons de lisibilité). Les performances des différentes méthodes ont été évaluées au regard du document guide de la Commission Européenne pour le développement et la validation des méthodes d'analyse des résidus de pesticides dans les aliments (SANTE/11945/2015). Les principaux critères de qualité sont le pourcentage de recouvrement (quantifié si $70\% < R < 120\%$) et la répétabilité, mesurée par l'écart type relatif (acceptable si $RSD < 20\%$). Pour chaque contaminant et chaque niveau de dopage, on peut ainsi caractériser les performances de la méthode et déterminer une limite de détection.

Deux techniques de traitement de l'échantillon ont été comparées pour les échantillons sous forme liquide (boisson) : la LLE et la « dilute & shoot » décrites plus haut (chapitre 1 partie 2.2.2). Pour les échantillons solides (feuilles), une approche par extraction directe par solvant a été mise au point. La méthode UHPLC-HRMS utilisée est commune aux deux états de matrice, et se distingue par l'utilisation d'une colonne de silice avec des greffons C18-PFP (pentafluorophényle), permettant une rétention améliorée des composés les plus polaires. Une méthode différente est utilisée pour chaque mode d'ionisation.

La publication présente tout d'abord le choix des conditions analytiques (conditions chromatographiques et MS), ainsi que les performances instrumentales (partie 2.3.1). Le choix de la colonne y est notamment argumenté, ainsi que la nécessité d'utiliser une méthode par mode d'ionisation. La sensibilité instrumentale de la méthode pour les 32 traceurs est ensuite discutée au regard des réglementations en vigueur et des niveaux attendus dans les aliments. Les résultats des expériences de

dopage/recouvrement sur le thé sous forme liquide (infusion) et solide (feuilles) sont ensuite présentés et discutés.

2. ARTICLE

Multi-class analysis for simultaneous determination of pesticides, mycotoxins, process-induced toxicants and packaging contaminants in tea

Mathieu Cladière*, Grégoire Delaporte, Even Le Roux, Valérie Camel.

UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300 Massy, France

Corresponding author : mathieu.cladiere@agroparistech.fr*

Published in *Food Chemistry*, 2018, 242, pp.113-121

Available at:

<https://doi.org/10.1016/j.foodchem.2017.08.108>

<https://hal.archives-ouvertes.fr/hal-01637734>

Abstract

This study attempts at uniting the analysis of four different classes of contaminants for both liquid and solid tea samples. A total of 32 compounds, classified as pesticides, mycotoxins, process-induced toxicants or packaging contaminants, were carefully chosen for their diversity of structures and physicochemical properties. The proposed method combines a sample treatment strategy coming from metabolomics with liquid chromatography analysis using a silica bonded C18-pentafluorophenyl column coupled to high resolution mass spectrometry. For tea brew, dilute and shoot method provides good quantification (70-120% recoveries and <20% RSD) for more than 80% of compounds. For tea leaves, strong matrix effects are observed, thus, matrix-matched calibration is required to reach good performances, i.e. 63% of compounds quantified and 81% detected at 10 µg/kg. Finally, method performances were evaluated against existing regulations, and it appears that 69% of contaminants are quantified and 91% detected at levels lower than their respective European regulation limits.

Keywords

Acrylamide; UHPLC-TOF-MS; C18-PFP; food safety; multi-residue analysis

2.1. INTRODUCTION

Food safety remains an everyday challenge toward the globalization of production and sometimes the lack of traceability of products. There are numerous sources of chemical contaminants in food products, ranging from the raw material itself (pesticides, mycotoxins and veterinary drugs in case of food of animal origin), its transportation, its processing (process-induced toxicants), and finally its packaging (migrants). Most analytical strategies, developed by laboratories and food safety authorities around the world, rely on the carrying out of several targeted analyzes in order to cover this broad range of contamination sources and quantify as many contaminants as possible. Such a multi-analysis approach is both very costly and time consuming; on top of that, the environmental footprint of the analysis is increased by the use of larger amounts of solvent and reagent compared with the single-analysis approach. As a result, recent reviews have underlined the need for multi-class methods capable of analyzing a large number of compounds in a single analysis (Antignac et al., 2011; Castro-Puyana & Herrero, 2013).

Indeed, the current technology based on liquid chromatography coupled to mass spectrometry (LC-MS) offers the feasibility of developing new approaches thanks to the increasing sensibility and possibility of full scan analysis using high resolution apparatus. As a consequence, few multi-class methods have emerged in recent years (Danezis, Anagnostopoulos, Liapis, & Koupparis, 2016; Jin et al., 2017; Petrarca, Fernandes, Godoy, & Cunha, 2016). Yet, although these methods are able to quantify hundreds of contaminants simultaneously, they generally focus on only one or two classes of contaminants with close physicochemical properties. To the best of our knowledge, only one study deals with the analysis of various classes of contaminants including migrants from packaging and process-induced toxicants, but it focuses on LC-MS optimization without assessing efficiency of sample treatment (Pérez-Ortega et al., 2016). The authors pointed out the difficulties of this approach since the structural diversity of targeted compounds leads to heterogeneous behaviors during the LC-MS analysis, both in terms of retention and matrix effects. Consequently, developing a multi-class method capable of analyzing contaminants having a wide-range of chemical structures remains a challenge, particularly in real food matrices due to their complexity.

To reach this objective, tea has been chosen as development matrix for four main reasons. Firstly, this is the most consumed manufactured beverage in the world with 4.8 million tons (of tea leaves) produced in 2013, increasing by 5% per year since 2008 (Chang, 2015); medium term outlooks suggest a slightly higher increasing of tea consumption (black or green) until 2023. Secondly, tea is produced in remote countries where contamination risks may be difficult to manage; as a consequence, monitoring and regulatory control analyzes regularly show the presence of chemical contaminants (especially pesticide residues) exceeding their European maximum limits (EFSA (European Food Safety Authority), 2016). Thirdly, tea can be analyzed in both solid and liquid states, through tea leaves or brew, which is an interesting feature when developing analytical methods for food matrices. Finally, tea is a complex food

product regarding its organic composition with a wide variety of phytochemicals (such as antioxidants, aroma compounds, xanthins and alkaloids) which commonly bring strong matrix effects during analysis, therefore making it an interesting real complex food matrix to consider.

This study reports the development and validation of a method for simultaneous extraction and quantification of multi-class contaminants (covering a wide range of chemical structures and properties) for both liquid and solid tea matrices. Until now, this is the first multi-class method covering four classes of food contaminants including pesticides, mycotoxins, process-induced toxicants and migrants from packaging. A total of 32 target compounds were carefully chosen in order to cover the diversity of classes and physicochemical properties encountered. As an illustration, for pesticides, organochlorinated, organophosphates, organosulfites, triazines, auxinic herbicides, neonicotinoids, benzoylureas, pyrazoles, dinitrophenols and carbamates have been considered; as far as we know, all these pesticide families are considered together for the very first time (Eitzer, Hammack, & Filigenzi, 2014; Hou et al., 2014). In addition, 4 mycotoxins, 2 process-induced toxicants (including acrylamide which has never been considered in any multi-residue method before) and 5 migrants from packaging have been examined. The challenge of our method lies in the different behaviors of these 32 compounds regarding extraction and chromatographic separation, in order to propose the best compromise and provide analytical performances in compliance with the European regulation on food contaminants (European Commission, 2005, 2006).

2.2. MATERIALS AND METHODS

2.2.1. TARGETED CONTAMINANTS

For developing our multi-class method, several target food contaminants were selected from different classes, namely process-induced toxicants (n=2), migrants from packaging (n=5), mycotoxins (n=4), herbicides (n=11), insecticides (n=6) and acaricides (n=4). The selection of these 32 compounds was based on two main criteria; their physicochemical properties to offer a broad range and a large diversity (i.e. hydrophobicity, aromaticity, functional groups) so as to be representative of other contaminants not considered here, and their relevance for tea. As an example, propargite (acaricide) has been quoted 11 times by the European Rapid Alert System for Food and Feed (RASFF) in 2016, leading to border rejections of black and green teas, while acetamiprid and imidacloprid (insecticides) were cited more than 40 times between 2012 and 2015.

The 32 compounds as well as their analytical information are listed in Table 2.1.

Analytical standards (100 µg/mL) for pesticides, mycotoxins, process-induced toxicants and labelled compounds acrylamide-d3, dimethoate-d6 and malathion-d6 (purity > 99%) were supplied by CIL Cluzeau (France). Ochratoxin-d5, bisphenol A, F and S, BADGE, BFDGE and bisphenol A-d14 (purity > 99%) were provided by Sigma Aldrich (France).

2.2.2. MATERIALS AND REAGENTS

Acetonitrile (ACN) (HPLC plus gradient, LC-MS), water, methanol (MeOH) and formic acid (FA) (all LC-MS grade) were purchased from Carlo Erba. Ultrapure water (Milli-Q[®]) was produced by an Integral 3 water purification system from Millipore[®]. The compound used for MS calibration was Leucine Enkephalin (LC-MS grade), purchased from Waters[®]. Magnesium sulphate (MgSO₄) and sodium citrate (NaCit) salts (analytical grade) were provided by VWR France.

Analyzes of trace contaminants have been performed on a Waters[®] Acquity H-Class UPLC[®] system, composed of a quaternary solvent manager pump (QSM), a refrigerated sample manager Flow-Through-Needle (SM-FTN) and a column oven, coupled to a Waters[®] high resolution mass spectrometer with a Time of Flight analyzer Xevo[®] G2-S ToF (UHPLC/MS-ToF). An electrospray ionization source was used in both positive (ESI⁺) and negative (ESI⁻) modes.

2.2.1. ANALYSIS CONDITIONS

2.2.1.1. CHROMATOGRAPHIC CONDITIONS

Chromatographic separation was done on a column made of silica based particles bonded with C18-pentafluorophenyl functions (C18-PFP) (dimensions were 150 x 2.1 mm; 2 μ m particles diameter, from ACE, provided by AIT, France). In addition to conventional hydrophobic interactions (provided by C18 chains), the PFP groups enable hydrogen bonds, π - π and dipole-dipole interactions, affording a higher capacity for retaining the highly polar compounds (such as acrylamide or acidic herbicides) than a regular C18-silica column.

2.2.1.2. MS ANALYSIS

Since different mobile phases were used for positive and negative ionization, analyzes were performed separately for both modes. All analyzes were done using the resolution mode (30,000 FWHM at 200 m/z) for a scan time of 0.5 s, with mass range between 60 and 800 m/z and data acquired in centroid. Internal calibration of ToF analyzer was performed with a continuous flow at 5 μ L/min of Leucine Enkephalin for one scan every 30 s during 0.1 s.

For ESI⁺ the mobile phase was composed of water (A), ACN (B), both acidified with 0.1% FA, and MeOH (C), flowing at 0.4 mL/min. The gradient started at 100% A and reached 100% B in 10 min, this composition being kept for 6 min before switching to 100% C to rinse the system in 1 min, being hold for 5 min, returning back to 100% A in 1 min and finally equilibrating for 3 min, with a total run duration of 26 min. For ESI⁻, the mobile phase was composed of water buffered with 10 mM of ammonium formate (A) and MeOH (B) flowing at 0.3 mL/min. The gradient started at 100% A and reached 100% B in 13 min, holding this condition for 7 min before turning back to 100% A in 1 min and finally equilibrating for 3 min, with a total run duration of 24 min. For both chromatographic methods the column was heated at 30°C.

TABLE 2.1 TARGETED COMPOUNDS ALONG WITH THEIR INSTRUMENTAL OPTIMIZED PARAMETERS AND METHOD PERFORMANCES.

N°	Name	Abbrev.	Class	Chemical family	log K _{ow}	RT (min)	ESI (+/-)	Ion	m/z	LOD (ng/mL)	LOQ (ng/mL)
1	Acrylamide	AA	Process-induced	Amide	-0.67	1.28	+	[M+H] ⁺	72.044	5	51.5
2	Hydroxymethylfurfural	HMF	Process-induced	Furan	-0.09	3.11	+	[M+H] ⁺	127.039	2.5	7
3	Deoxynivalenol	DON	Mycotoxin	Trichothecene	0.29	3.56 6.24	+ -	[M+H] ⁺ [M-H] ⁻	297.133 295.119	2.5	18
4	Imidacloprid	IMID	Pesticide (insecticide)	Neonicotinoid	0.57	5.05	+	[M+H] ⁺	256.059	0.5	19.5
5	Dimethoate	DIME	Pesticide (acaricide)	Organophosphate	0.7	5.07	+	[M+H] ⁺	230.006	0.1	0.5
6	Acetamiprid	ACET	Pesticide (insecticide)	Neonicotinoid	0.8	5.24	+	[M+H] ⁺	223.074	0.5	0.5
7	Fumonisin B1	FB1	Mycotoxin	Fumonisin	-0.5	5.57	+	[M+H] ⁺	722.395	0.5	6.4
8	Aldicarb	ALD	Pesticide (acaricide)	Carbamate	1.15	5.63	+	[M+Na] ⁺	213.0674	0.1	0.5
9	Ametryn	AME	Pesticide (herbicide)	Triazine	2.63	5.78	+	[M+H] ⁺	228.127	0.1	0.4
10	Dichlorvos	DIVO	Pesticide (acaricide)	Organochlorinated	1.9	6.11	+	[M+H] ⁺	220.953	0.5	3.9
11	Fumonisin B2	FB2	Mycotoxin	Fumonisin	1.2	6.17	+	[M+H] ⁺	706.4	0.5	1.8
12	Atrazine	ATZ	Pesticide (herbicide)	Triazine	2.7	6.61	+	[M+H] ⁺	216.101	0.5	1.6
13	Diuron	DION	Pesticide (herbicide)	Phenylurea	2.87	7.09	+	[M+H] ⁺	233.024	0.5	9.5
14	Ochratoxin A	OTA	Mycotoxin	Ochratoxin	4.74	7.55 10.85	+ -	[M+H] ⁺ [M-H] ⁻	404.089 402.075	0.5 0.5	19 4.5
15	BFDGE ¹	BFDGE	Migrant from packaging	Diglycidyl ether		8.17	+	[M+Na] ⁺	335.125	0.5	22.8
16	Malathion	MLT	Pesticide (insecticide)	Organophosphate	2.75	8.22	+	[M+H] ⁺	331.043	0.5	1.6
17	Diflubenzuron	DIFLU	Pesticide (insecticide)	Benzoylurea	3.89	8.46	+	[M+H] ⁺	311.039	0.5	0.8
18	BADGE ²	BADGE	Migrant from packaging	Diglycidyl ether		8.52	+	[M+Na] ⁺	363.157	4	21.2
19	Pirimiphos-methyl	PIRI	Pesticide (insecticide)	Organophosphate	3.9	8.72	+	[M+H] ⁺	306.103	0.1	0.6
20	2,4-Dichlorophenoxyacetic acid	2,4-D	Pesticide (herbicide)	Auxinic herbicide	-0.82	9.41	-	[M-H] ⁻	218.9621	6.9	19.4
21	Bisphenol S	BPS	Migrant from packaging	Bisphenol	2.91	9.42	-	[M-H] ⁻	249.022	0.1	0.1
22	Tolfenpyrad	TOLF	Pesticide (insecticide)	Pyrazole	5.61	9.43	+	[M+H] ⁺	384.147	0.1	0.9
23	2-Methyl-4-chlorophenoxyacetic acid	MCPA	Pesticide (herbicide)	Auxinic herbicide	-0.81	9.58	-	[M-H] ⁻	199.017	1	3.6
24	Propargite	PROP	Pesticide (acaricide)	Organosulfite	5.7	9.88	+	Frag ³	231.174	0.6	3.2
25	Dichloprop	DIPRO	Pesticide (herbicide)	Auxinic herbicide	2.29	10.17	-	[M-H] ⁻	232.977	1.7	18.8
26	Mecoprop	MCP	Pesticide (herbicide)	Auxinic herbicide	-0.19	10.21	-	[M-H] ⁻	213.032	0.7	1
27	2,4,5-Trichlorophenoxyacetic acid	2,4,5-T	Pesticide (herbicide)	Auxinic herbicide	2.88	10.61	-	[M-H] ⁻	252.9232	1	4.6
28	Bisphenol F	BPF	Migrant from packaging	Bisphenol	1.65	10.73	-	[M-H] ⁻	199.076	40	50
29	Fenoprop	2,4,5 TP	Pesticide (herbicide)	Auxinic herbicide	2.84	11.12	-	Frag ³	194.917	0.8	5.3
30	Dinoseb	DINO	Pesticide (herbicide)	Dinitrophenol	2.29	11.43	-	[M-H] ⁻	239.067	0.2	0.4
31	Bisphenol A	BPA	Migrant from packaging	Bisphenol	3.3	11.51	-	[M-H] ⁻	227.107	10	42
32	4-(2,4-Dichlorophenoxy)butyric acid	2,4-DB	Pesticide (herbicide)	Auxinic herbicide	1.35	11.62	-	Frag ³	160.957	6.9	19.4

BFDGE¹: Bisphenol F diglycidyl ether; BADGE²: Bisphenol A diglycidyl ether; Frag³: Fragment

The optimized parameters for ESI⁺ and ESI⁻ are presented in Supplementary data, Table 2.3.

Based on the instrument factory settings, detection of acrylamide remained unsuccessful. Indeed, the ion path dedicated to discard neutral molecules before analysis (called “StepWave” by Waters[®], which is the ionic path between the source and the mass analyzer) proved to be the limiting step for small molecules ($m < 90 m/z$). Therefore, its settings were set according to Waters[®] instructions for extending the range of mass detected in a single run (60 - 800 m/z) and ensuring acrylamide detection.

2.2.1.3. QUANTIFICATION AND QUALITY CONTROLS

In order to select the best compromise for quantification of targeted contaminants in tea samples, two main quantification methods were considered: solvent calibration and matrix-matched calibration. Each time, classical external calibration was considered and compared with labelled molecules correction.

For external calibration, 11 standard solutions were prepared in a ACN/water mix (20/80 v/v) acidified with 0.1% FA, with concentrations ranging from 0.1 to 120 ng/mL for most compounds, except for a few molecules exhibiting lower sensitivity (namely hydroxymethylfurfural (HMF), deoxynivalenol (DON), bisphenol A (BPA) and bisphenol F (BPF)) with concentrations five times higher (from 0.5 to 600 ng/mL), as well as acrylamide (AA) with concentrations 10 times higher (1 to 1,200 ng/mL).

For labelled molecules correction, deuterated isotopes were used in addition to external calibration, either in positive mode (acrylamide-d3, dimethoate-d6, ochratoxin A-d5 and malathion-d6) and/or in negative mode (ochratoxin A-d5 and bisphenol A-d14). These five molecules were associated to the non-labelled analytes they are supposed to mimic according to their similarities and the closeness of their retention time. All labelled molecules were added before sample treatment in order to reach a concentration in the final extract of 40 ng/mL for dimethoate-d6, ochratoxin A-d5 and malathion-d6, 200 ng/mL for bisphenol A-d14 and 400 ng/mL for acrylamide-d3.

Last but not least, for matrix-matched calibration, 8 calibration points were prepared directly in sample extract, with concentrations ranging from 2 to 80 ng/mL (with respectively the same multiplicative factors as for external calibration).

The analytical sequence was randomized and a mobile phase blank as well as a quality standard (standard solution containing all of 32 compounds at 16 ng/mL) were analyzed every 10 samples. No cross-contamination was detected during long sequences.

2.2.1.4. DATA TREATMENT

The entire data processing was done using open-source, freely available tools. Data files were first converted in mzXML with proteowizard (Chambers et al., 2012), and then processed with MZmine 2 (Pluskal, Castillo, Villar-Briones, & Orešič, 2010), using its “Targeted peak detection” module. When necessary, the instrumental drift along the analytical sequence was addressed by performing “All Loess Pool” correction using the Workflow4Metabolomics Galaxy interface (Giacomoni et al., 2015; Van Der

Kloet et al., 2009). Briefly, analytical drift was modelled based on the quality standard samples, and then the signal was corrected for all samples according to their injection order. Corrected data were then processed with Microsoft Excel.

2.2.2. TEA SAMPLES AND BREW PREPARATION

Tea leaves used for optimization of extraction procedures came from a bulk green tea bought in a tea shop in Paris (France). Tea brew was prepared in accordance with the ISO 3103:1980 norm ensuring a reproducibility of infusion (International Organization for Standardization, 1980). Briefly, 200 mL tap water were heated at 100°C and infusion was prepared using 2 g of tea leaves weighted in a tea filter and then put in contact with hot tap water for 6 min. Blank sample was performed under the same procedure but without tea leaves, in order to investigate possible contamination from tap water, tea filter, preparation or extraction protocols; no contamination was detected along the analytical process. Non-spiked tea samples were also systematically considered in order to check the absence of native target contaminants (only native HMF could be found in our tea samples), to avoid overestimation of recoveries for the spiking.

2.2.3. SAMPLE TREATMENT PROTOCOLS

2.2.3.1. TEA BREW

Two different sample preparation methods were evaluated for tea brew. These two methods were already developed by other authors as mentioned below, but they needed to be adapted here to match our initial chromatographic conditions (i.e. mobile phase: 100% water) and our tea matrix.

The first method is based on liquid-liquid extraction (LLE) commonly used for multi-residue analyzes in food products, and adapted from Mol *et al.* (Hans G. J. Mol et al., 2008). Briefly, 2 mL of tea brew were mixed with 4 mL of ultrapure water, and 3 mL of ACN were added with 2 g of MgSO₄ and 1 g of NaCit. The sample was then agitated during 1 h before centrifugation at 3,000 g for 5 min. Aliquot (1 mL) of supernatant was sampled and diluted with 4 mL of water. Finally, 0.5 mL was sampled and filtered at 0.2 µm with a syringeless filter (mini-uniprep G2, Whatmann) before analysis.

The second sample treatment tested, named dilute and shoot (D&S), comes from metabolomic approaches (Tengstrand, Rosén, Hellenäs, & Åberg, 2013). It is intended to favor the analysis of numerous targeted organic compounds in complex matrices. In our case, 2 mL of tea brew were diluted with 6 mL of pure water plus 2 mL of ACN, both acidified with 0.1% FA, and agitated during 5 min. Then 0.5 mL of the mixture was sampled and filtered at 0.2 µm with a syringeless filter (mini-uniprep G2) before analysis.

2.2.3.2. TEA LEAVES

Tea leaves were crushed with a pestle and a mortar in order to obtain a homogenous powder. Firstly the extraction protocol described by Mol et al. (Hans G. J. Mol et al., 2008) (using water and acetonitrile

containing formic acid - “proposed method A”) was chosen. This method derives from QuEChERS extraction protocols that proved their efficiency for multi-residue analyses in food samples (recoveries between 80 and 110%) (Dong & Xiao, 2017; Dzuman, Zachariasova, Veprikova, Godula, & Hajslova, 2015), except that the dispersive solid-phase clean-up step has been removed to avoid losses of targeted compounds. In addition, considering the low sensitivity and high matrix effects for some compounds, changes were required. In particular, an evaporation step was needed and acidified ACN was selected to recover the dry residue as inspired by Cotton et al. (Cotton et al., 2014). Since this solvent was unable to recover fumonisins B1 and B2, two more options were further investigated to improve the efficiency of the extraction: 1) increasing the percentage of FA or 2) adding MeOH to the mixture (Hans G. J. Mol et al., 2008). The first option led to strong matrix effects for acrylamide without improving fumonisins recovery. Therefore a mixture of ACN/MeOH (90/10 v/v) acidified with 0.1% FA was finally chosen as extracting solvent.

Finally, 5 mL of ACN/MeOH (90/10 v/v) mixture acidified with 0.1% FA were added to 1 g and agitated during 1 h before centrifugation at 3,000 g for 5 min. Then, the supernatant was collected and an aliquot (1 mL) was evaporated to dryness under a gentle stream of nitrogen. The extract was further reconstituted in 0.2 mL of ACN + 0.1% FA. Then 0.8 mL of ultrapure water with 0.1% FA was added in order to reconstitute 1 mL of final volume. Finally, 0.5 mL were sampled and filtered at 0.2 µm using syringeless filters (mini-uniprep G2) before analysis.

2.2.4. METHOD VALIDATION

Instrumental limits of detection and quantification (LOD, LOQ) were evaluated based on repeatability and reproducibility of analyzes of standard solutions at different low levels (from 0.1 to 100 ng/mL). For each level, 3 replicates were analyzed 5 times. Relative standard deviation (RSD) was plotted against concentration for each compound, and LODs or LOQs were then graphically assessed when RSDs reached 30% or 10% respectively according to IUPAC recommendation (Currie, 1995; Thompson, Ellison, & Wood, 2002). The overall method limits of quantification (MLOQs) were determined according to the guideline from the SANTE document of the European Commission – Directorate General for Health and Food Safety (European Commission - DG Health and Food Safety, 2015) by the lowest spiked level reaching the following criteria: recovery 70–120% and RSD <20%.

Trueness and precision were evaluated based on triplicates spiked with a mixture solution of all investigated contaminants. Spiking levels were established in accordance with the European regulation for some contaminants (such as maximum residue limits for pesticides) or the lowest concentration expected for other contaminants. Thus, for tea brew, spiking levels were 10 ng/mL for level 1 and 100 ng/mL for level 2 for most of compounds, and 50 ng/mL (level 1) and 500 ng/mL (level 2) for the less sensitive compounds (AA, HMF, DON, BPA, BPF). For tea leaves, three levels were set: 10, 50 and 100 µg/kg for most of compounds; 50, 250 and 500 µg/kg for HMF, DON, BPA and BPF; 100, 500 and 1,000 µg/kg for AA. Spiking was performed with 100 µL of standard solution in ACN followed by

equilibrium overnight. For non-spiked samples, 100 μ L of ACN was added. The trueness was considered acceptable for recoveries between 70 and 120%, and precision (assessed by RSD) lower than 20% was required (European Commission - DG Health and Food Safety, 2015; H. G J Mol, Reynolds, Fussell, & Štajnbaher, 2012).

2.3. RESULTS AND DISCUSSION

2.3.1. CHOICE OF ANALYTICAL CONDITIONS AND VALIDATION

2.3.1.1. CHROMATOGRAPHIC SEPARATION

The first challenge for the chromatographic separation was the retention of acrylamide, since it is a very small and highly polar molecule ($\log K_{ow} = -0.67$). This probably explains why acrylamide has never been considered in a multi-residue approach, as its physicochemical properties are too far from those of the other contaminants.

A preliminary test performed on a Hypersil C18 Gold column (150x2.1 mm and 2 μ m particles, Thermo Fisher Scientific) gave a retention factor (k) value of 0.33, showing insufficient acrylamide retention, probably due to the fraction of ACN in the extract (20%) that was needed to avoid severe losses of hydrophobic compounds upon extract filtration (PTFE filter). Facing the impossibility to change the fraction of ACN, the only choice was to change the stationary phase of the column. Therefore the C18-PFP phase, combining the strength of C18-chains (hydrophobicity) and PFP-groups (hydrogen bonds, dipole-dipole and π - π interactions) and affording possible enhanced retention of polar compounds, has been chosen. To date, while some studies combined C18 and PFP phases using two successive columns in LCxLC applications (Mena-Bravo, Priego-Capote, & Luque de Castro, 2016), only two applications dealing with a C18-PFP column have been published yet, for analyzes in rat plasma (Zgair et al., 2015) and markers of lung cancer in human serum (Klupczynska et al., 2017). Thus, the potential of this new stationary phase for the separation of multi-class contaminants in food samples is described here for the first time. Retention tests for acrylamide on C18-PFP with the previously mentioned conditions revealed a k value about 1.16, showing a significant enhancement of acrylamide retention on this stationary phase, still operating in a reversed phase mode. Such a retention factor above 1 opens new perspectives for the separation of acrylamide from the very first matrix compounds eluting with the dead volume (dead time near 0.6 min here – see Figure 2.1.1). Dealing with highly polar compounds is an important challenge for multi-class analysis as pointed out by Pérez-Ortega *et al.*, who faced the co-elution of most polar compounds (among hundreds of targeted contaminants) with the dead volume upon analysis using a conventional C18-silica stationary phase, resulting in poor analytical performances (Pérez-Ortega et al., 2016). New combined phases such as C18-PFP may represent key tools to solve retention problems of very polar compounds in multi-class contaminants analysis on reversed phase columns.

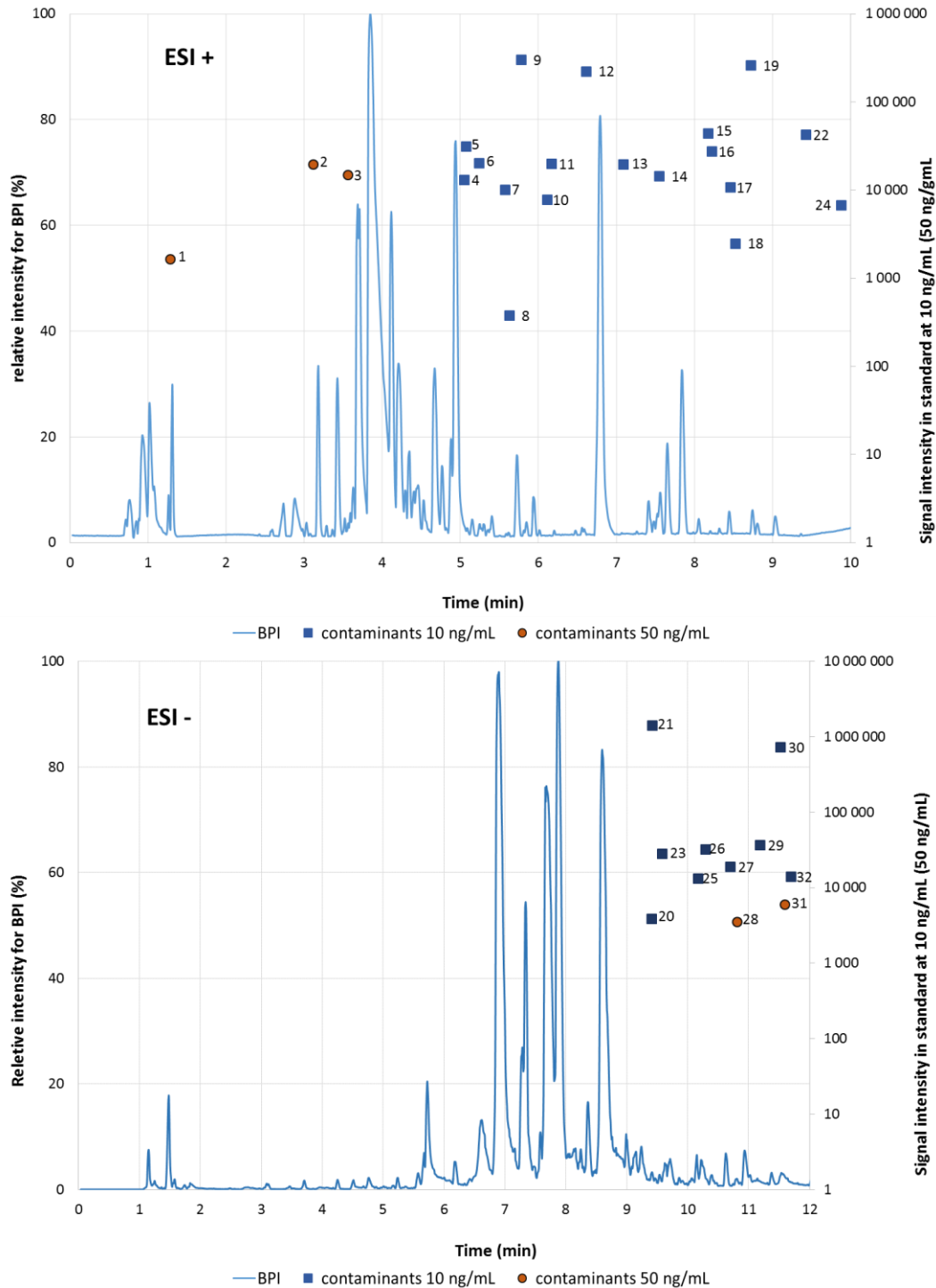


FIGURE 2.1.1 BASE PEAK ION CHROMATOGRAM (BPI) OF TEA LEAF EXTRACTS ALONG WITH SIGNAL INTENSITY AND RETENTION OF TARGETED CONTAMINANTS IN ESI + AND ESI-

The second challenge for the chromatographic separation is to ensure a good distribution of compounds along the analysis to reduce the probability of co-elution of compounds (with associated potential interferences) as highlighted by Pérez-Ortega *et al.* (Pérez-Ortega *et al.*, 2016). Such a good separation of compounds is particularly crucial for very complex matrices like food products since they contain several thousands of molecules. Figure 2.1.1 displays the distribution of compounds all along the chromatographic run, and the effective separation of targeted contaminants from the major ions of the

BPI of tea extracts. Limiting co-elution of targeted contaminants and major compounds (such as polyphenols, caffeine, amino acids) of tea may play a key role in reducing matrix effects as discussed below.

2.3.1.2. MS ANALYSIS

Table 2.1 displays some information about the 32 targeted compounds such as their class, log Kow, retention time and quantification ion selected. In ESI⁺, most of compounds were analyzed using [M+H]⁺ ion; in the case of aldicarb, BADGE and BFDGE, the most abundant ion was [M+Na]⁺ adduct. For propargite, although [M+Na]⁺ adduct was still the most abundant, a fragment was used instead for quantification since the [M+Na]⁺ adduct signal showed interfering ions. In ESI⁻, most of compounds were analyzed using [M-H]⁻ ion, except some acidic herbicides for which fragmentation occurred (in that case, the most abundant and specific fragment was selected for quantification). This fragmentation was suspected to occur in the StepWave which is a specificity of some Water[®] mass spectrometers, as the energy involved to curve the trajectory of ions is higher than the energy needed to fragment some molecules (3 eV) leading to in-instrument fragmentation. Therefore, the intensities of these fragments are specific of the instrument we used and may not be as intense on other HRMS analyzers set in full scan mode.

2.3.1.3. INSTRUMENTAL PERFORMANCES

Table 2.1 shows instrumental LODs and LOQs for our targeted contaminants using our LC-HRMS system. LOQs range from 0.1 ng/mL for BPS to 50 ng/mL for BPF. Generally speaking, for most investigated contaminants, our estimated LOQs are higher than LOQs mentioned in the literature for targeted analytical methods. For example a recent study reported LOQ as low as 5 ng/mL for acrylamide in standard solution while the current method only reaches 51.5 ng/mL (Troise & Fogliano, 2015). For pesticide compounds the gap is tighter since most of recent studies deal with multi-residue analysis, except for acidic herbicides because of the fragmentation of those compounds in our instrument (LOQ ranging from 4.6 to 19.4 ng/mL) (Hou et al., 2014). The slightly higher LOD and LOQ values for our method were expected since they are linked to the compromise required for multi-class analysis of contaminants exhibiting wide differences in their physicochemical properties, meaning that individual optimization of instrumental conditions (LC and MS) is impossible.

2.3.2. SAMPLE TREATMENT RECOVERIES

2.3.2.1. TEA BREW

For tea brew, two sample treatments commonly used for multi-residues analyzes were compared, namely LLE and D&S. Considering the moderate matrix effects observed, quantification was performed using labelled molecules correction here (matrix-matched calibration was not needed).

The percent of the 32 compounds in the different recovery ranges for LLE and D&S at two spiking levels (10 and 50 ng/mL for most compounds) are displayed in Figure 2.1.2, while the complete results for recoveries and RSDs are gathered in Table 2.4.

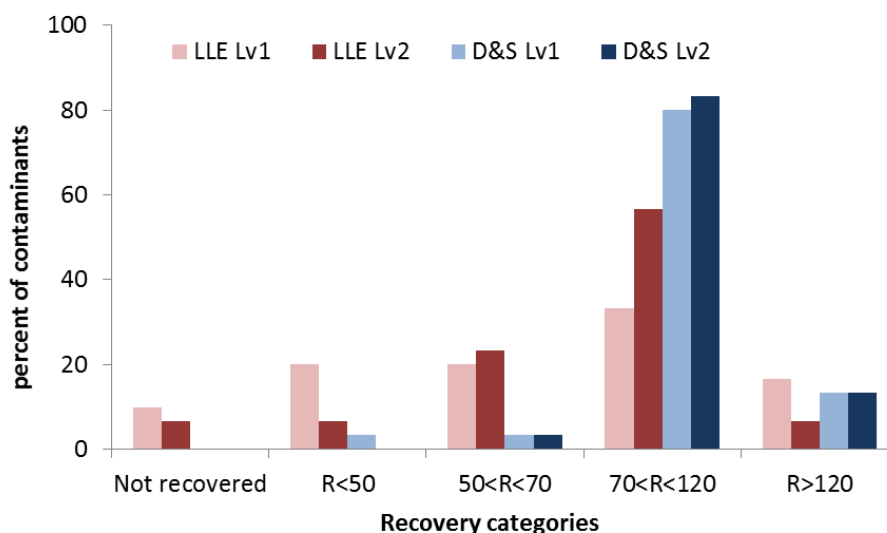


FIGURE 2.1.2 PERCENTAGE OF 32 TARGET CONTAMINANTS IN RECOVERY RANGE FOR SPIKED TEA BREW.

(LV1: 10 NG/ML FOR MOST COMPOUNDS EXCEPT 50 NG/ML FOR AA, HMF, BPA, BPF, DON;
LV2: 100 NG/ML FOR MOST COMPOUNDS EXCEPT 500 NG/ML AA, HMF, BPA, BPF, DON)

Based on guidelines for multi-residue analysis proposed by Mol *et al.* (H. G J Mol et al., 2012) and the European Commission (European Commission - DG Health and Food Safety, 2015), the recovery range should be 70 – 120% and the precision (RSD) lower than 20% in order to validate quantification of compounds. In the case of multi-residue analysis, recoveries lower than 70% are acceptable for some compounds provided that RSDs are low (European Commission - DG Health and Food Safety, 2015). With the D&S method, all compounds were detected at the lowest level, while 80 and 83% of target compounds for levels 1 and 2 respectively met the quantification criteria. The percentages were much lower for LLE method with 33% for level 1 and 60% for level 2. Surprisingly, whatever the level considered, tolfenpyrad and propargite were not recovered by LLE despite their hydrophobicity ($\log K_{ow} > 5$); on the opposite, both compounds reached 86 and 92% recoveries at 10 ng/mL using the D&S method. No similar results or explanation was found in the literature. Nonetheless, considering the European RASFF alerts for tea, these two pesticides are commonly cited so that their quantification at low concentrations is crucial. D&S method led to some overestimations ($R > 120\%$) especially for fumonisins B1 and B2 with respectively 178 and 161%, possibly due to matrix effects since similar

signal enhancement was found in tea leaves (Figure 2.1.3) and reported in the literature (Zhang et al., 2016).

The repeatability of extraction, based on triplicates per level, revealed that for level 1 using D&S preparation, 90% of compounds have an acceptable RSD (<20%), when this value dropped to 60% for LLE. For D&S, only HMF, acrylamide and 2,4-D showed RSDs higher than 20% owing to native contamination for the former and a signal close to the quantification limits of overall method for the latter. For level 2, 96% of compounds met the requirement of repeatability for D&S, but only 79% in the case of LLE.

At last, considering the overall results, the D&S method seems more appropriate for contaminants quantification in tea brew.

2.3.2.2. TEA LEAVES

- Matrix effects

Matrix effects were determined based on comparison between external calibration and matrix-matched calibration. The ratio of slopes from both calibration curves (external / matrix-matched), shown in Figure 2.1.3, is informative on the presence of matrix effects for ESI⁺ and ESI⁻.

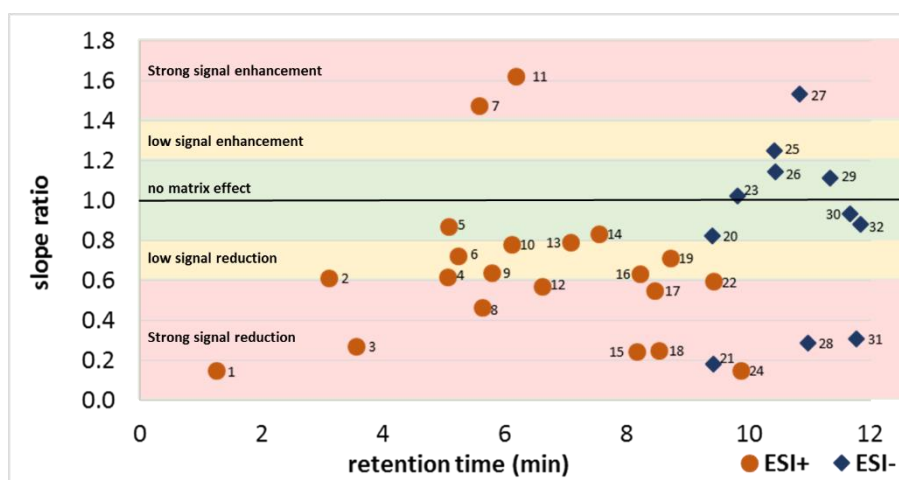


FIGURE 2.1.3 MATRIX EFFECTS OBSERVED UPON ANALYSIS OF TEA LEAVE EXTRACTS

Based on these results, 8 compounds (25% of target contaminants) do not undergo significant matrix effects (essentially in ESI⁻), 17 compounds (around 50%) present low matrix effects (ratios between 0.6 and 1.4), while for the remaining compounds strong matrix effects were observed, mainly with signal suppression. Only fumonisins B1 and B2 (ESI⁺) and 2,4,5-T (ESI⁻) revealed strong signal enhancement. According to our results no correlation between matrix effects and retention time was observed. These results are close to those reported by Mol *et al.* (Hans G. J. Mol et al., 2008) for other types of food (feed, maize, milk) on 172 analytes showing more intense signal reductions for methanol extraction than for acetonitrile. Signal enhancement for fumonisin B1 was also reported by Danezis *et al.* (Danezis et

al., 2016) for most of foods studied (between 163 and 175% of standard signals). Our results are in agreement with Dzuman, Zachariasova, Veprikova, Godula, & Hajslova (Dzuman et al., 2015) who reported 65% overall signal suppression for hundreds of contaminants in tea using a QuEChERS approach.

- Recoveries

The recoveries were determined based on three different calibration methods: external standard, labelled molecules correction and matrix-matched calibration.

With the use of labelled molecules, in order to correct the extraction losses and matrix effects, the 32 molecules were corrected by the most relevant labelled molecules of our mix (close retention time and related physicochemical properties). In that case, our results underline decreases of overall recoveries as compared to the use of external calibration. Hence, at 10 and 50 µg/kg only 38% of molecules meet the requirement for recovery (70-120%) for external calibration, and this percentage drops respectively to 28 and 31% at 10 and 50 µg/kg using labelled molecules correction. This drop is surprising since labelled correction (also called isotope dilution) is often used for quantification of contaminants in various food matrices (European Commission - DG Health and Food Safety, 2015). It has been attributed to strong matrix effects in ESI analyzes due to the absence of clean-up. Indeed, ochratoxin A-d5 undergoes a strong signal enhancement in negative mode (value ratio: 2.2) which affects quantification of compounds after labelled correction. Unsurprisingly, the correction of labelled molecules was only efficient for the corresponding non-labelled molecules and, therefore, for matrix-matched calibration we decided to avoid labelled correction for other molecules.

In the case of matrix-matched calibration, recoveries were in accordance with results reported by Kaczyński *et al.* for 300 pesticides in fish tissues (Kaczyński, Łozowicka, Perkowski, & Szabuńko, 2017) or by Mol *et al.* for 258 molecules in different types of food (Hans G. J. Mol et al., 2008), with around 63–78% of molecules within the recovery range 70-120% (see detailed results in Table 2.2). Only six molecules (19%) were not detected at the first spiking level: fumonisins B1 and B2 and propargite (10 µg/kg), BPA and BPF (50 µg/kg) and acrylamide (100 µg/kg), owing to a reduced sensitivity of the overall method for these contaminants. Fortunately, BPA, BPF and acrylamide were recovered with satisfying recoveries from spiking level 2 (250 and 500 µg/kg). On the opposite, fumonisins B1 and B2 as well as propargite exhibited recoveries lower or close to 50% even at the highest spiking level (Table 2.2). At last, the repeatability for the 32 molecules was acceptable since 72% of compounds have RSD lower than 20%. The high RSD values found for BADGE and BFDGE result from matrix compounds competition for Na⁺ adducts, and in the case of HMF it is attributed to the native presence of this molecule and its isomers at “high” concentration (≈ 265 µg/kg) in tea leaves (Table 2.2).

TABLE 2.2 RECOVERY, RSD, METHOD LIMIT OF QUANTIFICATION (MLOQ) AND MAXIMUM RESIDUE LIMIT (MRL) FOR THE 32 TARGETED COMPOUNDS IN TEA LEAVES QUANTIFIED WITH MATRIX-MATCHED CALIBRATION.

N°	Compound	Level 1 (10 µg/kg ^a = 50 µg/kg ^b = 100 µg/kg)		Level 2 (50 µg/kg ^a = 250 µg/kg ^b = 500 µg/kg)		Level 3 (100 µg/kg ^a = 500 µg/kg ^b = 1000 µg/kg)		MLOQ (µg/kg)	MRL (µg/kg)
		Recovery (%)	RSD (%)	Recovery (%)	RSD (%)	Recovery (%)	RSD (%)		
1	Acrylamide ^b	ND	-	86.4	19.0	110.2	21.1	500	-
2	Hydroxymethylfurfural ^a	89.4	77.5	51.3	41.1	51.0	23.3	-	-
3	Deoxynivalenol ^a	282.4	12.1	138.6	24.1	91.7	19.8	500	1,250
4	Imidacloprid	103.8	13.0	103.8	16.1	96.5	14.9	10	50
5	Dimethoate	113.8	15.8	96.6	15.8	107.0	2.4	10	50
6	Acetamiprid	115.4	17.4	115.4	14.8	100.4	16.9	10	50
7	Fumonisin B1	ND	-	13.4	58.5	22.4	5.0	-	-
8	Aldicarb	87.7	13.4	87.7	15.7	78.9	12.0	10	50
9	Ametryn	106.3	15.3	106.3	13.3	96.5	12.1	10	10
10	Dichlorvos	114.7	16.1	96.2	16.1	112.3	14.9	10	20
11	Fumonisin B2	ND	-	24.6	36.9	28.7	6.1	-	-
12	Atrazine	147.9	16.9	103.5	19.6	105.8	16.4	50	100
13	Diuron	101.8	11.6	88.0	7.1	91.0	8.3	10	50
14	Ochratoxin A	95.1	6.1	106.6	4.1	108.7	3.6	10	10
15	BFDGE ¹	77.4	26.4	73.1	12.0	78.9	17.0	50	-
16	Malathion	108.7	5.7	105.4	3.6	106.6	2.3	10	500
17	Diflubenzuron	76.3	15.9	77.0	6.8	80.0	7.3	10	100
18	BADGE ²	107.1	51.6	105.3	11.2	103.9	16.0	50	60,000
19	Pirimiphos-methyl	84.6	12.4	80.1	11.3	86.5	6.1	10	50
20	2,4-D	66.1	17.8	68.8	11.6	71.4	10.0	50	100
21	Bisphenol S	80.3	11.4	80.1	13.1	76.8	11.2	10	50
22	Tolfenpyrad	45.2	11.4	56.4	8.6	62.1	8.0	(50)	10
23	MCPA	73.6	8.1	94.8	10.2	71.1	10.3	10	100
24	Propargite	ND	-	38.5	18.5	52.6	20.3	(100)	100
25	Dichloprop	83.9	18.1	91.6	19.8	87.6	15.2	10	100
26	Mecoprop	79.9	6.0	98.1	12.3	74.1	11.5	10	100
27	2,4,5-T	82.9	8.8	91.3	8.8	88.3	18.9	10	50
28	Bisphenol F ^a	ND	-	75.8	25.8	72.3	21.4	(500)	-
29	Fenoprop	76.4	5.5	71.5	9.5	70.9	8.5	10	10
30	Dinoseb	62.1	6.1	66.3	15.9	60.5	11.6	(10)	100
31	Bisphenol A ^a	ND	-	78.2	36.7	74.7	22.5	(500)	600
32	2,4-DB	61.8	17.6	68.8	11.6	72.3	8.9	100	50

ND= Not detected. Values under bracket mean that validation criteria were not reached but were still acceptable according to the European commission guideline (European Commission - DG Health and Food Safety, 2015)

- Analytical performances

The MLOQs determined in accordance with the European guidance document (European Commission - DG Health and Food Safety, 2015) clearly show that our method performances are sufficient to ensure the compliance of a tea sample with the maximum residue limits (MRLs) fixed by the European Commission (European Commission, 2005). These results may open routine analysis field to broad-scale multi-class methods within the next years. Among the lowest MRLs set for ametryn, tolfenpyrad and fenoprop at 10 µg/kg of tea leaves, only tolfenpyrad do not meet the quantification requirement with a recovery around 45% and RSD of 11%. However, according to the European guidance document, such performances are still acceptable in the case of multi-residue methods (European Commission - DG Health and Food Safety, 2015). Finally, 69% of targeted molecules could be quantified at concentrations equal

to or lower than their MRLs or usual levels in tea for non-regulated compounds (European Commission, 2005, 2006), and additional 22% could be detected but not quantified without labelled correction (recoveries below 70% or above 120%). As an example, fenoprop and ametryne were correctly quantified at 10 µg/kg (respective recoveries: 76.4 and 89.7% and RSD <20%) while recovery for tolfenpyrad (about 45%) should require correction before quantification. For molecules with low concentration limits, improvement of our method would be to add their corresponding labelled molecule to ensure a correct quantification at low levels.

Finally, despite the large diversity of the 32 selected compounds coming from 4 different classes of contaminants, our analytical method displays equivalent performances than recent multi-residue methods dedicated for hundreds of compounds belonging to only 1 or 2 classes.

2.4. CONCLUSION

For the very first time, we have developed an analytical method dedicated to the simultaneous determination of food contaminants belonging to 4 different classes (pesticides, mycotoxins, process-induced toxicants and migrants from packaging) and exhibiting extremely different physicochemical properties. This method, based on metabolomics sample treatment and a new liquid chromatographic column stationary phase coupled to high resolution mass spectrometry, has been validated on 32 target contaminants, ranging from highly polar acrylamide ($\log K_{ow} = -0.67$) to less polar tolfenpyrad ($\log K_{ow} = 5.61$), in both tea brew and leaves. All these molecules could be detected and correctly quantified at trace levels (near µg/kg in tea leaves or µg/L in tea brew) generally below their respective regulation limits. Only fumonisins B1 and B2 could not be quantified at any concentrations assessed in tea leaves due to low recoveries for these molecules, while quantification was correct at 10 µg/L in tea brew. Matrix-matched calibration was required to hinder strong matrix effects for several molecules for tea leaves while only labelled correction was needed for tea brew (lower matrix effects). Our method can be considered as a new step beyond current multi-residue analyzes, and its applicability to other food matrices would be valuable to be investigated in a near future.

2.5. FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

2.6. REFERENCES

- 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>
- Castro-Puyana, M., & Herrero, M. (2013). Metabolomics approaches based on mass spectrometry for food safety, quality and traceability. *TrAC Trends in Analytical Chemistry*, 52, 74–87. <https://doi.org/10.1016/j.trac.2013.05.016>

- Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., ... Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, 30(10), 918–920. <https://doi.org/10.1038/nbt.2377>
- Chang, K. (2015). *World tea production and trade. Current and future development* (p. 17). Rome: food and agriculture organization of the united nations.
- Cotton, J., Leroux, F., Broudin, S., Marie, M., Corman, B., Tabet, J.-C., ... Junot, C. (2014). High-Resolution Mass Spectrometry Associated with Data Mining Tools for the Detection of Pollutants and Chemical Characterization of Honey Samples. *Journal of Agricultural and Food Chemistry*, 62(46), 11335–11345. <https://doi.org/10.1021/jf504400c>
- Currie, L. A. (1995). Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). *Pure and Applied Chemistry*, 67(10), 1699–1723. <https://doi.org/10.1351/pac199567101699>
- Danezis, G. P., Anagnostopoulos, C. J., Liapis, K., & Koupparis, M. A. (2016). Multi-residue analysis of pesticides, plant hormones, veterinary drugs and mycotoxins using HILIC chromatography – MS/MS in various food matrices. *Analytica Chimica Acta*, 942, 121–138. <https://doi.org/10.1016/j.aca.2016.09.011>
- Dong, H., & Xiao, K. (2017). Modified QuEChERS combined with ultra high performance liquid chromatography tandem mass spectrometry to determine seven biogenic amines in Chinese traditional condiment soy sauce. *Food Chemistry*, 229, 502–508. <https://doi.org/10.1016/j.foodchem.2017.02.120>
- 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, 29–40. <https://doi.org/10.1016/j.aca.2015.01.021>
- EFSA. (2016). *Chemicals in food 2016: Overview of selected data collection* (p. 40).
- Eitzer, B. D., Hammack, W., & Filigenzi, M. (2014). Interlaboratory Comparison of a General Method To Screen Foods for Pesticides Using QuEChERS Extraction with High Performance Liquid Chromatography and High Resolution Mass Spectrometry. *Journal of Agricultural and Food Chemistry*, 62(1), 80–87. <https://doi.org/10.1021/jf405128y>
- European Commission. Regulation (EC) No. 396/2005 of the European Parliament and of the council of 23 February 2005 on maximum residue levels of pesticides in or on food and feed of plant and animal origin and amending Council Directive 91/414/EC., 70 Official Journal of the European Union § (2005).
- European Commission. Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs, 364 Official Journal of the European Union § (2006).
- European Commission – Directorate General for Health and Food Safety. (2015). Guidance document on analytical quality control and method validation procedures for pesticides residues analysis in food and feed. *SANTE/11945/2015*.
- Giacomini, F., Le Corguille, G., Monsoor, M., Landi, M., Pericard, P., Petera, M., ... Caron, C. (2015). Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*, 31(9), 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>
- Hou, X., Lei, S., Qiu, S., Guo, L., Yi, S., & Liu, W. (2014). A multi-residue method for the determination of pesticides in tea using multi-walled carbon nanotubes as a dispersive solid phase extraction absorbent. *Food Chemistry*, 153, 121–129. <https://doi.org/10.1016/j.foodchem.2013.12.031>
- International Organization for Standardization. (1980). *ISO 3103:1980 - Tea -- Preparation of liquor for use in sensory tests* (p. 4).
- Jin, Y., Zhang, J., Zhao, W., Zhang, W., Wang, L., Zhou, J., & Li, Y. (2017). Development and validation of a multiclass method for the quantification of veterinary drug residues in honey and

- royal jelly by liquid chromatography–tandem mass spectrometry. *Food Chemistry*, 221, 1298–1307. <https://doi.org/10.1016/j.foodchem.2016.11.026>
- Kaczyński, P., Łozowicka, B., Perkowski, M., & Szabuńko, J. (2017). Multiclass pesticide residue analysis in fish muscle and liver on one-step extraction-cleanup strategy coupled with liquid chromatography tandem mass spectrometry. *Ecotoxicology and Environmental Safety*, 138, 179–189. <https://doi.org/10.1016/j.ecoenv.2016.12.040>
- Klupczynska, A., Dereziński, P., Garrett, T. J., Rubio, V. Y., Dyszkiewicz, W., Kasprzyk, M., & Kokot, Z. J. (2017). Study of early stage non-small-cell lung cancer using Orbitrap-based global serum metabolomics. *Journal of Cancer Research and Clinical Oncology*, 143(4), 649–659. <https://doi.org/10.1007/s00432-017-2347-0>
- Mena-Bravo, A., Priego-Capote, F., & Luque de Castro, M. D. (2016). Two-dimensional liquid chromatography coupled to tandem mass spectrometry for vitamin D metabolite profiling including the C3-epimer-25-monohydroxyvitamin D3. *Journal of Chromatography A*, 1451, 50–57. <https://doi.org/10.1016/j.chroma.2016.05.006>
- Mol, H. G. J., Reynolds, S. L., Fussell, R. J., & Štajnbaher, D. (2012). Guidelines for the validation of qualitative multi-residue methods used to detect pesticides in food. *Drug Testing and Analysis*, 4(S1), 10–16. <https://doi.org/10.1002/dta.1364>
- Mol, Hans 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>
- 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>
- Petrarca, M. H., Fernandes, J. O., Godoy, H. T., & Cunha, S. C. (2016). Multiclass pesticide analysis in fruit-based baby food: A comparative study of sample preparation techniques previous to gas chromatography–mass spectrometry. *Food Chemistry*, 212, 528–536. <https://doi.org/10.1016/j.foodchem.2016.06.010>
- Pluskal, T., Castillo, S., Villar-Briones, A., & Oresic, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *Bmc Bioinformatics*, 11, 395. <https://doi.org/10.1186/1471-2105-11-395>
- Tengstrand, E., Rosen, J., Hellenas, K.-E., & Aberg, 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, M., Ellison, S. L. R., & Wood, R. (2002). Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). *Pure and Applied Chemistry*, 74(5), 835–855. <https://doi.org/10.1351/pac200274050835>
- Troise, A. D., Fiore, A., & Fogliano, V. (2014). Quantitation of Acrylamide in Foods by High-Resolution Mass Spectrometry. *Journal of Agricultural and Food Chemistry*, 62(1), 74–79. <https://doi.org/10.1021/jf404205b>
- van der Kloet, F. M., Bobeldijk, I., Verheij, E. R., & Jellema, R. H. (2009). Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. *Journal of Proteome Research*, 8(11), 5132–5141. <https://doi.org/10.1021/pr900499r>
- Zgair, A., Wong, J. C. M., Sabri, A., Fischer, P. M., Barrett, D. A., Constantinescu, C. S., & Gershkovich, P. (2015). Development of a simple and sensitive HPLC–UV method for the simultaneous determination of cannabidiol and Δ^9 -tetrahydrocannabinol in rat plasma. *Journal of*

Pharmaceutical and Biomedical Analysis, 114, 145–151.
<https://doi.org/10.1016/j.jpba.2015.05.019>

Zhang, S., Lu, J., Wang, S., Mao, D., Miao, S., & Ji, S. (2016). Multi-mycotoxins analysis in *Pheretima* using ultra-high-performance liquid chromatography tandem mass spectrometry based on a modified QuEChERS method. *Journal of Chromatography B*, 1035, 31–41.
<https://doi.org/10.1016/j.jchromb.2016.09.022>

2.7. SUPPLEMENTARY MATERIALS

TABLE 2.3 IONIZATION PARAMETERS FOR ESI+ AND ESI-.

Parameter	ESI+	ESI-
Capillary (kV)	1.5	1.0
Sampling cone (V)	20	35
Source offset (V)	20	80
Source temperature (°C)	130	130
Desolvation temperature (°C)	500	500
Cone gas flow (L/h)	50	20
Desolvation gas flow (L/h)	1,200	600

TABLE 2.4 RECOVERIES, RSDs AND ESTIMATED MLOQs IN TEA BREW FOR COMPOUNDS SPIKED BETWEEN 10 AND 100 µG/KG EXTRACTED BY LLE OR D&S METHOD

N°	Name	Liquid- liquid extraction					Dilute and shoot				
		Level 1 10 µg/L a = 50 µg/L		Level 2 100 µg/L a = 500 µg/L		MLOQ (µg/L)	Level 1 10 µg/L a = 50 µg/L		Level 2 100 µg/L a = 500 µg/L		MLOQ (µg/L)
		Recovery (%)	RSD (%)	Recovery (%)	RSD (%)		Recovery (%)	RSD (%)	Recovery (%)	RSD (%)	
1	Acrylamide ^a	66	4.2	98	22.1	500	49	33.3	83	1.0	500
2	Hydroxymethylfurfural ^a	116	9.2	124	7.3	50	132	26.0	108	8.5	500
3	Deoxynivalenol ^a	134	22.6	102	8.9	500	83	7.9	74	15.7	50
4	Imidacloprid	125	40.7	116	4.5	100	121	5.2	125	8.0	(10)
5	Dimethoate	102	5.9	94	3.4	10	95	5.3	95	3.2	10
6	Acetamiprid	118	8.6	111	3.8	10	117	8.8	116	7.4	10
7	Fumonisin B1	48	20.1	49	47.9	-	109	2.9	178	9.7	10
8	Aldicarb	144	0.3	161	25.5	-	135	0.2	165	6.5	-
9	Ametryn	113	6.3	111	3.1	10	109	7.5	104	5.2	10
10	Dichlorvos	92	12.3	99	4.3	10	95	8.0	109	4.0	10
11	Fumonisin B2	85	4.8	32	113	10	109	7.5	161	8.7	10
12	Atrazine	136	4.8	110	7.3	100	114	7.5	119	5.4	10
13	Diuron	120	13.1	105	5.8	100	97	2.5	95	5.3	10
14	Ochratoxin A	106	12.0	107	2.5	10	96	8.5	100	2.9	10
15	BFDGE	ND	-	112.8	7.1	100	91	27	91.3	19.6	100
16	Malathion	97	17.8	99	1.6	10	96	4.5	96	1.9	10
17	Diflubenzuron	25	61.1	85	15.6	100	119	7.6	119	6.3	10
18	BADGE	ND	-	57.5	34	-	120.0	21	118.3	18.7	(10)
19	Pirimiphos-methyl	27	19.2	63	43.6	-	119	4.4	114	4.3	10
20	2,4-D	ND	-	99	20.9	100	117	32.5	114	7.5	100
21	Bisphenol S	50	11.4	57	16.6	-	58	1.6	84	4.0	100
22	Tolfenpyrad	ND	-	ND	-	-	86	5.4	92	3.3	10
23	MCPA	77	25.3	106	16.9	100	128	10.9	101	3.9	100
24	Propargite	ND	-	ND	-	-	91	4.6	92	7.5	10
25	Dichloprop	27	53.7	55	15.4	-	89	12.6	107	5.1	10
26	Mecoprop	67	21.2	109	16.9	100	98	7.8	97	4.8	10
27	2,4,5-T	15	30.1	66	10.5	(100)	78	8.8	90	4.3	10
29	Fenoprop	55	25.0	71	12.2	100	99	6.2	101	7.0	10
28	Bisphenol F	52	13.6	52	16.9	-	70	0.5	61	8.5	10
30	Dinoseb	64	23.3	75	5.9	100	113	3.7	109	5.8	10
31	Bisphenol A	62	9.3	62	5.4	(10)	71	14.8	71	28.3	10
32	2,4-DB	79	20.5	57	8.2	-	102	19.0	100	8.6	10

3. CONCLUSION

Les principaux résultats obtenus sont les suivants.

Pour le thé sous forme de boisson, les taux de recouvrement ont permis de comparer les deux méthodes de traitement de l'échantillon. La technique par « dilute & shoot » montre de meilleures performances en termes de recouvrements et de répétabilité, puisque plus de 80% des traceurs sont quantifiés dès 10 µg/kg, contre moins de 40% pour la LLE.

Pour le thé sous forme de feuilles, les effets de matrices ont d'abord été évalués. De forts effets de suppression d'ions ont été observés, notamment en mode positif, dans des proportions similaires par rapport à ce qu'avaient observé d'autres auteurs auparavant. Les recouvrements ont ensuite été calculés, avec trois méthodes de calibration différentes : une calibration externe en solvant, une calibration par l'utilisation interne de molécules marquées, et une calibration externe en matrice. Cette dernière a été identifiée comme offrant la meilleure correction des effets de matrice. Au total, 69% des traceurs ont pu être quantifiés à un niveau supérieur ou égal à leur limite réglementaire, et 91% détectés mais non-quantifiés à ce niveau.

En conclusion, la sensibilité observée sur les 32 contaminants « traceurs » est satisfaisante au regard des niveaux de contamination attendus et des limites maximales en vigueur au sein de l'UE. La méthode multi-classes développée doit pouvoir être implémentée dans des études non-ciblées, puisque de nombreuses classes de composés peuvent être analysées simultanément. Par ailleurs, cette diversité d'analytes, couplée à l'absence d'étape de purification dans la préparation d'échantillon et l'utilisation d'une colonne chromatographique à spectre plus large qu'une colonne de silice greffée C18 « classique », font que la détection en aveugle de contaminants, même différents de ceux ayant servi au développement, est favorisée.

CHAPITRE 3
DÉVELOPPEMENT DE L'APPROCHE NON-CIBLÉE

1. INTRODUCTION ET RÉSUMÉ DE L’ARTICLE

La méthode multi-classes développée précédemment a été implémentée pour proposer une preuve de concept d’approche non-ciblée.

Cet article a pour but de prouver la faisabilité de la détection non-ciblée de contaminants à l’état de traces dans un produit agro-alimentaire complexe, à savoir notre produit d’étude, le thé. Un processus de traitement de données LC-HRMS hautement automatisé et modulable a été proposé dans ce but.

Deux jeux de données ont été considérés. Le premier a servi au développement de l’approche ; il est constitué de quatre groupes d’échantillons de thé vert d’une même marque et du même lot (n=3) dopés à trois niveaux différents (10, 50 et 100 µg/kg) avec le mélange de 32 contaminants décrits au Chapitre 2 et rappelé ici dans le Table 3.3, plus un groupe de contrôle non dopé. Le deuxième, utilisé pour la validation de l’approche, est composé de quatre groupes d’échantillon de thé vert de deux marques et origines différentes (n=4). Pour chaque marque, un groupe de contrôle (non dopé) et un groupe dopé à 50 µg/kg avec 3 contaminants sélectionnés (OTA, BPS et tolfenpyrad) ont été considérés. Pour chaque jeu d’échantillon, des échantillons de contrôle qualité sont réalisés en mélangeant une quantité égale de chaque extrait.

Les échantillons ont été analysés grâce à la méthode décrite au Chapitre 2, et trois réplicats d’injection ont été réalisés par échantillon. L’instrument utilisé est de type UHPLC-ToF de marque Waters, et la source d’ionisation de type ESI est opérée en mode positif et négatif. Une légère modification a été apportée au gradient chromatographique pour limiter les effets d’une contamination de l’appareil par des tensioactifs (rajout d’une étape de rinçage au MeOH en fin de gradient). La stratégie d’analyse utilisée vient du champ de la métabolomique, et consiste à aléatoriser les ordres d’injection des échantillons et à injecter régulièrement (typiquement toute les 10 injections) un échantillon de contrôle qualité et un échantillon de solvant.

Un processus de traitement des données en cinq grandes étapes, présenté en Figure 3.1 a été mis en place. La première étape est la construction de la matrice de données à partir des fichiers de données brutes. Le paquet d’outils XCMS, implémenté sur la plateforme Workflow4Metabolomics a été utilisé. Les paramètres complets utilisés pour ces outils se trouvent en Table 3.4. Le processus de construction de la matrice des données est divisé en plusieurs étapes, à savoir l’extraction des pics chromatographiques, plusieurs phases d’alignement des m/z et RT, et enfin une étape de complétion des valeurs manquantes. Trois matrices de données sont obtenues à la fin de cette étape, une « sampleMetadata » contenant les informations jugés utiles sur les échantillons, une « variableMetadata » contenant les informations sur les variables (=ions) détectées (par exemple m/z et

RT mesuré) et une « dataMatrix » contenant les valeurs des aires mesurées pour chaque échantillon et chaque ion.

Ces trois matrices de données sont ensuite importées dans Matlab pour leur analyse. La première étape de l'analyse est d'enlever de ces matrices de données les ions ne présentant pas d'intérêt pour l'étude ou susceptibles d'interférer avec la détection des contaminants. Une série de tests statistiques univariés (test de Student) sont utilisés pour éliminer les ions ne variant pas significativement entre les blancs et les échantillons et entre les groupes d'échantillons. Le calcul du « fold change » (défini, pour chaque ion, comme le rapport entre l'aire médiane mesurée pour le groupe le plus intense et la médiane du groupe le moins intense) permet de ne sélectionner que les ions présentant un fort contraste entre les différents groupes. Plusieurs outils de prétraitement sont ensuite appliqués à cette matrice de données nettoyée. Les données sont mises à l'échelle grâce à un passage en log et une normalisation pareto, puis normalisées par la méthode PQN (Probabilistic Quotient Normalization). La matrice ainsi préparée est maintenant prête pour l'analyse à proprement parler. Une méthode statistique multivariée (Independent Components Analysis, ICA) a été utilisée pour visualiser les données et ainsi détecter les échantillons suspects. Cette méthode a été comparée avec PCA (Principal Components Analysis), qui est la méthode multivariée non-supervisée de référence. La dernière étape est l'étape d'annotation. Un premier algorithme détecte la présence de motifs isotopiques au sein de la matrice de données, et une recherche dans une base de données locale à spectre large est réalisée. Grâce à cela des hypothèses peuvent être formulées sur l'identité des potentiels contaminants présents dans les échantillons suspects, et une proposition d'annotation putative peut être faite.

2. ARTICLE

Untargeted food contaminant detection using UHPLC-HRMS combined with multivariate analysis: feasibility study on tea

Grégoire Delaporte, Mathieu Cladière, Delphine Jouan-Rimbaud Bouveresse, Valérie Camel*.

UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300 Massy, France

¹ *present address: UMR Physiologie de la Nutrition et du Comportement Alimentaire, AgroParisTech, Inra, Université Paris-Saclay, 75005 Paris, France*

** Corresponding author: AgroParisTech, 16 rue Claude Bernard, F-75005, Paris, France*

Phone: +33 1 44 08 17 25 – email: valerie.camel@agroparistech.fr

Food Chemistry, 2018

Available at: <https://doi.org/10.1016/j.foodchem.2018.10.089>

Abstract

The feasibility of untargeted detection of trace contaminants in complex food samples at low levels is assessed in this work. Powerful data pretreatment strategies inspired from the field of metabolomics were adapted to chemical food safety context to enable samples discrimination by multivariate methods. A highly automated workflow was produced, indicating the user the presence of suspect samples, and hypotheses on the potential contaminants which quickly gives the user two main feedbacks: (i) is there a discrimination between sample groups? (ii) are potential food contaminants responsible for this discrimination? The open-source XCMS package was used and efficient data filtration strategies were set up to reduce the data. Independent Components Analysis (ICA) was then performed, being applied here for the first time to mass spectrometry data for food safety issues. Data mining strategies were then set up to automatically detect and annotate suspect ions. Our method achieved the detection of contaminated tea samples (even at $10 \mu\text{g}\cdot\text{kg}^{-1}$), unexpected impurities in the spiking mix standard were highlighted. “Tracer” contaminants considered to develop the method cover a broad range of physicochemical properties with overall 66% annotated blindly. The methodology was finally successfully applied (100% annotated) to a data set exhibiting fewer contaminants and more product diversity as an even more challenging application.

Keyword

Independent Components Analysis; XCMS; ToF; Chemical food safety; Non-targeted approaches; Unexpected contaminants

2.1.INTRODUCTION

In recent years, a new field of food research called “Foodomics”, defined as “a discipline that studies the food and nutrition domains through the application of omics technologies”, has emerged (Cifuentes, 2009). Even though foodomics was first intended for the study of major constituents of food products, related for instance to their authenticity (Ortea et al., 2012), its potential contribution to trace analysis for chemical food safety assessment has also been raised (Antignac et al., 2011). However, chemical food safety assessment faces several challenges since food matrices are highly complex, chemical contaminants are numerous and diverse (pesticides, mycotoxins, process-induced toxicants or migrants from packaging), and often found at trace levels (down to $\mu\text{g}\cdot\text{kg}^{-1}$) (Antignac et al., 2011). Classical targeted analysis strategies are limited since unexpected or unknown contaminants remain non-detected (Tengstrand, Rosén, Hellenäs, & Åberg, 2013). Therefore, there is a rising interest (and demand) for developing untargeted (also referred as “non-targeted”) analyzes of contaminants in food products using a relevant instrumental / analytical strategy combination, as pointed out by numerous reviews (Antignac et al., 2011; Castro-Puyana, Pérez-Míguez, Montero, & Herrero, 2017; Knolhoff & Croley, 2016; Lehotay, Sapozhnikova, & Mol, 2015). There is a consensus on the fact that hyphenated techniques (especially high resolution instrumentation operated in full-scan mode like Ultra-High Pressure Liquid

Chromatography coupled to High Resolution Mass Spectrometry, UHPLC-HRMS) are key technologies for this application thanks to their sensitivity and broad analytical scope, as well as the formula information they provide on potential contaminants (Antignac et al., 2011; Castro-Puyana et al., 2017). Our previous study combining generic extraction and UHPLC-HRMS showed its efficiency on analysis of a wide variety of contaminants with a large range of physicochemical properties (Cladière, Delaporte, Le Roux, & Camel, 2018).

Among new analytical strategies proposed for global chemical food safety assessment, suspect-screening approaches (Gómez-Ramos, García-Valcárcel, Tadeo, Fernández-Alba, & Hernando, 2016; Gosetti, Mazzucco, Gennaro, & Marengo, 2016) and relevant chemical patterns detection using data mining tools (Cotton et al., 2014) show interesting performances in terms of sensitivity. They do not require any initial analysis of the non-contaminated food product but they both rely on *a priori* hypotheses on the chemical structures of contaminants. On the opposite, untargeted metabolomics-like strategies require the analysis of a reference food product (to compare signals between a control and a suspect group for differences detection), without any *a priori* hypotheses on the structure of potential contaminants. Hence, only untargeted strategies based on tools from the field of metabolomics might enable the real “blind” detection of unknown or unexpected trace molecules in complex food samples. The main characteristic of such untargeted approaches lies in the generation of a very high number of signals (several thousand for a single sample). Therefore, powerful data analysis strategies must be set up to increase the probability to detect contaminated samples. Early results showed their potential in detecting unexpected compounds in food products (Inoue et al., 2015; Knolhoff, Zweigenbaum, & Croley, 2016; Kunzelmann, Winter, Åberg, Hellenäs, & Rosén, 2018; Tengstrand et al., 2013). However, food matrices studied remain relatively simple (orange juice, milk) with either high levels of contamination (near mg.kg^{-1}) (Tengstrand et al., 2013) or low molecular diversity of chemical contaminants (Inoue et al., 2015; Knolhoff et al., 2016). The latest published paper (Kunzelmann et al., 2018) shows promising results in terms of sensitivity (contamination detection down to $25 \mu\text{g/kg}$), but only focused on pesticides. More work is therefore needed to develop such untargeted strategies in the food safety field, especially considering even lower contamination levels (down to $10 \mu\text{g.kg}^{-1}$ as frequently required by the European regulation) and a wider contaminants diversity (including migrants from packaging and process-induced toxicants). To that end, the method proposed here relies on the combination of three tools to take full advantage of UHPLC-HRMS data: (i) data filtration based on univariate statistics, (ii) separation of sample groups and highlighting of discriminating ions using Independent Components Analysis (ICA), an unsupervised multivariate method based on source signals decomposition (Rutledge & Jouan-Rimbaud Bouveresse, 2015), (iii) automated data mining-tools to help the annotation of discriminating ions. Thus, our data analysis strategy combines the use of XCMS open-source R package (Smith, Want, O'Maille, Abagyan, & Siuzdak, 2006) and ICA method: to the best of our knowledge, this combination for MS data analysis in untargeted food safety analysis is

successfully performed for the very first time. Unless previous untargeted approaches reporting the use of either vendor (Knolhoff et al., 2016) or in-house tools (Tengstrand et al., 2013), that often work as “black boxes”, our approach benefits from using a freely available package that exists for more than 10 years and is supported by a dynamic and worldwide scientific community, which made it become very versatile for MS data analysis. On top of that, it became user friendly thanks to the development of free web-based platforms like XCMS-Online (Tautenhahn, Patti, Rinehart, & Siuzdak, 2012) or Workflow4Metabolomics (Giacomoni et al., 2015).

Tea has been chosen as the development foodstuff, for it is the most consumed hot beverage in the world (Chang, 2015). Moreover, in its raw product form (tea leaves), it is classified as a difficult commodity by the European Commission (European Commission - DG Health and Food Safety, 2015) which makes it very interesting as a methodological development food sample. In addition, tea is frequently produced under remote areas where agricultural and production practices may be less controlled than in Europe. Therefore this food product is the subject of frequent alerts on the European Rapid Alert System for Food and Feed, relative to non-authorized pesticides or contaminant levels above regulated limits (i.e. minimum 10 $\mu\text{g}\cdot\text{kg}^{-1}$ for most pesticides). Finally, recent metabolomics-like approaches have been reported on tea but they only focus on quality and authenticity issues (Fraser et al., 2013; Pongsuwan et al., 2008), i.e. on major constituent. The methodology that we propose here focuses on trace compounds, with specific analytical methods and data treatment strategies to be set up to achieve their detection.

2.2. MATERIAL AND METHODS

2.2.1. CHEMICALS AND REAGENTS

Acetonitrile (ACN) (HPLC plus gradient, LC/MS), water, methanol (MeOH) and formic acid (FA) (all LC/MS grade) were purchased from Carlo Erba. Ultrapure water (Milli-Q[®]) was produced by an Integral 3 water purification system from Millipore[®]. The compound used for ToF-MS calibration was Leucine Enkephalin (LC/MS grade), purchased from Waters[®].

Analytical standards solutions (100 $\mu\text{g}\cdot\text{mL}^{-1}$ in ACN or MeOH) for 21 pesticides, 4 mycotoxins, 2 process-induced toxicants and labelled compounds acrylamide-d3, dimethoate-d6 and malathion-d6 were purchased at CIL Cluzeau France. Ochratoxin-d5, bisphenols A, F and S, bisphenol A diglycidyl ether (BADGE), bisphenol F diglycidyl ether (BFDGE) and bisphenol A-d14 (purity > 99%) were provided by Sigma Aldrich (Saint-Quentin Fallavier, France). Two pooled stock solutions containing respectively all non-labelled molecules (each at 1 $\mu\text{g}\cdot\text{mL}^{-1}$), and all labelled molecules (each at 1 $\mu\text{g}\cdot\text{mL}^{-1}$) were prepared in ACN and stored in the fridge. Regularly, target analyzes of these solutions were done to check for their stability.

2.2.2. SAMPLE COLLECTION AND STUDY SET-UP

The goal of this work is to assess the ability of a workflow based on UHPLC-HRMS and chemometrics methods (including multivariate analysis) to blindly detect an unexpected contamination in a food

sample. To that end, a study was designed so that this workflow would face two very different situations: (i) a quite homogeneous product (i.e. samples from one brand) contaminated at several levels by a large number of molecules (development data set) (ii) a heterogeneous product (i.e. samples from two brands) contaminated at a single level by only few molecules (validation data set). Raw data sets have been deposited to the EMBL-EBI MetaboLights database (DOI: 10.1093/nar/gks1004. PubMed PMID: 23109552) with the respective identifiers MTBLS752 and MTBLS754 for data set n°1 (development) and n°2 (validation of the approach) (Haug et al., 2013). The complete data sets can be accessed at <https://www.ebi.ac.uk/metabolights/MTBLS752> and <https://www.ebi.ac.uk/metabolights/MTBLS754>.

Green teas were purchased at local retailers (Paris, France) and ground in our laboratory using a mortar and a pestle. Green tea n°1 is a Japanese Bancha tea, and green tea n°2 a Chinese tea. They were used to generate the development data set (green tea n°1) and the validation data set (green teas n°1 & 2).

Two spiking mixes were prepared. The first one (mix n°1), intended for the development data set, consists in a pool of 32 chemical contaminants (a detailed list of compounds used can be found in Table 3.3). These target molecules, called “tracers”, were chosen to be representative of potential contaminants, both in terms of chemical structures, source types (mycotoxins, pesticides, process-induced toxicants and migrants from packaging) and analytical behavior (instrumental response, peak width, retention time and adduct / isotopic information). The second one (mix n°2) consists in selected three contaminants from the previous list, chosen for their chemical diversity, namely ochratoxin A (OTA), bisphenol S (BPS) and tolfenpyrad.

For the development data set, four samples were considered (each time three sub-samples were collected to obtain triplicates of preparation): three samples spiked with mix n°1 at 10, 50 or 100 $\mu\text{g.kg}^{-1}$, and a control sample (i.e. spiked only with the ACN solvent). Tea samples were initially analyzed using a classical multi-residue method in order to check for the absence of the “tracers” considered (Cladière et al., 2018).

For the validation data set, four samples were considered as well (again three sub-samples were collected each time to get triplicates of preparation): two control samples respectively made of unspiked green tea n°1 and unspiked green tea n°2, and two suspect samples respectively composed of green tea n°1 or green tea n°2 spiked at a level of 50 $\mu\text{g.kg}^{-1}$ with mix n°2.

Spiking levels were chosen in accordance to EU regulation No 396/2005 and 1881/2006. Some of the least sensitive compounds (namely deoxynivalenol, bisphenols A and F) were spiked with a magnification factor of 5, and for the same reason acrylamide was spiked with a factor of 10. In addition, all sample groups were systematically spiked (at 40 $\mu\text{g.kg}^{-1}$) with the pool of labelled molecules for analytical quality control purpose. For spiking, samples of 1 g were weighted in centrifuge polypropylene tubes (Corning, New York, USA), and spiking was performed using the lowest possible

volume of solution (maximum of 100 μL). After spiking, samples were homogenized using a vortex and allowed to equilibrate for 2 hours at room temperature.

The workflow employed (both for sample preparation and data treatment) is shown in Figure 3.1.

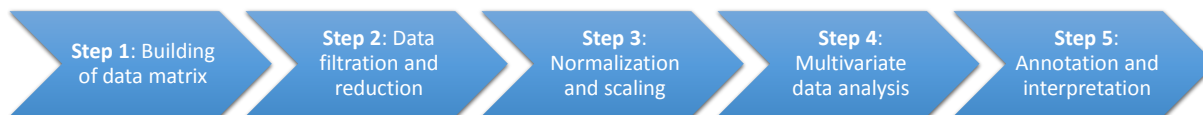


FIGURE 3.1 DATA PROCESSING STEPS

2.2.3. ANALYTICAL METHOD

The generic analytical method is based on previous work (Cladière et al., 2018). Tea samples were extracted using direct solvent extraction with 5 mL of an ACN/MeOH (90/10 v/v) mixture acidified with 0.1% FA, and tubes were agitated upside-down on an agitating plate during 1 h before centrifugation at 3,000 g for 10 min. The supernatant was then collected and an aliquot (2 mL) was evaporated to dryness under a gentle stream of nitrogen. The extract was further reconstituted in 0.2 mL of ACN acidified with 0.1% FA. Then 0.8 mL of ultrapure water with 0.1% FA was added in order to reconstitute 1 mL of final volume, and centrifuged at 12,000 g for 10 min. At the end, 0.5 mL of the final extract was sampled and filtered at 0.2 μm using a syringeless filter vial (mini-uniprep G2, Whatman) before analysis. A Quality Control (QC) sample for each data set was prepared by pooling together 0.2 mL of final extract from every sample of the set; an aliquot of 0.5 mL was then taken and filtrated at 0.2 μm using a syringeless filter vial.

Analyzes were performed on a Waters® Acquity UPLC® H-Class system, composed of a quaternary solvent manager pump, a refrigerated sample manager Flow-Through-Needle and a column oven, coupled to a Waters® high resolution Time-of-Flight mass spectrometer Xevo® G2-S ToF operated in centroid mode (UHPLC/HRMS-ToF). Separation was performed on a C18-PFP column (150 \times 2.1 mm, 2 μm particles diameter, ACE supplied by AIT France). An electrospray ionization source was used in both positive (ESI⁺) and negative (ESI⁻) modes. ESI⁺ and ESI⁻ modes were run separately. For ESI⁺, the mobile phase was composed of water (A) and ACN (B), both acidified with 0.1% FA, and MeOH (C), flowing at 0.4 mL.min⁻¹. Gradient started at 100% A and reached 100% B in 10 min, being kept for 6 min before switching to 100% C to rinse the system in 1 min, being hold for 5 min, returning back to 100% A in 1 min and finally equilibrating for 3 min, with a total run duration of 26 min. For ESI⁻, the mobile phase was composed of water buffered at pH 6.45 with 10 mM of ammonium formate (A) and MeOH (B) flowing at 0.3 mL.min⁻¹. The gradient started at 100% A and reached 100% B in 13 min, holding this condition for 7 min before turning back to 100% A in 1 min and finally equilibrating for 3 min, with a total run duration of 24 min. For both chromatographic methods the temperature of the column oven was kept at 30°C. Electrospray parameters have been fixed at the previously reported values (Cladière et al., 2018).

The analytical sequence started with injection of 10 mobile phase blanks in order to reach complete equilibrium of UHPLC-HRMS-ToF apparatus. Sample vials were randomized in the analytical sequence, and a blank as well as a QC sample were injected every 10 sample vials. For each ionization mode, all sub-samples were injected either in triplicate (development data set) or in quadruplicate (validation data set).

2.2.4. DATA TREATMENT

The data treatment workflow was set up to be as much automated as possible. Indeed, only few manual steps are remaining, the main one being the final curation of the automated annotation algorithm. Moreover, no information about the level or nature of contaminants are provided in the workflow, only the group information (ex: "vial n°1 belongs to group n°3"). The term "group" refers here to all injections related to the same sample (i.e. three replicates for sample preparation plus triplicates or quadruplicates of injections each time). In other word, for each sample several raw data are obtained, these being grouped together before data treatment.

- *Step 1: Building of data matrix from raw data*

Vendors (Waters®) raw data files were first converted to the open-source format mzXML using ProteoWizard (Chambers et al., 2012) and then uploaded onto the Workflow4Metabolomics (W4M) platform (Giacomini et al., 2015). Data matrix building was then achieved using open-source XCMS package (Smith et al., 2006) on this platform.

XCMS builds the data matrix from raw data files using the following workflow. First, "xcmsSet" with CentWave method (Tautenhahn, Bottcher, & Neumann, 2008) extracts peaks from the data files. Peaks are then grouped across the samples and aligned using "group", "retcor" and then "group" functions again. The final step of the algorithm, "fillpeaks", identifies for each sample the peaks for which this sample has no value: for these peaks, the tool integrates the signal noise in this area to avoid missing values at the end. XCMS parameter values for each step of the workflow were chosen as suggested for UHPLC-Q-ToF instruments by Patti *et al.* (Patti, Tautenhahn, & Siuzdak, 2013), except for the "peak width" parameter in the "xcmsSet" step which was chosen less stringent (5-60 s instead of 5-20 s) to limit data loss. A complete list of XCMS parameters can be found in Table 3.4. The XCMS peak extractions were performed separately for ESI⁺ and ESI⁻ sequences. Finally, the data matrix is a table gathering the different peak areas integrated by XCMS sorted by ions in row (combination of m/z and retention time) and by samples in column. Data matrix files (.txt) were then imported in Matlab using in-house scripts.

- *Step 2: Data filtration and reduction*

Since the number of output ions generated by XCMS is very high (between 10,000 and 30,000), the data matrices needed to be filtrated to remove as many irrelevant ions as possible. Data cleaning by

successive filtration steps is critical since an adequate filtration should enable to clean the data from irrelevant signals while avoiding or minimizing relevant chemical information loss. Data matrices were filtrated using only the group information (i.e. blind to the nature and levels of the spiked molecules). Therefore, ions that do not differ from the blanks or do not vary between samples were filtrated in order to keep only suspect ions and try to highlight a food contamination. This filtration strategy is commonly used in metabolomics approaches (Antignac et al., 2011) but applied here for the first time to non-targeted food safety analysis. It is generally based on statistical tests (t-test) designed to determine significant differences for each ion between samples at a commonly admitted p-value of 0.05 (Gika, Theodoridis, Plumb, & Wilson, 2014; Rubert et al., 2017; Thévenot, Roux, Xu, Ezan, & Junot, 2015). This strategy can be completed by using the fold change of each ion between samples: the common fold change value used is 2, but it is still under discussion for metabolomics purpose (Ortmayr, Charwat, Kasper, Hann, & Koellensperger, 2017).

Finally, filtration of data matrices composed initially of around 20,000 ions was done within three successive automated steps plus an automated pre-filtering step:

- a) Pre-filtering step: unusable and unreliable variables that exhibit a poor stability, meaning relative standard deviation (%RSD) on peak area above 100% in every sample group were discarded (about 50 to 100 ions discarded).
 - b) First step of filtration: removal of ions that show no significant difference (peak area) between blank runs and any of the sample groups using pairwise t-tests results (blanks vs. sample groups, about 19,000 ions remaining).
 - c) Second step of filtration: removal of ions that show no significant difference (peak area) between any sample groups using pairwise t-tests results (sample groups vs. sample groups, about 10,000 ions remaining).
 - d) Third step of filtration: removal of ions that show a low fold change among sample groups, to select only ions exhibiting high contrast between sample groups. For each ion, the median value of peak areas of each sample groups (n=9, extraction triplicates, each analyzed in triplicate) was considered. The fold change is then calculated by dividing the highest median value by the lowest one (assumed to be the most concentrated sample divided by the least concentrated one, about 1,000 ions remaining after this step).
- *Step 3: Normalization and scaling*

Missing values and algorithm artifacts such as zero, infinite and negative values in the data matrix were managed according to the guidelines given by Wherens et al. (Wehrens et al., 2016). Briefly, for each ion (named as “*m/z* – retention time” combination), any irrelevant value was replaced by the lowest value of this ion (blank excluded).

The data matrix was then log- and pareto- scaled (Antignac et al., 2011), and normalized using a median-based Probabilistic Quotient Normalization (PQN) (Dieterle, Ross, Schlotterbeck, & Senn, 2006) using the QC samples.

- *Step 4: Multivariate data analysis*

Principal Component Analysis (PCA) and Independent Component Analysis (ICA) were tested in order to discriminate contaminated samples. ICA showed better performance than PCA to resolve complex signal mixtures as already demonstrated for metabolomics data (Liu et al., 2016). Therefore, ICA was used to interpret and visualize data with respect to their source signals, and then try to evidence potential discrimination between sample groups. Indeed, ICA is a blind source separation method, which aims at extracting from mixed signals their original source signals as well as the weights in which they are mixed. Among the different few algorithms enabling to compute ICA models, the JADE algorithm was used here (Rutledge & Jouan-Rimbaud Bouveresse, 2015).

The determination of optimal number of Independent Components (ICs) to use is the key step during the building of an ICA model. This optimal number was determined using the random ICA method (Kassouf, Jouan-Rimbaud Bouveresse, & Rutledge, 2017), briefly summarized as follows: the data set is randomly split into two equivalent groups. ICA models with 1 to F components (here, $F = 20$) are calculated in each subset. For each model (i.e., each investigated number of ICs), correlations between all ICs from one subset and all ICs from the other subset are calculated. The idea underlying this procedure is that if an IC is significant, it should be extracted in each subset and therefore, strong correlations should be observed between ICs from each subset. Hence, one looks for the highest number of ICs for which each IC of one subset is highly correlated with one IC of the other subset. However, the repartition of samples into the two subsets being random, there is a possibility that the subsets are not representative, in which case a significant IC might be extracted from one subset only. This is the reason why this procedure has to be repeated, here, 50 times.

- *Step 5: Annotation and interpretation*

For automated annotation, an in-house broad-scale database was built combining data from several databases, namely the Toxin and Toxin-Target Database (T3DB, <http://www.t3db.ca/> (Wishart et al., 2015)), the literature (Gallart-Ayala, Núñez, & Lucci, 2013; Nielsen & Smedsgaard, 2003), and to a lesser extent, the Pesticides Properties Database (Lewis, Tzilivakis, Warner, & Green, 2016).

After evidencing a discrimination along one component of the ICA model, the signal matrix, which gives the weight of each ion (“ m/z – retention time” pair) in the component, was analyzed: ions were sorted by descending contribution value along the components explaining the group separation, and the annotation was performed according to the following strategy:

- a) In-house automated tools were developed for isotopic pattern detection (inspired from work by Cotton et al. (Cotton et al., 2014)) and then for in-house toxicant database annotation to highlight suspect ions;
- b) Manual curation of the annotation results was then performed for discriminating ions using information provided by step a) as well as online databases [such as Metlin (<https://metlin.scripps.edu/>), HMDB (<http://www.hmdb.ca/>), mzCloud (<https://www.mzcloud.org/>) and T3DB] and raw data visualization when necessary.

With mass spectrometry, and especially electrospray ionization, a single molecule usually produces several observed signals, either fragments, adducts or isotopic peaks. All ions assumed as coming from the same compound (i.e. retention time, correlation, known $\Delta m/z$: M+1, M+2 with relevant intensity ratio) were grouped in “features” during step a), each one representing a single compound. During the same step, adducts were annotated with the database search. Annotation levels nomenclature used is based on guidelines proposed by Sumner et. al. (Sumner et al., 2007).

2.3. RESULTS AND DISCUSSION

2.3.1. MULTIVARIATE ANALYSIS OF DEVELOPMENT DATA SET

All data treatments were developed and performed blindly, meaning without optimizing the parameters for our tracers. The objective of this approach is to evaluate the efficiency of a generic blind untargeted analysis based on multivariate tools to discriminate contaminated samples and annotate ions of potential contaminants.

A multivariate exploration of the data was first tried without any (pre)filtration, but it remained unsuccessful since no clear group separation could be observed (see Figure 3.2). The detailed filtration process was thus developed and applied. It appears (Supplementary material - Figure 3.4) that, even though PCA enables a discrimination for the filtrated data, the one given by ICA is superior both by its quality (better sample separation) and its ability to align chemical phenomenon on a single component. So, the detection of suspect samples and ions is eased thanks to ICA by simply sorting ions based on their weight on the discriminating component.

2.3.1.1. BLIND DISCRIMINATION OF SAMPLES BY ICA

Optimal number of ICs was determined as 4 for the development data set in both positive and negative modes with the random ICA method. For each ionization mode, score plots were drawn considering the different ICs prone to discriminate sample groups (an illustrative plot is given in Figure 3.2). Interestingly, all sample groups could be discriminated whatever the ionization mode, and each time IC1 was determined as the most probable meaningful component regarding group information. It is clear that group separation along IC1 is related to the level of contamination. It should be emphasized that tea samples contaminated at the low level ($10 \mu\text{g}\cdot\text{kg}^{-1}$) could be distinguished from control tea samples, even with a high chemical diversity of contaminants. This is the first time that contaminated food

samples are discriminated from control samples at this level: this opens new perspectives for food safety control, since $10 \mu\text{g.kg}^{-1}$ is the maximum level authorized for several regulated chemicals, especially pesticides.

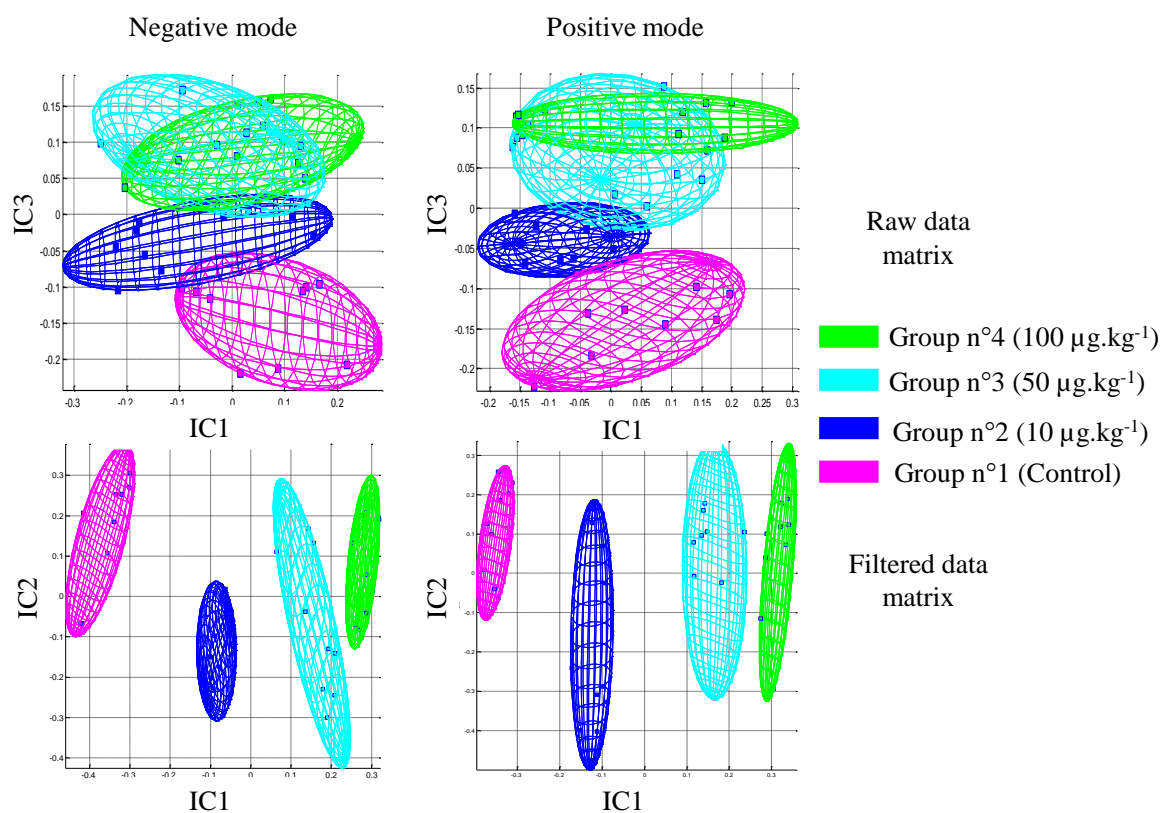


FIGURE 3.2 ICA SCORE PLOTS FOR FILTERED AND UNFILTERED DATA MATRIX FOR BOTH IONIZATION MODES ON THE DEVELOPMENT DATA SET

The next step is the annotation of discriminating ions, in order to assess if the discrimination observed is really due to the contaminants.

2.3.1.2. ANNOTATION AND INTERPRETATION OF ICA OUTPUT

Most of the annotation process was automated thanks to database search and data mining scripts. For each ionization mode, the filtered data matrix went through two automated steps: (i) isotopic peaks were first grouped together, and (ii) observed ions were searched through a broad-range toxicants database for testing their potential matching with different adducts. Then, for each ionization mode, ions were sorted along the discriminating component(s) and extracted with their information (m/z , retention time, presence of isotopes, and potential match in the database for different adducts). At the end, the results can be quickly curated by the user, who can then rapidly spot suspect samples and ions. These information can be completed, when needed, by a manual exploration of raw data files.

That way, for positive mode, the 69 first discriminating ions were putatively annotated or characterized and grouped into 20 “features” (see Table 3.1). Over those 20 “features”, 14 were attributed to our “tracers”. Three others (#13, 18 and 19) were not expected to be present in the samples (since they were

not found after a targeted analysis of the control samples, nor reported in the analysis certificate of standards used). Raw chemical formulas could be proposed, that show very strong similarities with some of our “tracers”, so they were putatively annotated as impurities from the initial standard solution (such result has been confirmed by an *a posteriori* classical targeted analysis of the standard solution). This clearly underlines the potential of our developed method to detect unexpected compounds at trace levels since those impurities were not expected before the analysis. Three other features remained unknown.

Putative annotation and characterization of the 69 first ions in negative mode was also achieved, and these ions were grouped into 14 “features” (see Table 3.2). For these features, raw molecular formulas hypotheses were made based on information given by the automated annotation step, enabling a putative characterization to be achieved. Thus, 12 of our “tracers” could be recovered. Two “features” (corresponding to eight ions) remained unresolved after annotation attempt (using both automated scripts and manual exploration of raw data), but they were characterized as being halogenated compounds thanks to the isotopic peaks found during the automated data mining step.

2.3.2. METHOD PERFORMANCE

Performance was assessed based on blind detection rates of “tracer” contaminants. In positive mode, 44% of our “tracers” were successfully putatively annotated, and 38% in negative mode. When considering both modes, the overall detection rate is 66% (since some molecules were detected in both modes, e.g. diuron and ochratoxin A). By comparing with detection performance of a dedicated targeted multi-residue method on the same samples (Cladière et al., 2018), it appears that the “tracers” not successfully annotated using the untargeted approach were also the most difficult to analyze with a targeted approach (i.e. showing high ion suppression and therefore low signal/noise ratio, high relative standard deviation of the signals and poor recoveries). A manual exploration of the raw chromatograms reveals that these molecules give very noisy peaks, which are not even extracted during the pretreatment step with XCMS. In fact, to date, it is likely that there is no algorithm that can achieve exhaustive peak extraction from raw data in untargeted LC-MS study (Coble & Fraga, 2014). In our case, XCMS managed to extract 75% of our “tracers” (i.e. 24 over 32) from the raw data files, which is still a good score even though the noisiest peaks are missed. We tested the only existing optimization algorithm for XCMS parameters (IPO) (Libiseller et al., 2015), but with no improvement. Unsurprisingly, the response factor of a compound has been found to be the main factor affecting its detectability.

Thus, the data treatment methodology applied after XCMS treatment (including filtration of the data matrix, preprocessing and multivariate analysis coupled with data mining) successfully annotated 21 tracers over the 24 extracted by XCMS (i.e. 88%). This is very satisfactory regarding the wide diversity of molecules studied, both in terms of chemical structure and response factor in LC-MS, as well as regarding the trace levels studied.

TABLE 3.1 DISCRIMINATING FEATURES FOR POSITIVE IONIZATION MODE AND RESPECTIVE PUTATIVE ANNOTATION

POSITIVE IONIZATION MODE									
Feature # ^a	Cumulative weight of the feature	Number of ions in the feature	Experimental exact mass of adduct ^b	Adduct	Proposed raw formula	Proposed putative annotation	Mono-isotopic mass ^c	Delta (ppm)	Estimated LOD ($\mu\text{g}\cdot\text{kg}^{-1}$)
1	23.76	8	331.0435	[M+H] ⁺	C ₁₀ H ₁₉ O ₆ PS ₂	Malathion	330.0361	0.48	1.1
2	20.44	12	607.2926	[M+H] ⁺	N/A ^d	Unknown	N/A	N/A	N/A
3	18.32	7	230.0076	[M+H] ⁺	C ₅ H ₁₂ NO ₃ PS ₂	Dimethoate	228.9996	3.04	0.9
4	13.83	5	233.0248	[M+H] ⁺	C ₉ H ₁₀ Cl ₂ N ₂ O	Diuron	232.0170	2.32	1.0
5	12.22	5	404.0894	[M+H] ⁺	C ₂₀ H ₁₈ ClNO ₆	Ochratoxin A	403.0823	-0.34	4.9
6	9.37	4	229.0416	[M+K] ⁺	C ₇ H ₁₄ N ₂ O ₂ S	Aldicarb	190.0776	4.41	15.1
7	9.23	3	311.0398	[M+H] ⁺	C ₁₄ H ₉ ClF ₂ N ₂ O ₂	Diflubenzuron	310.0321	1.45	1.9
8	9.09	3	220.9537	[M+H] ⁺	C ₄ H ₇ Cl ₂ O ₄ P	Dichlorvos	219.9459	2.47	3.6
9	8.56	3	384.1471	[M+H] ⁺	C ₂₁ H ₂₂ ClN ₃ O ₂	Tolfenpyrad	383.1401	-0.67	1.3
10	8.00	3	228.1283	[M+H] ⁺	C ₉ H ₁₇ N ₅ S	Ametryn	227.1205	2.59	1.5
11	7.86	3	216.1010	[M+H] ⁺	C ₈ H ₁₄ ClN ₅	Atrazine	215.0938	-0.14	2.9
12	6.01	2	306.1041	[M+H] ⁺	C ₁₁ H ₂₀ N ₃ O ₃ PS	Pirimiphos methyl	305.0963	1.56	0.4
13	4.69	2	263.0243	[M+H] ⁺	C ₁₁ H ₁₂ Cl ₂ O ₃	<i>2,4-D Isopropyl Ester</i> ^e	<i>262.0163</i>	<i>2.73</i>	<i>0.5</i>
14	4.69	1	256.0604	[M+H] ⁺	C ₉ H ₁₀ ClN ₅ O ₂	Imidacloprid	255.0523	3.19	16.3
15	4.36	2	621.2713	N/A	N/A	Unknown	N/A	N/A	N/A
16	4.29	2	623.2868	N/A	N/A	Unknown	N/A	N/A	N/A
17	2.73	1	251.0380	[M+H] ⁺	C ₁₂ H ₁₀ O ₄ S	Bisphenol S	250.0300	3.08	12.1
18	2.24	1	202.0855	[M+H] ⁺	C ₇ H ₁₂ ClN ₅	<i>Simazine</i> ^e	<i>201.0781</i>	<i>0.68</i>	<i>0.7</i>
19	2.03	1	182.1282	[M+NH ₄] ⁺	C ₉ H ₁₂ N ₂ O	<i>Fenuron</i> ^e	<i>164.0950</i>	<i>-3.57</i>	<i>0.9</i>
20	1.69	1	335.1254	[M+Na] ⁺	C ₁₉ H ₂₀ O ₄	Bisphenol F diglycidyl Ether	312.1362	0.12	57.0

^a Features sorted by descending cumulative weight of ions in the discriminating IC

^b Mass measured for the ion having the highest weight in the discriminating IC

^c Electron mass used: $5.485 \cdot 10^{-4}$ Da

^d Not Applicable

^e Found in the spiking mix, may be considered as impurities

TABLE 3.2 DISCRIMINATING FEATURES FOR NEGATIVE IONIZATION MODE AND RESPECTIVE PUTATIVE ANNOTATION

NEGATIVE IONIZATION MODE									
Feature # ^a	Cumulative weight of the feature	Number of ions in the feature	Experimental exact mass of adduct ^b	Adduct	Proposed raw formula	Proposed putative annotation	Mono-isotopic mass ^c	Delta (ppm)	Estimated LOD ($\mu\text{g.kg}^{-1}$)
1	27.50	10	266.9385	[M-H]-	C ₉ H ₇ Cl ₃ O ₃	Fenoprop	267.9461	-1.02	2.0
2	22.07	8	252.9227	[M-H]-	C ₈ H ₅ Cl ₃ O ₃	2,4,5-T	253.9304	-1.90	3.0
3	21.62	8	309.0249	[M-H]-	C ₁₄ H ₉ ClF ₂ N ₂ O ₂	Diflubenzuron	310.0321	0.40	2.6
4	16.59	6	231.0091	[M-H]-	C ₉ H ₁₀ Cl ₂ N ₂ O	Diuron	232.0170	-2.97	0.9
5	14.32	8	232.9771	[M-H]-	C ₉ H ₈ Cl ₂ O ₃	Dichlorprop	233.9850	-2.75	5.9
6	13.48	5	213.0313	[M-H]-	C ₁₀ H ₁₁ ClO ₃	MCPD	214.0397	-4.98	2.5
7	11.68	5	199.0152	[M-H]-	C ₉ H ₉ ClO ₃	MCPA	200.024	-7.72	5.2
8	8.05	6	204.9217	N/A	N/A	Unknown chlorinated	N/A	N/A	N/A
9	7.75	3	249.0223	[M-H]-	C ₁₂ H ₁₀ O ₄ S	Bisphenol S	250.0300	-1.53	6.0
10	5.37	2	239.0668	[M-H]-	C ₁₀ H ₁₂ N ₂ O ₅	Dinoseb	240.0746	-2.08	10.4
11	4.61	2	402.0749	[M-H]-	C ₂₀ H ₁₈ ClNO ₆	Ochratoxin A	403.0823	-0.13	10.3
12	3.88	2	254.0444	[M-H]-	C ₉ H ₁₀ ClN ₅ O ₂	Imidacloprid	255.0523	-2.29	25.9
13	2.62	2	218.9611	[M-H]-	C ₈ H ₆ Cl ₂ O ₃	2,4-D	219.9694	-4.74	15.5
14	2.36	2	771.1431	N/A	N/A	Unknown chlorinated	N/A	N/A	N/A

Last but not least, the detection and putative annotation of unexpected impurities coming from the spiking mix highlights the ability of our untargeted approach to detect potentially unknown or unexpected trace contaminants in food products and to propose the user annotation hypotheses. It should be spotted that a molecule generating a high number of ions (adducts, isotopes, fragments, etc.) will be more easily annotated than a molecule generating only few signals.

For each annotated compound, a limit of detection (LOD) was estimated based on the calculated fold change (calculated between the group with the highest level -100 $\mu\text{g.kg}^{-1}$ - and the control group). Briefly, a rule of three was made to figure out what concentration would lead to a fold change of 3, which is the most commonly used signal/noise ratio for LOD determination. For compounds annotated as impurities, assumption was made that they come from standards of similar families (i.e. atrazine for simazine, diflufenzuron or diuron for fenuron, and acid herbicides for 2,4-D isopropyl ester) and a LOD was then estimated for each by taking the standard purity into account. As shown in Table 3.1 and Table 3.2, estimated LODs are relevant against EU regulation since they are in the range 10 $\mu\text{g.kg}^{-1}$ or below for almost every annotated compound. Again, our untargeted approach proves to have quite similar performance in terms of sensitivity as compared to our dedicated targeted multi-residue method (Cladière et al., 2018), having the additional asset to detect unexpected molecules.

2.3.3. APPLICATION ON VALIDATION DATA SET

The developed methodology (including filtration parameters, pretreatment steps and multivariate method) was blindly applied to the validation data set, obtained based on the analysis of two different types of green tea, either non-spiked (controls) or spiked with a mix composed of only three contaminants in order to offer a much more challenging discrimination between blanks and contaminated samples.

Data from the two types of tea were treated simultaneously. Data matrix was filtrated using same parameters as for development set, and then the optimal number of ICs was determined for this data set. It was calculated as 6 both for positive and negative modes. Unsurprisingly, since the data set is more heterogeneous, the filtration led to a smaller reduction of ion number than for the development data set. Still, the number of ions dropped from 23,391 and 17,269 (respectively for positive and negative mode) to 9,789 and 9,409 thanks to the filtration. For each ionization mode, two discriminating ICs were clearly observed (Figure 3.3). IC1 separates samples based on their brand, and IC6 separates control and contaminated samples. It should be emphasized that in ICA, ICs are not ordered by descending contribution like in PCA, meaning in other words that IC1 does not necessarily explain more variability than IC6.

Annotation of the data matrix was done as described for the development data set. IC6 was determined as bearing the separation due to contaminants thank to information provided in the automated annotation step.

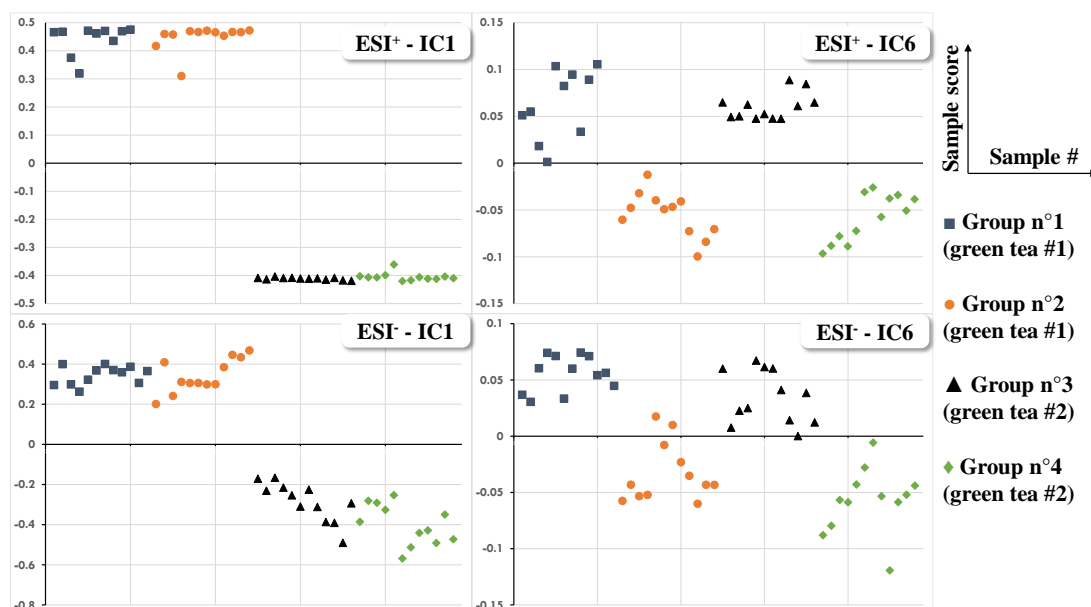


FIGURE 3.3 SCORE PLOTS OF DISCRIMINATING ICs (BOTH IONIZATION MODES) FOR THE VALIDATION DATA SET

The three “tracers” were successfully annotated within the first 10 ions of IC6 with this methodology in at least one ionization mode. This result is of prime interest in our case since ICA, as employed here, shows its main assets which is to separate independent phenomenon. Indeed, we can see that in our case, the “natural” variability of the product is well separated from the variability brought by the spiking (these two phenomenon are likely to be mixed in less powerful methods). Moreover, the number of ions generated in a complex data set such as the validation one is very high despite of filtration strategy applied (in our case ~9,000 per ionization mode). Therefore, the use of multivariate methods enables the reduction of the dimensionality of the data and the achievement of suspect samples and ions detection. Thanks to ICA, the annotation of only 10 ions per ionization mode was sufficient to underscore a contamination of tea.

This highlights the ability of ICA to resolve complex signal mixtures and simplify annotation of relevant ions, even in cases where the information is borne by few, low intensity ions. Our proposed approach thus has a strong potential in detecting food contaminants at low levels in complex and rather heterogeneous data sets. Its applicability to other food matrices should be feasible if reference samples are available.

2.4. CONCLUSION

This work shows some important methodological features for untargeted approach development for food chemical contaminants detection. It gives evidence that the blind untargeted detection of contaminants in complex food matrices is feasible thanks to high resolution methods coupled to powerful data analysis strategies. A widely spread, well-known, freely available and easy to use tool (i.e. XCMS run on W4M

platform) was used for peaks extraction from raw data. Then, an efficient automated strategy was set up for data filtration, using well-known easy to use tools (t-tests and fold change). Samples were separated using a multivariate method (ICA), and discriminating ions putatively annotated by the help of automated data mining methods.

Thanks to this strategy, 66% of the “tracers” considered were successfully putatively annotated. This detection rate rises to 88% if brought back to tracers actually in the data matrix (after the peak extraction step). This shows the power of our developed data treatment strategy to detect potential food contaminants in a data matrix. In addition to the known “tracers”, some unexpected molecules were detected and putatively annotated in the samples, which clearly highlights the potential of this approach. Method LODs were roughly estimated for each putatively annotated compound (both expected and unexpected), with values below or near 10 $\mu\text{g.kg}^{-1}$ for most of them, which compares favorably with a targeted multi-residue method. The approach, developed on a rather simple case, has been validated on a more complex and realistic situation, where the contamination is brought by a low number of molecules, and in which different brands of the same food product are considered simultaneously. This study opens new perspectives in the development of truly untargeted approaches based on tools and strategies from metabolomics (particularly HRMS and chemometrics) for food chemical safety assessment. Such approaches may constitute, in a near future, a major complement to targeted methods in a view of rapidly screening possibly contaminated food products. A next step towards a routine use of these approaches would be to implement them on even more complex cases like the following of a production batch. Interestingly, since our developed methodology is rather generic, it could be applied with only few development on any other UHPLC-HRMS data set, or even to other applications such as origin or authenticity issues to complement existing approaches. As a conclusion, no doubt that these results will encourage new developments on this analytical issue, both on the methodology and the tools, especially on the improvement of existing peak extraction methods or the handling of the product variability.

2.5.FUNDING

This work was supported by Paris Institute of Technology for Life, Food and Environmental Sciences (AgroParisTech), the French National Institute for Agricultural Research (INRA) and the French Ministry of Higher Education and Research.

2.6.ACKNOWLEDGEMENTS

The authors wish to thank Céline Dalle for her advices on W4M platform and Even Leroux for his technical help with MS instrumentation.

2.7. REFERENCES

- 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>
- Castro-Puyana, M., Pérez-Míguez, R., Montero, L., & Herrero, M. (2017). Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability. *TrAC - Trends in Analytical Chemistry*, 93, 102–118. <https://doi.org/10.1016/j.trac.2017.05.004>
- Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., ... Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, 30(10), 918–920. <https://doi.org/10.1038/nbt.2377>
- Chang, K. (2015). World tea production and trade Current and future development. Food and Agriculture Organisation.
- Cifuentes, A. (2009). Food analysis and foodomics. *Journal of Chromatography A*, 1216(43), 7109. <https://doi.org/10.1016/j.chroma.2009.09.018>
- Cladière, M., Delaporte, G., Le Roux, E., & Camel, V. (2018). Multi-class analysis for simultaneous determination of pesticides, mycotoxins, process-induced toxicants and packaging contaminants in tea. *Food Chemistry*, 242, 113–121. <https://doi.org/10.1016/j.foodchem.2017.08.108>
- 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. <https://doi.org/10.1016/j.chroma.2014.06.100>
- Cotton, J., Leroux, F., Broudin, S., Marie, M., Corman, B., Tabet, J. C., ... Junot, C. (2014). High-resolution mass spectrometry associated with data mining tools for the detection of pollutants and chemical characterization of honey samples. *Journal of Agricultural and Food Chemistry*, 62(46), 11335–11345. <https://doi.org/10.1021/jf504400c>
- Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Analytical Chemistry*, 78(13), 4281–4290. <https://doi.org/10.1021/ac051632c>
- European Commission - DG Health and Food Safety. (2015). Guidance document on analytical quality control and method validation procedures for pesticides residues analysis in food and feed (Vol. 11945). Retrieved from https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_mrl_guidelines_wrkdoc_11945.pdf
- Fraser, K., Lane, G. A., Otter, D. E., Hemar, Y., Quek, S. Y., Harrison, S. J., & Rasmussen, S. (2013). Analysis of metabolic markers of tea origin by UHPLC and high resolution mass spectrometry. *Food Research International*, 53(2), 827–835. <https://doi.org/10.1016/j.foodres.2012.10.015>
- Gallart-Ayala, H., Núñez, O., & Lucci, P. (2013). Recent advances in LC-MS analysis of food-packaging contaminants. *TrAC - Trends in Analytical Chemistry*, 42, 186–204. <https://doi.org/10.1016/j.trac.2012.09.017>
- Giacomini, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., ... Caron, C. (2015). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics*, 31(9), 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>
- Gika, H. G., Theodoridis, G. A., Plumb, R. S., & Wilson, I. D. (2014). Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *Journal of Pharmaceutical and Biomedical Analysis*, 87, 12–25. <https://doi.org/10.1016/j.jpba.2013.06.032>
- Gómez-Ramos, M. M., García-Valcárcel, A. I., Tadeo, J. L., Fernández-Alba, A. R., & Hernando, M. D. (2016). Screening of environmental contaminants in honey bee wax comb using gas

- chromatography–high-resolution time-of-flight mass spectrometry. *Environmental Science and Pollution Research*, 23(5), 4609–4620. <https://doi.org/10.1007/s11356-015-5667-0>
- Gosetti, F., Mazzucco, E., Gennaro, M. C., & Marengo, E. (2016). Contaminants in water: non-target UHPLC/MS analysis. *Environmental Chemistry Letters*, 14(1), 51–65. <https://doi.org/10.1007/s10311-015-0527-1>
- Haug, K., Salek, R. M., Conesa, P., Hastings, J., De Matos, P., Rijnbeek, M., ... Steinbeck, C. (2013). MetaboLights - An open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Research*, 41(D1), 781–786. <https://doi.org/10.1093/nar/gks1004>
- Inoue, K., Tanada, C., Sakamoto, T., Tsutsui, H., Akiba, T., Min, J. Z., ... Toyo'Oka, T. (2015). Metabolomics approach of infant formula for the evaluation of contamination and degradation using hydrophilic interaction liquid chromatography coupled with mass spectrometry. *Food Chemistry*, 181, 318–324. <https://doi.org/10.1016/j.foodchem.2015.02.117>
- Kassouf, A., Jouan-Rimbaud Bouveresse, D., & Rutledge, D. N. (2017). Determination of the Optimal Number of Components in Independent Components Analysis. *Talanta*, 179, 538–545. <https://doi.org/10.1016/j.talanta.2017.11.051>
- 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>
- Knolhoff, A. M., Zweigenbaum, J. A., & Croley, T. R. (2016). Nontargeted Screening of Food Matrices: Development of a Chemometric Software Strategy to Identify Unknowns in Liquid Chromatography-Mass Spectrometry Data. *Analytical Chemistry*, 88(7), acs.analchem.5b04208. <https://doi.org/10.1021/acs.analchem.5b04208>
- 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*, 1–10. <https://doi.org/10.1007/s00216-018-1028-4>
- Lehotay, S. J., Sapozhnikova, Y., & Mol, H. G. J. (2015). Current issues involving screening and identification of chemical contaminants in foods by mass spectrometry. *TrAC - Trends in Analytical Chemistry*, 69, 62–75. <https://doi.org/10.1016/j.trac.2015.02.012>
- Lewis, K. A., Tzilivakis, J., Warner, D. J., & Green, A. (2016). An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment*, 22(4), 1050–1064. <https://doi.org/10.1080/10807039.2015.1133242>
- Libiseller, G., Dvorzak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., ... Magnes, C. (2015). IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics*, 16(1), 118. <https://doi.org/10.1186/s12859-015-0562-8>
- Liu, Y., Smirnov, K., Lucio, M., Gougeon, R. D., Alexandre, H., & Schmitt-Kopplin, P. (2016). MetICA: Independent component analysis for high-resolution mass-spectrometry based non-targeted metabolomics. *BMC Bioinformatics*, 17(1), 114. <https://doi.org/10.1186/s12859-016-0970-4>
- 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](https://doi.org/10.1016/S0021-9673(03)00490-4)
- Ortea, I., Pascoal, A., Cañas, B., Gallardo, J. M., Barros-Velázquez, J., & Calo-Mata, P. (2012, August). Food authentication of commercially-relevant shrimp and prawn species: From classical methods to Foodomics. *Electrophoresis*. <https://doi.org/10.1002/elps.201100576>

- Ortmayr, K., Charwat, V., Kasper, C., Hann, S., & Koellensperger, G. (2017). Uncertainty budgeting in fold change determination and implications for non-targeted metabolomics studies in model systems. *The Analyst*, 142(1), 80–90. <https://doi.org/10.1039/C6AN01342B>
- Patti, G. J., Tautenhahn, R., & Siuzdak, G. (2013). Meta-Analysis of Untargeted Metabolomic Data: Combining Results from Multiple Profiling Experiments. *Nature Protocols*, 7(3), 508–516. <https://doi.org/10.1038/nprot.2011.454>. Meta-Analysis
- Pongsuwan, W., Bamba, T., Harada, K., Yonetani, T., Kobayashi, A., & Fukusaki, E. (2008). High-throughput technique for comprehensive analysis of Japanese green tea quality assessment using ultra-performance liquid chromatography with time-of-flight mass spectrometry (UPLC/TOF MS). *Journal of Agricultural and Food Chemistry*, 56(22), 10705–10708. <https://doi.org/10.1021/jf8018003>
- Rubert, J., Righetti, L., Stranska-Zachariasova, M., Dzuman, Z., Chrpova, J., Dall'Asta, C., & Hajslova, J. (2017). Untargeted metabolomics based on ultra-high-performance liquid chromatography–high-resolution mass spectrometry merged with chemometrics: A new predictable tool for an early detection of mycotoxins. *Food Chemistry*, 224, 423–431. <https://doi.org/10.1016/j.foodchem.2016.11.132>
- Rutledge, D. N., & Jouan-Rimbaud Bouveresse, D. (2015). Corrigendum to “Independent Components Analysis with the JADE algorithm.” *TrAC - Trends in Analytical Chemistry*, 67, 220. <https://doi.org/10.1016/j.trac.2015.02.001>
- Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), 779–787. <https://doi.org/10.1021/ac051437y>
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., ... Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3(3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- Tautenhahn, R., Bottcher, C., & Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, 9, 16. <https://doi.org/10.1186/1471-2105-9-504>
- Tautenhahn, R., Patti, G. J., Rinehart, D., & Siuzdak, G. (2012). XCMS online: A web-based platform to process untargeted metabolomic data. *Analytical Chemistry*, 84(11), 5035–5039. <https://doi.org/10.1021/ac300698c>
- 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>
- Thévenot, E. A., Roux, A., Xu, Y., Ezan, E., & Junot, C. (2015). Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *Journal of Proteome Research*, 14(8), 3322–3335. <https://doi.org/10.1021/acs.jproteome.5b00354>
- Wehrens, R., Hageman, J. A., van Eeuwijk, F., Kooke, R., Flood, P. J., Wijnker, E., ... de Vos, R. C. H. (2016). Improved batch correction in untargeted MS-based metabolomics. *Metabolomics*, 12(5), 88. <https://doi.org/10.1007/s11306-016-1015-8>
- Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., ... Rappaport, S. M. (2015). T3DB: The toxic exposome database. *Nucleic Acids Research*, 43(D1), D928–D934. <https://doi.org/10.1093/nar/gku1004>

2.8. SUPPLEMENTARY MATERIALS

TABLE 3.3 INFORMATION ON "TRACER" CONTAMINANTS USED IN THIS STUDY

N°	Name	Class	Chemical family	log K _{ow}	Ionization mode	Adduct - ESI+/ESI-
1	(2,4-dichlorophenoxy)acetic acid <i>2,4-D</i>	Pesticide (herbicide)	Auxinic herbicide	-0.82	+/-	[M+H] ⁺ /[M-H] ⁻
2	(4-chloro-2-methylphenoxy)acetic acid <i>MCPA</i>	Pesticide (herbicide)	Auxinic herbicide	-0.81	-	ND/[M-H] ⁻
3	2-(4-chloro-2-methylphenoxy)propanoic acid <i>MCPP</i>	Pesticide (herbicide)	Auxinic herbicide	-0.19	-	ND/[M-H] ⁻
4	2,4,5-trichlorophenoxyacetic acid <i>2,4,5-T</i>	Pesticide (herbicide)	Auxinic herbicide	2.88	-	ND*/[M-H] ⁻
5	4-(2,4-dichlorophenoxy)butanoic acid <i>2,4-DB</i>	Pesticide (herbicide)	Auxinic herbicide	1.35	+/-	[M+H] ⁺ /Frag**
6	Acetamiprid	Pesticide (insecticide)	Neonicotinoid	0.8	+	[M+H] ⁺ /ND
7	Acrylamide	Process-induced	Amide	-0.67	+	[M+H] ⁺ /ND
8	Aldicarb	Pesticide (acaricide)	Carbamate	1.15	+	[M+Na] ⁺ /ND
9	Ametryn	Pesticide (herbicide)	Triazine	2.63	+	[M+H] ⁺ /ND
10	Atrazine	Pesticide (herbicide)	Triazine	2.7	+	[M+H] ⁺ /ND
11	Bisphenol A	Migrant from packaging	Bisphenol	3.3	-	ND/[M-H] ⁻
12	Bisphenol A diglycidyl ether <i>BADGE</i>	Migrant from packaging	Diglycidyl ether	3.84	+	[M+Na] ⁺ /ND
13	Bisphenol F	Migrant from packaging	Bisphenol	1.65	-	ND/[M-H] ⁻
14	Bisphenol F diglycidyl ether <i>BFDGE</i>	Migrant from packaging	Diglycidyl ether	Not available	+	[M+Na] ⁺ /ND
15	Bisphenol S	Migrant from packaging	Bisphenol	2.91	+/-	[M+H] ⁺ /[M-H] ⁻
16	Deoxynivalenol	Mycotoxin	Trichothecene	0.29	+/-	[M+H] ⁺ /[M-H] ⁻
17	Dichloprop	Pesticide (herbicide)	Auxinic herbicide	2.29	-	ND/[M-H] ⁻
18	Dichlorvos	Pesticide (acaricide)	Organochlorinated	1.9	+	[M+H] ⁺ /ND
19	Diflubenzuron	Pesticide (insecticide)	Benzoylurea	3.89	+/-	[M+H] ⁺ /[M-H] ⁻

Chapitre 3 – Développement de l'approche non-ciblée

20	Dimethoate	Pesticide (acaricide)	Organophosphate	0.7	+	[M+H] ⁺ /ND
21	Dinoseb	Pesticide (herbicide)	Dinitrophenol	2.29	-	ND/[M-H] ⁻
22	Diuron	Pesticide (herbicide)	Phenylurea	2.87	+/-	[M+H] ⁺ /[M-H] ⁻
23	Fenoprop 2,4,5-TP	Pesticide (herbicide)	Auxinic herbicide	2.84	-	ND/Frag**
24	Fumonisin B1	Mycotoxin	Fumonisin	-0.5	+	[M+H] ⁺ /ND
25	Fumonisin B2	Mycotoxin	Fumonisin	1.2	+	[M+H] ⁺ /ND
26	Hydroxymethylfurfural	Process-induced	Furan	-0.09	+	[M+H] ⁺
27	Imidacloprid	Pesticide (insecticide)	Neonicotinoid	0.57	+/-	[M+H] ⁺ /[M-H] ⁻
28	Malathion	Pesticide (insecticide)	Organophosphate	2.75	+	[M+H] ⁺ /ND
29	Ochratoxin A	Mycotoxin	Ochratoxin	4.74	+/-	[M+H] ⁺ /[M-H] ⁻
30	Pirimiphos-methyl	Pesticide (insecticide)	Organophosphate	3.9	+	[M+H] ⁺ /ND
31	Propargite	Pesticide (acaricide)	Organosulfite	5.7	+	Frag**/ND
32	Tolfenpyrad	Pesticide (insecticide)	Pyrazole	5.61	+/-	[M+H] ⁺ /[M-H] ⁻

*ND: Not Detected

** Frag=Fragment

TABLE 3.4 FULL PARAMETERS AND THEIR CORRESPONDING VALUES FOR PEAK EXTRACTION USING XCMS

Step	Parameter	Value
xcmsSet	scanrange	180-2400
	nSlaves	1
	method	centWave
	ppm	15
	peakwidth	5-60
	mzdiff	-0.001
	snthresh	10
	integrate	1
	noise	0
	prefilter	0
group - A	method	density
	minfrac	0.5
	bw	2
	mzwid	0.015
	sleep	0.001
retcor	method	peakgroups
	smooth	loess
	extra	1
	missing	1
	span	0.2
	family	gaussian
group - B	method	density
	minfrac	0.5
	bw	2
	mzwid	0.015
	sleep	0.001
	max	50
fillPeaks	method	chrom
	convertRTMinute	FALSE
	numDigitsMZ	4
	numDigitsRT	2
	intval	into
annotatediff	nSlaves	4
	sigma	6
	perfwhm	0.6
	ppm	15
	mzabs	0.015
	maxcharge	1
	maxiso	4
	minfrac	0.5
	quick	TRUE
	convertRTMinute	FALSE
	numDigitsMZ	4
	numDigitsRT	0
	intval	into

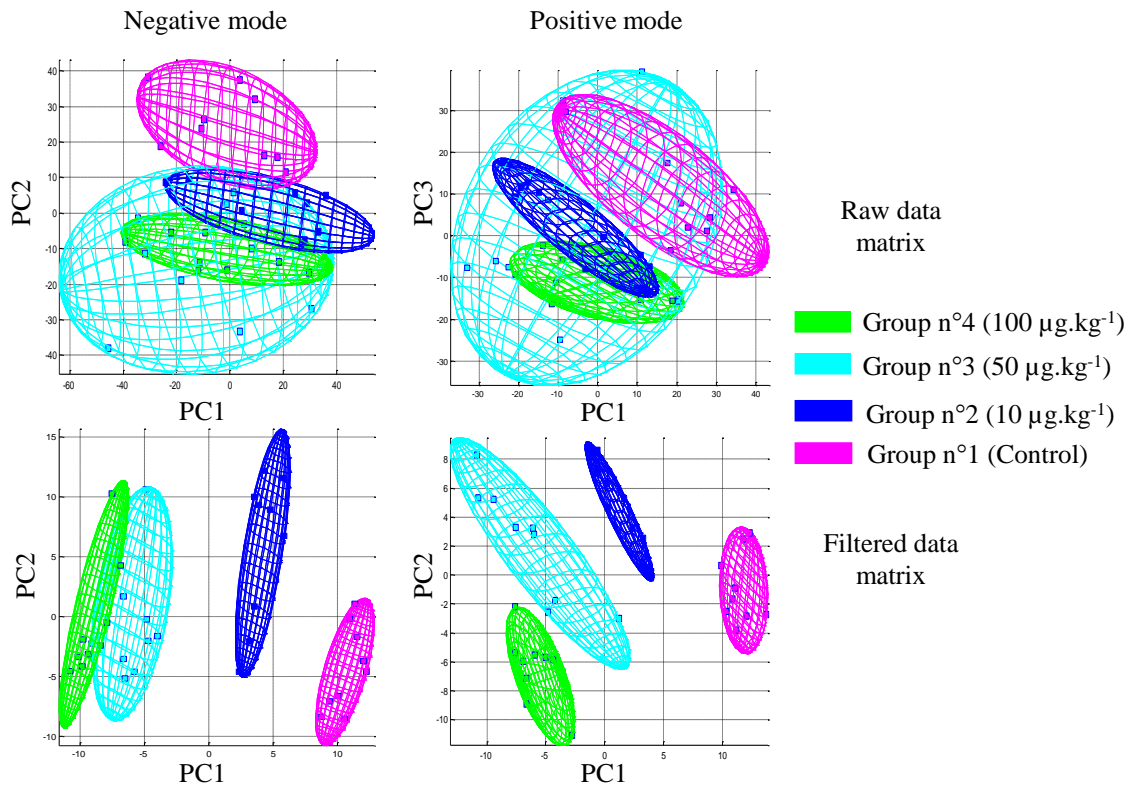


FIGURE 3.4 PCA SCORE PLOTS FOR FILTERED AND UNFILTERED DATA MATRIX FOR BOTH IONIZATION MODES ON THE DEVELOPMENT DATA.

3. CONCLUSION

L’approche proposée permet de détecter la présence d’échantillons contaminés dès 10 µg/kg (ce qui correspond au niveau de contamination de « sécurité » dans l’UE) dans le jeu de développement. Au total, 66% des contaminants dopés ont été détectés avec succès dans ce jeu de données, ce pourcentage montant à 88% si on ne considère que les composés effectivement présents dans les données brutes. Pour le jeu de validation, la totalité des contaminants dopés ont été détectés, et ce malgré la plus grande variabilité inter-échantillons due à la présence de plusieurs marques et origines dans le jeu de données.

Cette publication présente une première preuve de concept, qui va permettre une étude méthodologique plus poussée de certains points. Plusieurs étapes dans le processus de traitement des données ont été abordées avec les outils immédiatement disponibles, et mériteraient, au vu des manques de la littérature sur le sujet, une étude plus approfondie. C’est le cas notamment pour l’étape de filtration, pour laquelle des méthodes alternatives pourraient être appliquées.

CHAPITRE 4
ÉTUDE MÉTHODOLOGIQUE DU PRÉTRAITEMENT
DES DONNÉES

1. INTRODUCTION ET RÉSUMÉ DE L'ARTICLE

Dans le chapitre précédent, un processus de traitement de données a été mis en place. Dans ce processus, deux étapes en particulier (la gestion des valeurs manquantes et la filtration des données) ont été identifiées comme particulièrement sensibles, avec plusieurs outils disponibles et aucune littérature existante dans le domaine de l'évaluation de la qualité sanitaire chimique des aliments. Dans un premier temps, des outils simples d'utilisation et sur lesquels nous disposions d'un minimum de retours ont été sélectionnés, respectivement `xcms.fillPeaks` pour la gestion des valeurs manquantes, et une stratégie de filtration de données basée sur l'utilisation de tests statistiques univariés et le calcul du « fold change » (t-tests et FC à seuils fixes), la sélection se faisant avec des valeurs seuils fixes ($p\text{-value} < 0,05$ pour les tests statistiques et $FC > 2$).

La nécessité d'étudier ces deux étapes plus en détails part des constats suivants. Tout d'abord, bien que l'outil `xcms.fillPeaks` est censé donner des valeurs très proches de la réalité et fonctionne très bien sur les jeux de données acquis sur la plateforme ToF, des difficultés étaient attendues lors de l'analyse de fichiers de données possédant des faibles niveaux de bruits. `xcms.fillPeaks` générerait alors beaucoup de zéros dans la matrice de données et la présence d'une transformation log dans le processus en empêcherait l'analyse, de même que l'impossibilité de réaliser certains tests statistiques si trop de zéros étaient reportés. Il a donc été nécessaire de rechercher une nouvelle méthode de gestion des valeurs manquantes pour venir compléter `xcms.fillPeaks`. Le cahier des charges pour cette méthode est le suivant : elle doit être facilement implémentable en ligne avec le processus existant et pouvoir également fonctionner seule, c.à.d. sans `xcms.fillPeaks`. En ce qui concerne la filtration de la matrice de données, la méthode proposée a l'inconvénient d'utiliser des seuils de filtration fixes identiques pour tous les ions, alors que la qualité du signal (notamment la distribution de la mesure des aires) peut varier d'un ion à l'autre. La parution d'un article dans le domaine de la métabolomique, portant sur le calcul d'un « fold change » minimum pertinent noté FC_{\min} (Ortmayr et. al, 2017), a apporté une nouvelle piste pour la sélection des ions.

Le but de ce chapitre est d'évaluer l'influence, sur les performances de notre processus de traitement, de trois méthodes de gestion des valeurs manquantes (`xcms.fillPeaks`, mean-LOD et SVD-QRILC) et de deux méthodes de filtration (t-test et FC avec seuils fixes, FC_{\min}), incluant à chaque fois les méthodes originellement proposées dans le processus (soit un total de six combinaisons). Cette évaluation se fera par l'étude de deux jeux de données connus, à savoir ceux ayant servi au développement de la preuve de concept au chapitre précédent. Les performances des six combinaisons seront évaluées sur leur capacité à sélectionner les ions d'intérêt, les performances obtenues lors de l'analyse complète (incluant l'analyse multivariée et l'annotation), ainsi que sur leur facilité d'implémentation dans le processus existant.

2. ARTICLE

Missing values imputation and data cleaning in untargeted food chemical safety assessment by LC-HRMS

Grégoire Delaporte, Mathieu Cladière*, Valérie Camel

UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300 Massy, France

** Corresponding author: AgroParisTech, 16 rue Claude Bernard, F-75005, Paris, France*

Phone: +33 1 44 08 37 01 – email: mathieu.cladiere @agroparistech.fr

Submitted to *Chemometrics and Intelligent Laboratory Systems*

Abstract

Untargeted food safety assessment by the use of LC-HRMS instrumentation combined to chemometric tools is a rather new field. As a consequence there is a lack of methodological assessment of the different steps of the data treatment workflow. Thus, we propose a comparison of different methods applied to two major steps of data matrix pretreatment, namely missing value imputation and ion selection. To that end, a missing value classification method has been proposed for the first time for MS data. Several metrics have also been proposed to assess pretreatment step performance as well as to investigate global untargeted approach efficiency for all method combinations considered. Different contaminants were considered as “tracers” to address their detection rates. Pretreatment methods were applied here on two data sets, aiming at illustrating either a simple contamination case to detect or a more complicated application. The data sets used in this study were from the EML-EBI Metabolights data exchange platform (MTBLS752 and MTBLS754), offering other research groups the opportunity to develop and compare their own data treatment strategies with the combinations discussed in this work.

Keywords

LC-HRMS; food contaminants; non-targeted; variable selection; filtration

2.1. INTRODUCTION

Due to the complexification of food production chain and market, and the growing demand of consumers for safer food products, the development of new untargeted analytical strategies for food chemical safety assessment emerged over the last years (Cotton et al., 2014; Delaporte, Cladière, Jouan-Rimbaud Bouveresse, & Camel, 2019; Knolhoff, Zweigenbaum, & Croley, 2016; Kunzelmann, Winter, Åberg, Hellenäs, & Rosén, 2018; Tengstrand, Rosén, Hellenäs, & Åberg, 2013). To that end, high resolution hyphenated instruments such as UHPLC-HRMS combined with chemometrics methods were identified as highly promising tools, since they had already been applied to detect and characterize unknown or

unexpected compounds in metabolomics studies (Dunn et al., 2014; Thévenot, Roux, Xu, Ezan, & Junot, 2015). However, their adaptation to food chemical safety assessment raises many challenges due to the complexity of food samples and the trace levels of contaminants (Antignac et al., 2011; Castro-Puyana, Pérez-Míguez, Montero, & Herrero, 2017; Knolhoff & Croley, 2016).

Untargeted analyzes do generate highly complex signal mixtures, often composed of several thousand ions after peak extraction and alignment (Knolhoff & Croley, 2016). In chemical food safety applications, the user is often interested in only few dozens of those signals, related to chemical contaminants or residues. Those signals of interest are most of the time of much lower intensity compared with other signals present (especially those related to food constituents), meaning that strong data filtration approaches (Delaporte et al., 2019; Knolhoff & Croley, 2016; Knolhoff et al., 2016) coupled to powerful data exploration strategies (Cotton et al., 2014; Roullier et al., 2016) and multivariate methods (Antignac et al., 2011; Delaporte et al., 2019; Tengstrand et al., 2013) must be set up to detect potential contaminants. Inappropriate filtration methods may lead to either false negative results (compounds of interest are removed from the data matrix) or unusable data matrix (too much interfering compounds remain in the data matrix). In that view, strategies based on univariate statistics coupled to the use of a fixed fold change (FC) threshold have been proposed (Knolhoff et al., 2016). Another approach of data filtration of metabolomics-like LC-MS data sets has been proposed recently (Ortmayr, Charwat, Kasper, Hann, & Koellensperger, 2017), based on the calculation of a minimum relevant FC (FC_{\min}) from which a signal difference can be considered as significant for each peak. Thus, the comparison between this new approach and the strategy based on the t-test / fixed FC combination should bring interesting outcomes in untargeted chemical food safety assessment studies.

Moreover, despite the performances of analytical methods and peak extraction algorithms, MVs are frequently found in final data matrices (Wei et al., 2018); they are of great concern in untargeted approaches since they may represent around 20% of all values in MS-based data sets (Hrydziuszko & Viant, 2012). MVs are generally classified into three categories (Lazar, Gatto, Ferro, Bruley, & Burger, 2016): (i) Missing Completely at Random (MCAR) that occur randomly and independently to other variables, (ii) Missing at Random (MAR) that occur randomly but for which the probability of missing is influenced by other variables, (iii) Missing Not at Random (MNAR) for peaks below the detection capability of the instrument or below minimum criteria of the peak extraction algorithm. In MS-based data sets, MCAR and MAR cannot be distinguished since they are due to errors in the measurement or peak extraction process (Lazar et al., 2016; Wei et al., 2018); therefore, they will be considered as a unique MAR category in this study.

Bad handling of MVs is known to lead to poor outcome of the data process (Di Guida et al., 2016; Hrydziuszko & Viant, 2012; Wei et al., 2018). Comparison of MV imputation methods has been recently reported for LC-MS metabolomics data sets (Di Guida et al., 2016; Hrydziuszko & Viant, 2012; Wei et al., 2018): imputing a single value (for example zero or the median of measurements) to all MVs gave

poor outcomes; another approach is to use data analysis tools and multivariate methods to predict MVs. Last but not least, MVs can be imputed by a forced peak integration of the raw data: this strategy is implemented within the XCMS R package (“xcms.fillPeaks” module) (Smith, Want, O’Maille, Abagyan, & Siuzdak, 2006). Compared with previously described methods, the values provided by this latter approach should be closer to reality; however, with HRMS technologies, missing peaks may generate a total absence a signal (i.e. a flat baseline) and further a high amount of zero values in the raw data set, with subsequent numerous MNAR values. While efficiency of single value and multivariate imputation methods have already been discussed for metabolomics studies (Di Guida et al., 2016; Hrydziusko & Viant, 2012; Wei et al., 2018), xcms.fillPeaks has never been compared to the other approaches. Also, several works suggest that MNAR and MAR should be implemented by different methods for LC-MS data sets (Lazar et al., 2016; Wei et al., 2018), which is not the case for reported studies (Di Guida et al., 2016; Hrydziusko & Viant, 2012).

Two approaches can be reported for missing values study. The first, used by Wei et al (Wei et al., 2018), consists in using a complete data set in which missing values are artificially generated, their distribution being controlled. This offers the advantage of easily making a fine assessment of missing value imputation methods, but the distribution of missing values in the data set may be different than for “native” ones. The second approach, used by Di Guida et al (Di Guida et al., 2016) relies on the use of benchmark data sets, on which several data treatment processes featuring various missing value imputation methods are applied. In this case, the performance assessment is more difficult and rely on global performance index of the approach (e.g. detection rate) or intermediate metrics. However, this latter approach enables the implementation of the methods in “real-life” cases and should give a more realistic, even though less fine, overview of the method performances. So, our work is based on real data sets that contain native missing values.

As spotted by Di Guida et. al. (Di Guida et al., 2016) for metabolomics studies, an assessment of the whole workflow and of the influence of each step on its outcome is complementary to the study of the tools themselves to propose guidelines, since the quality of each step is highly linked to the one of the previous. Nowadays, even though the global workflow for untargeted food safety assessment using a metabolomics-like approach seems to be more or less established (Antignac et al., 2011; Knolhoff & Croley, 2016), there is a lack of vision on the influence of the different tools used for each step on the performance of the whole process. So, this paper aims at giving an overview of the influence of two important steps in the data treatment, imputation of missing values and filtration of data matrix. For the first time to the best of our knowledge, a missing value classification method was proposed for MS data. This classification method was used to set up missing value imputation approaches by combining existing imputation methods (namely “mean-LOD” and “SVD-QRILC”), which were compared to the fillPeaks tool of the XCMS package (which is a classical reference missing value imputation method for LC-MS data). For data filtration, a method commonly used in untargeted food safety studies based

on t-test and fold change calculation with fixed filtration thresholds was compared with a rather new one coming from the field of metabolomics, based on the calculation of a minimum relevant fold change for each ion (Ortmayr et al., 2017). Resulting data treatment processes were applied on different UHPLC-HRMS data sets related to untargeted food chemical safety assessment and their respective performances presented and discussed.

2.2. MATERIAL AND METHODS

The influence of three MV imputation and two filtration methods (leading to six different combinations) has been assessed as part of an existing data treatment workflow developed for untargeted food contaminants detection (Delaporte et al., 2019).

2.2.1. DATA SETS

Unlike metabolomics studies, there is currently no data set on untargeted food contaminants detection available online excepted two data sets recently deposited by our team on the EMBL-EBI MetaboLights database (DOI: 10.1093/nar/gks1004. PubMed PMID: 23109552 (Haug et al., 2013)) with the identifiers MTBLS752 (data set #1 <https://www.ebi.ac.uk/metabolights/MTBLS752>) and MTBLS754 (data set #2 <https://www.ebi.ac.uk/metabolights/MTBLS754>) (Delaporte et al., 2019). Each data set is composed of two sub-sets, one for each ionization mode. These two house data sets were selected for the present study since the lack of others available online makes impossible the discussion on other data sets.

Both data sets are based on tea samples. Green tea leaves (*camellia sinensis*) samples from two brands were bought from local stores: green tea n°1 is an organic Bancha tea from Japan and green tea n°2 is a conventional farming tea from China. Tea samples were spiked at several levels (from 10 to 100 µg/kg) (3 preparation replicates per level) with two mixes of contaminants (plus a mix of isotopically labelled molecules to check the quality of the analysis) (see Table 4.1). They were further analyzed using a generic sample treatment (direct solvent extraction and concentration) followed by broad range UHPLC-HRMS method (Cladière, Delaporte, Le Roux, & Camel, 2018) (Waters H-Class UPLC system coupled with a Waters Xevo G2-S ToF mass spectrometer equipped with and electrospray ion source in positive and negative ion centroid mode, m/z range from 60 to 800) described in supplementary materials. Each sample preparation replicate was injected three or four times (depending on the data set), data files originating from samples of same brands and same spiking levels being called “group” (n=9 or 12; 3 sample replicates analyzed 3 - 4 times each). Injection orders were randomized, and each data set also includes blank (solvent) injections as well as quality control samples (QC, pooled extracts) injected regularly (every 10 or 15 injections depending of the data set).

TABLE 4.1 MAIN CHARACTERISTICS OF STUDIED DATA SETS (DELAPORTE ET AL., 2019)

Data set	Number of brands	Spiking mix*	Spiking levels ($\mu\text{g}/\text{kg}$)
#1	1	32 contaminants	0; 10; 50; 100
#2	2	3 contaminants	0; 50 (for each brand)

* Details on spiked contaminants can be found in Table 4.7

Each data set presents a different challenge. The contamination is expected to be easy to detect in data set #1, due to the presence of numerous molecules in the spiking mix. For this data set, the main question will be on the detection rate obtained with each method combination. In data set #2, to distinguish between the variability due to the spiking and the one caused by the sample is likely to be the main challenge.

An additional data set has also been used to discuss the behavior of fillPeaks algorithm on data exhibiting flat baselines for some ions. It is also related to tea samples spiked with several food contaminants at low levels, but in that case analyzes were conducted on a LC-Orbitrap platform. Only data files acquired in positive ionization mode were used. Experimental details on this data set, as well as raw data files can be found on Metabolights data repository with the identifier MTBLS771 (<https://www.ebi.ac.uk/metabolights/MTBLS771>).

2.2.2. DATA TREATMENT WORKFLOW

The data treatment workflow is described in Figure 4.1. It can be divided into four main steps: A - building the data matrix, B – preparation and pretreatment of the built matrix (i.e. handling of missing values and ions filtration), C - Scaling and normalization, D - multivariate analysis and suspect ions annotation. Data files were firstly converted to mzXML format using Proteowizard (Chambers et al., 2012), and then uploaded on the Galaxy/Workflow4Metabolomics (W4M) platform (Giacomoni et al., 2015) where the data matrix was built using the CentWave algorithm of the XCMS package (Smith et al., 2006; Tautenhahn, Bottcher, & Neumann, 2008) (a full list of XCMS parameters can be found in Table 4.8). The data matrix is composed of the peak areas for the different replicates for every ion (i.e. variable) characterized by its retention time (RT) and m/z . At this point (between step A and B in Figure 4.1), metrics on missing values (detailed in 2.2.3) were calculated. Missing values (MV) were then imputed either on W4M, RStudio (Version 1.1.383, R version 3.4.1) or in Matlab (Matlab 7.5.0, 2007b, The MathWorks) depending on the imputation method used.

All steps further were done in Matlab. After the filtration, the data matrix undergoes a normalization and scaling step (step C in Figure 4.1: log, Pareto and Probabilistic Quotient Regression –PQN– were applied). Finally a multivariate method (Independent Component Analysis, ICA (Delaporte et al., 2019; Rutledge & Jouan-Rimbaud Bouveresse, 2015)) was implemented to highlight a potential separation of groups. Thanks to ICA, group separations could be linked to corresponding ions which were then

automatically annotated using a data mining method to detect isotopic patterns (Cotton et al., 2014; Delaporte et al., 2019) followed by a broad range in-house built database search. At the end, the annotation of discriminating ions was manually curated and the found contaminants compared with the ones spiked (called “tracers”), enabling a detection rate of our “tracers” to be estimated.

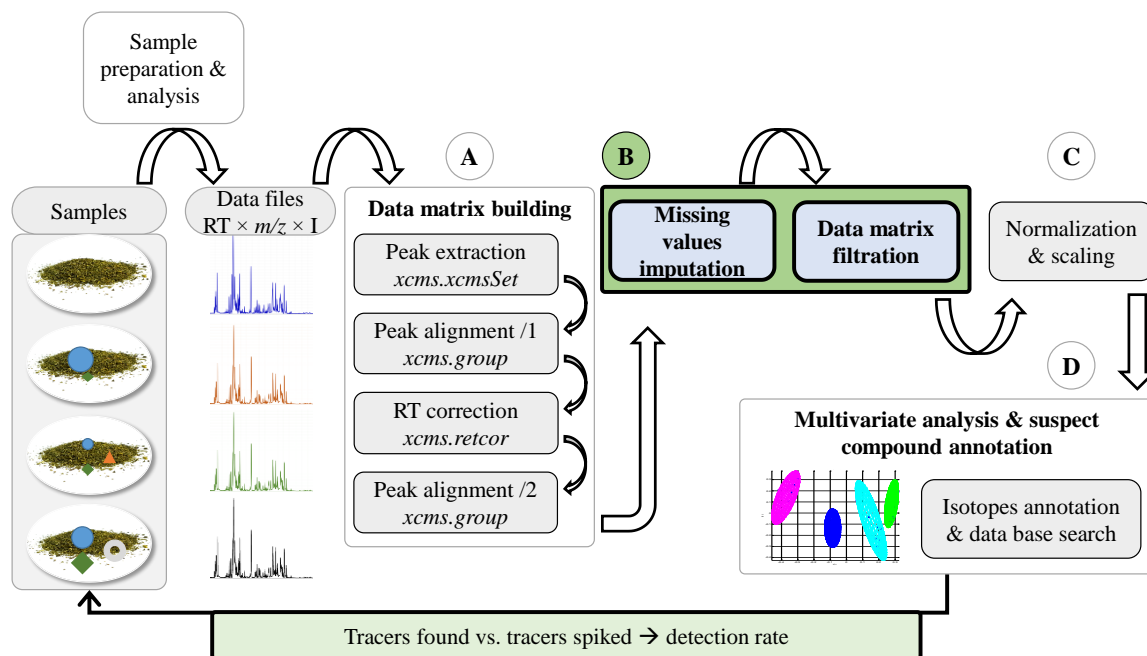


FIGURE 4.1 WORKFLOW IMPLEMENTED

2.2.3. MV CLASSIFICATION AND METRICS

In most MS-based metabolomics studies (Hrydziuszko & Viant, 2012; Di Guida et al., 2016), missing values are all imputed using the same method, either simple (e.g. all missing values are imputed with zero or the median of all measurements), or more complex (e.g. missing values are predicted using multivariate statistical methods). Even though the multiplicity of nature of missing values is well known by statisticians for a long time, its implication in MS data sets has been only raised in 2016 in the field of proteomics (Lazar et al., 2016), and even more recently in the field of metabolomics (Wei et al., 2018). However, until now, there is no methodology to classify missing values in MS data sets. Yet, as spotted by Wei et al. (Wei et al., 2018), it is important to differentiate missing values depending on their nature. Thus, we used injection replicates for each sample preparation replicate to determine the nature of the missing values, according to the following (for each ion and sample):

- Missing at Random (MAR) when, for one sample preparation replicate, there is only one missing value among the 3 or 4 injection replicates;

- Missing not at random (MNAR) when, for one sample preparation replicate, there is more than one missing value among the injection replicates (i.e. value near or below the detection capability of the overall method).

The proposed classification methodology for missing values is represented in Figure 4.2.

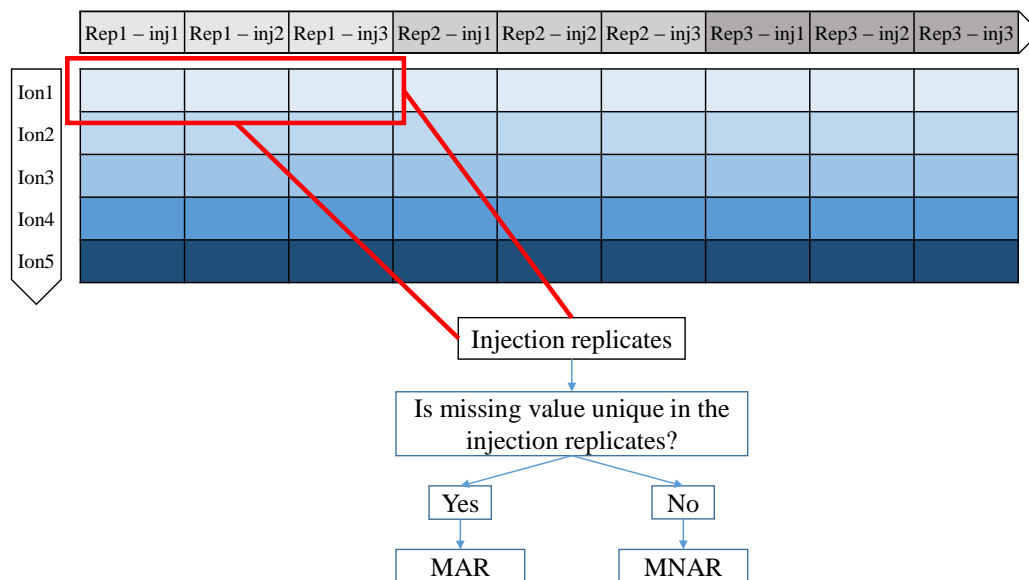


FIGURE 4.2 DATA MATRIX LAYOUT AND REPRESENTATION OF THE CLASSIFICATION METHODOLOGY FOR MISSING VALUES

Although these classification criteria can surely be improved and discussed, they have the advantage to be consistent with the performance of the instrumentation used in terms of stability and repeatability, and also to be easily applied to large data sets. Thanks to this methodology, it is now easy to pick the best method for each category (MNAR or MAR). Another advantage is the easy combination with any existing missing value imputation methods, simple or complex, including new ones.

This classification was done after the peak extraction and alignment step (step A in Figure 4.1), and several metrics were then calculated on each data set: global, group-wise and category-wise missing value rates were calculated. To assess the distribution of missing values in data sets, Pearson correlation coefficients between the frequency of missing values and m/z , RT or mean areas were calculated. Missing values frequencies were also plotted against m/z , RT or mean areas to assess any potential trend which could not be detected only by correlation coefficients (Di Guida et al., 2016; Hrydziuszko & Viant, 2012).

2.2.4. MV IMPUTATION STRATEGIES

In this work, three different methods were picked for missing values (MV) imputation: one imputes all missing values at once by forced integration of the raw chromatogram while two impute separately MAR and MNAR.

The first method replaces all MV by values estimated upon signal integration in the RT window of the missing peak in the raw data files. This was automatically performed using in-line implementation of the fillPeaks method on W4M platform.

The second method (named “mean-LOD”) imputes MV separately with simple methods. MAR are imputed by the mean of the non-missing replicates of the concerned ion and a noise component, with a random relative standard deviation (RSD) between -20% and +20% around the mean value (approximately corresponding to the observed RSD on reliable peaks on pool samples in the data set), is added to limit its influence on the following steps of the process. MNAR are imputed by the Limit of Detection (LOD) of the instrumental method, calculated here as the mean of the 3% lowest non-missing values (Libiseller et al., 2015), while adding the same noise component as for MAR.

The third method (named SVD-QRILC) imputes MAR and MNAR separately with methods based on statistical tools, respectively singular value decomposition (SVD) for MAR (Wei et al., 2018) and quantile regression imputation of left-censored data (QRILC) for MNAR. For SVD method, MV are firstly initialized to 0 and then estimated through an iteratively application of an eigen-values decomposition: here, the R wrapper based on the function “pcaMethod” has been used (Stacklies, Redestig, Scholz, Walther, & Selbig, 2007). Another method (Random Forest, RF), possibly better than SVD for MAR (Wei et al., 2018), has been assessed but either the size of our data sets (more than 20,000 ions for around 40 samples) or the MV distribution was such that computation time was too high (no convergence was observed after 24 h of computing against a convergence achieved in a few dozen of seconds for SVD) for its application here. Hence, we were not able to compare the results of this algorithm based on a learning method to the other proposed methods due to insufficient computing power despite the use of a computational server. So, the possible contribution of learning based algorithms for MV imputation on such complex data sets should deserve further studies with a more powerful computational server. On the other hand, the QRILC method was invented for left-censored data imputation: MV are imputed by a random value generated by a truncated normal distribution. This method has been reported to better handle MNAR in metabolomics data sets than others (Wei et al., 2018); the R wrapper based on “imputeLCMD” function has been retrieved from previous reported work (Lazar, 2015).

2.2.5. FILTRATION STRATEGIES

The first filtration method assessed (named “t-test / fixed FC”) relies on univariate statistic tests followed by the calculation of FC, for each ion. The Student test (t-test) is used since it is easy-to-use, it can handle rather small sample sets and moreover it has been already successfully implemented for

untargeted food safety analysis (Delaporte et al., 2019; Knolhoff et al., 2016). Two successive t-tests are made, one between each group and the blank injections, the second between each group. The FC value for each ion is further calculated as the ratio between the median peak area (blanks and QC excluded) of the highest group over the median of the lowest. For each step (t-tests, FC), a fixed threshold is used for filtration (p-value < 0.05 for t-test, FC > 2 for fold change) whatever the ions considered.

The second approach (named “FC_{min}”) is based on the calculation, for each ion, of the uncertainty on the FC (Ortmayr et al., 2017) (U_{FC}), thanks to an error propagation estimation. U_{FC} enables then the determination of a relevant minimum fold change (FC_{min}) from which a significant effect can be distinguished from the overall method variability thanks to the equation: $FC_{min} = \frac{1}{1-U_{FC}}$. A peak is then selected if the corresponding FC is superior to the FC_{min}. Detailed calculation of FC_{min} can be found in the paper published by Ortmayr et al (Ortmayr et al., 2017).

2.2.6. STUDY DESIGN AND METHODS PERFORMANCE

The different combinations of MV imputation/filtration methods (i.e. 6 different combinations, see Table 4.2) were tested on the previously described data sets, and the performance of each combination assessed.

TABLE 4.2 COMBINATION OF PRETREATMENT METHODS TESTED

Combination n°	MV imputation method	Filtration method
1	fillPeaks	t-test/fixed FC
2	fillPeaks	FC _{min}
3	Mean-LOD	t-test/fixed FC
4	Mean-LOD	FC _{min}
5	SVD-QRILC	t-test/fixed FC
6	SVD-QRILC	FC _{min}

Considering the study design and the fact that missing values were natively present in the tested data sets, limited quality metrics were available. Therefore, three indicators (2 quantitative, 1 qualitative) have been proposed to discuss the performance of the different combinations:

1. At the end of step B (Figure 4.1): total number of remaining ions after each combination, as well as number of ions of interest (i.e. “tracers”) recovered. Venn diagrams have been used to spot the similarities and differences of selected ions between combinations;
2. At the end of step D (Figure 4.1): after multivariate analysis and annotation of suspect ions, group separation can be visually assessed and the detection percentage of the spiked contaminants (“tracers”) can be calculated in both polarity modes. A global detection rate combining both ionization modes is determined as well.

3. Ease of implementation in the workflow (e.g. can the tool be implemented in-line with XCMS or the Matlab script or is a change of calculation platform needed? Is the method easy to handle?).

2.3. RESULTS AND DISCUSSION

2.3.1. STUDY OF MVs IN DATA SETS

First of all, descriptive metrics on missing values were calculated on each data set for positive and negative mode, and their distribution in each data set visualized. These metrics include the global percentage of MV, group-wise MV percentages, the respective rates of MAR and MNAR, Pearson correlation coefficients between MV rate for each ion and their m/z , RT or mean peak area. They are presented in Table 4.3. For the positive mode, a higher MV percentage is observed in data set #2 as compared to data set #1 (53.6 vs. 40.7%, respectively), but this phenomenon is not observed for the negative ionization mode.

To assess the presence of a trend within the distribution of missing values, a MV percentage is firstly calculated for each ion. Then, median of MV percentages is calculated for each percentile of relevant observed quantities (m/z , RT and median peak area of the ion), and the corresponding plots are drawn. Illustrations for data set #1 in positive mode and data set #2 in negative mode as displayed in Supplementary material Figure 4.4.

TABLE 4.3 PROPERTY SUMMARY OF DATA SETS (FOR BOTH IONIZATION MODES)

	Data set #1		Data set #2	
	POS	NEG	POS	NEG
Number of extracted ions (XCMS)	29,755	24,543	23,891	17,269
Number of data files	48	48	57*	59
Global MV rate	40.7	30.1	53.6	23.3
Group-wise MV rates	Blanks	92.7	95.3	94.6
	QC	29.7	17.0	43.7
	Group 1	32.4	19.9	50.1
	Group 2	31.9	19.3	49.2
	Group 3	32.8	19.3	51.2
	Group 4	31.2	18.2	51.3
MNAR % in MV	86.3	83.0	93.5	84.0
MAR % in MV	13.7	17.0	6.5	16.0
Pearson correlation MV / m/z	0.02	-0.01	0.13	0.02
Pearson correlation MV / RT	0.08	-0.14	0.07	-0.07
Pearson correlation MV / mean area	0.01	-0.06	-0.18	-0.07

* Injections outliers were visually detected in data set #2 for positive mode, and thus discarded

Strong similarities can be observed among our data sets. First of all, no correlation nor graphical trend could be established between the rate of MV and the measured m/z and RT, which might suggest that

MV are distributed randomly regarding m/z and RT in our data set (Pearson correlation coefficients between -0.14 and 0.13). Besides, even though the Pearson's correlation coefficients between MV rates and mean ion intensities are not significant (respectively 0.01 and -0.07), clear trends can be observed in the plot, with MV rates decreasing with the median peak area. Interestingly, when classifying MV with our approach, MNAR were predominant (83.0-93.5% of total MV) in all data sets, which is relevant with this trend, since, in MS data, MNARs often account for ions close to the limit of detection of the instrument (Lazar et al., 2016). A group-by-group study shows that they are distributed very evenly within the different sample groups, while the rate is lower in the QC samples (except for data set #2 in negative mode), and much higher in blank injections (this being expected). A slightly higher between-groups variability can be noticed for data set #2 in negative ionization mode, which cannot be explained. Overall, the properties of both data sets, acquired using either negative or positive ionization modes, are very similar regarding MV, even though less MV are observed in negative mode.

2.3.2. IONS SELECTION AFTER FILTRATION

For all combinations of MV imputation and filtration methods two figures were monitored:

1. The total number of ions selected after filtration (meaning ions that pass pretreatment steps and will be used for multivariate data analysis);
2. The number of ions of interest selected.

Ions of interest for spiked contaminants (defined as $[M+H]^+$ and $[M+Na]^+$ forms for positive mode, $[M-H]^-$ for negative mode, and their corresponding M+1 and M+2 isotopic peaks) were *a posteriori* searched in the data matrices to assess any information loss during data treatments. Based on our spiking conditions (either 32 or 3 contaminants), a targeted screening of the initial data matrices reported a total of 57 ions of interest for our "tracers" in the data matrix built for data set #1, and 8 in the one built for data set #2 for positive mode (respectively 36 and 4 in negative mode). Over the 57 ions of interest in positive mode, 54 have at least one MV needing MV imputation to enable statistical selection (for negative mode: 36 over 36).

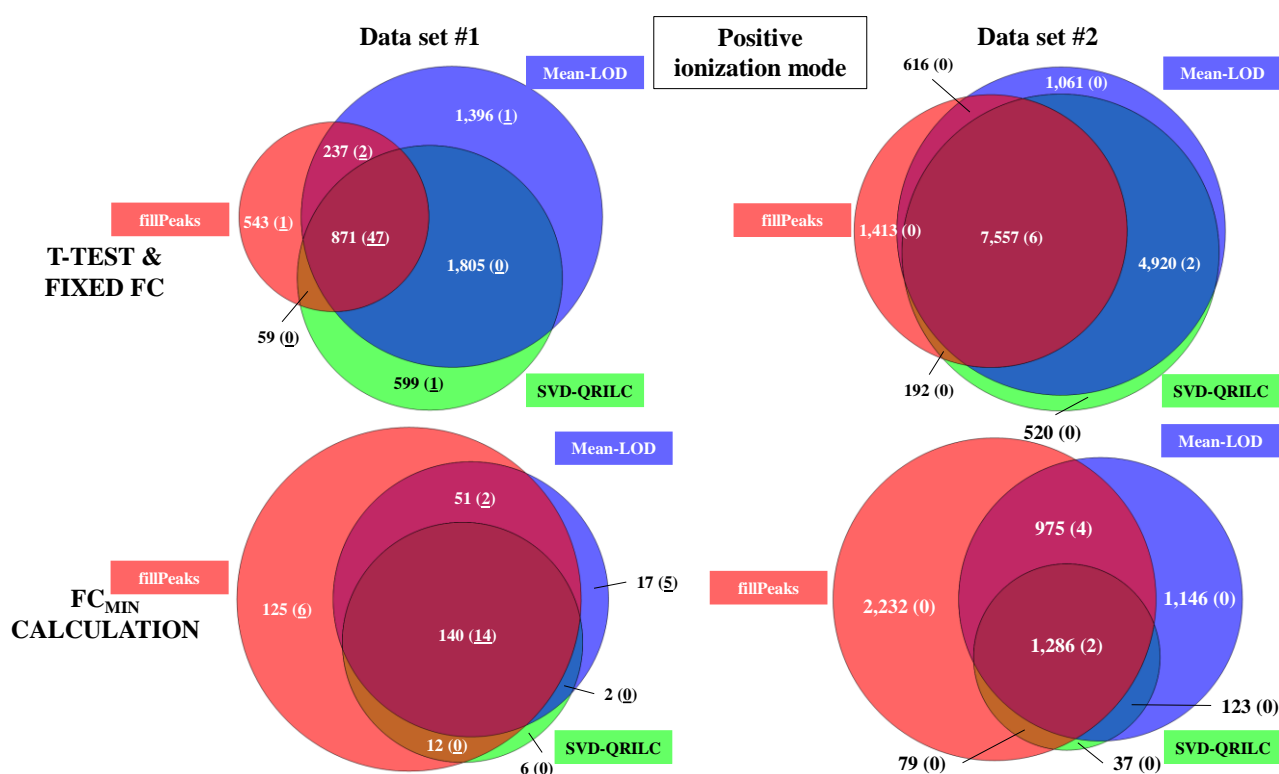
The effects of each filtration and MV imputation method on ions selection are visualized using Venn diagrams to spot common selected ions between method combinations (see in Figure 4.3).

For data set #1 in both polarity, with the t-test/fixed FC filtration, a common core of ions has been selected (871 for positive mode and 579 for negative mode) among which the majority of ions of interest (47/57 for positive mode, 36/36 for negative mode). This result highlights the ability of all MV imputation methods to allow the selection of relevant ions when combined with the t-test/fixed FC filtration method on these rather simple data sets (since 54/57 needed MV imputation for ESI⁺ and 36/36 for ESI⁻). The total number of ions selected in these data sets using FC_{min} filtration method is generally lower than with t-test/fixed FC (~100 vs. ~1,000, except for the negative mode with fillPeaks imputation). With FC_{min} filtration, fewer ions of interest are selected (14/57 for positive ionization mode

and 29/36 for the negative mode); its influence on the final detection rate of the whole process will be discussed in 3.3.

The implementation of both filtration strategies on data set #2 (positive and negative ionization modes) leads to the selection of more ions than on data set #1. This is due to the higher between-samples variability, with about ten times more ions selected each time. As for data set #1, the common core of ions selected with the combinations containing the t-test/fixed FC filtration method (7,557 ions for ESI⁺ and 8,964 for ESI⁻) contains the majority of ions of interest (6/8 for positive mode and 4/4 for negative mode). The application of FC_{min} filtration on these data sets leads to the selection of less ions than for t-test/fixed FC as generally observed for data set #1. The common core of ions of interest is more reduced in positive mode (2/8) with combinations containing FC_{min} filtration. On this data set #2, combination containing SVD-QRILC method failed to recover the 4 ions of interest due to a too stringent filtration (about 1,500 ions selected against about 4,000 respectively). In negative mode, 3 ions of interest (out of 4) are selected with all combinations, and one extra-ion is picked by the fillPeaks method.

To conclude, considering the t-test/fixed FC filtration method, a common core of selected ions gathered the most part of ions of interest, meaning that all MV imputation method are efficient to enable their selection during filtration.



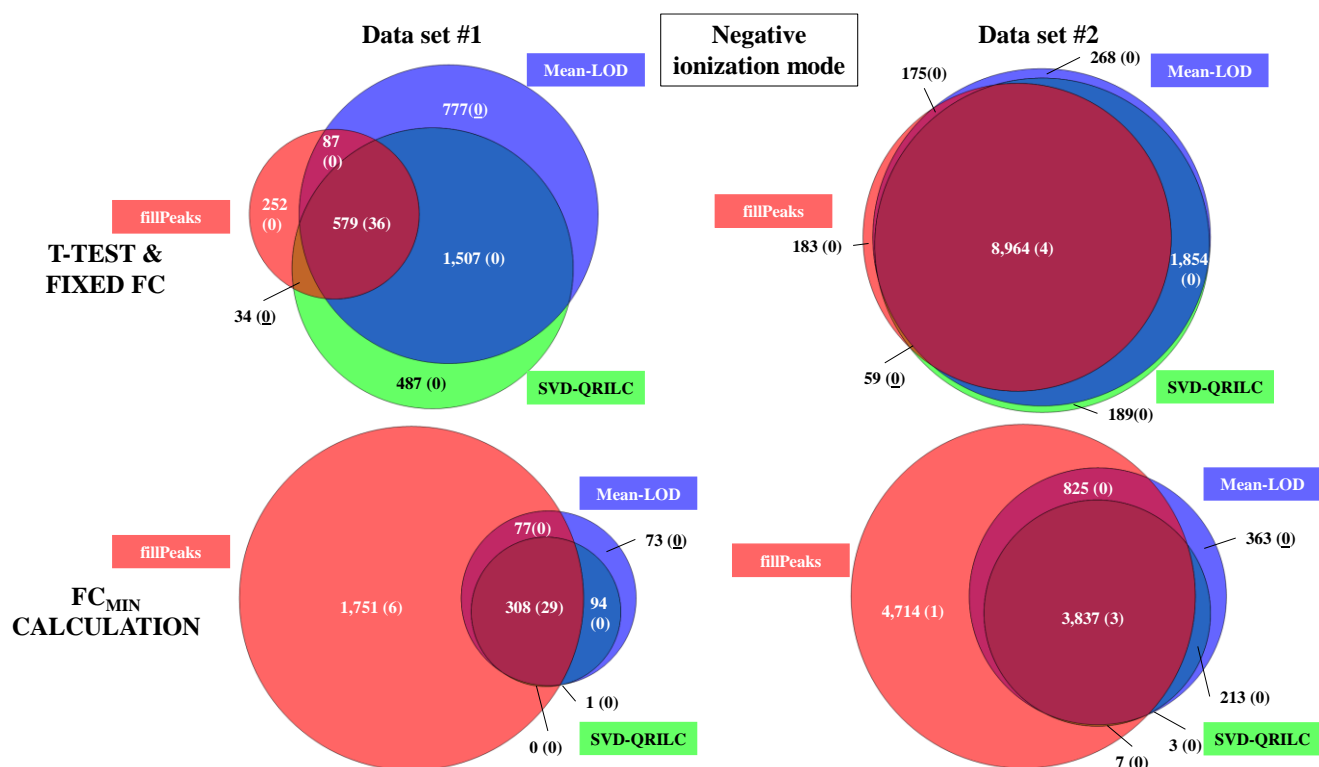


FIGURE 4.3 VENN DIAGRAMS FOR NUMBER OF IONS* SELECTED BY EACH PRETREATMENT COMBINATION ON BOTH DATA SETS FOR POSITIVE AND NEGATIVE IONIZATION MODES

*number of ions relative to our "tracers" are indicated within parenthesis

Comparison between MV imputation methods lies also in the number of total ions selected as this is indicative of the strength of the filtration. For FC_{\min} filtration method, the conclusions are different since the common core of selected ions regroup less than 50% of ions of interest. In that case, fillPeaks and mean-LOD methods were more efficient for the selection of ions of interest, but in the meantime they led to high numbers of total ions selected.

2.3.3. GLOBAL PERFORMANCE OF THE APPROACH

The whole workflow (including the final multivariate and annotation steps) was considered to figure out which pretreatment method(s) offers(s) the best performances for untargeted food contaminants detection. Results for positive and negative ionization modes are presented in Table 4.4. A global detection rate of the method (obtained by the combination of results from both ionization modes) is displayed as well.

TABLE 4.4 PERFORMANCES OF THE WHOLE WORKFLOW DEPENDING ON THE MV IMPUTATION / FILTRATION METHODS COMBINATION AND DATA SETS

Data set	Combination	Positive mode		Negative mode		Global
		Number of ions after filtration	Detection rate (%)	Number of ions after filtration	“Tracers” detection (%)	Detection rate (%)
#1	fillPeaks + t-test/fixed FC	1,710	44	952	38	66
	fillPeaks + FC_{\min}	328	50	2,136	38	72
	mean-LOD + t-test/fixed FC	4,309	50	2,950	41	75
	mean-LOD + FC_{\min}	210	44	552	34	66
	SVD-QRILC + t-test/fixed FC	3,336	53	2,607	38	78
	SVD-QRILC + FC_{\min}	160	38	403	38	66
#2	fillPeaks + t-test/fixed FC	9,778	100	9,381	67	100
	fillPeaks + FC_{\min}	4,572	100	9,383	67	100
	mean-LOD + t-test / fixed FC	14,142	0	11,261	67	67
	mean-LOD + FC_{\min}	3,530	100	5,238	67	100
	SVD-QRILC + t-test / fixed FC	13,188	0	11,066	67	67
	SVD-QRILC + FC_{\min}	1,524	67	4,060	0	67

Firstly, for data set #1, whatever the pretreatment method combinations, all sample groups could be discriminated with our untargeted approach (see in Figure 4.5 for positive ionization mode score plots and Figure 4.6 for negative ionization mode). It means that the multivariate method used can successfully separate the “unnecessary” ions in the data matrix from the common core of ions of interest observed both for t-test/fixed FC and FC_{\min} (see in Table 4.4). Percentages of detection for our “tracers”

ranged from 38 to 53% in positive mode and from 34 to 41% for negative mode (leading to global detection rates between 66 and 78% when combining both modes) for this data set that mimics a quite simple case (one brand, three different levels of contamination plus a control group). The influence of MV imputation method on the detection rates seems minor since all method combinations give acceptable performance. Interestingly, no clear link can be established between the number of ions of interest selected and the detection rate of the method, meaning that, even though FC_{min} filtration method selected less ions of interest than t-test/fixed FC (see in Figure 4.3), it seems to select the most important ones (i.e. monoisotopic ions) with a minor impact on the detection rates observed. Interestingly, for both positive and negative ionization modes, mean-LOD method coupled to FC_{min} filtration strategy seems to lead to lower relative intra-group variances (see in Figure 4.5 and Figure 4.6). This may be the consequence of the use of the injection replicate information to fill MV with this imputation methods, and also of the stronger data reduction brought by FC_{min} compared with t-test/fixed FC method.

On the other hand, all combinations do not seem suitable for the more complex data set #2. In positive ionization mode, mean-LOD & SVD-QRILC coupled with t-test/fixed FC filtration do not manage to detect the three contaminants spiked, and SVD-QRILC coupled with FC_{min} filtration only achieved the detection of 2 contaminants out of 3. Interestingly, all contaminants were detected using fillPeaks coupled with t-test/fixed FC or FC_{min} (combinations n°1 & 2 as displayed in Table 4.2) and mean-LOD coupled with FC_{min} . The performance of the methods are more homogeneous in negative ionization mode since only SVD/QRILC coupled with FC_{min} filtration failed to detect the spiked contaminants. At the end, when considering simultaneously both polarities, fillPeaks appears as the only MV imputation method that enables the annotation of all contaminants whatever the filtration method used.

Based on those results, as well as the easiness of implementation of each tool, the main characteristics of imputation and filtration methods were proposed (Table 4.5 and Table 4.6).

We observed that only combinations relying on fillPeaks successfully enabled the detection of spiked contaminants (or a majority of them) in all data sets. In addition, this MV imputation method does not need to classify missing values as MNAR or MAR, and it is easily implemented in-line after peak extraction since it is part of the XCMS package (being already implemented on every XCMS-based platforms). Practically speaking, fillPeaks is very user-friendly, with easy-to-use graphical interfaces developed by the community (e.g. W4M and XCMS Online). Yet, it relies on a complex algorithm, so that inconsistent results may be difficult to troubleshoot, especially for unexperienced users even though graphical outputs are available. Hopefully, this tool benefits from a very dynamic and open scientific community that brings help and technical support. As stated before, another drawback of this method, based on forced integration, lies in ions presenting flat baselines (cut-off during the acquisition) where MVs are imputed as zeros and should be handled afterward since they may prevent the use of some critical pretreatment methods (e.g. log normalization and univariate statistics). This is illustrated by our results on data set MTBLS 771: a total of 48% of the values in the data matrix were missing before the

fillPeaks step, while, after fillPeaks the data matrix contains 8% of zeros (i.e. 17% of initial missing values). This clearly shows that a significant number of zeros may be present after the fillPeaks completion, and suggests the advantage of combining fillPeaks with other MV imputation methods such as mean-LOD for example. The results obtained by the combination of both methods (fillPeaks and mean-LOD) can be found in the dedicated publication (Delaporte et al., 2019b).

Methods needing MV classification suffer from the absence of established methodologies to classified MV in MS-based data sets; under our experience, both methods (mean-LOD and SVD-QRILC) did not always enable the detection of contaminants in the most complicated case studied. Since MAR and MNAR are not imputed with the same algorithm, a MV misclassification may lead to an inconsistent imputation. The effect of such misclassification is expected to be higher with SVD-QRILC (based on statistical methods) than with the simple mean-LOD method, in agreement with the lower performance of SVD-QRILC compared with mean-LOD observed in this work. Our results pointed out the classification method for MAR and MNAR as a possible limiting step for the efficiency of these MVs imputation methods and further investigations discussing this first proposed methodology are needed. Despite this drawback, both methods have the advantage to fully complete the data matrix since no zeros are obtained at the end.

Regarding the mean-LOD method, the noise component set as a random value between -20% and +20% around the estimated values (mean or LOD) may sometimes over- or under-estimate the “real” standard deviation of the data. This over- or under-estimation may disturb the calculation of the FC uncertainty U_{FC} and influence the filtration method FC_{min} . However, the detection rate of the combination mean-LOD + FC_{min} indicated in Table 4.4 proved that the over- or under-estimation is not a critical issue for the tested data sets, but more tests are needed to confirm a larger applicability.

TABLE 4.5 MAIN FEATURES OF MV IMPUTATION METHODS

MV imputation method	Main characteristics
fillPeaks	<p><i>Pros</i></p> <p>No need for MV classification Easy in-line implementation within XCMS Gives good results with all filtration methods on every data sets Benefits from the support of a dynamic scientific community Can be easily combined with other statistical MV imputation methods</p> <p><i>Cons</i></p> <p>May generate a lot of zeros on flat baseline with no signal (i.e. issues with log scaling and univariate statistics) Relies on a complex algorithm that can be difficult to troubleshoot, especially for unexperienced users even though graphical outputs are available</p>
Mean-LOD	<p><i>Pros</i></p> <p>Simple tools, understandable by all No zeros at the end of the process</p> <p><i>Cons</i></p> <p>Needs MV classification (= more complex to implement and may be subjected to discussion) May lead to over-fitting of the data Does not enable the detection of all “tracers” when combined with t-test / fixed FC on the most heterogeneous data set</p>
SVD-QRILC	<p><i>Pros</i></p> <p>Best detection rate on the simple data set No zeros at the end of the process</p> <p><i>Cons</i></p> <p>Needs MV classification More complex methods than mean-LOD, may be difficult to troubleshoot for unexperienced users Performs badly on the most heterogeneous data set</p>

TABLE 4.6 MAIN FEATURES OF FILTRATION METHODS

Filtration method	Main characteristics
t-test/fixed FC	<p><i>Pros</i></p> <p>Easy to implement More ions of interest selected</p> <p><i>Cons</i></p> <p>Lower detection rates on heterogeneous data sets with imputation methods other than fillPeaks Filtration thresholds may be subjected to discussion</p>
FC _{min}	<p><i>Pros</i></p> <p>Easy to implement No parameter to set (filtration threshold determined by the quality of measurement for each ion) Greater reduction of selected ion numbers Goes well with all imputation methods used</p> <p><i>Cons</i></p> <p>Fewer ions of interest selected</p>

The t-test/fixed FC method relies on the use of fixed, generic threshold for each step (p -value < 0.05 for t-tests and $FC > 2$), which can be a limit since all ions do not necessarily have the same characteristics in terms of distribution and variance. On the opposite, the FC_{\min} method adjusts the threshold to the measurement quality of each ion, which may enable a better quality of filtration, with the selection of peaks exhibiting lower relative standard deviations, and therefore potentially less artifacts. However, the t-test/fixed FC strategy offers more flexible applications than FC_{\min} since one can use any univariate statistical test to better fit to the data structure, or apply one or two filtration steps (for example by omitting the fixed FC step) if too much data of interest seem to be lost. Even though FC_{\min} leads to the selection of fewer ions of interest than t-test/fixed FC, the global detection rates obtained are very similar (see in Table 4.4). On more complex data set such as data set #2, the greater reduction of ions number generally observed with the FC_{\min} filtration method can also be an asset since it makes the computation easier and faster. In the meantime, the risk of discarding a potential contaminant is also higher, especially with molecules having a signal close to the limit of detection of the instrument. Consequently, it could be recommended to implement both filtration methods in parallel to increase the detection probability of potential contaminants.

2.4. CONCLUSION

Several pretreatment methods (three missing value imputation methods - one based on the forced integration of raw data, two based on the classification of MV as MAR or MNAR - coupled to two filtration methods, leading to six combinations) were tested on two LC-MS data sets dedicated to untargeted food chemical safety. They were integrated in a general workflow, and the final detection rate calculated for each data set and method combination. In addition to this global performance assessment, the ions selected by each combination were more deeply investigated.

As expected initially, the total number of ions selected varies a lot between pretreatment methods. Interestingly the ions of interest (corresponding to spiked contaminants) were selected by most methods. Considering the whole workflow, all combinations were able to detect the spiked contaminants on the data sets corresponding to a simple contamination scenario (positive and negative ionization modes), with different success rates (from 66 to 78%). The more heterogeneous data set was more problematic since several combinations did not enable the detection of the spiked contaminants. In fact, the only imputation method that enables the detection of our tracers for this contamination scenario, whatever the filtration approach used, is fillPeaks based on the re-analysis of raw data. This tool has also the advantage to be easily implemented in-line with the peak extraction step if this one is carried out with the wide-spread, user friendly package XCMS or its online implementations XCMS-online or Workflow4Metabolomics. However, data exhibiting a flat baseline with no signal in case of no peak, it can generate an important amount of zeros. In that case, they should be handled as MV to avoid any problematic issues in the workflow afterwards. We suggest that mean-LOD method should be used to

complement fillPeaks on remaining zeros since it is very easy to implement and still shows satisfactory results.

Unlike existing missing value imputation approaches, two methods presented here rely on a classification of missing values according to their nature. This very simple methodology is based on instrumental replicates, thereby authorizing a quick classification; in addition, it can be easily combined with any MV imputation method. Yet, it seems to face some limits when dealing with heterogeneous data sets, so that more work is needed to better address MV classification for MS-based data sets. In this work, this classification-based approach has been used with either a simple method (mean-LOD) or a more sophisticated one chosen for its performances on respective missing values types (SVD-QRILC). The results presented here constitute a good proof of concept of the potential of such classification-based approaches to help missing value imputation. It would surely benefit from its implementation with other imputation methods such as ones based on machine learning algorithms, for example Artificial Neural Networks [ANN, (Armitage et al., 2015)] or genetic algorithm (Aydilek & Arslan, 2013). Such work would be a natural extension of the present publication and would provide highly interesting results for the scientific community, even outside the field of untargeted food safety assessment. In addition, studying the proposed missing value imputation strategy on a simulated data set (i.e. on better controlled, even though less realistic, situation) could lead to interesting contribution to the understanding of the missing value imputation process.

In applied fields such as untargeted food chemical safety assessment, the user mainly focus on the final outcome of the approach, but our understanding of the process should be improved in order to build better tools and workflows. This work shows a first attempt in that direction but more work and more data sets dealing with untargeted food safety are needed to get a critical point of view on all the steps of the workflow and their influence on the detection rates.

2.5. FUNDING

This work was supported by Paris Institute of Technology for Life, Food and Environmental Sciences (AgroParisTech), the French National Institute for Agricultural Research (INRA) and the French Ministry of Higher Education and Research.

2.6. REFERENCES

- 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>
- Armitage, E.G., Godzien, J., Alonso-Herranz, V., López-González, Á., Barbas, C., Missing value imputation strategies for metabolomics data, *Electrophoresis*. 36 (2015) 3050–3060. doi:10.1002/elps.201500352.
- Aydilek, I.B., Arslan, A., A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm, *Inf. Sci. (Ny)*. 233 (2013) 25–35.

doi:10.1016/j.ins.2013.01.021.

- Castro-Puyana, M., Pérez-Míguez, R., Montero, L., & Herrero, M. (2017). Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability. *TrAC - Trends in Analytical Chemistry*, *93*, 102–118. <https://doi.org/10.1016/j.trac.2017.05.004>
- Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., ... Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, *30*(10), 918–920. <https://doi.org/10.1038/nbt.2377>
- Cladière, M., Delaporte, G., Le Roux, E., & Camel, V. (2018). Multi-class analysis for simultaneous determination of pesticides, mycotoxins, process-induced toxicants and packaging contaminants in tea. *Food Chemistry*, *242*, 113–121. <https://doi.org/10.1016/j.foodchem.2017.08.108>
- Cotton, J., Leroux, F., Broudin, S., Marie, M., Corman, B., Tabet, J. C., ... Junot, C. (2014). High-resolution mass spectrometry associated with data mining tools for the detection of pollutants and chemical characterization of honey samples. *Journal of Agricultural and Food Chemistry*, *62*(46), 11335–11345. <https://doi.org/10.1021/jf504400c>
- Delaporte, G., Cladière, M., Jouan-Rimbaud Bouveresse, D., & Camel, V. (2019). Untargeted food contaminant detection using UHPLC-HRMS combined with multivariate analysis: feasibility study on tea. *Food Chemistry*, doi:10.1016/j.foodchem.2018.10.089.
- Delaporte, G., Cladière, M., Camel, V., Untargeted food chemical safety assessment: A proof-of-concept on two analytical platforms and contamination scenarios of tea, *Food Control*. *98* (2019b) 510–519. doi:10.1016/j.foodcont.2018.12.004.
- Di Guida, R., Engel, J., Allwood, J. W., Weber, R. J. M., Jones, M. R., Sommer, U., ... Dunn, W. B. (2016). Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, MV imputation, transformation and scaling. *Metabolomics*, *12*(5). <https://doi.org/10.1007/s11306-016-1030-9>
- Dunn, W. B., Lin, W., Broadhurst, D., Begley, P., Brown, M., Zelena, E., ... Kell, D. B. (2014). Molecular phenotyping of a UK population: defining the human serum metabolome. *Metabolomics*, *11*(1), 9–26. <https://doi.org/10.1007/s11306-014-0707-1>
- Franceschi, P., Masuero, D., Vrhovsek, U., Mattivi, F., & Wehrens, R. (2012). A benchmark spike-in data set for biomarker identification in metabolomics. *Journal of Chemometrics*, *26*(1), 16–24. <https://doi.org/10.1002/cem.1420>
- Giacomini, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., ... Caron, C. (2015). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics*, *31*(9), 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>
- Haug, K., Salek, R. M., Conesa, P., Hastings, J., De Matos, P., Rijnbeek, M., ... Steinbeck, C. (2013). MetaboLights - An open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Research*, *41*(D1), 781–786. <https://doi.org/10.1093/nar/gks1004>
- Hrydziuszkó, O., & Viant, M. R. (2012). MVs in mass spectrometry based metabolomics: An undervalued step in the data processing pipeline. *Metabolomics*, *8*, 161–174. <https://doi.org/10.1007/s11306-011-0366-4>
- 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>
- Knolhoff, A. M., Zweigenbaum, J. A., & Croley, T. R. (2016). Nontargeted Screening of Food Matrices: Development of a Chemometric Software Strategy to Identify Unknowns in Liquid Chromatography-Mass Spectrometry Data. *Analytical Chemistry*, *88*(7), acs.analchem.5b04208. <https://doi.org/10.1021/acs.analchem.5b04208>
- 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*, 1–10. <https://doi.org/10.1007/s00216-018-1028-4>
- Lazar, C. (2015). imputeLCMD: A collection of methods for left-censored missing data imputation v.2.0. Retrieved from <https://cran.r-project.org/package=imputeLCMD>
- Lazar, C., Gatto, L., Ferro, M., Bruley, C., & Burger, T. (2016). Accounting for the Multiple Natures of MVs in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *Journal of Proteome Research*, 15(4), 1116–1125. <https://doi.org/10.1021/acs.jproteome.5b00981>
- Libiseller, G., Dvorzak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., ... Magnes, C. (2015). IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics*, 16(1), 118. <https://doi.org/10.1186/s12859-015-0562-8>
- Ortmayr, K., Charwat, V., Kasper, C., Hann, S., & Koellensperger, G. (2017). Uncertainty budgeting in fold change determination and implications for non-targeted metabolomics studies in model systems. *The Analyst*, 142(1), 80–90. <https://doi.org/10.1039/C6AN01342B>
- Roullier, C., Guittou, Y., Valery, M., Amand, S., Prado, S., Robiou Du Pont, T., ... Pouchus, Y. F. (2016). Automated Detection of Natural Halogenated Compounds from LC-MS Profiles- Application to the Isolation of Bioactive Chlorinated Compounds from Marine-Derived Fungi. *Analytical Chemistry*, 88(18), 9143–9150. <https://doi.org/10.1021/acs.analchem.6b02128>
- Rutledge, D. N., & Jouan-Rimbaud Bouveresse, D. (2015). Corrigendum to “Independent Components Analysis with the JADE algorithm”, [Analytical Chemistry, 50, (2013) 22-32, doi:10.1016/j.trac.2013.03.013]. *TrAC - Trends in Analytical Chemistry*, 67, 220. <https://doi.org/10.1016/j.trac.2015.02.001>
- Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), 779–787. <https://doi.org/10.1021/ac051437y>
- Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). pcaMethods - A bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, 23(9), 1164–1167. <https://doi.org/10.1093/bioinformatics/btm069>
- Tautenhahn, R., Bottcher, C., & Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, 9, 16. <https://doi.org/10.1186/1471-2105-9-504>
- 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>
- Thévenot, E. A., Roux, A., Xu, Y., Ezan, E., & Junot, C. (2015). Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *Journal of Proteome Research*, 14(8), 3322–3335. <https://doi.org/10.1021/acs.jproteome.5b00354>
- Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., & Ni, Y. (2018). MV Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Scientific Reports*, 8(1), 663. <https://doi.org/10.1038/s41598-017-19120-0>

2.7. SUPPLEMENTARY MATERIALS

Analytical method

Samples extraction is done using direct solvent extraction with 5 mL of an ACN/MeOH (90/10 v/v) mixture acidified with 0.1% formic acid (FA). Tubes are first agitated upside-down on an agitating plate during 1 h before centrifugation at 3,000 g for 10 min. Supernatant is then collected and an aliquot (2

mL) then evaporated to dryness under a gentle stream of nitrogen. The extract is then reconstituted in 0.2 mL of ACN acidified with 0.1% FA. Then 0.8 mL of ultrapure water with 0.1% FA is added in order to reconstitute 1 mL of final volume, and centrifuged at 12,000 g for 10 min. At the end, 0.5 mL of the final extract is sampled and filtered at 0.2 μm using a syringeless filter vial (mini-uniprep G2, Whatman) before analysis. A Quality Control (QC) sample is prepared for each data set by pooling together 0.2 mL of final extract from every sample of the set; an aliquot of 0.5 mL is then taken and filtered at 0.2 μm using a syringeless filter vial.

Sample extracts are then analyzed on a UHPLC-HRMS system (Acquity H-Class coupled with a Xevo G2-S ToF analyzer, Waters). Separation is achieved on a C18-PFP column (150 \times 2.1 mm, 2 μm particles diameter, ACE supplied by AIT France). An electrospray ion source is used in both positive (ESI+) and negative (ESI-) modes. Both modes are run separately. For ESI+, the mobile phase is composed of water (A) and ACN (B), both acidified with 0.1% FA, and MeOH (C), flowing at 0.4 mL.min⁻¹. Gradient starts at 100% A and reaches 100% B in 10 min, being kept for 6 min before switching to 100% C to rinse the system in 1 min, being held for 5 min, returning back to 100% A in 1 min and finally equilibrating for 3 min, with a total run time of 26 min. For ESI-, the mobile phase is composed of water buffered at pH 6.45 with 10 mM of ammonium formate (A) and MeOH (B) flowing at 0.3 mL.min⁻¹. The gradient starts at 100% A and reaches 100% B in 13 min, holding this condition for 7 min before turning back to 100% A in 1 min and finally equilibrating for 3 min, with a total run time of 24 min. For both chromatographic methods the temperature of the column oven is kept at 30°C. ToF mass analyzer is operated in resolution mode (R~30,000) and detected m/z ranges from 60 to 800.

TABLE 4.7 INFORMATION ON SPIKING CONTAMINANTS USED IN THIS STUDY

N°	Name	Class	Chemical family	log K _{ow}	Ionization mode	Used in data set
1	(2,4-dichlorophenoxy)acetic acid 2,4-D	Pesticide (herbicide)	Auxinic herbicide	-0.82	+/-	1
2	(4-chloro-2-methylphenoxy)acetic acid MCPA	Pesticide (herbicide)	Auxinic herbicide	-0.81	-	1
3	2-(4-chloro-2-methylphenoxy)propanoic acid MCPA	Pesticide (herbicide)	Auxinic herbicide	-0.19	-	1
4	2,4,5-trichlorophenoxyacetic acid 2,4,5-T	Pesticide (herbicide)	Auxinic herbicide	2.88	-	1
5	4-(2,4-dichlorophenoxy)butanoic acid 2,4-DB	Pesticide (herbicide)	Auxinic herbicide	1.35	+/-	1
6	Acetamiprid	Pesticide (insecticide)	Neonicotinoid	0.8	+	1
7	Acrylamide	Process-induced	Amide	-0.67	+	1
8	Aldicarb	Pesticide (acaricide)	Carbamate	1.15	+	1
9	Ametryn	Pesticide (herbicide)	Triazine	2.63	+	1
10	Atrazine	Pesticide (herbicide)	Triazine	2.7	+	1
11	Bisphenol A	Migrant from packaging	Bisphenol	3.3	-	1
12	Bisphenol A diglycidyl ether BADGE	Migrant from packaging	Diglycidyl ether	3.84	+	1
13	Bisphenol F	Migrant from packaging	Bisphenol	1.65	-	1
14	Bisphenol F diglycidyl ether BFDGE	Migrant from packaging	Diglycidyl ether	Not available	+	1
15	Bisphenol S	Migrant from packaging	Bisphenol	2.91	+/-	1; 2
16	Deoxynivalenol	Mycotoxin	Trichothecene	0.29	+/-	1
17	Dichloprop	Pesticide (herbicide)	Auxinic herbicide	2.29	-	1
18	Dichlorvos	Pesticide (acaricide)	Organochlorinated	1.9	+	1
19	Diflubenzuron	Pesticide (insecticide)	Benzoylurea	3.89	+/-	1

Chapitre 4 – Étude méthodologique du prétraitement des données

20	Dimethoate	Pesticide (acaricide)	Organophosphate	0.7	+	1
21	Dinoseb	Pesticide (herbicide)	Dinitrophenol	2.29	-	1
22	Diuron	Pesticide (herbicide)	Phenylurea	2.87	+/-	1
23	Fenoprop 2,4,5-TP	Pesticide (herbicide)	Auxinic herbicide	2.84	-	1
24	Fumonisin B1	Mycotoxin	Fumonisin	-0.5	+	1
25	Fumonisin B2	Mycotoxin	Fumonisin	1.2	+	1
26	Hydroxymethylfurfural	Process-induced	Furan	-0.09	+	1
27	Imidacloprid	Pesticide (insecticide)	Neonicotinoid	0.57	+/-	1
28	Malathion	Pesticide (insecticide)	Organophosphate	2.75	+	1
29	Ochratoxin A	Mycotoxin	Ochratoxin	4.74	+/-	1; 2
30	Pirimiphos-methyl	Pesticide (insecticide)	Organophosphate	3.9	+	1
31	Propargite	Pesticide (acaricide)	Organosulfite	5.7	+	1
32	Tolfenpyrad	Pesticide (insecticide)	Pyrazole	5.61	+/-	1; 2

TABLE 4.8 FULL PARAMETERS AND THEIR CORRESPONDING VALUES FOR PEAK EXTRACTION USING XCMS

Step	Parameter	Value
xcmsSet	scanrange	180-2400
	nSlaves	1
	method	centWave
	ppm	15
	peakwidth	5-60
	mzdiff	-0.001
	snthresh	10
	integrate	1
	noise	0
	prefilter	0
group - A	method	density
	minfrac	0.5
	bw	2
	mzwid	0.015
	sleep	0.001
retcor	method	peakgroups
	smooth	loess
	extra	1
	missing	1
	span	0.2
	family	gaussian
group - B	method	density
	minfrac	0.5
	bw	2
	mzwid	0.015
	sleep	0.001
	max	50
fillPeaks	method	chrom
	convertRTMinute	FALSE
	numDigitsMZ	4
	numDigitsRT	2
annotatediff	intval	into
	nSlaves	4
	sigma	6
	perfwHM	0.6
	ppm	15
	mzabs	0.015
	maxcharge	1
	maxiso	4
	minfrac	0.5
	quick	TRUE
	convertRTMinute	FALSE
	numDigitsMZ	4
	numDigitsRT	0
intval	into	

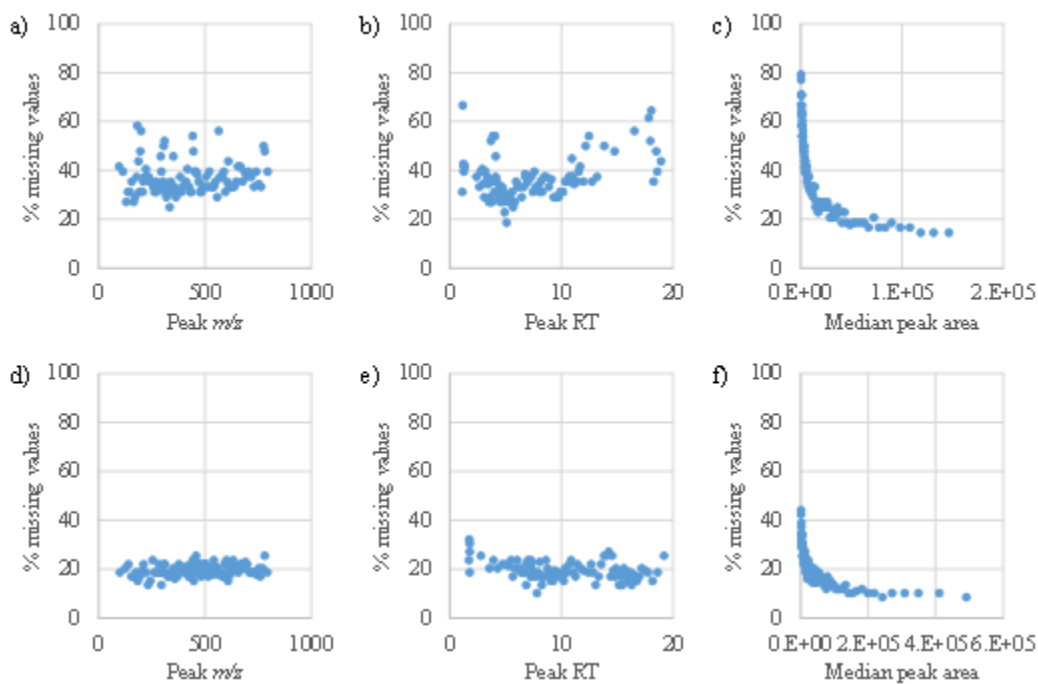


FIGURE 4.4 REPRESENTATION OF MV RATE AGAINST m/z , RT AND MEDIAN PEAK AREAS OF IONS.

Figures a, b and c are for data set #1 in positive mode, figures d, e and f for data set #2 in negative mode. For figure c and f, the five highest percentiles were discarded for plotting purpose.

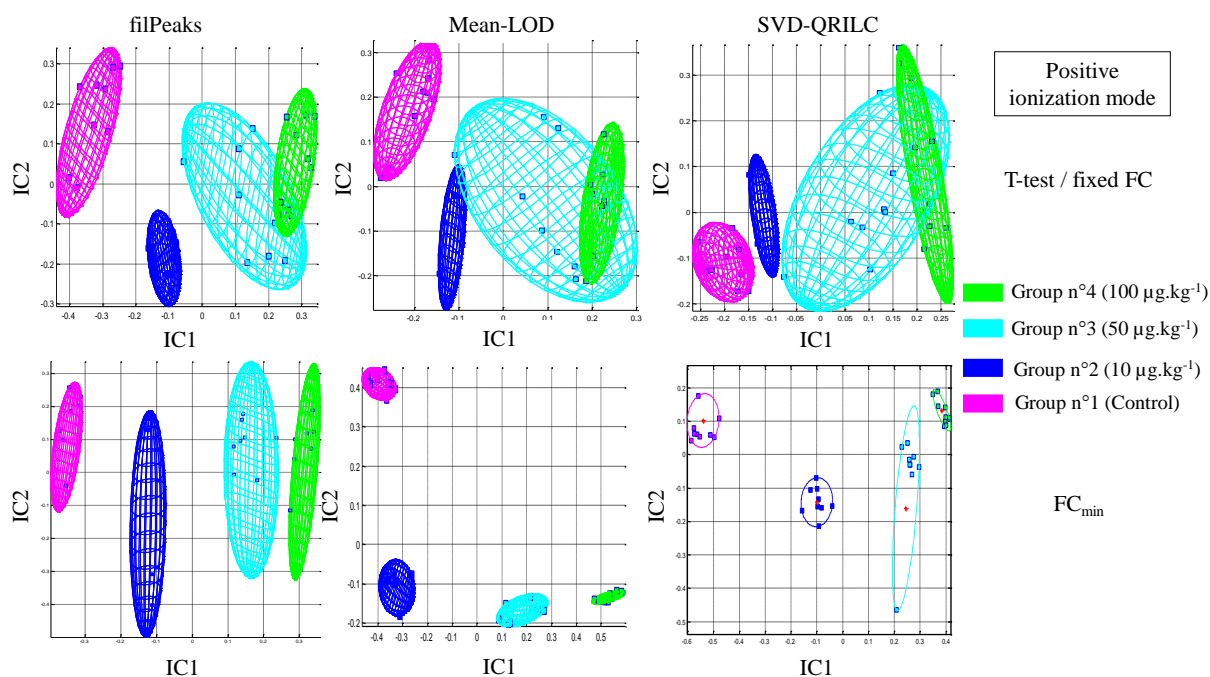


FIGURE 4.5 SCORE PLOTS OF ICA OUTPUT FOR DATA SET #1 ACQUIRED IN POSITIVE IONIZATION MODE

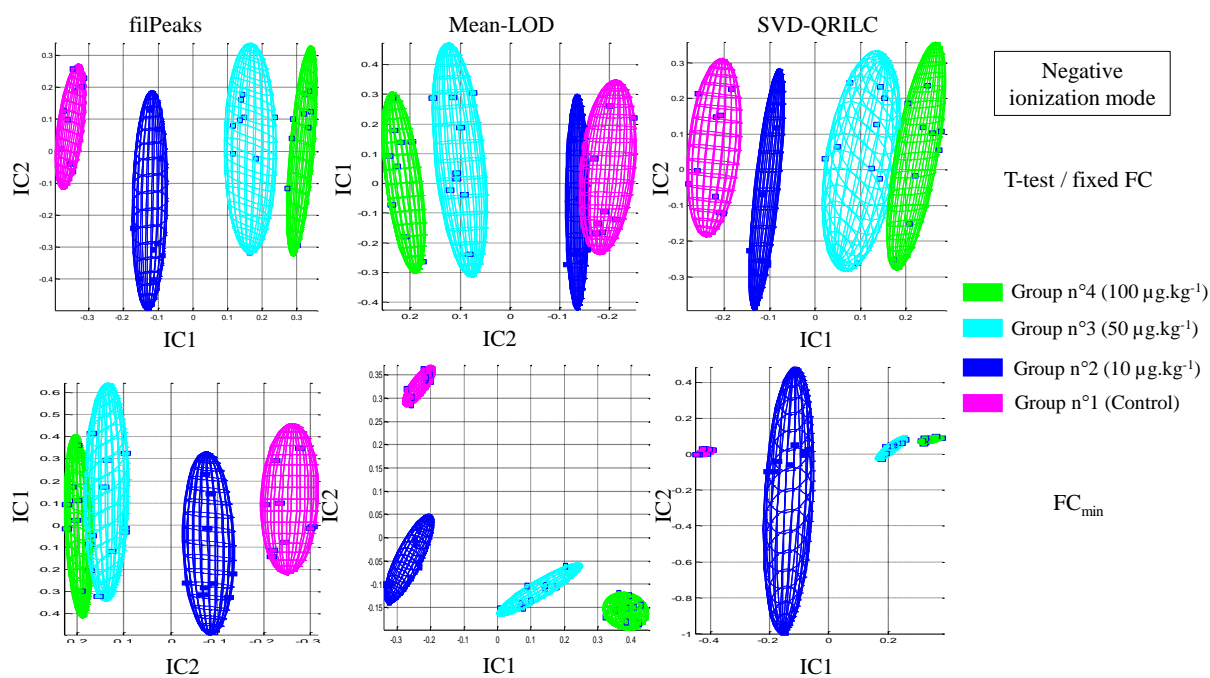


FIGURE 4.6 SCORE PLOTS OF ICA OUTPUT FOR DATA SET #1 ACQUIRED IN NEGATIVE IONIZATION MODE

3. CONCLUSION

Des résultats obtenus avec les six combinaisons de méthodes, on peut en tirer les conclusions suivantes. Premièrement, sur les trois méthodes de gestion des valeurs manquantes testées, seules deux présentent des performances satisfaisantes (`xcms.fillPeaks` et `mean-LOD`). La dernière (`SVD-QRILC`) montre de sérieuses limites lors de l'analyse du jeu de données plus complexe. Elle est aussi plus difficile à mettre en œuvre, nécessitant un changement de plateforme de calcul, alors que `xcms.fillPeaks` est implémentée sur W4M en ligne avec l'extraction des pics, et `mean-LOD` peut facilement être implémentée au cœur du processus dans Matlab. Ces résultats confirment la pertinence du choix de `xcms.fillPeaks` pour la complétion des valeurs manquantes, et soulignent le potentiel de `mean-LOD` pour compléter `xcms.fillPeaks`.

Les deux méthodes de filtration donnent des résultats satisfaisants, FC_{\min} permettant une plus forte réduction du nombre d'ions que `t-test/FC` (avec environ un facteur 10). Ces méthodes sont très simples à implémenter, il peut donc être pertinent de les implémenter en parallèle.

Une version finalisée du processus de traitement des données a donc été mise en place, avec, en plus de `xcms.fillPeaks`, une étape de gestion des zéros en tant que valeurs manquantes par `mean-LOD`. De plus, les deux méthodes de filtration sont implémentées en parallèle dans le but de favoriser la détection d'une contamination.

CHAPITRE 5
VALIDATION ET APPLICATION DE LA PREUVE DE
CONCEPT

1. INTRODUCTION ET RÉSUMÉ DE L'ARTICLE

L'approche preuve de concept mise au point au Chapitre 3 est maintenant dans sa forme finale, après les derniers développements présentés dans le Chapitre 4. Ainsi, une étape de gestion des valeurs aberrantes (zéros notamment) a été ajoutée en plus de celle déjà existante pour s'assurer que la matrice de données ne contienne pas de valeurs empêchant l'analyse. De plus, une deuxième stratégie de filtration des données (FC_{\min}) a été ajoutée en parallèle de celle déjà existante.

Pour sa validation, l'approche est appliquée sur deux scénarios de contamination du thé et deux plateformes LC-HRMS de technologies différentes (LC-ToF et LC-Q-Orbitrap). Pour cette étude, l'analyse des échantillons et des données est réalisé en aveugle, c.à.d. que les informations concernant les niveaux et la nature du dopage ne sont pas connues de l'opérateur.

Le premier scénario considéré a été conçu avec plusieurs groupes d'échantillons de thé vert dopés à deux niveaux (10 et 25 $\mu\text{g}/\text{kg}$) avec un mélange de dopage d'une dizaine d'herbicides plus un groupe d'échantillons de contrôle. Il vise à reproduire la situation utilisée pour le développement de l'approche avec des composés n'ayant jamais été analysés avec la méthode analytique (et à ce titre réellement inconnus) dopés à des niveaux plus bas que ceux déjà analysés (10 et 30 $\mu\text{g}/\text{kg}$ contre 10, 50 et 100 $\mu\text{g}/\text{kg}$ pour le jeu d'échantillons initial).

Le deuxième consiste en trois groupes d'échantillons de thé noir : un de contrôle, un dopé à 10 $\mu\text{g}/\text{kg}$ avec le mélange d'herbicide décrit ci-dessus, et un dopé à 10 $\mu\text{g}/\text{kg}$ avec un mélange de 3 contaminants sélectionnés parmi les 32 initiaux. Ce scénario est plus complexe et pensé pour être plus proche d'une situation de contamination « réelle », avec plusieurs contaminations différentes présentes simultanément dans les différents échantillons.

Les extraits des échantillons ont été successivement analysés sur une plateforme LC-ToF, puis sur un instrument de type LC-Q-Orbitrap avec le même gradient chromatographique. Les données provenant des deux instruments suivent le même processus de traitement de données décrit au Chapitre 3, enrichi des apports décrits au Chapitre 4.

2. ARTICLE

Untargeted food chemical safety assessment: a proof-of-concept on two analytical platforms and contamination scenarios of tea

Grégoire Delaporte, Mathieu Cladière, Valérie Camel*

UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300 Massy, France

** Corresponding author: AgroParisTech, 16 rue Claude Bernard, F-75005, Paris, France*

Phone: +33 1 44 08 17 25 – email: valerie.camel@agroparistech.fr

Submitted to Food Control

Abstract

This study aims at assessing the capability of comparing and combining different instrumental platforms in an untargeted approach with a view of detecting chemical contaminants in food matrices at low levels. A strategy based on liquid chromatography-high resolution mass spectrometry (LC-HRMS) and chemometrics has been applied on two different complex food contamination scenarios, with tea as study product. The first scenario aimed at mimic the presence of a dozen of contaminants at levels just above regulatory limits (i.e. 10 and 30 $\mu\text{g}/\text{kg}$); the second scenario, more complex, aimed at simulate the presence of several different contaminations at levels close to regulatory limits (10 $\mu\text{g}/\text{kg}$) in different samples. This work was carried on two LC-HRMS platforms (with respectively ToF and Orbitrap mass analyzer technologies), and a highly automated data treatment workflow was implemented to deal with data acquired on both platforms. The untargeted approach performed well on all scenarios (even the most complex) and analytical platforms. Performance comparison between LC-HRMS technologies was made possible thanks to a vendor-neutral data treatment process.

Keywords

Food safety; Independent Components Analysis; LC-HRMS; Orbitrap; ToF; XCMS

2.1. INTRODUCTION

Recent food safety crises like the presence of fipronil in European eggs (summer 2017) underline the limits of current targeted analytical approaches and the need for new untargeted methods able to point out such non-expected contaminants. To that end, mass spectrometry (MS)-based untargeted approaches were identified as having the strongest potential (Antignac et al., 2011; Castro-Puyana & Herrero, 2013; Lommen et al., 2007), giving promising results on relatively simple contamination scenarios: high levels of contaminants (around mg/kg) and rather simple matrices [orange juice (Tengstrand, Rosén, Hellenäs,

& Åberg, 2013) or infant formulas (Inoue et al., 2015)]. Only a few recent studies have reported the detection of contaminants at levels down to 10 µg/kg (Cotton et al., 2014; Delaporte, Cladière, Jouan-Rimbaud Bouveresse, & Camel, 2019; Knolhoff, Zweigenbaum, & Croley, 2016; Kunzelmann, Winter, Åberg, Hellenäs, & Rosén, 2018) in more complex matrices such as honey (Cotton et al., 2014) or tea (Delaporte et al., 2019). Further developments are required, especially to assess the ruggedness of the proposed approaches (especially concerning the data treatment part) regarding both the instrument used and contamination scenarios studied. This research field is facing a huge interest worldwide (Antignac et al., 2011; Castro-Puyana, Pérez-Míguez, Montero, & Herrero, 2017).

The general workflow (see Figure 5.1) for untargeted food chemical safety assessment can be established based on reviews (Antignac et al., 2011; Castro-Puyana et al., 2017; Knolhoff & Croley, 2016). As a general rule, the first step consists in a broad-range sample treatment method followed by UHPLC-HRMS analysis (Cotton et al., 2014; Delaporte et al., 2019; Inoue et al., 2015; Knolhoff et al., 2016; Kunzelmann et al., 2018; Tengstrand et al., 2013). For the next step (data treatment), dedicated tools should be implemented. The use of in-line proprietary tools, often supplied with the instrument, has been reported (Inoue et al., 2015; Knolhoff et al., 2016), with the advantage of being user-friendly and to “fit-for-purpose” the data files generated by the instrument. However, such tools lack flexibility and versatility, making it impossible to analyze data from various instruments or to implement in-house data analysis methods. Other strategies are based on open-source tools (such as the XCMS R package) (Cotton et al., 2014; Delaporte et al., 2019). On top of that, easy-to-use user interfaces and free-to-use online calculation platforms were developed recently, such as web interfaces for the XCMS package, namely Workflow4Metabolomics (Giacomoni et al., 2015) and XCMS-online (Tautenhahn, Patti, Rinehart, & Siuzdak, 2012). These tools are often designed to be versatile and modular, meaning that they may handle data from multiple instruments and be adapted to new fields and problematics.

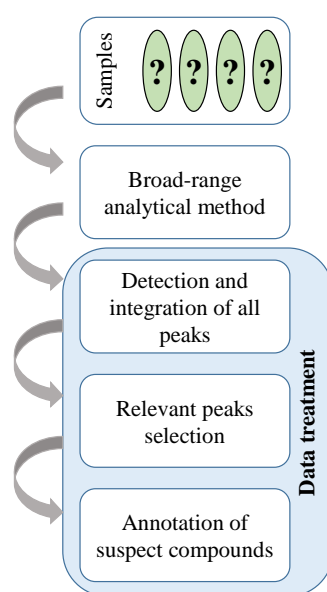


FIGURE 5.1 GENERIC WORKFLOW USED FOR UNTARGETED FOOD SAFETY ASSESSMENT

Until now, the implementation of multiple LC-HRMS platforms has been assessed only on targeted food safety approach (Saito-Shida, Hamasaka, Nemoto, & Akiyama, 2018). Even though the comparison of multiple analytical platforms for untargeted approaches has been reported in metabolomics studies (Glauser, Veyrat, Rochat, Wolfender, & Turlings, 2013), it has never been done in the field of untargeted food safety analysis, existing untargeted approaches being developed on a single instrument. As a consequence, there is a need to test data treatment capability when considering different instrumental platforms since each data generated has its own characteristics (m/z resolution and accuracy, signal/noise ratio among others). In addition, untargeted methods were mostly developed on a single data set, with the notable exceptions of the method proposed by Tengstrand et al. (Tengstrand et al., 2013) later successfully applied on a more complex case (Kunzelmann et al., 2018), and of our developed method applied on two contamination scenarios (Delaporte et al., 2019).

To assess ruggedness of those methods, several contamination scenarios should be considered that mimic more complex and realistic food safety applications (e.g. multi-class contaminations, levels and molecules involved, variable contamination between samples).

So, this work aims at assessing the implementation and ruggedness of an untargeted approach for detection of non-expected food contaminants. It is based on open-source tools and methods, previously developed on green tea samples using a single LC-HRMS platform (Delaporte et al., 2019). Here, two complex contamination scenarios are considered (with green and black tea leaves), and all analysis were conducted on two different platforms, i.e. UHPLC-Q-Orbitrap and UHPLC-ToF which are the most used HRMS technologies for untargeted contaminants detection thanks to their mass accuracy and resolution (Castro-Puyana et al., 2017; Knolhoff & Croley, 2016).

2.2.MATERIAL & METHODS

2.2.1. REAGENTS AND SAMPLE COLLECTION

Acetonitrile (ACN) (HPLC plus gradient, LC/MS), water, methanol (MeOH) and formic acid (FA) (all LC/MS grade) were purchased from Carlo Erba. Ultrapure water (Milli-Q®) was produced by an Integral 3 water purification system from Millipore®. The ToF mass spectrometer was calibrated using Leucine Enkephalin (LC/MS grade, Waters®) and the Orbitrap with the Pierce™ calibration solutions (Thermo Fisher Scientific, caffeine 2 $\mu\text{g.mL}^{-1}$, MRFA 1 $\mu\text{g.mL}^{-1}$, Ultramark 1621 0.001% and *n*-butylamine 0.0005% for positive mode; SDS 2.9 $\mu\text{g.mL}^{-1}$, taurocholate 5.4 $\mu\text{g.mM}^{-1}$ and Ultramark 1621 0.001% for negative mode).

Analytical standards solutions (100 $\mu\text{g/mL}$ in ACN or MeOH, purity >97%) of malathion, ochratoxin A and bisphenol S or individual labelled compounds (acrylamide-d3, bisphenol A-d14, dimethoate-d6 or malathion-d6) were purchased from CIL Cluzeau France. All other standard solutions (100 $\mu\text{g/mL}$ in ACN or MeOH), including herbicide mix (100 $\mu\text{g/kg}$ in ethyl acetate, purity >98%) were provided by

Sigma Aldrich (Saint-Quentin Fallavier, France). A pooled stock solution containing all labelled molecules (each at 1 µg.mL⁻¹) was prepared in ACN and stored in the fridge.

Loose tea samples (green and black) were purchased at local retailers. The green tea was an organic Bancha tea from Japan, and the black tea an organic Keemun tea from China.

A standard solution (called stability mix) containing 32 known contaminants from different families at 16 ng/mL (pesticides, mycotoxins, migrants from packaging and process-induced toxicants) was used to check the stability of the instrument before each analytical sequence (Cladière, Delaporte, Le Roux, & Camel, 2018).

2.2.2. STUDY SET-UP

The study was carried out by two different persons: a study designer who set up the experimental design and spiked the samples, and an operator who did the chemical and data treatment analyses. In particular, the composition and levels of contaminants of the different sample groups were not known by the operator, from the beginning of sample extraction to final compounds annotation. In addition, most of compounds studied were analyzed for the very first time using our method and therefore nor the operator, neither the study designer had any background regarding the overall method efficiency for the analyzed compounds (only that they are LC-MS amenable). After all the analyses, the results were sent back by the operator to the study designer for detection performance assessment.

Two different contamination scenarios have been established (Table 5.1) using two spiking mix. Spiking mix n°1 is composed of 11 herbicides in ACN, and spiking mix n°2 of three food contaminants from three different classes (malathion for pesticides, ochratoxin A for mycotoxins and bisphenol S for migrants from packaging) in ACN. Details on spiked contaminants such as raw formula, physico-chemical properties and respective European Maximum Residue Limits (MRLs) when applicable can be found in Table 5.7. Spiking levels were chosen in accordance with EU regulations n° 396/2005 (pesticides) and 1881/2006 (other contaminants). Each sample was spiked using the following procedure: 1 g of sample was firstly weighted in a centrifuge polypropylene tube (Corning, New York, USA) and the spiking mix added using the smallest volume possible (below 100 µL); 100 µL of ACN was added to the unspiked samples. For quality control purpose, 40 µL of the labelled spiking mix were added to all samples (leading to a concentration of 40 µg/kg of labelled compounds in the samples). Each sample was then mixed using a vortex and left for equilibration during 2 hours at room temperature, and then put in the fridge overnight.

While the operator just had information about the group's number (group n°1, 2 or 3) for black and green teas (to ensure real blind analysis relative to the contaminants), two different contamination studies were designed as indicated in Table 5.1. The first scenario, considered as "simple", consists in three groups (of three samples each) from the same green tea; two of these groups were spiked with the mix n°1 (one at 10 µg/kg, which corresponds to the default safety MRL in the EU legislation, and one

at 30 µg/kg), the third being used as a control group. The second scenario, considered more complex, consists in three groups (of three samples each) from the same black tea; one group was spiked at 10 µg/kg with the mix n°1, another one at the same level with the mix n°2, and the last one used as a control.

TABLE 5.1 CONTAMINATION SCENARIOS CONSIDERED

Group number	Scenario n°1 <i>Green tea</i>	Scenario n°2 <i>Black tea</i>
1	10 µg/kg with mix n°1	10 µg/kg with mix n°1
2	Unspiked	10 µg/kg with mix n°2
3	30 µg/kg with mix n°1	Unspiked

Here we aimed to test the capability of our untargeted approach to discriminate low levels of the same contamination, enabling detection of non-conformities (scenario n°1), and also to detect unexpected contaminants in samples that do comply with the regulation and face different contamination patterns (scenario n°2).

2.2.3. SAMPLE TREATMENT

Samples were extracted using a method based on previous work (Cladière et al., 2018): 5 mL of an ACN/MeOH (90/10 v/v) mixture acidified with 0.1% FA were added to each sample for extraction. Tubes were then agitated upside-down on an agitator plate for 1 hour, and centrifuged at 3,000 g for 10 minutes. Then 2 mL of the supernatant were collected and evaporated to dryness under a gentle nitrogen stream at 35°C. The extract was reconstituted in 0.2 mL of ACN acidified with 0.1% FA, and completed by 0.8 mL of H₂O acidified with 0.1% FA, leading to a total volume of 1 mL. The reconstituted extract was then centrifuged at 12,000 g for 10 minutes, 0.5 mL collected and filtered at 0.2 µm using a syringeless filter vial (mini-uniprep G2, Whatman) before analysis. For each type of tea (green and black), a quality control sample (QC) was made by pooling together in a glass tube 0.2 mL of each final extract considered in the analytical sequence, from which a 0.5 mL aliquot was sampled and filtered using a syringeless filter vial.

2.2.4. UHPLC-HRMS METHODS

Samples were analyzed on two UHPLC/HRMS platforms. The first is a Waters® Acquity UPLC® H-Class system, composed of a quaternary solvent manager pump, a refrigerated sample manager Flow-Through-Needle and a column oven, coupled to a Waters® high resolution Time-of-Flight mass spectrometer Xevo® G2-S ToF operated in centroid mode (resolution of 30,000 FWHM at m/z 200, m/z range from 60 to 800, 2 scans/s) using an electrospray ion source (ESI). The second is a Thermo Scientific UltiMate 3000 UHPLC system composed of quaternary pumps, refrigerated auto-sampler and a column oven coupled to a Q-Exactive Orbitrap mass spectrometer operated in centroid mode (resolution of 70,000 FWHM at m/z 200, m/z range from 60 to 800) with a heated electrospray ion source

(HESI). For the Orbitrap MS method “AGC target” and “Maximum IT” parameters were set to 10^6 and 200 ms respectively.

For each platform, the injection volume was 10 μ L. Separation was made on a C18-PFP column (150 \times 2.1 mm, 2 μ m particles diameter, ACE supplied by AIT France), and the same chromatographic gradients were used on both instruments. For ESI+, the mobile phase was composed of water (A) and ACN (B), both acidified with 0.1% FA, and MeOH (C), flowing at 0.4 mL.min⁻¹. Gradient started at 100% A and reached 100% B in 10 min, being kept for 6 min before switching to 100% C to rinse the system in 1 min, being hold for 5 min, returning back to 100% A in 1 min and finally equilibrating for 3 min, with a total run duration of 26 min. For ESI-, the mobile phase was composed of water buffered at pH 6.45 with 10 mM of ammonium formate (A) and MeOH (B) flowing at 0.3 mL.min⁻¹. The gradient started at 100% A and reached 100% B in 13 min, holding this condition for 7 min before turning back to 100% A in 1 min and finally equilibrating for 3 min, with a total run duration of 24 min. For both chromatographic methods the temperature of the column oven was kept at 30°C. Parameters for electrospray ion sources can be found in Table 5.8 & Table 5.9.

Each sequence started with 10 mobile phase injections and 5 stability mix injections to equilibrate the instrument (Dunn et al., 2011). Sample injection orders were randomized, and QC samples as well as blanks injected every 15 samples. For each ionization mode and instrument, each sample was analyzed in triplicate.

2.2.5. DATA TREATMENT WORKFLOW

Before any untargeted analysis, the overall method quality (extraction + UHPLC-MS analysis) was visually assessed for each injection replicate thanks to the Total Ion Current (TIC) and the peak intensity of each labelled molecule. This step only aimed at the early detection of analytical outliers and did not replace the quality control procedure that will be implemented afterwards.

A highly automated data treatment workflow based on previous work (Delaporte et al., 2019) has been implemented here (Figure 5.2). Vendors raw data files were first converted to an open-source format (.mzXML) with the help of ProteoWizard module “MSConvert” (Chambers et al., 2012) using a noise threshold of 100 for ToF data, and then uploaded onto the Workflow4Metabolomics (W4M) computation platform (Giacomini et al., 2015). Data matrices were built using XCMS R package (Smith, Want, O’Maille, Abagyan, & Siuzdak, 2006). Parameters for XCMS algorithm were inspired by those suggested by Patti (Patti, Tautenhahn, & Siuzdak, 2013) for high resolution UHPLC-Q-ToF and UHPLC-Orbitrap instruments. The full list of parameters for both Orbitrap and ToF data files can be found in Table 5.10. In-line with XCMS, data sets were quickly visualized using the “Quality Metrics” module of W4M. A LOESS (“Locally Estimated Scatterplot Smoothing”) analytical drift correction on ion intensities (Van Der Kloet et al., 2009) was applied when needed as spotted in Figure 5.5.

In-house tools and methods were then implemented to achieve the detection of contaminants. As shown in Figure 5.2, remaining zero values were first imputed (especially for Orbitrap data) since some methods used afterwards are sensitive to zeros in the data matrix. A first missing value completion method, fillPeaks, was applied within XCMS as shown in Figure 5.2. It relies on the forced integration of missing ions. However, especially in low-noise data, we observed that this method tends to generate a lot of zero values when no peaks could be found. Therefore, we implemented a method to impute those remaining missing values, which appeared as zeros in our data matrices. The presence of injection replicates was used to impute them by the best value possible (Delaporte, Cladière, & Camel, submitted). Briefly, for each ion (identified in the data matrix by their combination “retention time- m/z ”), if a zero is found in only one injection replicate, it is imputed by the mean of the replicates. If more than one replicate has a zero value, they are imputed by the estimated limit of detection of the instrument [i.e. mean of the 3% lowest values, zeros excluded (Libiseller et al., 2015)].

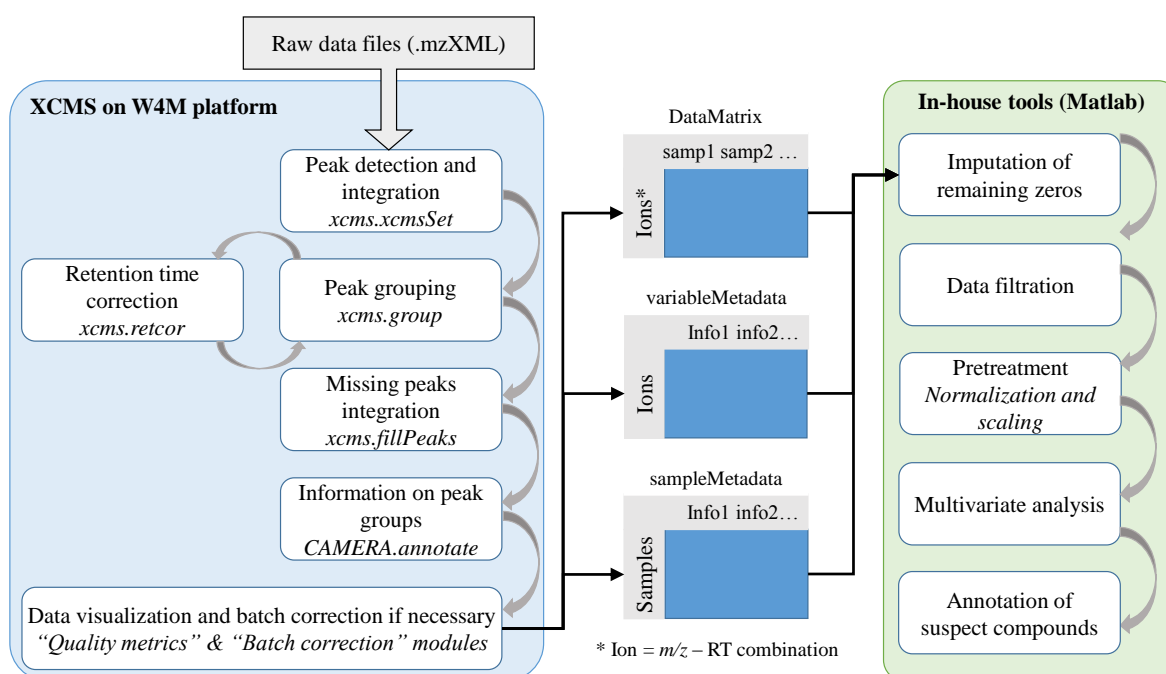


FIGURE 5.2 UNTARGETED DATA TREATMENT WORKFLOW

It was then necessary to reduce the number of ions present in the data matrix (several thousands). To that end, two complementary filtration strategies were implemented in parallel. The first (which will be called “t-test”) relies on the implementation of univariate t-tests between sample groups and blank injections to remove ions already present in blanks, and then between groups to remove ions showing no significant differences between groups. These tests are followed by the calculation of the fold change (FC, calculated for each ion as the ratio of the medians of the highest group over the lowest). Fixed thresholds are used to filter the data matrix, respectively <0.05 for p-values (of “t-test”) and >2 for FC. The alternative filtration strategy relies instead on the calculation, for each ion, of a relevant minimum FC (FC_{\min}) from which one can assume that a significant signal difference is observed between groups

(Ortmayr, Charwat, Kasper, Hann, & Koellensperger, 2017). FC_{\min} calculation relies on the estimation of the uncertainty on FC based on the relative standard deviations observed for the considered samples. Only ions exhibiting a FC higher than the FC_{\min} are kept.

Data matrices then go through several preprocessing methods. They were log- and pareto-scaled (Antignac et al., 2011; Delaporte et al., 2019) and normalized with a median-based Probabilistic Quotient Normalization (PQN) based on QC samples (Delaporte et al., 2019; Dieterle, Ross, Schlotterbeck, & Senn, 2006). PQN aims at limiting the influence of potential dilution effects occurring in the study, whereas the previously used LOESS method aims at correcting a signal intensity drift, caused for example by the fouling of the ion source during the study.

At this point, the data matrix is most often composed of a hundred to few thousands ions (with the notable exception of ToF data in positive ionization mode, for which only five ions were remaining). As a consequence, to visualize the data and assess the presence of trends and patterns, it is necessary to implement data analysis methods to reduce its dimensionality. In metabolomics studies, this step is typically done using multivariate methods (Gorrochategui et al., 2016). Independent Component Analysis (ICA) was the method selected here to visualize the data based on the remaining ions (Delaporte et al., 2019). This method was implemented using the JADE algorithm (Rutledge & Jouan-Rimbaud Bouveresse, 2015) and the optimal number of ICs to compute determined with Random-ICA method (Kassouf, Jouan-Rimbaud Bouveresse, & Rutledge, 2017). ICA is a blind source signals decomposition method, that gives two main outputs, respectively signals (representing the pure source signals found in the data matrix) and scores (representing the weight of each sample in the different signals). The scores make it possible to visualize a potential separation of samples, and then the signals enable to link a sample group separation to specific ions. For example, if a sample group separation is observed along component $n^{\circ}1$ (IC1) (with sample group $n^{\circ}1$ having higher scores than sample group $n^{\circ}2$), signals constituting IC1 will be sorted in descending order and the corresponding discriminating ions will be selected.

All discriminating ions were then annotated using automated in-house data-mining tools to detect isotopic patterns and adducts based on previous work (Cotton et al., 2014), and they were matched against a broad-range in-house database containing around 2,000 known toxicants and their most probable adducts in LC-HRMS (Delaporte et al., 2019). The presence of the spiked contaminants has not been assessed in the database before the study to maintain their blind character. Finally, ions selected thanks to the multivariate analysis were manually curated to highlight suspect compounds.

2.2.6. PERFORMANCE ASSESSMENT

As already mentioned the operator carried out all the instrumental analysis and the data treatment process without any knowledge on the possible contamination of the samples (real blind procedure). At the end of the process, this operator issued a summary with two main information: which molecules are

suspected to cause a group separation (with their respective m/z , retention time and putative annotation), and for each, in which group the concentration is the highest. Percentages of molecules detection for each experiment were then assessed, by returning back the results to the study designer who spiked the samples. Two detection percentages were computed from these results by the designer of the study as shown in. The first was calculated as the ratio between the number of blindly annotated contaminants and the number of contaminants actually detected in the spiked samples by a manual peak detection made by the study designer: it represents the success rate of the blind data treatment approach to detect and characterize a suspect signal in the data matrix. The second was calculated between the number of blindly annotated molecules against the number of LC-ESI-MS amenable molecules, defined as the number of molecules detected by a targeted analysis (conducted by the study designer) of a highly concentrated standard solution (500 ng/mL) with our method. This percentage characterizes the sensitivity of the global methodology (analytical and data treatment methods) to detect a contamination.

2.3. RESULTS & DISCUSSION

Raw data sets have been deposited to the EMBL-EBI MetaboLights database (DOI: 10.1093/nar/gks1004. PubMed PMID: 23109552) with the respective identifiers MTBLS771 and MTBLS772 (Haug et al., 2013). The complete data sets can be accessed at <https://www.ebi.ac.uk/metabolights/MTBLS771> and <https://www.ebi.ac.uk/metabolights/MTBLS772>.

2.3.1. DATA SETS METRICS AND PRETREATMENT

After peak extraction from raw data files by XCMS, the main metrics (number of ions and percentage of zero values) were calculated for each data set (see Table 5.2). All data sets appeared to be rather similar in terms of number of ions or percentage of zeros, regardless of the instrument used. An analytical drift on peak intensities was clearly visible on ToF data sets (an example is displayed in Figure 5.5) and was corrected using the LOESS method. No drift was observed on measured retention times and m/z whatever the data set considered.

TABLE 5.2 DATA SETS AVAILABLE AND THEIR RESPECTIVE METRICS

Matrix	UHPLC-HRMS platform	Ionization mode	Analytical drift correction	Number of ions before filtering	% zeros
Green tea contamination mix n°1	ToF	ESI+	Yes (LOESS)	15,548	6.45
		ESI-		28,275	4.76
	Q-Orbitrap	ESI+	No	16,053	8.33
		ESI-		17,646	9.69
Black tea contamination mixes n°1 & 2	ToF	ESI+	Yes (LOESS)	11,912	6.57
		ESI-		21,461	4.45
	Q-Orbitrap	ESI+	No	12,736	4.46
		ESI-		12,976	5.62

For ToF data, about twice as many ions were detected in negative ionization mode as compared to positive mode. However, it does not seem to have an impact on the percentage of zeros, which remains very similar between all data sets. A group-wise study of percentages of zeros reveals that, as expected, blank injections contain a higher proportion of zeros (between of 20 and 50% of values) than spiked tea samples (between 1 and 5 % of values). Still, the presence of zeros may lead to errors in the following steps of the process (especially for univariate statistics and log transformation), as a consequence they must be imputed. Zeros, i.e. missing values (see in 2.5) were first classified according to their nature thanks to the methodology described in 2.5 based on the presence of injection replicates, and then imputed either by the mean of replicates or by an estimated instrumental noise.

2.3.2. CONTAMINATION CASE N°1

Sample groups were successfully separated thanks to the multivariate analysis by ICA for both polarities and instruments. Results for ToF data are displayed in Figure 5.3 (and in Figure 5.6 for Orbitrap data). A clear concentration trend can be observed between sample groups on the discriminating component.

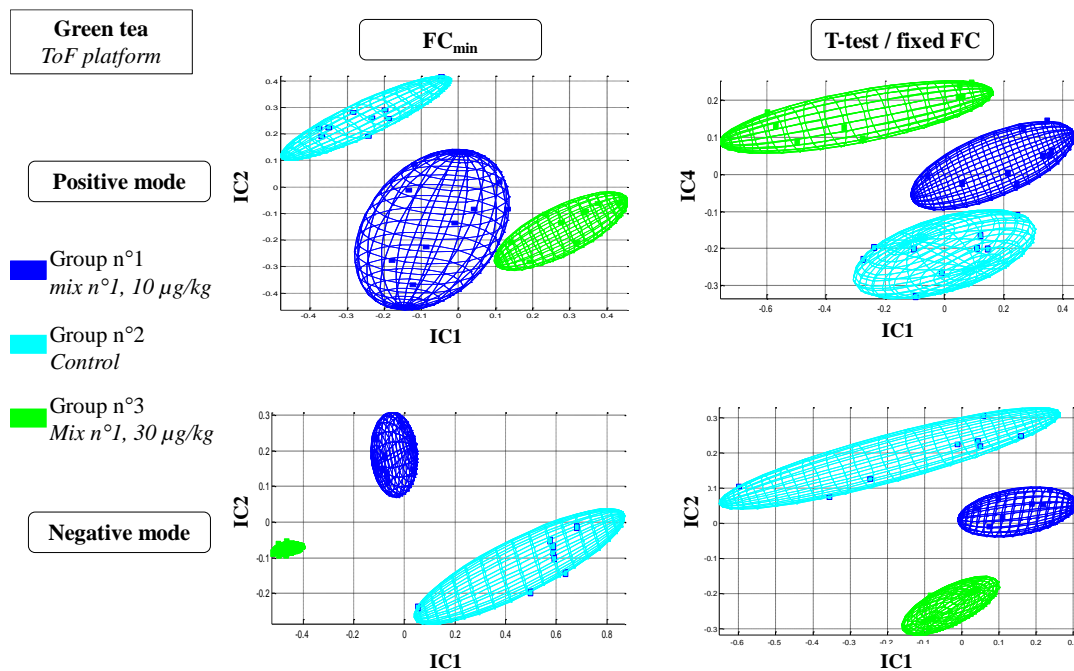


FIGURE 5.3 SCORE PLOT OF ICA OUTPUT FOR GREEN TEA SAMPLES ANALYZED ON TOF PLATFORM

For each filtration method applied, one or several discriminating components were found, the corresponding ions annotated and detection rates calculated (see Table 5.3 for annotated contaminants and Table 5.4 for detection rates).

TABLE 5.3 ANNOTATED CONTAMINANTS IN GREEN TEA SAMPLES

Measured mono-isotopic mass of adduct	Measured retention time (min)	Ionization mode	Most intense adduct	Proposed raw formula	Proposed mono-isotopic mass of compound	Mass error (ppm)	Proposed putative annotation	Detected in groups	Relative intensities	Filtration method
Orbitrap										
216.1009	7.83	POS	[M+H] ⁺	C ₈ H ₁₄ ClN ₅	215.0938	-0.61	Atrazin	3 & 1	gp3 > gp1	FC _{min} / t-test
216.1414*	10.22	POS	[M+H]⁺	C₁₁H₂₁NOS	215.1344	-1.38	Cycloate	3 & 1	gp3 > gp1	FC_{min} / t-test
253.1655	6.72	POS	[M+H] ⁺	C ₁₂ H ₂₀ N ₄ O ₂	252.1586	-1.53	Hexazinone	3 & 1	gp3 > gp1	FC _{min} / t-test
215.0960	7.30	POS	[M+H] ⁺	C ₈ H ₁₄ N ₄ OS	214.0888	-0.37	Metribuzin	3 & 1	gp3 > gp1	FC _{min} / t-test
188.1105	8.94	POS	[M+H] ⁺	C ₉ H ₁₇ NOS	187.1031	0.50	Molinate	3 & 1	gp3 > gp1	FC _{min} / t-test
261.0229 / 259.0085**	7.01 / 12.34	POS / NEG	[M+H] ⁺ / [M-H] ⁻	C ₉ H ₁₃ BrN ₂ O ₂	260.0160	-1.73 / 3.18	Bromacil	3 & 1	gp3 > gp1	FC _{min} / t-test
215.0587	12.69	NEG	[M-H] ⁻	C ₉ H ₁₃ ClN ₂ O ₂	216.0666	2.37	Terbacil	3 & 1	gp3 > gp1	FC _{min} / t-test
ToF										
216.1016	6.52	POS	[M+H] ⁺	C ₈ H ₁₄ ClN ₅	215.0938	2.78	Atrazin	3 & 1	gp3 > gp1	FC _{min} / t-test
282.2797***	10.32	POS	[M+H] ⁺	C ₁₈ H ₃₅ NO	281.2719	1.91	Dodemorph	1; 2 & 3	gp1 > gp2 > gp3	t-test
253.1669	5.44	POS	[M+H] ⁺	C ₁₂ H ₂₀ N ₄ O ₂	252.1586	4.01	Hexazinone	3 & 1	gp3 > gp1	FC _{min} / t-test
215.0967	6.02	POS	[M+H] ⁺	C ₈ H ₁₄ N ₄ OS	214.0888	2.90	Metribuzin	3 & 1	gp3 > gp1	FC _{min} / t-test
188.1119	7.57	POS	[M+H] ⁺	C ₉ H ₁₇ NOS	187.1031	8.21	Molinate	3 & 1	gp3 > gp1	FC _{min} / t-test
259.0085	10.39	NEG	[M-H] ⁻	C ₉ H ₁₃ BrN ₂ O ₂	260.0160	3.13	Bromacil	3 & 1	gp3 > gp1	FC _{min} / t-test
215.0586	10.73	NEG	[M-H] ⁻	C ₉ H ₁₃ ClN ₂ O ₂	216.0666	1.99	Terbacil	3 & 1	gp3 > gp1	FC _{min} / t-test

* Detected only with Orbitrap mass analyzer

** Detected in both positive and negative ionization mode

*** False positive, not present in the spiking solution

TABLE 5.4 DETECTION RATES FOR CONTAMINATION CASE N°1.

UHPLC-HRMS platform	Mode	Rate			Rate
		[found compounds / added compounds]	<i>t-test / fixed FC</i>	<i>FC_{min}</i>	<i>Global</i>
ToF	ESI+	36%	36%	55%	86%
	ESI-	18%	18%		
Q-Orbitrap	ESI+	55%	55%	64%	100%
	ESI-	18%	18%		

Hence, UHPLC-ToF platform enabled the successful detection of 4 contaminants (corresponding to 36% of the spiked molecules) in positive mode and of two others (corresponding to 18% of the spiked molecules) in negative mode. Similar results were obtained with *t-test / fixed FC* and *FC_{min}* filtration strategies. Results from both polarity modes and data filtration strategies were gathered and a global detection rate of 55% was calculated for ToF platform. A targeted search of spiked contaminants in the raw data files shows that only 7 of them can actually be seen (in which only 5 exhibit a signal/noise ratio above 3 for the 10 µg/kg level). It means that the main loss of information came from the analytical step, the data treatment being able to detect and annotate 6 contaminants out of 7 (i.e. 86%) actually visible in the raw chromatograms. A contaminant (dodemorph, see Table 5.3) was initially detected with the *t-test* based filtration approach, but appeared to be a false positive. This annotation was unsure from the start since it had not been reported with *FC_{min}* filtration.

On the UHPLC-Q-Orbitrap platform, 6 contaminants (i.e. 55% of the spiked molecules) were blindly detected in positive mode and 2 others in negative mode, leading to a total detection rate of 64%, which is slightly better than with the ToF platform. Interestingly, one contaminant could be annotated only in the Orbitrap data set as displayed in Table 5.3. All annotated spiked contaminants can be detected at levels as low as 10 µg/kg in samples, which is relevant regarding existing EU MRLs. Interestingly, the data treatment success rates rise to 100% for Orbitrap platform, meaning that all compounds present in the raw data files were successfully annotated by our methodology.

The sample diversity considered here is rather low since all samples come from the same brand and production batch. This topic has been discussed previously (Delaporte et al., 2019; Knolhoff et al., 2016), but the need for a control, unspiked sample for relevant signals selection is an issue that has not been solved yet despite of its critical aspect for untargeted food safety assessment. Moreover, this study relies on spiked samples, which also may lead to better detection rates than if natively contaminated (i.e. “real”) samples were analyzed. In fact, the extraction of contaminants trapped within the sample matrix (and not solely present on its surface, as in the case of spiked samples) may induce more matrix effects and lower recoveries of compounds of interest that will probably make the implementation of the approach more difficult. The solving of these issues would surely imply further adjustments concerning the analytical method (namely the sample treatment step) and possibly the data treatment process. In a

final step, ring studies will be necessary as it is done for traditional targeted approaches and in the field of metabolomics studies. Yet, our results are still highly promising since the detection rates obtained are within the same order of magnitude of the ones presented in similar studies (Kunzelmann et al., 2018). On top of that, several contaminants studied here have never been analyzed with our method, so that their potential analytical responses and detectability were truly “unknown”.

2.3.3. CONTAMINATION CASE N°2

In this more complex contamination scenario, group separation was also achieved. However, this time the group separation pattern observed in all cases suggests the presence of different contaminations among sample groups, and not a single contamination present at several levels as in scenario n°1. This highlights the interest of using ICA here, since it produces independent components, meaning that the ions combination used to build IC1 should be different from the combination used to build IC2. This is particularly visible on data from Orbitrap platform (in Figure 5.4).

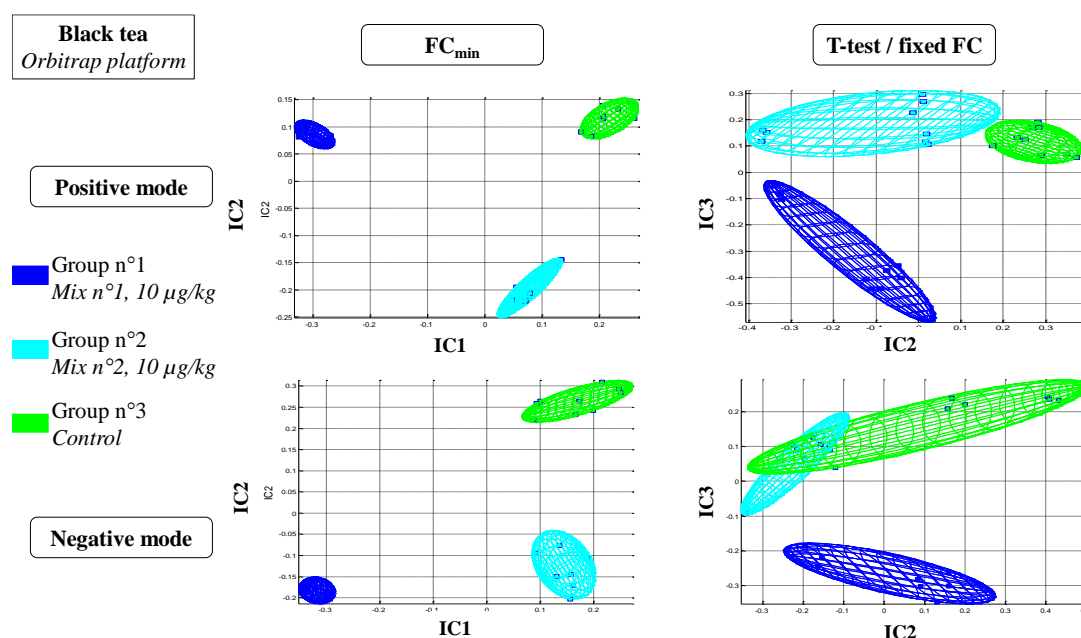


FIGURE 5.4 SCORE PLOTS OF ICA OUTPUT ON BLACK TEA SAMPLES ANALYZED ON ORBITRAP PLATFORM

Again, the two previously described filtration methods (t-test/ fixed FC and FC_{min}) were implemented, but some combinations of instrument/polarity/filtration method do not enable a clear sample separation, and some even lead to the selection of too few ions to perform a multivariate analysis (see Figure 5.4 for Orbitrap data and Figure 5.7 for ToF data). For example, for positive ionization mode data acquired on the ToF platform, the FC_{min} filtration method leads to the selection of 5 ions, thus multivariate analysis was not needed and each ion was putatively annotated individually.

The annotation of discriminating ions is completed by the visualization of ion intensity differences among sample groups thanks to an in-house Matlab script to assess the relative levels of annotated contaminants. The final outcome of this process is presented in Table 5.5. A separate detection rate has been then calculated for each spiking mix (see Table 5.6).

The contamination pattern of group n°1 was first investigated thanks to ICA decomposition. On the UHPLC-ToF platform, 3 contaminants can be tentatively annotated in positive ionization mode, and 2 others in negative mode, corresponding to detection rates of 27% and 18% respectively, leading to 45% when considering both modes which is lower than previously observed on green tea. In fact, the detection rate in scenario n°1 is favored by the presence of a group with a “high” spiking level (30 $\mu\text{g}/\text{kg}$), not present in the scenario n°2. Consequently, one compound (molinate) is missed in case n°2 (but found in case n°1) since it exhibits a signal / noise ratio close to the limit of detection of the ToF platform. A similar decrease of detection rate was reported by Kunzelmann (Kunzelmann et al., 2018) from 89% detection at 25 $\mu\text{g}.\text{kg}^{-1}$ to less than 35% at 5 $\mu\text{g}.\text{kg}^{-1}$. Interestingly, more contaminants can be annotated using the UHPLC-Q-Orbitrap platform.

Overall, same performances were achieved as for contamination scenario n°1, with 6 contaminants annotated in positive mode (i.e. 55% of spiked molecules) and 2 others in negative mode (i.e. 18% of spiked molecules) in sample group n°1. Globally, 7 contaminants (out of 11, corresponding to a detection rate of 64%) were successfully annotated in this sample group, and 100% of compounds present in the raw data files were successfully annotated by our methodology.

Then group n°2 was proposing a completely different contamination pattern as shown in Table 5.5. On the ToF platform, this contamination has been detected only in negative mode, with two contaminants putatively annotated (ochratoxin A and bisphenol S). The missing contaminant (malathion, normally amenable in positive mode) was not detected due to a high analytical drift on signal intensity that cannot be corrected satisfyingly. Interestingly, both LC-MS platforms show a real complementarity since malathion can be annotated in positive mode using Orbitrap data. However, this time bisphenol S and ochratoxin A cannot be annotated, even using a targeted approach (see in Table 5.5). This is not due to an issue in the data treatment process or to an analytical drift as for ToF data, but to a lack of sensitivity of the method (the peak cannot be seen at all in the raw data files) maybe caused by strong matrix effect inherent to the tea matrix. These results indicate that the approach benefits from the presence of numerous contaminants since the detection of the contamination with 3 compounds was more difficult than with 11 compounds. With fewer contaminants, the outcome of the approach seems indeed more vulnerable to potential failures in the analytical process. It is the first time this fact is highlighted since it did not occur in previous studies focusing on few contaminants (Delaporte et al., 2019, Knolhoff et al., 2016). Ensuring their ruggedness seems to be one of the key stake of those approaches, especially since untargeted food safety studies could imply the detection of a single contaminant, for which the analytical response may not be optimal.

These results are very encouraging since despite the presence of two different contaminations at very low levels (only 10 $\mu\text{g.kg}^{-1}$), satisfactory detection rates were achieved on both analytical platforms. Even with this complex situation, annotation rates are similar with those found for already existing, simpler cases (Kunzelmann et al., 2018).

TABLE 5.5 ANNOTATED CONTAMINANTS IN BLACK TEA SAMPLES

Measured mono-isotopic mass of adduct	Measured retention time (min)	Ionization mode	Most intense adduct	Proposed raw formula	Proposed mono-isotopic mass of compound	Mass error (ppm)	Proposed putative annotation	Detected in groups	Relative intensities	Filtration method
Orbitrap										
353.0243*	9.54	POS	[M+Na]⁺	C₁₀H₁₉O₆PS₂	330.0361	-2.96	Malathion	2	N/A**	FCmin
216.1009	7.83	POS	[M+H] ⁺	C ₈ H ₁₄ ClN ₅	215.0938	-0.61	Atrazine	1	N/A	FCmin / t-test
216.1414*	10.22	POS	[M+H]⁺	C₁₁H₂₁NOS	215.1344	-1.38	Cycloate	1	N/A	FCmin / t-test
253.1655	6.72	POS	[M+H] ⁺	C ₁₂ H ₂₀ N ₄ O ₂	252.1586	-1.53	Hexazinone	1	N/A	FCmin / t-test
215.0960	7.30	POS	[M+H] ⁺	C ₈ H ₁₄ N ₄ OS	214.0888	-0.37	Metribuzin	1	N/A	FCmin
188.1105*	8.94	POS	[M+H]⁺	C₉H₁₇NOS	187.1031	0.50	Molinate	1	N/A	FCmin / t-test
261.0229 / 259.0084***	7.01 / 12.34	POS / NEG	[M+H] ⁺ / [M-H] ⁻	C ₉ H ₁₃ BrN ₂ O ₂	260.0160	-1.60 / - 1.30	Bromacil	1	N/A	FCmin / t-test (POS) FCmin (NEG)
215.0587	12.69	NEG	[M-H] ⁻	C ₉ H ₁₃ ClN ₂ O ₂	216.0666	-2.80	Terbacil	1	N/A	FCmin / t-test
ToF										
402.0733*	11.03	NEG	[M-H]⁻	C₂₀H₁₈ClNO₆	403.0823	-4.07	Ochratoxin A	2	N/A	FCmin
249.0220*	9.56	NEG	[M-H]⁻	C₁₂H₁₀O₄S	250.0300	-2.80	Bisphenol S	2	N/A	FCmin
216.1016	6.52	POS	[M+H] ⁺	C ₈ H ₁₄ ClN ₅	215.0938	2.78	Atrazine	1	N/A	FCmin
253.1669	5.44	POS	[M+H] ⁺	C ₁₂ H ₂₀ N ₄ O ₂	252.1586	4.01	Hexazinone	1	N/A	FCmin
215.0967	6.02	POS	[M+H] ⁺	C ₈ H ₁₄ N ₄ OS	214.0888	2.90	Metribuzin	1	N/A	FCmin
259.0084	10.39	NEG	[M-H] ⁻	C ₉ H ₁₃ BrN ₂ O ₂	260.0160	-1.32	Bromacil	1	N/A	FCmin / t-test
215.0585	10.73	NEG	[M-H] ⁻	C ₉ H ₁₃ ClN ₂ O ₂	216.0666	-3.27	Terbacil	1	N/A	FCmin / t-test

* Compounds detected on only one instrument

** N/A = Not applicable

*** Detected in both positive and negative ionization mode

TABLE 5.6 DETECTION RATES FOR CONTAMINATION CASE N°2.

A) CONTAMINANTS SPIKED IN GROUP N°1. OVERALL (TOF + ORBITRAP) DETECTION RATE: **64%**

UHPLC-HRMS platform	Mode	Rate			Rate [found compounds / compounds visible in raw data]
		$\frac{\text{[found compounds / added compounds]}}{t\text{-test / fixed FC}}$	<i>FCmin</i>	Global	
ToF	ESI ⁺	0%	27%	45%	71%
	ESI ⁻	18%	18%		
Q-Orbitrap	ESI ⁺	36%	55%	64%	100%
	ESI ⁻	9%	18%		
	ESI ⁻	9%	18%		

B) CONTAMINANTS SPIKED IN GROUP N°2. OVERALL (TOF + ORBITRAP) DETECTION RATE: **100%**

UHPLC-HRMS platform	Mode	Rate			Rate [found compounds / compounds visible in raw data]
		$\frac{\text{[found compounds / added compounds]}}{t\text{-test / fixed FC}}$	<i>FCmin</i>	Global	
ToF	ESI ⁺	0%	0%	67%	67%
	ESI ⁻	67%	67%		
Q-Orbitrap	ESI ⁺	33%	33%	33%	100%
	ESI ⁻	0%	0%		
	ESI ⁻	0%	0%		

2.3.4. TOF / ORBITRAP COMPARISON

The two instruments possess different characteristics, both in terms of ionization source (geometry, ESI vs. HESI) and mass analyzer performances (resolution, accuracy, detection technology). Therefore, their performance in terms of sensitivity and scope are expected to be different. Interestingly, our developed methodology can be applied on data collected from both instruments with promising results as indicated above. Unexpected contamination could be detected and characterized by the use of different LC-HRMS platforms, even at levels as low as 10 µg/kg and with samples exhibiting different contaminations within the same analytical sequence. This is a major advance for the development and implementation of untargeted approach for chemical food safety applications, since existing approaches only focus on a single contamination scenario and one analytical platform.

The majority of contaminants were detected on both platforms, giving rather similar annotation success rates. Both platforms performed well on the simplest case (n°1), with close detection rates (55% for ToF analyzer vs. 64% for Orbitrap). The main difference observed came from the false detection of a contaminant (dodemorph, see Table 5.3) when using the t-test based filtration on the ToF data. This false positive cannot be explained since this peak was absent from Orbitrap data, as well as from the spiking mix. On the rest of the data (both ToF and Orbitrap), both filtration strategies gave similar results.

However, several contaminants were missed by both instruments for case n°1. Based on Table 5.3 and Table 5.7, the missed contaminants are the following: butylate, EPTC, isopropalin and pebulate (plus cycloate for ToF platform). For cycloate, it is most likely due to the mass spectrometry detection, and more particularly the geometrical and technological differences between the two ionization sources, two critical parameters in complex matrices such as tea. A closer look at the respective physico-chemical properties of the remaining compounds, and especially vapor pressures, enables to give hypothesis on their non-detection. First of all, two of them (butylate and EPTC) have vapor pressures over 100 mPa (see Table 5.7), which may lead to a loss during the evaporation/concentration step. However, other contaminants have similar vapor pressures (for example molinate) and are still detected. The explanation for the non-detection of butylate and EPTC therefore should be a combination between significant losses during sample treatment and strong matrix effects in mass spectrometry. For the two last ones (pebulate and isopropalin), only few papers were reported (Mayer-Helm et al., 2006; Liu et al., 2004) with levels analyzed rather high compared with our study (between 0.1 and 1 mg/kg), suggesting that those compounds may be difficult to analyzed using LC-ESI-MS. This hypothesis must be taken with care given the technological developments in LC-MS between the mid-2000s and late 2010s. Still, the signal/noise ratios obtained with our method for those two compounds are rather low, even in a clean solvent matrix, so that it is likely that even minor loss due to either the sample treatment or the ionization process would lead to their non-detection in a complex matrix such as tea.

Differences between the two instruments appeared more clearly during the analysis of the contamination case n°2, which is more complex. In fact, as displayed in Table 5.5 (in bold), several contaminants were annotated in data from only one instrument, which clearly highlights the interest of implementing multiple LC-HRMS technologies for untargeted studies. As indicated in Table 5.6, for contamination mix n°1 (11 herbicides) same detection rates were achieved than in case n°1, and non-detects should then have the same causes. For contamination mix n°2 (mix of OTA, BPS and malathion), the detection rate difference along with the complementarity observed between the two instruments suggest that non-detects probably come from the mass spectrometer technology (including the ionization sources and ion paths) and not from the upstream analytical process (sample treatment and chromatography methods). In fact, the ToF instrument used appeared to be more sensitive to source fouling (and so to ion intensity drift), leading to the non-detection of malathion in positive mode.

Differences were also observed between the two filtration methods, since FC_{\min} strategies gave overall better results than t-test strategy on our data. This observation may be explained by the computing, in the FC_{\min} strategy, of the discarding criteria as a function of the repeatability of the measurement for each ion, whereas t-test strategy uses fixed thresholds for all ions.

However, the achievement of such performances required the development of a data process adjusted to specificities of both LC-HRMS platforms. In particular, the process already developed (Delaporte et al., 2019) for ToF platform was modified to handle zeros generated by the low noise of the data produced by

the Orbitrap platform. Moreover, analytical drift correction was required for data from ToF platform only. These issues were easily handled using easy-to-use, open-source and freely available tools which highlight the advantages of using those tools to take full advantage of the use of multiple instruments.

In addition, some specificity of each instrument can be spotted. Hence malathion and cycloate were only detected on the Orbitrap platform, while ochratoxin A and bisphenol S were recovered using the ToF instrument. Thus, the future of food safety control using untargeted approaches will benefit from the implementation of multiple analytical platforms in the process, which will be eased by the use of open-source, flexible tools.

2.4. CONCLUSION

In this work, an untargeted strategy to detect trace contaminants at relevant levels in complex food samples has been blindly implemented on two complex contamination scenarios in a model food product (green and black tea leaves) using two LC-HRMS platforms (respectively LC-ToF and LC-Q-Orbitrap). To the best of our knowledge, this is the first time that two different platforms are compared for untargeted food chemical safety assessment. The comparison of the performances of the two instruments has been eased by the use of freely available and open-source tools which enabled the implementation of the same data treatment workflow for all data sets. Most contaminants have been detected with both instruments, despite the complexity of the scenarios investigated (unexpected contaminants, different contamination between samples of the same data set and low contamination levels, i.e. down to 10 µg/kg). Most contaminants spiked have never been analyzed with our analytical method before and so were truly unknown to the methodology, which is rather new in this kind of study. As far as we know, this is the first time that several contamination patterns have been investigated simultaneously at such levels (down to 10 µg/kg). This constitutes a major step towards an implementation of untargeted approaches in routine analysis. Different information could be obtained between two LC-HRMS systems, and this is especially critical in complex matrices, in which strong signal suppression phenomenon are often observed. Our data treatment approach performed well on all data sets, which is highly promising with a view of its implementation on other studies and LC-HRMS platforms for blind contaminants detection. However, the cases investigated in this study differ from real food safety applications. First, the sample diversity considered here is much lower than a real case, which would involve samples from several production batches, geographical origins and processes. A first attempt has been made to increase the sample diversity in our previous work with promising results (Delaporte et al., 2019), but still with limited variability and applicability as an unspiked control sample was always needed to detect a contamination. No doubt that future works would imply the study of a greater, “real-life” sample variability, by taking into consideration the different factors mentioned above. Moreover, contaminants may be more difficult to recover from a natively contaminated sample than from a spiked one; a water soaking step might help to recover native contaminants in the case of dried tea samples (Martínez-Domínguez et al, 2015). The solving of these two

issues (sample variability and the gap between spiking studies and real samples) would certainly constitute major scientific breakthrough in the field of chemical food safety assessment.

The development of untargeted approaches benefits from its application on multiple data sets and contamination scenarios. For that purpose, data sets used in this publication will be made publicly available on the MetaboLights data exchange platform.

2.5. FUNDING

This work was supported by Paris Institute of Technology for Life, Food and Environmental Sciences (AgroParisTech), the French National Institute for Agricultural Research (INRA) and the French Ministry of Higher Education and Research.

2.6. ACKNOWLEDGEMENTS

The authors would like to thank Even Leroux, Brigitte Pollet and Pascale Lieben for their most helpful help on the LC-MS instrumentation.

2.7. REFERENCES

- 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>
- Castro-Puyana, M., & Herrero, M. (2013). Metabolomics approaches based on mass spectrometry for food safety, quality and traceability. *TrAC - Trends in Analytical Chemistry*, 52, 74–87. <https://doi.org/10.1016/j.trac.2013.05.016>
- Castro-Puyana, M., Pérez-Míguez, R., Montero, L., & Herrero, M. (2017). Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability. *TrAC - Trends in Analytical Chemistry*, 93, 102–118. <https://doi.org/10.1016/j.trac.2017.05.004>
- Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., ... Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, 30(10), 918–920. <https://doi.org/10.1038/nbt.2377>
- Cladière, M., Delaporte, G., Le Roux, E., & Camel, V. (2018). Multi-class analysis for simultaneous determination of pesticides, mycotoxins, process-induced toxicants and packaging contaminants in tea. *Food Chemistry*, 242, 113–121. <https://doi.org/10.1016/j.foodchem.2017.08.108>
- Cotton, J., Leroux, F., Broudin, S., Marie, M., Corman, B., Tabet, J. C., ... Junot, C. (2014). High-resolution mass spectrometry associated with data mining tools for the detection of pollutants and chemical characterization of honey samples. *Journal of Agricultural and Food Chemistry*, 62(46), 11335–11345. <https://doi.org/10.1021/jf504400c>
- Delaporte, G., Cladière, M., & Camel, V. (2018). Missing value imputation and data cleaning in untargeted food chemical safety assessment by LC-HRMS. *Chemometrics and Intelligent Laboratory Systems*, Submitted.
- Delaporte, G., Cladière, M., Jouan-Rimbaud Bouveresse, D., & 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>
- Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ¹H NMR

- metabonomics. *Analytical Chemistry*, 78(13), 4281–4290. <https://doi.org/10.1021/ac051632c>
- Dunn, W.B., Broadhurst, D., Begley, P., Zelena, E., Francis-Mcintyre, S., Anderson, N., ... Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, *Nature Protocols*, 6, 1060–1083. <https://doi:10.1038/nprot.2011.335>
- Giacomini, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., ... Caron, C. (2015). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics*, 31(9), 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>
- Glauser, G., Veyrat, N., Rochat, B., Wolfender, J. L., & Turlings, T. C. J. (2013). Ultra-high pressure liquid chromatography-mass spectrometry for plant metabolomics: A systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers. *Journal of Chromatography A*, 1292, 151–159. <https://doi.org/10.1016/j.chroma.2012.12.009>
- Gorrochategui, E., Jaumot, J., Lacorte, S., & Tauler, R. (2016). Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, *TrAC - Trends in Analytical Chemistry*, 82, 425–442. <https://doi:10.1016/j.trac.2016.07.004>
- Haug, K., Salek, R. M., Conesa, P., Hastings, J., De Matos, P., Rijnbeek, M., ... Steinbeck, C. (2013). MetaboLights - An open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Research*, 41(D1), 781–786. <https://doi.org/10.1093/nar/gks1004>
- Inoue, K., Tanada, C., Sakamoto, T., Tsutsui, H., Akiba, T., Min, J. Z., ... Toyo'Oka, T. (2015). Metabolomics approach of infant formula for the evaluation of contamination and degradation using hydrophilic interaction liquid chromatography coupled with mass spectrometry. *Food Chemistry*, 181, 318–324. <https://doi.org/10.1016/j.foodchem.2015.02.117>
- Kassouf, A., Jouan-Rimbaud Bouveresse, D., & Rutledge, D. N. (2017). Determination of the Optimal Number of Components in Independent Components Analysis. *Talanta*, 179, 538–545. <https://doi.org/10.1016/j.talanta.2017.11.051>
- 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>
- Knolhoff, A. M., Zweigenbaum, J. A., & Croley, T. R. (2016). Nontargeted Screening of Food Matrices: Development of a Chemometric Software Strategy to Identify Unknowns in Liquid Chromatography-Mass Spectrometry Data. *Analytical Chemistry*, 88(7), acs.analchem.5b04208. <https://doi.org/10.1021/acs.analchem.5b04208>
- 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*, 1–10. <https://doi.org/10.1007/s00216-018-1028-4>
- Libiseller, G., Dvorzak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., ... Magnes, C. (2015). IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics*, 16(1), 118. <https://doi.org/10.1186/s12859-015-0562-8>
- Liu, H., Ding, C., Zhang, S., Liu, H., Liao, X., Qu, L., Zhao, Y., & Wu, Y. (2004). Simultaneous residue measurement of pendimethalin, isopropalin, and butralin in tobacco using high-performance liquid chromatography with ultraviolet detection and electrospray ionization/mass spectrometric identification, *Journal of Agricultural and Food Chemistry*, 52, 6912–6915. <https://doi:10.1021/jf0488965>
- Lommen, A., van der Weg, G., van Engelen, M. C., Bor, G., Hoogenboom, L. A. P., & Nielen, M. W. F. (2007). An untargeted metabolomics approach to contaminant analysis: Pinpointing potential unknown compounds. *Analytica Chimica Acta*, 584(1), 43–49. <https://doi.org/10.1016/j.aca.2006.11.018>
- Martínez-Domínguez, G., Nieto-García, A.J., Romero-González, R., & Frenich, A.G. (2015). Application

- of QuEChERS based method for the determination of pesticides in nutraceutical products (*Camellia sinensis*) by liquid chromatography coupled to triple quadrupole tandem mass spectrometry, *Food Chemistry*, 177, 182–190. <https://doi.org/10.1016/j.foodchem.2015.01.032>
- Mayer-Helm, B., Hofbauer, L., & Müller, J. (2006). Development of a multi-residue method for the determination of 18 carbamates in tobacco by high-performance liquid chromatography/positive electrospray ionisation tandem mass spectrometry, *Rapid Communication in Mass Spectrometry*, 20, 529–536. <https://doi.org/10.1002/rcm.2334>
- Ortmayr, K., Charwat, V., Kasper, C., Hann, S., & Koellensperger, G. (2017). Uncertainty budgeting in fold change determination and implications for non-targeted metabolomics studies in model systems. *The Analyst*, 142(1), 80–90. <https://doi.org/10.1039/C6AN01342B>
- Patti, G. J., Tautenhahn, R., & Siuzdak, G. (2013). Meta-Analysis of Untargeted Metabolomic Data: Combining Results from Multiple Profiling Experiments. *Nature Protocols*, 7(3), 508–516. <https://doi.org/10.1038/nprot.2011.454>. Meta-Analysis
- Rutledge, D. N., & Jouan-Rimbaud Bouveresse, D. (2015). Corrigendum to “Independent Components Analysis with the JADE algorithm.” *TrAC - Trends in Analytical Chemistry*, 67, 220. <https://doi.org/10.1016/j.trac.2015.02.001>
- Saito-Shida, S., Hamasaka, T., Nemoto, S., & Akiyama, H. (2018). Multiresidue determination of pesticides in tea by liquid chromatography-high-resolution mass spectrometry: Comparison between Orbitrap and time-of-flight mass analyzers. *Food Chemistry*, 256(October 2017), 140–148. <https://doi.org/10.1016/j.foodchem.2018.02.123>
- Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), 779–787. <https://doi.org/10.1021/ac051437y>
- Tautenhahn, R., Patti, G. J., Rinehart, D., & Siuzdak, G. (2012). XCMS online: A web-based platform to process untargeted metabolomic data. *Analytical Chemistry*, 84(11), 5035–5039. <https://doi.org/10.1021/ac300698c>
- 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>
- Van Der Kloet, F. M., Bobeldijk, I., Verheij, E. R., Jellema, R. H., Kloet, F. M. Van Der, Bobeldijk, I., ... Jellema, R. H. (2009). Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *Journal of Proteome Research*, 8(11), 5132–5141. <https://doi.org/10.1021/pr900499r>

2.8. SUPPLEMENTARY MATERIALS

TABLE 5.7 MAIN CHARACTERISTICS OF STUDIED COMPOUNDS

Contaminant	Formula	Molecular weight	log Kow	pKa	Authorized substance for tea	MRL (mg/kg) in tea leaves
Spiking mix n°1						
Atrazine <i>Pesticide</i>	C ₈ H ₁₄ ClN ₅	215.0938	2.7	1.7	No	0.1
Bromacil <i>Pesticide</i>	C ₉ H ₁₃ BrN ₂ O ₂	260.0160	1.88	9.27	No	0.01
Butylate <i>Pesticide</i>	C ₁₁ H ₂₃ NOS	217.1500	4.1	N/A ¹	No	0.05
Cycloate <i>Pesticide</i>	C ₁₁ H ₂₁ NOS	215.1344	4.11	N/A	No	0.01
EPTC <i>Pesticide</i>	C ₉ H ₁₉ NOS	189.1187	3.2	N/A	No	0.05
Hexazinone <i>Pesticide</i>	C ₁₂ H ₂₀ N ₄ O ₂	252.1586	1.17	2.2	No	0.01
Isopropalin <i>Pesticide</i>	C ₁₅ H ₂₃ N ₃ O ₄	309.1689	5.29	N/A	No	0.01
Metribuzin <i>Pesticide</i>	C ₈ H ₁₄ N ₄ OS	214.0888	1.26	0.99	Yes	0.1
Molinate <i>Pesticide</i>	C ₉ H ₁₇ NOS	187.1031	2.86	N/A	No	0.05
Pebulate <i>Pesticide</i>	C ₁₀ H ₂₁ NOS	203.1344	4	N/A	No	0.01
Terbacil <i>Pesticide</i>	C ₉ H ₁₃ ClN ₂ O ₂	216.0666	1.89	9.5	No	0.01
Spiking mix n°2						
Malathion <i>Pesticide</i>	C ₁₀ H ₁₉ O ₆ PS ₂	330.0361	2.75	N/A	Yes	0.5
Bisphenol S <i>Migrant from packaging</i>	C ₁₂ H ₁₀ O ₄ S	250.0300	2.91	N/A	N/A	N/A
Ochratoxin A <i>Mycotoxin</i>	C ₂₀ H ₁₈ ClNO ₆	403.0823	4.74	3.2 / 7.9	N/A	N/A

¹ N/A = Not Applicable

TABLE 5.8 SOURCE PARAMETERS FOR TOF PLATFORM

Parameter	ESI⁺	ESI⁻
Capillary voltage (kV)	1.5	1.0
Sample cone (V)	20	35
Source offset (V)	20	80
Source temperature (°C)	130	130
Desolvation temperature (°C)	500	500
Cone gas flow (L/h)	50	20
Desolvation gas flow (L/h)	1,200	600

TABLE 5.9 SOURCE PARAMETERS FOR ORBITRAP PLATFORM

Parameter	ESI⁺	ESI⁻
Capillary voltage (V)	3,000	2,500
Capillary temperature (°C)	350	300
Desolvation gas flow (A.U)	60	35
Auxiliary gas flow (A.U)	20	10

A.U = Arbitrary Unit

TABLE 5.10 PARAMETERS FOR PEAK EXTRACTION BY XCMS

Module	Parameter	ToF	Orbitrap
xcmsSet	scanrange	NEG: 180-2400 POS: 120-2060	All
	nSlaves	1	1
	method	centWave	centWave
	ppm	15	3
	peakwidth	5-60	5-20
	mzdiff	-0.001	-0.001
	snthresh	10	10
	integrate	1	1
	noise	0	0
	prefilter	0	3,5000
group	method	density	density
	minfrac	0.5	0.5
	bw	2	2
	mzwid	0.015	0.015
	sleep	0.001	0.001
retcor	method	peakgroups	peakgroups
	smooth	loess	loess
	extra	1	1
	missing	1	1
	span	0.2	0.2
	family	gaussian	gaussian
fillPeaks	method	chrom	chrom
	plottype	mdevden	mdevden
CAMERA.annotate	nSlaves	4	4
	sigma	6	6
	perfwhm	0.6	0.6
	ppm	15	3
	mzabs	0.015	0.015
	maxcharge	1	1
	maxiso	4	4
	minfrac	0.5	0.5
	quick	TRUE	TRUE
	convertRTMinute	FALSE	FALSE
	numDigitsMZ	4	4
	numDigitsRT	0	0
	intval	into	into

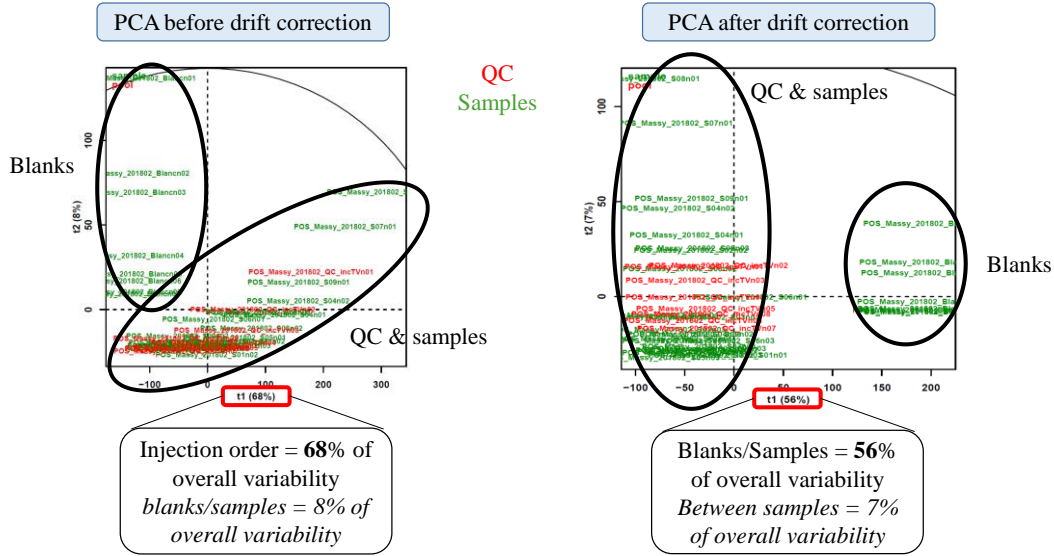


FIGURE 5.5 PCA OF RAW DATA BEFORE AND AFTER DRIFT CORRECTION (GREEN TEA, TOF DATA ACQUIRED IN POSITIVE IONIZATION MODE)

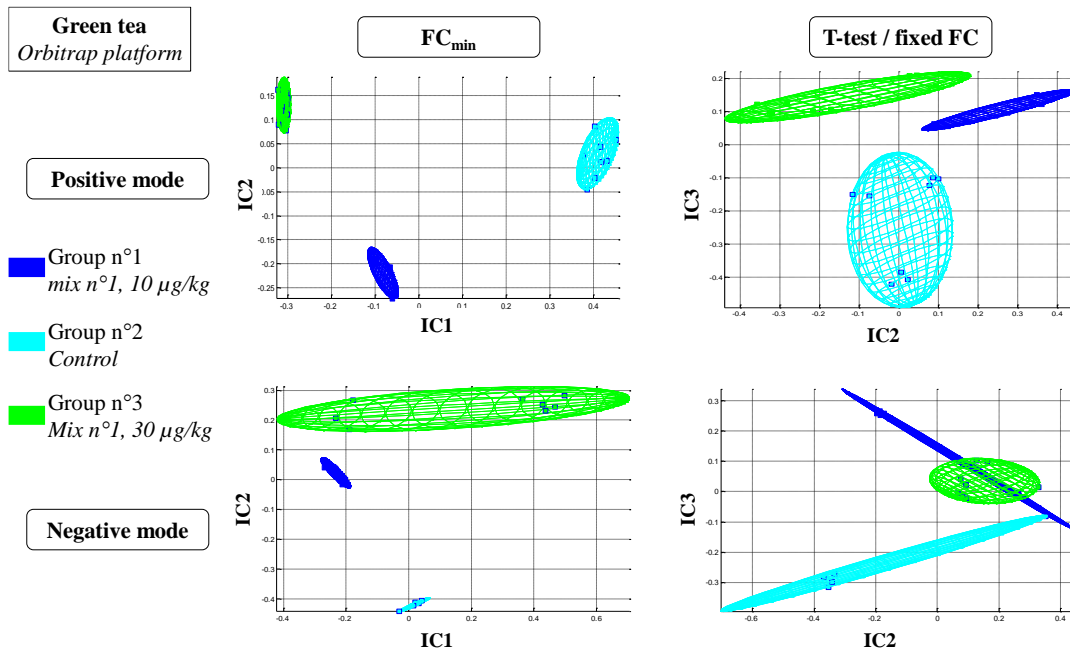


FIGURE 5.6 SCORE PLOTS OF ICA OUTPUT ON GREEN TEA SAMPLES ANALYZED ON ORBITRAP PLATFORM

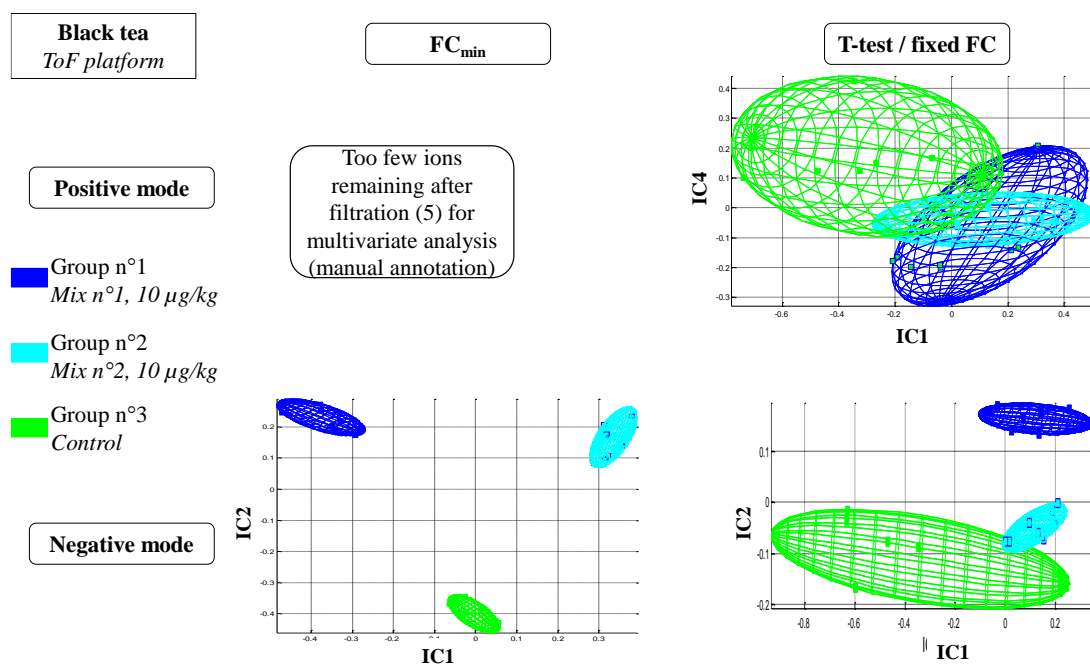


FIGURE 5.7 SCORE PLOTS OF ICA OUTPUT ON BLACK TEA SAMPLES ANALYZED ON TOF PLATFORM

3. CONCLUSION

Pour chacun des scénarios de contamination étudiés, notre approche a permis de détecter la présence d'échantillons contaminés, et ce avec les deux instruments utilisés.

Pour le premier scénario, 55% et 64% des molécules dopées ont respectivement été annotées dans les données ToF et Orbitrap, ce qui est pertinent avec les niveaux d'annotation reportés dans la littérature lors d'études similaires (Kunzelman et. al, 2018). Dans chaque cas, les niveaux relatifs des contaminants détectés ont pu être donnés avec succès. Un faux-positif a été reporté lors de l'analyse des données ToF, avec l'annotation d'un contaminant (dodemorph) qui n'était pas présent dans le mélange de dopage.

Lors de l'analyse du deuxième scénario, notre approche a réussi à mettre en évidence la présence de deux contaminations distinctes. Une différence plus marquée entre les deux plateformes UHPLC-HRMS a été notée. Pour la première contamination, 45% et 64% des contaminants dopés ont été détectés respectivement dans les données ToF et Orbitrap, grâce à une meilleure sensibilité de cette dernière. Pour la deuxième contamination, 67% et 33% des composés dopés ont été détectés dans les données ToF et Orbitrap, avec un taux d'annotation global (les deux instruments combinés) de 100%. Cependant, le faible taux de détection de l'Orbitrap semble pouvoir être expliqué par un problème de stabilité des extraits lors de leur stockage et transport, les deux composés manqués ayant déjà été analysés avec succès à ces niveaux (de façon ciblée) sur cette plateforme.

CHAPITRE 6
DISCUSSION GÉNÉRALE

L'objectif principal de cette thèse est de proposer une preuve de concept d'approche combinant une stratégie analytique, une méthode d'analyse et un processus de traitement de données automatisé permettant la détection non-ciblée de contaminants de l'aliment. Les approches existantes dans le domaine de la métabolomique ont été identifiées comme particulièrement prometteuses (Antignac et al., 2011; Castro-Puyana, Pérez-Míguez, Montero, & Herrero, 2017). Cependant, comme nous le rappelions au Chapitre 1 (partie 4), les problématiques scientifiques et analytiques sont différentes entre les études métabolomiques et les études portant sur la détection non-ciblée des contaminants de l'aliment. Ainsi, une étude métabolomique vise souvent la découverte de biomarqueurs fiables d'un phénomène (et non nécessairement tous les biomarqueurs), alors que la recherche de contaminants non-attendus dans un aliment implique de mettre en évidence tout ce qui peut différencier des échantillons avant d'isoler ensuite dans les signaux obtenus ceux dus aux contaminants. De plus, lors de l'étude d'un métabolisme, on observe généralement une réponse métabolique globale qui peut s'exprimer par l'augmentation et/ou la diminution de la concentration d'un certain nombre de composés, tandis que lors de l'étude des contaminants de l'aliment, les composés d'intérêt ont davantage un comportement de type absent/présent (ou bien « bruit de fond »/dépassement). Par ailleurs, les contaminants de l'aliment possèdent certaines caractéristiques chimiques récurrentes (par exemple la présence de composés halogénés) qui peuvent faciliter leur détection comme cela a déjà été réalisé pour des problématiques agro-alimentaires (Cotton et al., 2014) ou environnementales (Cariou, Omer, Léon, Dervilly-Pinel, & Le Bizec, 2016; Roullier et al., 2016).

Le développement de la preuve de concept passe par un certain nombre d'objectifs intermédiaires. Une méthode analytique LC-HRMS à spectre large pour l'analyse d'une grande diversité de contaminants du thé a tout d'abord été développée. Cette méthode a été ensuite implémentée pour proposer une première preuve de concept sur une situation de contamination connue. Dans ce but, un processus de traitement de données automatisé a été mis en place. Plusieurs étapes clés de ce processus ont fait l'objet d'études méthodologiques plus poussées dans l'optique d'acquérir des connaissances sur le comportement des différents outils disponibles et ainsi de rendre l'approche plus performante, modulable et robuste. L'approche dans sa version finale a enfin été implémentée en aveugle sur plusieurs situations de contamination et plateformes LC-HRMS pour évaluer sa robustesse et son applicabilité dans des situations plus proches de la réalité.

La méthode analytique a été développée pour 32 contaminants dits « traceurs », choisis pour leur diversité de structures chimiques dans des échantillons de thé sous forme liquide (infusion) ou solide (produit fini sous forme de feuilles). Quatre classes de contaminants ont été incluses dans cette étude (9 familles de pesticides, 3 familles de mycotoxines, des contaminants de l'emballage et des contaminants issus de procédés), certaines molécules étant pour la première fois prises en compte dans une méthode multi-résidus (comme par exemple l'acrylamide). Des expériences de dopage/recouvrement ont été menées à trois niveaux de concentration (10, 50 et 100 µg/kg) choisis en regard des niveaux de

contaminants usuels et de la réglementation en vigueur. La méthode proposée repose sur une préparation d'échantillon générique (« dilute & shoot » pour la forme liquide, extraction directe par solvant pour la forme solide) et une méthode LC-HRMS à spectre large utilisant une colonne de type silice greffée C18-PFP. Les effets de matrice induits par le produit d'étude (le thé), et la comparaison de plusieurs méthodes d'étalonnage (calibration en matrice, calibration par des composés marqués et calibration externe en solvant) ont permis d'établir qu'il est nécessaire de corriger les phénomènes de suppression d'ion observés pour obtenir des résultats quantitatifs, et que la calibration en matrice semble la meilleure méthode pour cela. Les limites de quantification et de détection obtenues sont pertinentes vis-à-vis des niveaux de référence que nous nous étions fixés, sur la base des teneurs usuelles rencontrées ou des teneurs maximum définies par la réglementation. En effet, 69% de ces « traceurs » peuvent être quantifiés et 91% détectés à un niveau inférieur ou égal à leur niveau de référence. Ces résultats confirment la capacité des méthodes à spectre large à analyser simultanément plusieurs classes de contaminants de structures variées à des niveaux pertinents, en y incluant pour la première fois certaines classes et molécules. Elles montrent également que, dans le cas du thé, l'étape de purification de l'extrait peut être omise lors de l'analyse de composés traces par LC-HRMS. Enfin, au vu de la diversité de contaminants considérée, cette méthode semble adaptée pour l'analyse non-ciblée des contaminants ce qui est le but principal recherché dans ces premiers travaux.

La stratégie analytique utilisée pour l'analyse non-ciblée s'inspire de celle employée en métabolomique (aléatorisation des injections, injection régulière de blancs et d'échantillons « QC » constitués d'un mélange de tous les extraits). Un processus de traitement des données pour la détection non-ciblée des contaminants a été mis en place. Il est basé sur le paquet de R « XCMS » implémenté sur la plateforme de calcul Workflow4Metabolomics (W4M) et sur un script Matlab développé pour l'occasion. Les ions sont d'abord extraits depuis les fichiers bruts sur la plateforme W4M et la matrice de données ainsi créée est importée dans Matlab. Le processus comprend ensuite trois grandes étapes : un prétraitement pour préparer les données, une analyse multivariée pour les visualiser, et enfin une étape d'interprétation permettant de donner des hypothèses sur la présence et l'identité des potentiels contaminants. L'annotation des composés suspects s'inspire de travaux ultérieurs (Cotton et al., 2014) et comporte notamment une détection automatisée de motifs isotopiques particuliers dans les données de spectrométrie de masse.

Pour évaluer et discuter les performances obtenues, cinq études sont disponibles dans la littérature. Cependant, cette évaluation peut être délicate car la conception de l'étude peut être différente [pas de dopage avec calcul de taux de détection (Cotton et al., 2014)] ou bien les niveaux étudiés non comparables avec notre étude [niveaux de l'ordre du mg/kg voire de la dizaine de mg/kg (Inoue et al., 2015; Tengstrand, Rosén, Hellenäs, & Åberg, 2013)]. Deux autres études sont plus pertinentes pour la comparaison (Knolhoff, Zweigenbaum, & Croley, 2016; Kunzelmann, Winter, Åberg, Hellenäs, & Rosén, 2018), car ciblant des niveaux de contamination similaires et donnant des taux de détection par

rapport aux composés dopés. Dans la première (Knolhoff et al., 2016), les 4 contaminants dopés (3 alcaloïdes naturels et un opiacé de synthèse donc de structure proche des alcaloïdes) sont détectés dès 10 µg/kg. Le faible nombre de composés, de même que la relative similarité structurale et de propriétés physico-chimiques des molécules incitent à prendre la comparaison avec nos données avec prudence. Kunzelman et ses collaborateurs (Kunzelmann et al., 2018) traitent quant à eux spécifiquement de problème de détectabilité en fonction de la concentration de dopage d'un mélange de 19 pesticides. Leur méthode permet de détecter 6 composés sur 19 (soit 32%) dès 5 µg/kg, 17 sur 19 (soit 89%) à 25 µg/kg, et 100% à 100 µg/kg. C'est la publication que nous jugeons la plus pertinente vis-à-vis de nos données.

Pour résumer les travaux existants, des premières approches ont été proposées, et montraient des résultats prometteurs à des niveaux de l'ordre de dizaine de µg/kg. Cependant, ces travaux ont été menés sur une relativement faible diversité de contaminants au regard du panel de molécules existant, avec la plupart du temps une seule famille de composés considérée simultanément. De plus, la variabilité inter-échantillons est très faible dans ces études. Enfin, dans toutes les études citées, une seule contamination est étudiée à la fois (à l'exception des travaux de Tengstrand mais lors desquels les concentrations étudiées sont de l'ordre du mg/kg).

En résumé, dans ce travail une première preuve de concept a été développée, puis appliquée sur quatre cas de contamination différents, récapitulés en Figure 6.1.

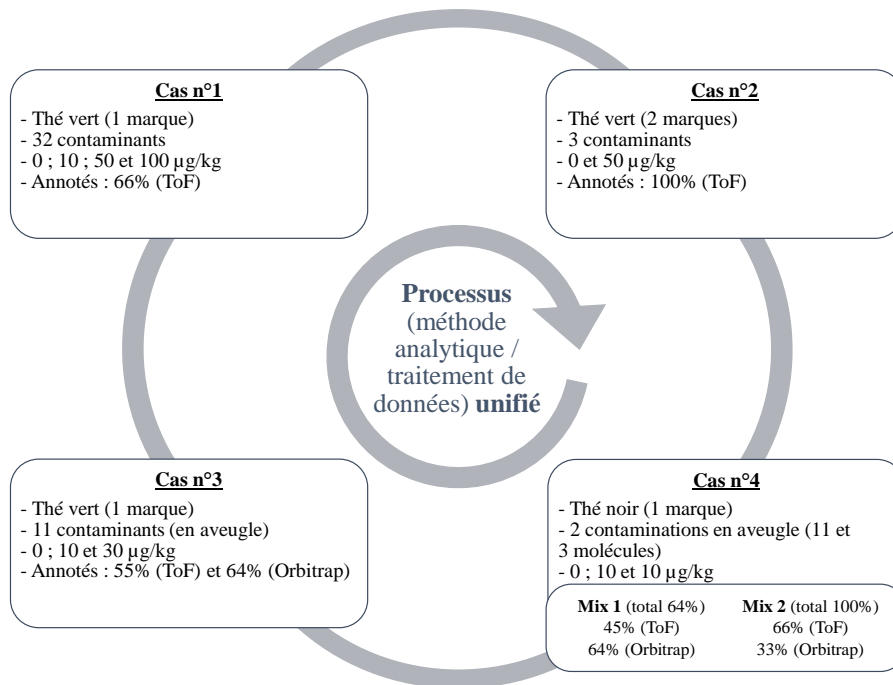


FIGURE 6.1 RÉCAPITULATIF DES EXPÉRIENCES NON-CIBLÉES RÉALISÉES

On peut diviser les quatre cas étudiés en trois types de scénarios de contamination, discutés ici par difficulté de traitement croissante.

Scénario 1 : analyse de contrôle d'un lot de production, et recherche de non-conformités

Le premier, regroupant les cas n°1 et n°3, consiste en un produit avec une faible variabilité inter-échantillons, dopé à différents niveaux par un mélange de contaminants (respectivement 32 et 11 composés). Cette situation vise à mimer un cas de non-conformité d'un échantillon (dépassement de la teneur maximale réglementée). La principale interrogation avant d'aborder cette situation portait sur la proportion de composés dopés que la méthodologie allait permettre d'annoter à l'aveugle. L'étude de ces cas (et particulièrement du cas n°1) a permis de développer des outils opérants en terme d'annotation des composés, avec notamment la construction d'une base données de contaminants et l'adaptation d'outils de fouille de données existants pour la détection des massifs isotopiques.

Cette situation est celle étudiée dans la littérature par plusieurs auteurs, soit avec divers niveaux de dopage (Kunzelmann et al., 2018), soit avec un seul niveau (Knolhoff et al., 2016; Tengstrand et al., 2013). Comme précédemment indiqué, l'étude qui nous parait la plus pertinente pour comparer et discuter nos résultats est celle de Kunzelman et al., où plusieurs niveaux de dopage (5, 25, 100 et 400 µg/kg) d'un mélange de 19 pesticides ont été étudiés. Dans le cas n°1 (Figure 6.1) étudié dans notre travail, la diversité de molécules cibles est plus élevée (4 classes de contaminants provenant de plus de 15 familles chimiques différentes), et le produit d'étude plus complexe chimiquement (thé vs. lait pasteurisé UHT). Le taux d'annotation obtenu par notre méthode dans ce cas est de 66%, ce qui est un peu plus faible que le taux obtenu dès 25 µg/kg par Kunzelman et al. Au vu de la diversité des composés chimiques testés dans notre cas, ce taux est satisfaisant, d'autant qu'une investigation plus poussée permet d'identifier l'origine des pertes d'informations au cours du processus (voir Figure 6.2.a). Ainsi, 3 composés sont perdus par manque de sensibilité de la méthode analytique (ce qui est attendu eu égard à la diversité des « traceurs » considérés), et 5 molécules ne sont pas détectées par l'algorithme d'intégration des pics. Cela est cohérent avec les résultats rapportés par Coble (Coble & Fraga, 2014) sur les performances des algorithmes de détection de pics, puisque les auteurs montrent qu'aucun outil existant n'est capable d'extraire la totalité de l'information d'un fichier de données. Par ailleurs, il s'avère que les 5 composés manqués par XCMS sont ceux présentant des rapports signal/bruit médiocres (<10, voire aux alentours de 3), les données provenant du ToF s'avérant très bruitées. Il existe un algorithme d'optimisation des paramètres de XCMS pour maximiser la détection des ions appelé IPO (Libiseller et al., 2015). Cet outil a été essayé sans succès sur nos données, le programme n'ayant pas réussi à converger vers une proposition de paramètre. Parallèlement, une impureté non-attendue de la solution de dopage a pu être annotée dans les échantillons, ce qui illustre la capacité de notre méthodologie à caractériser une contamination non attendue.

Le cas n°3 est similaire, mais avec des difficultés supplémentaires par rapport au n°1. Premièrement cette étude a été réalisée en aveugle, l'analyse des données se faisant sans aucune connaissance de la composition des échantillons et des molécules dopées « inconnues » de la méthode analytique. Le plus haut niveau de dopage investigué dans ce cas est 30 µg/kg (vs. 100 µg/kg dans le cas n°1). Par ailleurs,

les analyses ont été réalisées sur deux plateformes LC-HRMS de technologie différentes (ToF et Orbitrap), et le même processus de traitement de données utilisé à chaque fois, ce qui à notre connaissance n'a jamais étudié pour des approches non-ciblées de détection des contaminants. Là encore, la comparaison avec les travaux de Kunzelman et al. montre que les résultats obtenus par notre méthode sont honorables car 55% des contaminants sont détectés dans les données ToF et 64% pour les données Orbitrap. Par ailleurs, il est intéressant de voir que, pour chaque plateforme, les contaminants manqués le sont par manque de sensibilité de la méthode analytique (4 composés perdus, voir Figure 6.2.b et 6.2.c), et beaucoup moins par le traitement des données, étape d'extraction des pics par XCMS incluse (1 composé perdu pour les données ToF, aucun pour celles Orbitrap). La principale difficulté de ce cas était que les contaminants ajoutés, leurs niveaux et leur nombre étaient inconnus du manipulateur lors des expériences. Nous ne disposons donc d'aucune information sur leur sensibilité, leurs adduits et leur fragmentation potentielle.

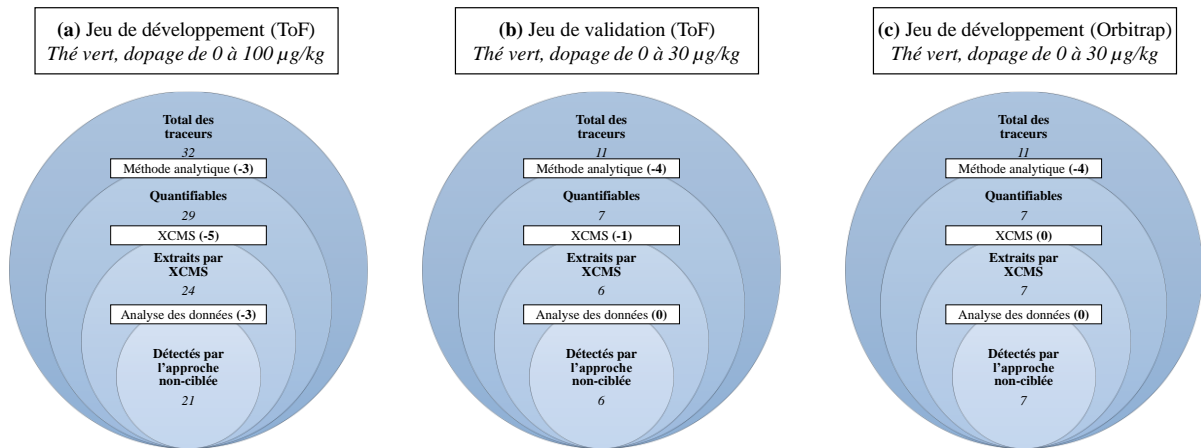


FIGURE 6.2 REPRÉSENTATION DES PERTES EN « TRACEURS » AU FIL DES ÉTAPES DE LA PREUVE DE CONCEPT POUR TROIS JEUX DE DONNÉES

Scénario 2 : analyse simultanée d'échantillons de provenances différentes

Le deuxième type de situation (cas n°2) implique une forte variabilité inter-échantillons dans le produit d'étude et vise à simuler l'analyse simultanée d'échantillons de provenances différentes. Cette situation n'a jamais été étudiée avec succès dans la littérature ; on peut relever la tentative de Knolhoff et al. (Knolhoff et al., 2016) sur une matrice peu variable (jus d'orange et formule infantile), tentative qui ne sera pas couronnée de succès. Nous ne disposons donc pas d'étude pour comparer les performances de notre approche. La première interrogation au moment d'aborder cette situation portait sur la capacité de notre méthodologie à distinguer la variabilité inter-échantillons « naturelle » (c.à.d. due à la variation de la matrice en elle-même) de la variabilité induite par le dopage. Au vu du faible nombre de molécules

dopées (3, toutes avec une bonne réponse en LC-MS), nous nous attendions à deux résultats extrêmes : détecter l'intégralité, ou n'aboutir à rien.

L'expérience a donc été conçue avec deux marques de thé vert, et pour chacune un groupe d'échantillons de contrôle et un groupe d'échantillons dopés (soit un total de quatre groupes d'échantillons). L'avantage de l'utilisation d'ICA apparaît dans ce cas car cette méthode permet la séparation de « signaux sources » correspondant à des phénomènes (physiques, chimiques, biologiques) indépendants au sein des échantillons (ici, le dopage et la variabilité naturelle du thé). Les trois contaminants dopés ont été annotés avec succès dans le jeu de données complet. Un essai a également été réalisé en enlevant le groupe d'échantillon dopé d'une des deux marques, et là encore, la contamination a été détectée avec succès. En revanche, en ne considérant qu'un groupe de contrôle d'une marque et le groupe dopé de l'autre, la contamination n'a pas pu être détectée, signe que des études plus poussées sur comment supprimer l'influence de la variabilité du produit sur l'analyse sont nécessaires.

Scénario 3 : analyse de plusieurs échantillons conformes, diversement contaminés

Le dernier type de situation (cas n°4) est caractérisé par la présence de plusieurs contaminations différentes dans les échantillons d'un même jeu de données, et ce à un très faible niveau (10 µg/kg). Cette situation vise à reproduire la présence de plusieurs échantillons ayant des contaminations différentes, sans pour autant dépasser les teneurs maximales autorisées. À notre connaissance elle n'a jamais été abordée dans les études existantes. En comparaison avec les deux situations précédentes (cas n°1 à 3), la difficulté ici était à la fois de réussir à séparer de façon satisfaisante les groupes d'échantillons, mais également d'annoter le plus possible de composés en un temps raisonnable.

Cette étude a été effectuée en aveugle, l'analyse des échantillons et des données se faisant sans aucune connaissance du plan de dopage. Par ailleurs nous avons proposé, comme dans le cas n°3, la comparaison de deux plateformes UHPLC-HRMS (ToF et Orbitrap) en utilisant le même procédé de traitement de données. Notre méthode montre de bonnes performances pour ce cas puisque respectivement 64% et 100% des composés ont été détectés et annotés avec succès pour chacun des dopages. Dans ce cas, la complémentarité entre l'Orbitrap et le ToF a été plus marquée que pour le cas n°3, puisque plusieurs molécules (5 sur un total de 10 molécules annotées toutes plateformes confondues) n'ont été annotées que dans les données d'une des deux technologies. Ce résultat est cependant à nuancer car il est en partie dû à des problèmes de répétabilité observés pour une séquence donnée (séquence en mode positif sur le ToF) et une suspicion de dégradation des extraits lors de leur stockage ou leur transport (bisphénol S et ochratoxine A sur l'Orbitrap). Cela nous amène donc à deux composés uniques sur l'Orbitrap grâce à une meilleure sensibilité (cycloate et molinate), et trois autres dont la cause de non-détection pour une des plateformes est incertaine (malathion, ochratoxine A et bisphénol S). Ces résultats mettent en avant l'intérêt d'utiliser deux instruments de technologies différentes en parallèle dans une étude non-ciblée des contaminants de l'aliment. Cela maximise la

détection de potentiels analytes inconnus ou non-attendus, les différences entre les données des deux instruments (adduits ou fragments par exemple) permettant de renforcer ou d'infirmer une hypothèse d'annotation.

De nombreux ajustement du procédé de traitement de données utilisé pour la preuve de concept initiale ont été nécessaires pour pouvoir traiter tous ces scénarios, et notamment ceux impliquant l'utilisation d'un analyseur Orbitrap. En particulier, deux étapes du prétraitement sont apparues comme sujettes à discussion, avec plusieurs outils et approches existants, pour la plupart inédits dans l'analyse des contaminants de l'aliment. Ces étapes sont l'imputation des valeurs manquantes dans la matrice de données et la filtration des données avant analyse. Plusieurs outils et approches ont été intégrés dans le processus existant, et leur influence sur les résultats finaux (séparation des échantillons contaminés des contrôles et taux de détection des « traceurs ») évaluée. Ces méthodes ont été implémentées sur des jeux de données connues, à savoir ceux utilisés pour le développement de la preuve de concept (cas n°1 et n°2 de la Figure 6.1), puis les outils sélectionnés utilisés en « routine » sur les cas n°3 et 4 (données Orbitrap et ToF).

Les combinaisons respectives de trois méthodes d'imputation des valeurs manquantes (une, utilisée dans la preuve de concept initiale, basée sur l'intégration « forcée » des pics manquants, et deux nouvelles reposant sur la classification des valeurs manquantes) et de deux méthodes de filtration des données (l'une basée sur des tests statistiques univariés avec une valeur seuil de sélection fixe, et l'autre basée sur le calcul d'un changement de signal significatif pour chaque ion) ont été comparées. Au vu des résultats, il a été décidé d'implémenter dans le processus de traitement des données existant une méthode d'imputation des valeurs manquantes basée sur la classification de ces dernières en plus de celle initiale afin de gérer d'éventuelles valeurs manquantes laissées par `xcms.fillPeaks` sous forme de zéros. En effet notre processus d'analyse de données comporte des étapes ne tolérant pas les zéros (notamment un passage en log). Enfin, il a été décidé d'implémenter les deux méthodes de filtration évaluées en parallèle. Ces travaux méthodologiques, inspirés de ceux existants dans le domaine de la métabolomique (Di Guida et al., 2016) sont à notre connaissance inédits pour l'analyse des contaminants de l'aliment, et constituent à ce titre une base de travail intéressante pour de futures études.

Par ailleurs, les résultats obtenus pour les cas n°3 et n°4 soulignent l'intérêt de réaliser les analyses sur plusieurs instruments en parallèle, en élargissant la gamme de molécules analysables et augmentant ainsi la probabilité de détection d'éventuels composés non-attendus. L'approche non-ciblée que nous avons développée a montré sa capacité à traiter des données provenant de plateforme différentes, permettant ainsi une comparaison fiable de leurs performances.

Ces travaux ont aussi permis de mettre en évidence un certain nombre de limites des stratégies non-ciblées existantes. Tout d'abord, bien que notre approche ait été appliquée avec succès à des jeux d'échantillons assez hétérogènes, il reste indispensable de disposer d'un échantillon de référence non-

contaminé pour la détection des contaminants. Cette limite est commune à toutes les approches non-ciblées rapportées jusqu'à présent, à l'exception notable de celle proposée par Cotton (Cotton et al., 2014), qui est néanmoins restreinte aux composés halogénés. Une piste ayant émergé au cours de cette thèse pour la résolution de ce problème (notamment lors des discussions avec les membres du comité de suivi de thèse) est d'utiliser une base de données expérimentale de composés « attendus » dans un produit (c.à.d. normalement présents et ne présentant pas d'intérêt pour l'évaluation de la sécurité sanitaire), qui permettrait ainsi d'exclure ces derniers de l'analyse des données. Cette base de données inclurait en particulier les composés possédant une très grande variabilité en fonction de divers paramètres, comme le procédé de fabrication ou l'origine géographique entre autres. Des essais dans ce sens ont été réalisés au cours de la thèse avec la construction d'une base de données de la composition du thé basée sur la littérature et les bases de données existantes, mais cela n'a pas été concluant. La construction expérimentale d'une telle base de données est d'un très fort intérêt mais constituait des travaux supplémentaires trop importants dans le cadre de cette thèse. Par ailleurs, l'approche actuelle repose sur l'analyse au cours de la même séquence analytique d'un échantillon de référence (c.à.d. connu comme sain) et d'un échantillon suspect (celui dont on veut évaluer la qualité sanitaire). Cette stratégie demande un temps d'analyse important (plusieurs jours par séquence), ce qui peut constituer un obstacle à l'implémentation de ces approches pour les analyses de routine. Un moyen de réduire ce temps d'analyse serait de se passer de l'analyse systématique de l'échantillon de référence grâce à la stratégie de base de données évoquée plus haut. Enfin, pour une réelle évaluation non-ciblée fiable de la qualité sanitaire chimique des produits alimentaires, une réflexion devra être menée sur l'implémentation simultanée de technologies analytiques complémentaires (par exemple, LC-ESI-HRMS, LC-APCI-HRMS et GC-CI-HRMS) pour augmenter le niveau de confiance dans les résultats obtenus et étendre la gamme de composés détectables. Cela passe par la conception de processus de traitement de données capables de gérer des données de provenances multiples et de les recouper. Le processus proposé ici, avec sa modularité et ses polyvalences peut se voir comme une ébauche dans cette direction.

Pour encourager le développement futur de nouveaux outils et approches, toutes les données utilisées au cours de cette thèse pour développer et tester les approches non-ciblées seront disponibles sur la plateforme d'échange de données MetaboLights, à l'image de ce qui se fait dans le champ des études métabolomiques.

RÉFÉRENCES

- 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>
- Cariou, R., Omer, E., Léon, A., Dervilly-Pinel, G., & Le Bizec, B. (2016). Screening halogenated environmental contaminants in biota based on isotopic pattern and mass defect provided by high

- resolution mass spectrometry profiling. *Analytica Chimica Acta*, 936, 130–138. <https://doi.org/10.1016/j.aca.2016.06.053>
- Castro-Puyana, M., Pérez-Míguez, R., Montero, L., & Herrero, M. (2017). Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability. *TrAC - Trends in Analytical Chemistry*, 93, 102–118. <https://doi.org/10.1016/j.trac.2017.05.004>
- 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. <https://doi.org/10.1016/j.chroma.2014.06.100>
- Cotton, J., Leroux, F., Broudin, S., Marie, M., Corman, B., Tabet, J. C., ... Junot, C. (2014). High-resolution mass spectrometry associated with data mining tools for the detection of pollutants and chemical characterization of honey samples. *Journal of Agricultural and Food Chemistry*, 62(46), 11335–11345. <https://doi.org/10.1021/jf504400c>
- Di Guida, R., Engel, J., Allwood, J. W., Weber, R. J. M., Jones, M. R., Sommer, U., ... Dunn, W. B. (2016). Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics*, 12(5). <https://doi.org/10.1007/s11306-016-1030-9>
- Inoue, K., Tanada, C., Sakamoto, T., Tsutsui, H., Akiba, T., Min, J. Z., ... Toyo'oka, T. (2015). Metabolomics approach of infant formula for the evaluation of contamination and degradation using hydrophilic interaction liquid chromatography coupled with mass spectrometry. *Food Chemistry*, 181, 318–324. <https://doi.org/10.1016/j.foodchem.2015.02.117>
- Knolhoff, A. M., Zweigenbaum, J. A., & Croley, T. R. (2016). Nontargeted Screening of Food Matrices: Development of a Chemometric Software Strategy to Identify Unknowns in Liquid Chromatography-Mass Spectrometry Data. *Analytical Chemistry*, 88(7), acs.analchem.5b04208. <https://doi.org/10.1021/acs.analchem.5b04208>
- 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*, 1–10. <https://doi.org/10.1007/s00216-018-1028-4>
- Libiseller, G., Dvorzak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., ... Magnes, C. (2015). IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics*, 16(1), 118. <https://doi.org/10.1186/s12859-015-0562-8>
- Roullier, C., Guitton, Y., Valery, M., Amand, S., Prado, S., Robiou Du Pont, T., ... Pouchus, Y. F. (2016). Automated Detection of Natural Halogenated Compounds from LC-MS Profiles-Application to the Isolation of Bioactive Chlorinated Compounds from Marine-Derived Fungi. *Analytical Chemistry*, 88(18), 9143–9150. <https://doi.org/10.1021/acs.analchem.6b02128>
- 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>

CONCLUSION GÉNÉRALE ET PERSPECTIVES

Ces travaux de thèse avaient pour but de proposer une approche utilisant les outils et stratégies de la métabolomique (principalement la LC-HRMS et la chimiométrie) pour la détection non-ciblée des contaminants de l'aliment. Bien que les publications existantes sur ce sujet au début de ces travaux fassent état de performances prometteuses en termes de sensibilité et de capacité à détecter des composés potentiellement non-attendus, seul un scénario de type dépassement des teneurs maximales réglementaires avait été reporté dans la littérature, et ce pour des matrices alimentaires relativement simples et peu variables. De plus, les approches existantes n'avaient encore jamais été validées sur des plateformes d'analyse de marques ou de technologies différentes de celle de développement.

Dans le présent travail, le produit d'étude choisi a été le thé, sous forme de feuilles transformées (produit fini), et dans une moindre mesure d'infusion. Les trois questions centrales auxquelles nous avons essayé d'apporter des éléments de réponse sont : Quelles performances (en termes de sensibilité et d'exhaustivité) des approches non-ciblées, notamment dans des situations de contamination complexes ? Quelle méthodologie pour adapter les outils de traitement de données de la métabolomique à la détection non-ciblée de contaminants alimentaires ? Quelle robustesse de ces approches vis-à-vis des technologies LC-HRMS employées et des scénarios de contaminations étudiés ?

En premier lieu, une méthode analytique adaptée aux exigences propres à cette thématique (sensibilité, facilité d'implémentation, répétabilité) a été mise en place et validée par des essais de dopage/recouvrement. Elle repose sur un traitement de l'échantillon générique par solvant (mélange ACN/MeOH acidifié avec 0,1% d'acide formique) suivi d'une analyse par UHPLC-HRMS utilisant une phase stationnaire relativement innovante (silice greffée C18-PFP), offrant une meilleure rétention des composés polaires et aromatiques qu'une silice greffée C18 classique. Cette méthode a été validée sur une sélection de 32 molécules provenant de 4 classes de contaminants différentes et représentatives de la diversité potentielle de molécules d'intérêt analysables par LC-MS. Les niveaux de concentration considérés (entre 10 et 1 000 µg/kg) sont en accord avec les niveaux usuels et réglementaires existants. La diversité et les concentrations des molécules étudiées ici permettent de se montrer relativement confiants sur l'applicabilité de cette méthode à des situations de contamination variées et complexes.

Cette méthode a ensuite été utilisée pour proposer une première preuve de concept d'approche non-ciblée sur un cas de contamination relativement simple, avec les mêmes concentrations (de 10 à 1 000 µg/kg) et les 32 mêmes composés, avec un analyseur de type « ToF ». La stratégie adoptée s'inspire de celles existantes dans le domaine de la métabolomique. Sur le plan analytique, elle repose sur la conception de la séquence d'analyse en vue de minimiser le biais analytique (ordres d'injection aléatoires, utilisation d'échantillons de contrôle entres autres). Du point de vue du traitement de données, elle repose sur l'utilisation de la suite d'outils XCMS opérée sur la plateforme W4M pour l'extraction

des pics, suivie de plusieurs étapes de prétraitement (filtration des données, mise à l'échelle et normalisation) et enfin d'une analyse multivariée par ICA pour déterminer les ions d'intérêt qui sont ensuite annotés de façon semi-automatique. Des résultats prometteurs ont été obtenus sur le jeu test, avec 66% des contaminants dopés détectés, les molécules manquées étant celles présentant la moins bonne réponse lors de l'analyse.

Pour améliorer la robustesse et les performances de notre méthodologie, deux aspects importants du traitement de données (la gestion des valeurs manquantes et la filtration des données) ont été étudiés plus en détails pour aboutir à une version finale du processus de traitement de données. Ces travaux ont permis de valider le choix de la méthode de gestion des valeurs manquantes utilisée, méthode qui a été complétée avec une méthode développée pour l'occasion pour être à même de gérer toutes les situations. Ils ont également permis de mettre en évidence l'intérêt d'utiliser deux stratégies de filtration en parallèle pour favoriser la détection de potentiels contaminants.

La preuve de concept dans sa version finale a ensuite été appliquée avec succès sur trois types de scénarios d'évaluation de la qualité sanitaire chimique : (i) une contamination ponctuelle dans un produit à forte variabilité, (ii) un dépassement des teneurs maximales réglementaires (i.e. non-conformité), et (iii) la présence de contaminations différentes dans un produit conforme. Deux de ces scénarios (forte variabilité des échantillons et contaminations multiples sans dépassement) n'ont jamais été décrits dans la littérature pour des approches non ciblées ; de plus, les scénarios (ii) et (iii) ont pu être étudiés sur deux plateformes LC-HRMS de technologie différente (ToF et Orbitrap), ce qui n'a jamais été reporté auparavant. L'analyse du scénario présentant une forte variabilité entre échantillons (2 marques de thé, dopage à 50 µg/kg avec 3 contaminants, analyses sur plateforme ToF) a permis la détection de la totalité (100%) des contaminants présents, en dépit des fortes différences de composition entre les échantillons. Dans le cas du scénario de non-conformité (11 contaminants à 10 et 30 µg/kg), des taux de détection de 55% (ToF) et 64% (Orbitrap) ont été obtenus, en accord avec les résultats préliminaires obtenus auparavant. Pour ce cas relativement simple, la différence entre les deux plateformes est donc assez minime. L'étude du scénario de contaminations multiples sans dépassement (deux mélanges de dopage à 10 µg/kg sur différents échantillons) est plus informative sur les différences entre les deux plateformes instrumentales. En effet, dans ce cas, aucune technologie ne semble supérieure en termes de taux de détection (contamination n°1 : ToF 45% et Orbitrap 64%, contamination n°2 : ToF 67% et Orbitrap 33%). Pour autant, les deux appareils montrent une bonne complémentarité, puisque la combinaison des résultats de chaque plateforme permet de détecter 64% des contaminants du mélange n°1 et 100% du deuxième mélange.

Ce travail constitue donc une avancée dans le développement d'approches non ciblées pour la détection de contaminants chimiques alimentaires. Cependant de nombreuses interrogations demeurent et un long

chemin reste à parcourir jusqu'à l'implémentation de ces approches en routine. Les limites qui seront peut-être les plus dures à repousser sont les limites méthodologiques inhérentes à la stratégie employée.

La première est que l'approche existante nécessite d'avoir un produit de « référence » pour comparer l'échantillon « suspect » et ainsi pouvoir déterminer quels sont les contaminants présents. Dans le cas idéal, le produit de référence possède la même composition chimique que le produit à analyser, à l'exception des contaminants. Ce cas est assez improbable dans une situation réelle, car un produit agro-alimentaire a nécessairement une certaine variabilité, même à marque et procédé constant. Cette variabilité peut être plus ou moins importante suivant l'échelle de la production (artisanale ou industrielle), ou encore selon la nature même du produit, mais elle sera toujours présente et risque de perturber l'analyse. Une manière de gérer ce problème pourrait être la construction, pour la matrice agro-alimentaire d'intérêt (ici le thé), d'une base de données des composés naturellement présents dans le produit et non concernés par les problématiques de sécurité sanitaire chimique. L'idée est d'acquérir une connaissance la plus exhaustive possible de la composition du produit d'intérêt pour retirer des données analysées les composés connus, et ainsi augmenter la probabilité de détection des contaminants. Cette connaissance pourrait être générée par une étude de la littérature et des bases de données existantes (comme foodDB par exemple). Une telle stratégie a l'avantage d'être relativement simple à mettre en œuvre mais son exhaustivité est fonction de la connaissance de la matrice étudiée (certaines sont mieux connues et/ou moins complexes que d'autres) et elle ne permet pas de disposer d'informations factuelles d'identification (e.g. le temps de rétention des molécules). Implémenter une telle stratégie nécessite donc de disposer d'un spectromètre de masse à très haute résolution (>100 000) pour pouvoir enlever les composés non-pertinents de façon fiable. Cette stratégie a brièvement été évaluée sur nos données sans fournir de résultats probants ; des tests plus poussés sont nécessaires pour savoir à quel point cette stratégie peut être implémentée avec succès. Cette base de données peut également être générée expérimentalement, mais cela soulève de nombreuses questions et défis. Il semble peu envisageable de réaliser un inventaire exhaustif de la composition d'un aliment au regard de la complexité des produits agro-alimentaires, et en particulier des matrices comme le thé. Ces matrices sont composées en effet de plusieurs centaines (voire de milliers) de molécules avec une forte variabilité en fonction du procédé de fabrication et de l'origine géographique. On peut donc envisager deux manières d'aborder ce problème. Il est concevable, grâce à l'analyse d'un large panel d'échantillons représentatifs de la diversité de produits rencontrés (origine, procédé, cultivar, lots de fabrication par exemple), de déterminer quels sont les composés ayant la plus grande variabilité en fonction de ces différents paramètres et ainsi de pouvoir les retirer de l'analyse des données. Cette approche, basée sur une connaissance *a priori* de la matrice, aurait l'avantage d'être extrêmement simple à implémenter en routine une fois le panel d'échantillons analysé et les composés responsables de la variabilité identifiés. Cependant, la conception initiale du panel constitue un point critique et nécessite l'analyse d'un grand nombre d'échantillons. L'autre approche consisterait à travailler avec une base de données intégrant les empreintes de tout nouvel

échantillon au fil de l'eau. Des outils adéquats pourraient permettre ensuite d'automatiser lors du traitement de données la distinction entre une variabilité naturelle normale et une variabilité anormale (et donc potentiellement causée par des contaminants). Cette approche ne nécessiterait pas de collecte d'échantillons exhaustive comme la précédente, celle-ci se faisant au fur et à mesure de l'analyse des différents échantillons. Elle requiert cependant des compétences en matière de traitement de données allant au-delà des outils chimiométriques classiques (analyse multivariée, etc.), alors que la première approche peut se faire avec les outils classiquement utilisés.

La deuxième limite méthodologique porte sur le choix d'un échantillon de contrôle qualité (QC) stable dans le temps pour la comparaison de résultats venant de séquences analytiques distinctes. C'est un point critique pour les applications en routine de ces approches puisqu'elles reposent sur la comparaison de l'échantillon à analyser avec un (ou des) échantillons de référence, ces derniers pouvant avoir été acquis au cours de séquences analytiques antérieures. Or, la réponse d'un instrument évolue au cours du temps : les QC (c.à.d. des échantillons systématiquement analysés au cours de chaque séquence) sont donc nécessaires pour pouvoir comparer des données provenant de séquences différentes. Durant cette thèse, toutes les études ont été effectuées en une séquence analytique par mode d'ionisation, le QC employé était donc un mélange de tous les extraits obtenus au cours de l'étude en cours. Ce mode de fonctionnement n'est pas envisageable pour des études couvrant plusieurs séquences analytiques espacées temporellement. Les extraits étant souvent peu stables dans le temps (mauvaise tolérance aux cycles congélation/décongélation par exemple), il semble plus pertinent de choisir comme QC un échantillon de référence conservé à -80°C et dont un aliquot serait extrait à chaque séquence. Deux pistes de choix d'échantillon QC sont apparues durant cette thèse et mériteraient d'être discutées dans des futurs travaux. Tout d'abord, la solution la plus simple serait de prendre un échantillon arbitraire, de le considérer comme « échantillon zéro » et de l'utiliser systématiquement pour corriger la variation de réponse de l'instrument. L'avantage de cette solution est sa simplicité de mise en œuvre. Cependant, cet échantillon serait forcément non-représentatif de la diversité des échantillons analysés et la qualité de la correction potentiellement réduite. Une autre solution serait de constituer un échantillon « moyen » par le mélange d'un panel d'échantillons représentatifs de la diversité de la matrice. Cette solution serait plus compliquée à mettre en place, mais probablement à même d'assurer une meilleure correction de la dérive grâce à une meilleure représentativité de la composition chimique potentielle de la matrice.

Des limites techniques ont également été identifiées, peut-être plus faciles à surmonter que les limites méthodologiques. Tout d'abord, pour avoir une vision réellement exhaustive de la qualité sanitaire d'un aliment, il est nécessaire d'utiliser plusieurs techniques analytiques complémentaires en parallèle. Les travaux réalisés ici sont basés sur la LC-HRMS avec une source d'ionisation électrospray ; il pourrait être pertinent d'élargir cette approche à d'autres techniques basées sur la spectrométrie de masse (telles que la LC-HRMS avec une source APCI, ou encore la GC-HRMS avec une source d'ionisation chimique). L'élargissement de cette approche à d'autres techniques devra être accompagné par

l'adaptation depuis le domaine de la métabolomique d'outils de traitement de données permettant le recoupement des informations entre les données d'instrument différents. Un moyen d'encourager et de faciliter ces développements passerait par la publication des données brutes sur des plateformes d'échanges de données à l'image de ce qui a été fait pendant cette thèse.

Un besoin qui a aussi très rapidement émergé au cours de ces travaux est la nécessité de disposer d'une base de données de contaminants de l'aliment la plus complète possible et facilement interrogeable au sein des outils de traitement de données (à l'image des bases de données METLIN et HMDB dans le domaine de la métabolomique). Un pas dans cette direction a déjà été effectué par la publication de la « Toxin and toxin-target database » (T3DB, contenant des composés considérés comme toxiques) et par la présence dans METLIN et HMDB de certains contaminants de l'aliment. Ces trois bases de données contiennent notamment des spectres de masse simulés, ou plus rarement expérimentaux, qui permettent dans certains cas d'accroître le niveau de confiance dans l'annotation. Dans ce travail, face à l'absence de base de données spécifiquement dédiées à la contamination des aliments, les données de la T3DB ont été recoupées avec celles d'autres bases de données pour constituer une liste d'ions suspects. En 2017 l'EFSA a publié une base de données des contaminants de l'aliment contenant de nombreuses informations (OpenFoodTox <http://www.efsa.europa.eu/en/microstrategy/openfoodtox>), mais malheureusement sans spectres de masse à l'heure actuelle. Il est possible de télécharger cette base de données et d'extraire les formules brutes des composés pour générer des listes d'ions « suspects » pouvant aider l'annotation.

Ainsi, si au cours de cette thèse nous avons montré le potentiel des approches non-ciblées de type métabolomique pour la détection des contaminants de l'aliment (y compris à des très faibles niveaux et dans des situations complexes et variées), des travaux sont encore nécessaires avant leur implémentation en routine, les principaux défis à relever venant de la variabilité et de la complexité des matrices agro-alimentaires.

Titre : Développement d'une approche non-ciblée par empreinte pour caractériser la qualité sanitaire chimique de matrices agro-alimentaires complexes

Mots clés : Empreintes ; Non-ciblée ; Chimométrie; Chromatographie ; Spectrométrie de masse

Résumé : L'assurance de la sécurité sanitaire des aliments vis-à-vis des contaminants chimiques est un enjeu en constante évolution en raison des sources multiples de contamination (pesticides, mycotoxines, néoformés indésirables, et migrants des matériaux au contact entre autres). Actuellement, l'évaluation complète de la qualité sanitaire d'un aliment nécessite la multiplication de méthodes analytiques dites « ciblées » ayant un coût important. De plus, malgré la multiplication des méthodes ciblées, tout contaminant non recherché ne sera pas détecté. Il apparaît nécessaire aujourd'hui de faire évoluer ces méthodes vers des approches analytiques « non-ciblées » susceptibles, via l'analyse d'empreintes chimiques, d'évaluer la présence d'une gamme aussi large que possible des contaminants dans une matrice alimentaire. Les travaux de thèse ont porté sur l'utilisation de la spectrométrie de masse haute résolution (LC-HRMS) et de la chimométrie pour développer une méthode capable de caractériser la qualité sanitaire des aliments. La matrice de développement est le thé, choisi pour sa complexité d'analyse, sa large consommation et les alertes sanitaires récurrentes à son sujet.

Une première preuve de concept de la méthode a été mise en place sur un thé vert de référence et un panel de 32 contaminants choisis pour leur diversité de sources et structures chimiques, puis des situations plus complexes ont été investiguées : application à d'autres types de thé, analyse simultanée d'échantillons de marques et d'origines géographiques distinctes, et enfin application en aveugle à des situations de contamination complexes avec la présence de plusieurs schémas de contamination au sein du même jeu d'échantillons. L'utilisation d'outils de traitement de données libres et ouverts a permis de développer un processus de traitement des données unifié pour deux plateformes analytiques LC-HRMS de technologies et de marques différentes (ToF et Orbitrap), ce qui n'a jamais été réalisé pour l'étude de la sécurité sanitaire chimique des aliments. Par ailleurs, le développement de ce processus a été l'occasion de réaliser une étude méthodologique du comportement de certains outils pour les approches non-ciblées de détection des contaminants de l'aliment.

Title : Development of a non-targeted fingerprinting approach to assess the chemical safety of complex food matrices

Keywords : Fingerprinting; Untargeted; Chemometrics; Chromatography; Mass spectrometry

Abstract: Ensuring food safety, especially toward chemical contaminants, is an issue in constant evolution due to multiple sources of contamination (pesticides, mycotoxins, neoformed contaminants, migrants from packaging among others). Currently, several targeted analyses are needed to fully assess the chemical safety of foods, generating high cost. Moreover, despite the number of analyses performed, a contaminant not targeted is not detected. Therefore, it is necessary to develop new methods based on non-targeted approaches able to assess, through analysis of chemical fingerprints, the occurrence of as many contaminants as possible in a food matrix. The main purpose of this work lies in the use of high resolution mass spectrometry (LC-HRMS) and chemometrics in order to develop a method capable of assessing food safety.

Tea has been chosen as a development product for its analytical complexity, its broad consumption and its safety issues. A first proof-of-concept of the method has been set up on a reference green tea with a pool of 32 representative food contaminants, chosen for their diversity in terms of sources and chemical structures. More complex situations were further investigated with different types of tea, several brands considered at once and, last but not least, with the application to blind detection of contaminants in complex cases. Free and open-source data analysis tools were used to build a unified data treatment process to analyze data from two LC-HRMS analytical platforms of different technologies (ToF and Orbitrap), which is new for food safety studies. The development of this process also enabled a methodological study of the behavior of several tools used in untargeted approaches for food safety.