



HAL
open science

Empreinte volatile des produits agroalimentaires : un nouvel outil pour tracer l'authenticité et suivre des procédés

Marine Reyrolle

► **To cite this version:**

Marine Reyrolle. Empreinte volatile des produits agroalimentaires : un nouvel outil pour tracer l'authenticité et suivre des procédés. Chimie analytique. Université de Pau et des Pays de l'Adour, 2022. Français. NNT : 2022PAUU3019 . tel-04147737

HAL Id: tel-04147737

<https://theses.hal.science/tel-04147737>

Submitted on 1 Jul 2023

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

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

THÈSE

UNIVERSITE DE PAU ET DES PAYS DE L'ADOUR

École doctorale des Sciences Exactes et de leurs Applications

Présentée et soutenue le 09 décembre 2022

par **Marine REYROLLE**

pour obtenir le grade de docteur
de l'Université de Pau et des Pays de l'Adour

Spécialité : Chimie Analytique

EMPREINTES VOLATILES DE PRODUITS AGROALIMENTAIRES : UN OUTIL POUR TRACER L'AUTHENTICITE ET SUIVRE DES PROCÉDES

MEMBRES DU JURY

RAPPORTEURS

- Frédéric THEVENET
- Jean-Luc LE QUERE

Professeur / IMT Nord Europe
Directeur de recherches / INRAE Dijon

EXAMINATEURS

- Valérie SIMON
- Hélène TORMO
- Florence PANNIER
- Thierry PIGOT

Maître de Conférences (HDR) / INP-ENSIACET, Toulouse
Enseignante Chercheuse / Ecole d'Ingénieur de Purpan, Toulouse
Professeure / Université de Pau et des Pays de l'Adour
Professeur / Université de Pau et des Pays de l'Adour

DIRECTEURS

- Valérie DESAUZIERS
- Mickaël LE BECHEC

Professeure / IMT Mines Alès, Pau
Ingénieur de Recherches / CNRS



REMERCIEMENT

La thèse est un travail de longue haleine nécessitant l'aide et la collaboration de plusieurs personnes. Je souhaite donc exprimer ma reconnaissance envers toutes les personnes qui m'ont épaulée et accompagnée au cours de ces trois années.

J'adresse mes sincères remerciements à Frédéric Thevenet et à Jean-Luc Le Quéré de l'honneur qu'ils me font d'accepter d'être rapporteurs et d'évaluer ce travail en y consacrant une partie de leur temps. Je tiens également à remercier les membres du jury, Valérie Simon, Héléne Tormo, Florence Pannier et Thierry Pigot, d'avoir répondu présent pour examiner ce travail.

Je remercie tout d'abord le laboratoire IPREM de m'avoir accueillie et de m'avoir permis de réaliser ces travaux de thèse. Je remercie également la région Nouvelle Aquitaine pour son soutien financier.

Je tiens principalement à remercier Valérie Desauziers et Mickael Le Behec, mes directeurs de thèse, d'avoir accepté de diriger ma thèse et de m'avoir offert l'opportunité de réaliser ce travail de recherche. Je leur dis merci pour leurs conseils toujours avisés, leur disponibilité et leur confiance. Merci de m'avoir donné les moyens de réaliser cette thèse dans les meilleures conditions, et de m'avoir aidé à surmonter toutes les difficultés qui se sont présentées.

J'adresse aussi mes remerciements aux personnes que je nomme « ressources » dans ma thèse pour leurs précieux conseils : Germain Vallverdu pour ses compétences en langage de programmation, Noëlle Bru pour son expertise statistique, Gilles Bareilles pour la compréhension des résultats, Mylène Ghislain pour ses connaissances sur l'instrument.

Merci aux partenaires agroalimentaires qui ont cru en ce projet et m'ont fait confiance : la chambre d'agriculture 64, l'INRA de Lusignan, les plateformes technologiques Errecart et Actalia, la société AIA ainsi que Lydia Gauthier.

Cette thèse est le fruit d'une collaboration au sein d'une équipe. Qu'elle en soit remerciée pour sa joie, sa bonne humeur, son soutien et ses conseils.

De ces 3 années est née une belle amitié avec mes doctorants (prédécesseurs ou successeurs). Je les remercie particulièrement pour leur bienveillance et leur soutien inconditionnel mais surtout pour tous ces bons moments et ces fous rires avec une pensée profonde pour la « Team Barbeuk ».

Un énorme merci à ma relectrice depuis des années qui a pris le temps de scruter l'intégralité de ce manuscrit à l'affût des fautes, tatie Jo.

J'aimerais remercier mon comité de soutien inconditionnel : mes amis (Kelly, Manon, Océane, Marie et Dimitri) et ma famille, qui me soutiennent et croient en moi depuis toujours.

Du fond de mon cœur, mille mercis à mes parents pour leur amour, les valeurs qu'ils m'ont transmises, leur compréhension et leur soutien inconditionnel tout au long de mes études. Merci d'avoir toujours cru en moi.

Ces remerciements ne peuvent s'achever sans une pensée pour mes animaux (Pisco, Game, Rico et Mandoline) qui n'ont probablement pas idée du réconfort qu'ils m'ont apportée.

Table des matières

INTRODUCTION	9
CHAPITRE 1 : Etat de l'art	13
I. Composés Organiques Volatils et aliments	13
A. Contexte général	13
B. Cas particulier du thé et du fromage.....	20
C. Importance des COV dans le domaine agroalimentaire	37
D. Conclusion	42
II. Les méthodes analyse des COV	43
A. Méthodes usuelles d'analyses des COV dans les matrices alimentaires	43
B. Nouvelle technologie de spectrométrie de masse à injection directe : SIFT-MS (Selected Ion Flow Tube Mass Spectrometer)	63
C. Conclusion	74
III. Chimiométrie.....	77
A. Prétraitement des données de volatolomique	77
B. Traitements statistiques des données	82
C. Conclusion	86
IV. Conclusion générale	87
CHAPITRE 2 : Développements analytiques	111
I. Description des échantillons étudiés	111
A. Thé	111
B. Fromage Ossau-Iraty	111
II. Conditions d'échantillonnage.....	112
A. Mise en œuvre de l'échantillon.....	113
B. Conditions d'espace de tête pour l'analyse SIFT-MS	115
C. Conditions HS-SPME pour l'analyse GC-MS/FID.....	122
D. Conclusion	129
III. Conditions analytiques	130
A. GC-MS/FID	130
B. SIFT-MS.....	131
C. Etalonnage.....	131
IV. Traitement des signaux	136
A. Données HS-SPME-GC-MS/FID.....	136

B.	Données HS-SIFT-MS	137
V.	Traitement statistique	145
A.	Analyse non supervisée exploratoire : la sparse ACP.....	145
B.	Analyse supervisée exploratoire et confirmatoires : la sparse PLS-DA.....	146
C.	Représentation graphique.....	146
D.	Tests statistiques confirmatoires	147
VI.	Conclusion	147
	CHAPITRE 3 : Empreinte volatile des produits alimentaires utilisant des données SIFT-MS non ciblées couplées à des méthodes mixOmics pour la discrimination des profils : application au fromage	151
	CHAPITRE 4 : Discrimination de la zone géographique et de la couleur de 26 thés à partir des empreintes volatiles obtenues par deux méthodes : HS-SPME-GC-MS/FID et HS-SIFT-MS.	175
	CHAPITRE 5 : Authenticité du thé par signatures multi-élémentaires, rapport isotopique et empreintes volatiles	189
	CONCLUSIONS ET PERSPECTIVES	207
	ANNEXE 1 : Résultats cinétique Ossau (VG-GC-FID).....	209
	ANNEXE 2 : Résultats cinétique Ossau (HS-SPME-GC-MS/FID).....	210
	ANNEXE 3 : Explication de l’algorithme de traitement des données SIFT-MS Full scan.....	211

Table des figures

Figure 1 : Voies métaboliques mise en jeu dans la formation des composés volatils odorants d'après Cadwaller (2019) ⁵	14
Figure 2 : Schéma résumé de la formation des COV odorants par la réaction de Maillard d'après Jousse (2002) ¹⁰	15
Figure 3 : Perception olfactive.....	17
Figure 4 : Courbe d'évolution de l'intensité odorante perçue en fonction de la concentration du composé d'après Medjkoune (2018) ³²	18
Figure 5 : Les différentes étapes de fabrication selon la couleur de thé d'après Lydia Gautier ⁵⁵	23
Figure 6 : Formation de la β -ionone par dégradation enzymatique du β -carotène	24
Figure 7 : Formation du nérolidol et de l' α -farnésène par photo-oxydation du phytofluène	24
Figure 8 : Formation du n-hexanal et n-hexanol par dégradation de l'acide linoléique.....	25
Figure 9 : Formation du Géraniol et du Linalool à partir du pyrophosphate de géranyle	25
Figure 10 : Formation de composés soufrés par la dégradation de strecker de la Méthionine d'après Ho et al (2015) ⁵³	26
Figure 11 : Les grandes familles de fromages avec des exemples	29
Figure 12 : les 6 grandes étapes de production du fromage.....	31
Figure 13 : Schéma résumé des trois voies biochimiques impliquées dans la formation des composés aromatiques du fromage d'après Andiç et al. (2015) ⁷⁵	32
Figure 14 : (A) Schéma synthétique des différentes voies de formation des composés aromatiques par le catabolisme des acides aminés (B) Exemple du catabolisme de la leucine pour former du 3-methylbutanal ou du 3-methylbutanol d'après Smit et al. (2005) ⁷⁸	32
Figure 15 : Schéma synthétique des différentes voies de formation des composés aromatiques à partir des voies cataboliques des acides gras d'après Collins et al (2003) ⁷⁹	33
Figure 16 : Formation du 2,3-butanediol par la voie métabolique du citrate ⁸⁰	33
Figure 17 : Principales voies catabolique du lactate d'après Andiç et al (2015) ⁷⁵	34
Figure 18 : Schéma de la micro-extraction en phase solide d'après Starowicz (2021) ²	46
Figure 19 : Représentation schématique du principe de l'HS-SPME.....	49
Figure 20 : Evolution de la quantité d'analyte n adsorbée en fonction du produit « $C_a \times t$ ».....	51
Figure 21 : Principe du SIFT-MS.....	64
Figure 22 : Thé vert de Ceylan de la marque Ethiquable utilisé pour le développement analytique.	113
Figure 23 : Thé sec, humide et infusât	113
Figure 24 : Chromatogrammes obtenus pour le thé sec (A), humide (B) et l'infusât (C) (acquisition en mode scan 10-250 u.m.a et tracés en mode extraction d'ions (EIC) de 35 à 250 u.m.a)	114
Figure 25 : Photo des morceaux de fromage reçu (A) et de la découpe (B)	115
Figure 26 : Schéma du flacon d'espace de tête utilisé pour l'analyse SIFT-MS.....	116
Figure 27 : Cinétiques de transfert des COV du thé vers l'espace de tête à différentes températures et avec différentes quantités de thé.	117
Figure 28 : Cinétique d'émission des COV par SIFT-MS, suivi avec l'ion H ₃ O ⁺ (A) et NO ⁺ (B)	121
Figure 29 : Aires des pics chromatographiques obtenues à partir du signal FID de quelques composés extraits du thé avec différentes fibres	125
Figure 30 : Effet de la température sur l'extraction des COV présents dans le thé par HS-SPME-GC-FID avec la fibre PDMS/DVB/CAR (aires de pics obtenues en FID).....	126
Figure 31 : Cinétique d'émission des COV (aires de pics obtenues en FID) présents dans le thé en faisant varier le temps d'incubation.....	127

Figure 32 : Cinétique d'extraction des COV (aires de pics obtenues en FID) présents dans le thé par HS-SPME-GC-MS/FID avec une fibre PDMS/DVB/CAR, un temps d'incubation de 12 min et une température de 60°C.....	128
Figure 33: Schéma du protocole d'échantillonnage du thé pour l'analyse des COV par HS-SIFT-MS et HS-SPME-GC-MS/FID	129
Figure 34 : Schéma du protocole d'échantillonnage du fromage pour l'analyse des COV par HS-SIFT-MS et HS-SPME-GC-MS/FID.....	130
Figure 35 : Concentration expérimentale de C ₆ F ₆ en fonction de la concentration théorique	132
Figure 36 : Histogramme représentant la concentration théorique attendue et la moyenne des concentration mesurées en C ₆ F ₆	133
Figure 37 : Histogramme représentant la concentration de l'étalon interne C ₆ F ₆ introduit dans l'espace de tête des flacons contenant des échantillons de fromages. La barre représente la concentration injectée dans les flacons.....	134
Figure 38 : Chromatogrammes (signal FID) obtenus pour l'analyse HS-SPME-GC-MS/FID de l'hexafluorobenzène seul (A) et dans l'espace de tête d'un échantillon de fromage (B)	135
Figure 39 : Graphique des cartes de contrôle pour l'ion H ₃ O ⁺ 22 (A) et H ₃ O ⁺ 52 (B)	139
Figure 40 : Histogramme des valeurs de fréquence obtenus pour l'ion H ₃ O ⁺ 52	140
Figure 41 : ACP sur les données des thés dont les fréquences ont été corrigées par l'hypothèse 4 (A), 5 (B) et 9 (C). Les codes couleurs correspondent aux dates des analyses (AnnéeMoisJour).....	143
Figure 42 : courbes de la valeur de l'ICF pour les 8 ions précurseurs en fonction du rapport masse sur charge des ions détectés.....	144

Table des tableaux

Tableau 1 : Quelques exemples de composés aromatiques identifiés dans le thé	28
Tableau 2 : Quelques exemple de composés aromatiques du fromage.....	37
Tableau 3 : Principales caractéristiques des phases absorbantes et adsorbantes des fibres SPME ²⁵⁸ . 52	
Tableau 4 : Exemple d'applications de l'HS-SPME dans l'analyse des composés volatils de produits agroalimentaires.....	57
Tableau 5 : Réaction d'ionisation chimique possible pour les 8 ions précurseurs.....	65
Tableau 6 : Ions produits lors des réactions entre les huit ions réactifs et l'eau, le dioxyde de carbone et l'oxygène. Les valeurs des rapports masse sur charge m/z des différents ions sont en italiques. D'après l'étude de Ghislain et al. (2020) ³⁴⁶	67
Tableau 7 : Constantes expérimentales de réaction pour chaque ion précurseur avec l'eau, l'oxygène et le dioxyde de carbone d'après l'étude de Ghislain et al. (2020) ³⁴⁶	67
Tableau 8 : Composition du mélange de calibration du SIFT-MS.....	68
Tableau 9 : exemple de conflits d'ions pour la quantification de 2-Hexen-1-ol, de 2-Hexanone et de 2,3-pentanedione	70
Tableau 10 : Exemple de conflits d'ions pour des composés isomériques avec l'acétone et du propanal	70
Tableau 11: Exemple d'applications de la technologie SIFT-MS dans l'analyse des composés volatils de produits agroalimentaires	75
Tableau 12 : Description de quelques logiciels libres d'accès pour l'exploitation non ciblée d'analyses GC/LC-MS	80
Tableau 13 : Principales méthodes statistiques utilisées pour interpréter des données obtenues par l'analyse des COV dans des études agroalimentaires.....	85
Tableau 14 : Protocole d'alimentation du producteur E1.....	112
Tableau 15 : Protocole d'alimentation du producteur E2.....	112
Tableau 16 : Liste des ions produits suivis et leur composé associé.....	120
Tableau 17 : Identification et détection des composés extraits avec les différentes fibres étudiées	124
Tableau 18 : Composition du mélange de gaz d'étalonnage	131
Tableau 19 : Gamme d'étalonnage C ₆ F ₆	132
Tableau 20 : Hypothèse de calculs pour corriger la fréquence.....	141

Abbreviations

ACP : Analyse en Composante Principale

ADH : Alcool DésHydrogénase

AOC : Appellation d'Origine Contrôlée

AOP : Appellation d'Origine Protégée

APCI-MS : Atmospheric Pressure Chemical Ionization Mass Spectrometry

ASE : Accelerated Solvent Extraction

Br : Branching ratio

CAH : Classification Ascendante Hiérarchique

CAR : CARboxen

COV : Composé Organique Volatil

DART-MS : Direct Analysis in Real Time Mass Spectrometry

DIMS : Direct Injection Mass Spectrometry

DI-SPME : Direct Injection Solid Phase Micro Extraction

DVB : DiVinylBenzène

E-nose : Electronic nose

ESI-MS : Electro Spray Ionization Mass Spectrometry

FPD : Flame Photometric Detector

FID : Flame Ionization Detector

GC : Gaz Chromatography

GC-O : Chromatographie Gazeuse couplée à détecteur olfactométrique

GC-FID : Chromatographie Gazeuse couplée à un Détecteur à Ionisation de Flamme

GC-IMS : Chromatographie Gazeuse couplée à la Spectrométrie de Mobilité Ionique

GC-MS : Chromatographie Gazeuse couplée à la Spectrométrie de masse

GC-MS/FID : Chromatographie Gazeuse couplée à un Spectromètre de Masse et un Détecteur à Ionisation de Flamme

HS : HeadSpace

HPL : HydroPéroxyde Lyase

HAP : Hydrocarbure Aromatique Polycyclique

HPI : High Precision Inlet

ICP-MS : Inductively Coupled Plasma Mass Spectrometry

IGP : Indication Géographique Protégée

LLE : Liquide-Liquide Extraction

LC-MS : Chromatographie Liquide couplée à la Spectrométrie de Masse

LOD : Limit of Detection

LOQ : Limit of Quantification

LOX : LipOxygénase

MAE : Microwave Accelerated Extraction

MC-ICP-MS : Multi-Collector coupled with Inductively Coupled Plasma Mass Spectrometry

MIM : Multi Ion Monitoring

MS : Mass Spectrometry

MW : Molecular weight ou MM : masse molaire

NPD : Nitrogen Phosphorus Detector

NIST : National Institute of Standards and Technology

OAV : Odour Activity Value

OT : Odour Threshold

PA : PolyAcrylate

PDMS : PolyDiMéthylSiloxane

PEG : PolyEthylène Glycol

PLE : Pressurized Liquid Extraction

PLS-DA : Partial Least Square Discriminant Analysis

PTR-MS : Proton-transfer-reaction mass spectrometry

P&T : Purge and Trap

RSD : Relative Standard Deviation

SBSE : Stir-Bar Sorptive Extraction

SIFT-MS : Selected Ion Flow Tube-Mass Spectrometry

SIM : Single Ion Monitoring

SLE : Solid-Liquid Extraction

SPE : Solid Phase Extraction

SPME : Solid Phase Micro Extraction

TOF : Time Of Flight

UAE : Ultrasound Assisted Extraction

XML : eXtensible Markup Language

INTRODUCTION

La France est un pays reconnu mondialement pour la qualité et la diversité de sa nourriture qui participe à l' « art de vivre à la Française ». Ainsi en 2010, l'UNESCO classait « le repas gastronomique des Français » au patrimoine culturel immatériel de l'humanité¹. De nombreux labels assurent aux consommateurs la qualité et la provenance de produits agroalimentaires. Parmi ces labels rencontrés dans notre quotidien, sont retrouvés les labels d'indications géographiques (Appellation d'Origine Protégée ou Contrôlée, AOP, AOC, Indication Géographique Protégée IGP...), les sigles garantissant une qualité spécifique (Spécialité Traditionnelle Garantie STG, Label rouge...) et les certifications garantissant des pratiques agricoles vertueuses (Agriculture Biologique AB, commerce équitable...). Tous ces labels doivent obligatoirement respecter un cahier des charges rigoureux qui précise les textes référentiels ou normes permettant une certification agréée par un organisme certificateur accrédité par le Comité Français d'Accréditation (COFRAC).

Malheureusement, de nombreux scandales dans la filière agroalimentaire (vache folle, grippe aviaire, cas de la viande de cheval dans les plats préparés en 2013²) ont ébranlé la confiance des consommateurs. De plus, ces dernières années une augmentation des fraudes alimentaires a été observée, se traduisant par des étiquetages erronés, la falsification de documents, l'adultération d'un produit par substitution d'un ingrédient de moindre qualité, ou encore l'émergence de pseudo-labels, utilisant la notoriété de certains produits et induisant la confusion chez les consommateurs. Rien de bien rassurant pour le consommateur qui n'arrive plus à savoir ce qui se retrouve dans son panier ou encore dans son assiette. Les consommateurs sont donc de plus en plus exigeants en matière de sécurité alimentaire, de qualité nutritionnelle, de diversité alimentaire et deviennent « informativores »³ par la recherche d'authenticité et de transparence des processus de production tout au long de la chaîne alimentaire, sans oublier pour autant l'impact environnemental. En effet, les consommateurs deviennent éco-responsables et montrent un intérêt important aux modes de production, au respect d'une agriculture responsable, au bien-être animal et à la rémunération équitable des agriculteurs. Ils privilégient de plus en plus les circuits courts permettant de manger des produits de saison tout en soutenant les producteurs de leurs territoires.

Face à ce constat, l'Union Européenne et l'industrie agro-alimentaire ont dû s'adapter pour répondre aux besoins des consommateurs afin de regagner leur confiance. La traçabilité est l'outil de gestion des risques qui suit le parcours des aliments, allant du producteur jusqu'au consommateur, en passant par les transformateurs et les distributeurs, c'est-à-dire « de la fourche à la fourchette »^{4,5}. Au-delà du parcours d'un aliment, son authenticité est un facteur clé pour le consommateur. En effet, déterminer l'authenticité des aliments signifie être capable de détecter si leur description est correcte ou erronée (ne répondant pas aux exigences d'un nom légal), si des ingrédients ont été remplacés par d'autres de moindre qualité, ou encore si la matière première est conforme (origine géographique, espèce...).

De nombreux outils sont utilisés par les filières agroalimentaires pour suivre la traçabilité et l'authenticité des produits, qu'ils soient déclaratifs avec la fourniture de documents administratifs ou de contrôles (sanitaires, qualité...). En effet, il existe de nombreuses techniques analytiques et émergentes qui permettent d'assurer la traçabilité et l'authenticité des aliments. Ces différentes méthodes peuvent être classées en trois grandes catégories : les techniques biologiques qui utilisent par exemple l'analyse de l'ADN et l'analyse de la flore microbienne, les techniques physiques mesurant

par exemple la couleur, le calibre ou la fermeté d'un produit et les techniques chimiques à la recherche de « traceurs » comme par exemple des isotopes ou bien des composés organiques volatils⁶.

Dans le secteur agroalimentaire, les composés organiques volatils (COV) émis par les aliments participent à leurs qualités organoleptiques et définissent l'odeur d'un aliment, premier sens mis à contribution lors de l'acceptation d'un produit par les consommateurs. Cette qualification organoleptique est principalement assurée par des analyses sensorielles dont les performances sont parfois affaiblies par des paramètres affectant la perception du jury. L'analyse de l'ensemble des COV est donc un point clé pour les acteurs de la filière puisqu'il caractérise la qualité et la typicité d'un produit, et apporte des informations essentielles quant à son évolution au cours de sa fabrication, son stockage et son vieillissement. La caractérisation de l'ensemble des COV peut constituer une empreinte volatile identifiant un produit et pouvant ainsi répondre à des questions de traçabilité, de qualification d'un procédé ou de l'évolution du produit⁷.

La méthode principalement utilisée pour l'analyse des COV se base sur une technique séparative : la chromatographie en phase gazeuse couplée à la spectrométrie de masse. Cependant cette approche est relativement longue (plusieurs dizaines de minutes) et ne répond pas complètement aux besoins des filières agroalimentaires qui requièrent des analyses à haut débit, sur un nombre élevé d'échantillons, si possible sans modification ou destruction du produit et sans impacter avec le procédé de transformation. Par contre, l'apparition de nouvelles technologies de spectrométrie de masse à injection directe (DIMS)^{8,9} s'avèrent très prometteuses puisqu'elles permettent des analyses relativement rapides (de l'ordre de la minute).

L'objectif principal de ce travail de thèse est donc de développer une nouvelle approche globale de mesure d'empreintes volatiles d'un produit agroalimentaire par spectrométrie de masse à injection directe. La technologie SIFT-MS (Selected Ion Flow Tube Mass Spectrometry) a été utilisée. Cette nouvelle méthode a pour but d'apporter des informations utiles aux producteurs sur l'authenticité, sur les procédés de production et de conservation de leurs produits.

Ce travail s'articule autour de 3 objectifs :

1. Le développement analytique de la mesure non spécifique des COV émis par une matrice alimentaire par la technologie SIFT-MS et la démonstration de son application au secteur agro-alimentaire.
2. La comparaison de cette nouvelle approche avec des méthodes traditionnellement utilisées pour l'analyse des COV dans les produits alimentaires comme la chromatographie en phase gazeuse (GC). Dans cette étude, la chromatographie en phase gazeuse équipée d'une double détection par spectrométrie de masse et ionisation de flamme (GC-MS/FID) sera utilisée. Les méthodes de préparation d'échantillon associées aux deux méthodes seront également développées.
3. Le couplage de cette approche par empreinte volatile avec d'autres techniques utilisées pour étudier la traçabilité et l'authenticité des produits alimentaires comme des méthodes d'analyse multi-élémentaires et de rapports isotopiques du strontium. Ces outils complémentaires peuvent être extrêmement puissants pour renforcer les systèmes de contrôle sur l'ensemble d'une chaîne d'approvisionnement.

Ce manuscrit est structuré en 5 chapitres :

Le CHAPITRE 1 présente un état de l'art avec une première partie présentant le contexte général et le rôle des COV dans l'agroalimentaire. Les méthodes de prélèvement et d'analyse des COV sont ensuite abordées et une dernière partie aborde les outils chimiométriques nécessaires à l'exploitation des données d'empreintes volatiles.

Le CHAPITRE 2 décrit le développement des méthodes d'échantillonnage, d'analyse et de traitement des données effectués pour mettre en place la caractérisation des empreintes volatiles spécifiques de deux matrices alimentaires (le thé et le fromage) par la technique SIFT-MS et la technique GC-MS/FID.

Le CHAPITRE 3 présente un premier cas d'application de l'approche des empreintes volatiles SIFT-MS dans des échantillons de fromage afin de discriminer la typicité des producteurs et étudier l'impact du changement alimentaire des animaux sur le produit final.

Le CHAPITRE 4 concerne l'étude comparative d'analyses ciblées et non ciblées de COV par SIFT-MS et GC-MS/FID, effectuées sur des échantillons de thé.

Dans le CHAPITRE 5, deux méthodes analytiques, ICP-MS et SIFT-MS permettant respectivement de déterminer la signature inorganique et les empreintes volatiles de produits alimentaires seront étudiées. L'objectif est d'évaluer, sur des échantillons de thé, les performances et la complémentarité de ces méthodes pour tracer la chaîne d'approvisionnement du thé.

Enfin, une conclusion générale reprend l'ensemble des travaux réalisés au cours de cette thèse tant sur le développement analytique que sur l'application au domaine agroalimentaire au travers de deux produits, le fromage et le thé. De nouveaux axes de travail émanant des résultats ou observations obtenus seront proposés.

Cette thèse, cofinancée par la région Nouvelle Aquitaine et l'Université de Pau et des Pays de l'Adour (UPPA), a été réalisée au laboratoire IPREM (Institut des Sciences Analytiques et de Physico-Chimie pour l'Environnement et les Matériaux). Des contrats de collaboration ont été établis avec différents partenaires du projet : la chambre d'agriculture 64, l'INRA de Lusignian, les plateformes technologiques Errecart et Actalia et Lydia Gauthier, expert en thé.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. UNESCO. Convention for the Safeguarding of the Intangible Cultural Heritage 2003. *Int. J. Cult. Prop.* **12**, 447–458 (2010).
2. Sammut, J., Gopi, K., Saintilan, N. & Mazumder, D. Facing the challenges of food fraud in the global food system. in *Food Authentication and Traceability* 35–63 (Elsevier, 2021). doi:10.1016/B978-0-12-821104-5.00009-X.
3. Halawany-Darson, R. Le Traitement de l'information dans le processus de prise de décision du consommateur : le cas de la traçabilité des produits alimentaires. (2012).
4. ISO/TC34/SC17. *Traçabilité de la chaîne alimentaire — Principes généraux et exigences fondamentales s'appliquant à la conception du système et à sa mise en oeuvre.* 8 (2007).
5. Bosona, T. & Gebresenbet, G. Food traceability as an integral part of logistics management in food and agricultural supply chain. *Food Control* **33**, 32–48 (2013).
6. Danezis, G. P., Tsagkaris, A. S., Camin, F., Brusica, V. & Georgiou, C. A. Food authentication: Techniques, trends & emerging approaches. *TrAC Trends Anal. Chem.* **85**, 123–132 (2016).
7. Cajka, T. & Hajslova, J. Volatile Compounds in Food Authenticity and Traceability Testing. in *Food Flavors* vol. 20116950 355–412 (CRC Press, 2011).
8. Biasioli, F., Yeretian, C., Märk, T. D., Dewulf, J. & Van Langenhove, H. Direct-injection mass spectrometry adds the time dimension to (B)VOC analysis. *TrAC Trends Anal. Chem.* **30**, 1003–1017 (2011).
9. Lebron-Aguilar, R., Soria, A. C. & Quintanilla-Lopez, J. E. Comprehensive evaluation of direct injection mass spectrometry for the quantitative profiling of volatiles in food samples. *Philos. Trans. R. Soc. -Math. Phys. Eng. Sci.* **374**, 20150375 (2016).

CHAPITRE 1 : Etat de l'art

I. Composés Organiques Volatils et aliments

Un composé organique volatil (COV) est un composé organique constitué de carbone et d'hydrogène, pouvant être substitué par des hétéroatomes (oxygène, azote, soufre, halogènes...) et dont la pression de vapeur saturante est supérieure à 10 Pa à 20°C et à pression atmosphérique (article 2 de la directive 1999/13/CE du Conseil européen du 11 mars 1999).

Les aliments sont des produits complexes dans lesquels on retrouve un grand nombre de COV dans une large gamme de concentrations¹. Parmi ceux-ci, on distingue principalement les arômes, composés odorants qui jouent un rôle essentiel dans la perception de l'aliment². En effet, bien que ces molécules n'apportent aucune qualité nutritive, elles contribuent à l'odeur qui est la première caractéristique déterminant l'acceptabilité d'un aliment¹.

Dans ce chapitre, un premier paragraphe présentera des généralités sur l'origine, la nature de ces arômes dans les aliments, leur mode de perception sensorielles associées. Une deuxième partie sera consacrée spécifiquement au thé et au fromage qui sont les matrices alimentaires principalement étudiées dans cette thèse. Dans la troisième partie, l'importance des COV dans l'agro-alimentaire sera présentée. Puis, pour conclure, une synthèse des enjeux de l'évaluation des COV pour la production alimentaire sera présentée.

A. Contexte général

1. Origine et nature des arômes dans les aliments

Les composés organiques volatils responsables de l'arôme des aliments présentent une très grande diversité physico-chimique. Les principales familles auxquelles ils appartiennent sont : les acides, les alcools, les amines, les esters, les cétones, les aldéhydes, les hétérocycles, les terpènes, les pyrazines, les furanes et les composés soufrés³. La diversité des molécules chimiques retrouvées dans un aliment montre l'implication de nombreuses réactions dans la formation des arômes. Un grand nombre de ces réactions sont ubiquitaires et produisent des composés aromatiques similaires même dans des aliments aux arômes différents tels que les fruits et la viande cuite par exemple.

Dans la formation des composés aromatiques on distingue les « arômes primaires » et les « arômes secondaires ». Les « arômes primaires » sont naturellement présents dans les aliments (notamment dans les fruits et légumes crus) et sont formés au cours des voies anaboliques et/ou cataboliques de la plante. Les « arômes secondaires » sont formés au cours de la transformation du produit, générés par des modifications microbiologiques, physiques, chimiques, biochimiques ou thermiques à partir de précurseurs non volatils (glucides, acides aminés, lipides...)⁴.

Les principales voies de biosynthèse impliquées dans la formation des arômes sont les voies métaboliques de dégradation des lipides, des acides aminés, des carbohydrates, des terpènes (Figure

1). Le métabolisme des acides aminés va former des composés aromatiques, aliphatiques et soufrés selon la nature de l'acide aminé.

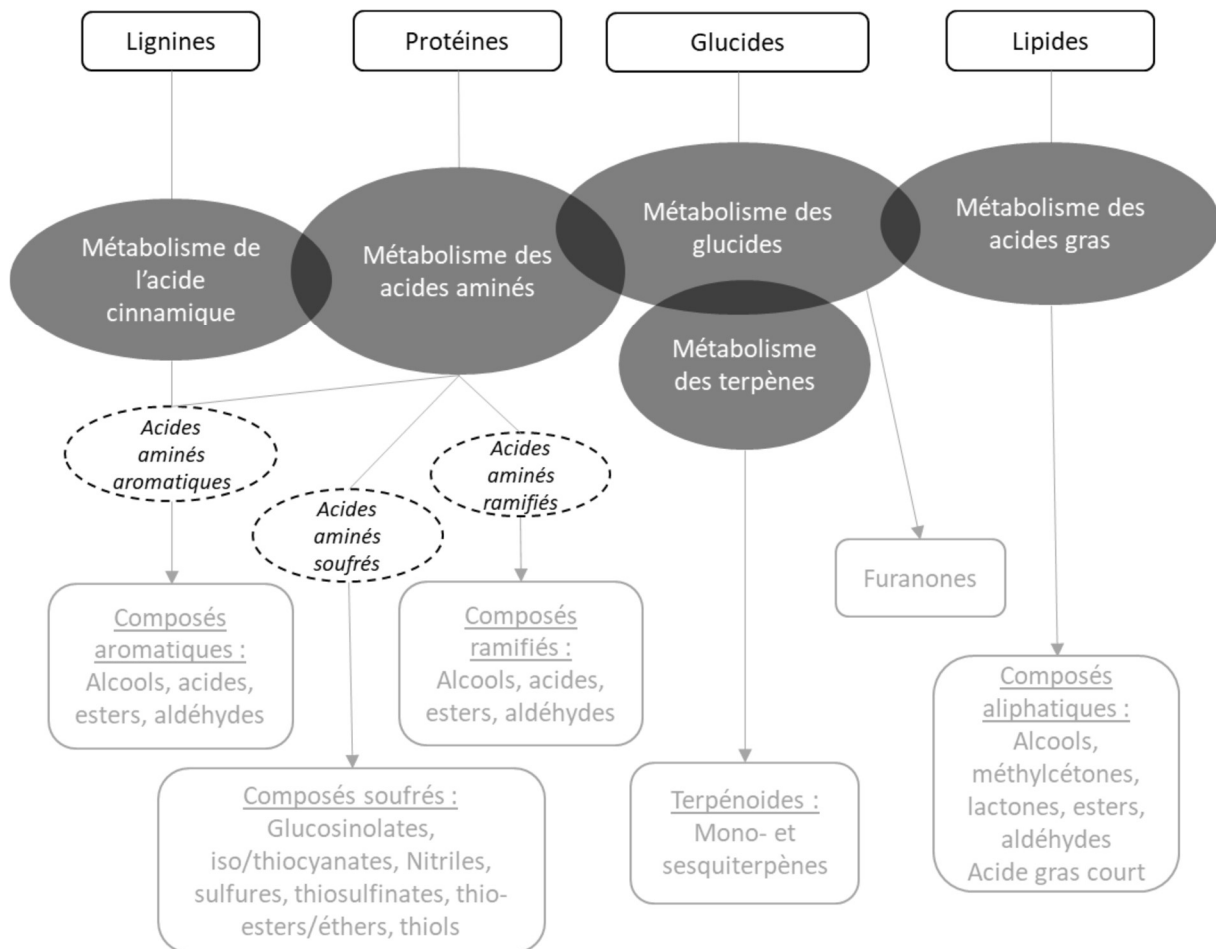


Figure 1 : Voies métaboliques mises en jeu dans la formation des composés volatils odorants d'après Cadwallar (2019) ⁵

La dégradation des lipides est associée à deux voies métaboliques principales : la bêta-oxydation (processus oxydatif) et la lipolyse médiée par les lipoxygénases (processus hydrolytique)⁶. L'hydrolyse lipide entraîne la libération d'acides gras libres à chaîne courte (volatils) ou de lactone (par exemple, dans les produits laitiers)⁵. L'oxydation des lipides sera source d'aldéhydes et de cétones volatils. Ces composés sont généralement instables et peuvent être transformés en d'autres composés volatils par d'autres réactions chimiques.

La voie métabolique des acides aminés conduit à la formation de composés aliphatiques, ramifiés, aromatiques et soufrés de la famille des alcools, aldéhydes, cétones, acides et esters. La formation de ces composés dépend de la nature de l'acide aminé. Au cours du métabolisme des acides aminés, une désamination de l'acide aminé catalysée par l'aminotransférase conduit à l' α -cétoacide correspondant. Cet α -cétoacide va subir une décarboxylation sous l'influence de décarboxylases pour former l'aldéhyde correspondant. Par la suite, l'aldéhyde formé peut subir plusieurs réactions, par exemple, la conversion en alcools correspondants en présence d'ADH (Alcool déshydrogénase) ou en acides correspondants, en présence d'aldéhyde oxydase⁴.

La voie métabolique des glucides conduit à la formation de deux groupes de composés aromatiques : les terpènes et les furanones. La biosynthèse des terpènes est associée à la voie des isoprénoïdes

tandis que les furanones sont produites à partir d'intermédiaires du métabolisme des glucides, comme le cycle des pentoses phosphates⁷.

A l'ensemble de ces voies métaboliques s'ajoute la réaction de Maillard, une réaction universelle dans le traitement des aliments ; elle est responsable du développement de l'arôme pouvant être souhaité ou indésirable dans les aliments traités thermiquement⁸. La réaction de Maillard produit un brunissement non enzymatique impliquant la réaction de composés carbonylés (les sucres réducteurs) avec des composés possédant un groupe amino libre (acides aminés, amines, protéines)⁸. Cette réaction complexe conduit à la formation de centaines de composés volatils qui peuvent être regroupés en plusieurs classes chimiques (Figure 2).

Nursten et al⁹ ont classé les composés volatils produits par la réaction de Maillard en trois groupes :

- Les produits « simples » de déshydratation/fragmentation des sucres (furanones, pyranones, cyclopentènes, composés carbonylés, acides).
- Produits "simples" de dégradation des acides aminés : aldéhydes, composés soufrés (sulfure d'hydrogène, méthanthiol), composés azotés (ammoniac, amines).
- Les produits formés par d'autres interactions : pyrroles, pyridines, pyrazines, imidazoles, oxazoles, thiazoles, thiophènes, dithiolanes, trithiolanes, dithianes, trithianes, furanethiols.

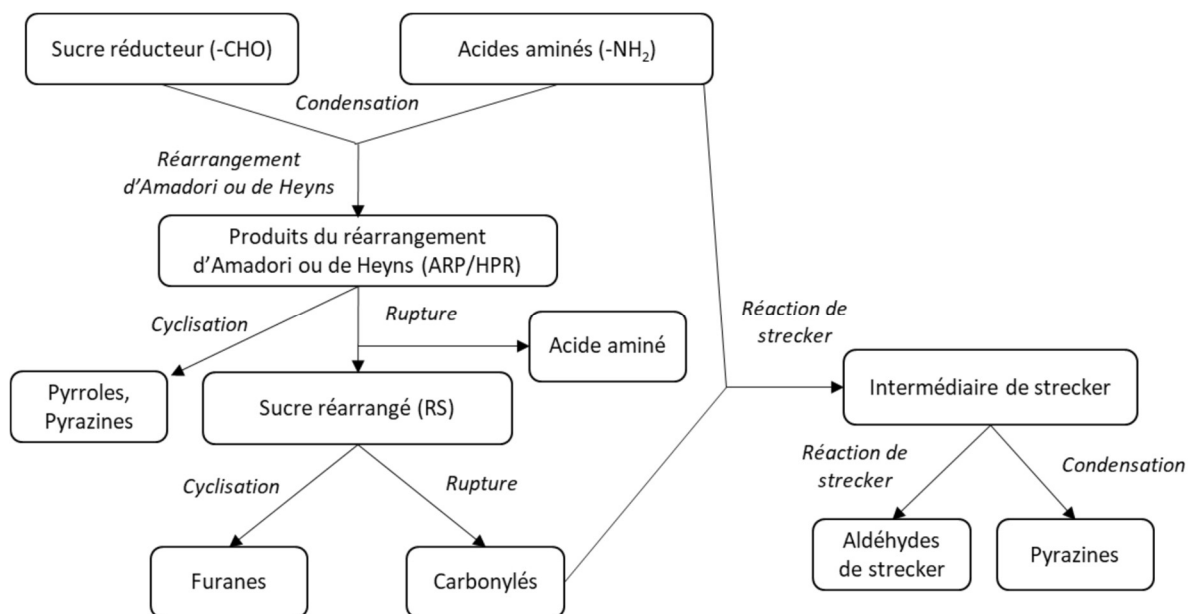


Figure 2 : Schéma résumé de la formation des COV odorants par la réaction de Maillard d'après Jousse (2002)¹⁰

Les acides représentent l'une des plus importantes classes de composés odorants³ mais du point de vue de l'odeur perçue, les alcools confèrent des caractéristiques importantes aux profils aromatiques des aliments. Les amines définissent généralement une odeur alcoolique fruitée ou vernie. Ces composés sont issus de la décarboxylation des acides aminés suite à une protéolyse. Les amines tertiaires possèdent des seuils de détection beaucoup plus faibles (exemple : triéthylamine 0,47 ppb) que les amines primaires (exemple : éthylamine 0,83 à 3,63 ppm) et secondaires (exemple : méthylamine 182 ppm)^{3,11}. Les aldéhydes ont une odeur, pour la plupart, fruitée ou florale, mais elles peuvent aussi conférer une note grasse³. Les cétones confèrent aux produits une note fruitée ou moisie et une saveur de beurre pour les plus petites³. Les notes odorantes attribuées aux esters sont généralement fruitées ou florales. Les composés soufrés possèdent des seuils de perception assez bas

(exemple : Méthanethiol 2 ppb)¹¹ et sont associés à une odeur correspondant à celle du chou, du chou-fleur et de l'ail cru et cuit³. Les terpènes et les terpénoïdes peuvent être fabriqués par les plantes ou produits par biotransformation microbienne ou par synthèse chimique. Ils représentent une catégorie importante d'ingrédients aromatiques¹². Les lactones sont des esters cycliques qui donnent principalement une note fruitée³. Les composés hétérocycliques (Furanones et Pyranones) sont associés à des notes caramélisées, sucrées, fruitées, de noisette ou brûlées³. Les pyrazines, issus de la réaction de Maillard, confèrent aux aliments des notes aromatiques torrifiées, grillées ou de noix. Dans les légumes crus, les pyrazines contribuent à leur note verte³. Les furanes sont une famille de molécules chimiques cycliques (hétérocycle aromatique). Ils se forment lors de la réaction de Maillard, de la caramélisation des sucres ou au cours du traitement thermique des aliments¹³.

2. Mauvaise odeur et « altération »

Dans les aliments, les mauvaises odeurs (« off-flavors ») ou les altérations (« taints ») peuvent causer de graves problèmes entraînant des pertes économiques dues aux rappels de produits, une diminution de la confiance des consommateurs et potentiellement une image de marque ternie¹⁴. Elles se caractérisent par l'apparition d'une odeur atypique résultant de composés formés par la détérioration interne de l'aliment (réactions chimiques et altération microbienne) ou résultant de la contamination par des produits étrangers provenant de l'environnement ou d'autres sources externes (emballage, produits chimiques synthétiques ou transformés par une action microbienne...). Les altérations et les mauvaises odeurs peuvent provenir de n'importe quelle partie de la chaîne d'approvisionnement, de la matière première au stockage et à la conservation du produit fini^{5,15}. La plupart des réactions impliquées dans la formation des mauvaises odeurs sont identiques ou analogues à celles déjà abordées dans les sections précédentes. Par exemple, la dégradation des lipides causée par l'autoxydation et/ou la lipolyse est une source courante d'arômes désirables ou de mauvaises odeurs dans les aliments¹⁶. Certaines amines formées lors de la dégradation des acides aminés au cours de la protéolyse peuvent conférer aux aliments des notes putrides ou de poisson qui peuvent être associées à une dégradation du produit. Les matières premières peuvent également entraîner la formation de mauvaises odeurs. Par exemple, l'odeur de sanglier associée à la viande provenant de porcs mâles intacts a été attribuée au stéroïde "androstérone", qui a une odeur ammoniacale et urinaire. Cette odeur est parfois accompagnée d'un métabolite du tryptophane à l'odeur fécale, le skatole¹⁷. Les altérations peuvent également résulter de la contamination de l'aliment par des produits d'origine externe. Par exemple, le goût de bouchon dans le vin est causé par les composés phénoliques halogénés présent dans le bouchon (par exemple, le trichloroanisole)^{18,19}, ainsi que la géosmine et le 2-méthylisoborneol à l'odeur de moisi ou de terre dans les produits de l'aquaculture^{20,21}.

De plus, certains composés possédant un caractère aromatique positif dans un cas donné peuvent provoquer une mauvaise odeur lorsqu'ils sont en interaction avec d'autres composés présents dans un mélange ou présents à des niveaux élevés⁵. Cela illustre la complexité de la perception sensorielle qui varie selon la composition des molécules volatiles et leurs concentrations.

3. Perception sensorielle d'arômes

La perception sensorielle est la somme des caractéristiques physiques et chimiques perçues par les organes des sens relayées par des réactions chimiques et neurologiques et provoquant chez la

personne des réactions émotionnelles d'intensités variables. Parmi les cinq sens, l'olfaction est le mécanisme moléculaire le plus complexe dû à des centaines de récepteurs²². Ce sens permet de discriminer des milliers de molécules odorantes aux structures chimiques très diverses à des concentrations très variables²³.

Les substances odorantes émises dans l'air par un aliment sont perçues par les sites récepteurs d'odeurs de l'organe olfactif. La fixation de molécules odorantes sur leurs récepteurs provoque l'émission d'un signal transmis par les neurones jusqu'au cerveau où ces signaux sont interprétés. Le système olfactif est composé de deux voies de perception des arômes²² : 1) la voie orthonasale, c'est-à-dire lorsqu'un individu sent les aliments avant de les manger et que les molécules pénètrent dans le nez par les narines pour atteindre les récepteurs ; 2) la voie rétronasale, c'est-à-dire lorsqu'un individu mastique et avale les aliments pendant le repas (Figure 3). La mastication libère des substances aromatiques qui sont transportées vers le nez par l'expiration pour atteindre les récepteurs olfactifs.

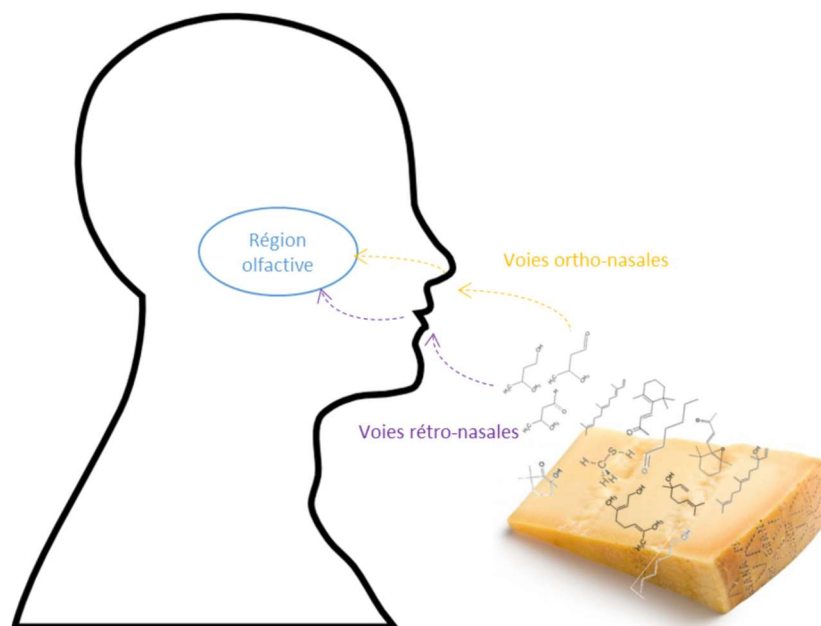


Figure 3 : Perception olfactive

La compréhension des composants des arômes alimentaires est devenue plus importante que jamais avec la transformation des systèmes alimentaires et l'innovation accrue dans l'industrie alimentaire. L'analyse sensorielle joue un rôle essentiel en fournissant des informations objectives sur la compréhension d'un produit par le consommateur, l'acceptation ou le rejet, la reconnaissance des odeurs ou saveurs et la description des émotions évoquées²⁴. Dans le secteur agroalimentaire, il existe plusieurs normes qui font appel à l'analyse sensorielle, à l'olfaction et à la reconnaissance des odeurs. Par exemple la norme ISO 16779 :2015²⁵ qui évalue la durée de conservation des produits alimentaires au moyen d'essais sensoriels. L'évaluation se porte sur les changements d'aspect, d'odeur, de saveur, de goût et de texture à des durées de conservations données. A noter que cette norme n'évalue pas les éventuels problèmes relatifs à la sécurité liée à l'utilisation ou à la consommation de l'aliment. La norme ISO 13302 :2003²⁶ qui décrit des méthodes pour évaluer les modifications de saveurs des aliments pouvant être causées par l'emballage (papier, carton, plastique, films, bois). Cette norme peut être étendue à tous les objets destinés à entrer en contact avec des produits alimentaires (ustensiles de cuisine, revêtements ou pièces d'équipement comme des joints ou tuyauterie) dans le but de contrôler la compatibilité alimentaire d'un point de vue sensoriel.

a. Qualité de l'odeur

La qualité d'une odeur est définie par sa note aromatique et son caractère hédonique. La note aromatique va permettre de décrire l'odeur perçue, la tonalité hédonique concerne la perception agréable ou désagréable. L'apprentissage olfactif, la mémoire et la sensibilité d'une personne jouent un rôle central dans la perception de la qualité odorante d'une substance aromatique²⁷. Généralement, les personnes arrivent facilement à faire la distinction entre différentes odeurs mais éprouvent des difficultés à les nommer²⁸. La qualité olfactive perçue d'une molécule est principalement liée à sa structure chimique. Cependant, une odeur peut correspondre à des centaines, voire des milliers, de composés volatils, dont certains sont présents à l'état de traces : il est donc difficile d'étudier les molécules qui produisent les arômes et de faire le lien entre la chimie et l'odeur.

b. Intensité d'odeur et seuils de perception

L'intensité d'une odeur est caractérisée par la force ou puissance avec laquelle est ressentie l'odeur. L'intensité est un paramètre mesurable qui répond à la norme NFX 43-103²⁹. Cette mesure consiste à comparer à une échelle de référence, constituée de solutions aqueuses à différentes concentrations d'un composé standard (n-butanol ou pyridine), la molécule odorante à évaluer. L'intensité d'une odeur étant une perception, une relation a été établie entre l'échelle de concentration de l'odorant et l'échelle d'intensité de perception olfactive : la loi de Stevens³⁰. La relation entre la concentration et l'intensité perçue n'est pas linéaire mais sigmoïde. Chaque odeur se caractérise par une courbe entre l'intensité olfactive et sa concentration. Cette courbe (Figure 4) présente trois parties. La première partie correspond à la phase inodore c'est-à-dire que pour des concentrations inférieures au seuil de perception, l'intensité de perception est nulle. La deuxième partie de la courbe correspond à une relation linéaire entre le logarithme de la concentration et le logarithme de l'intensité. Et enfin la dernière partie est la phase de saturation qui correspond à la valeur maximale de l'intensité d'odeur perceptible malgré une augmentation de la concentration du composé³¹.

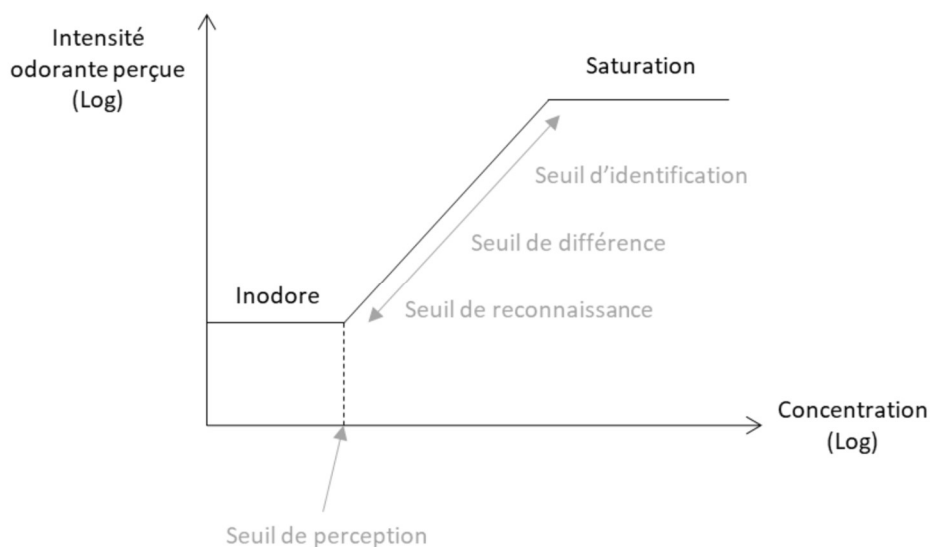


Figure 4 : Courbe d'évolution de l'intensité odorante perçue en fonction de la concentration du composé d'après Medjkoune (2018)³²

A l'intensité de l'odeur s'ajoute la notion de seuil, considérée comme l'intensité du stimulus qui définit la limite inférieure de sensibilité du système sensoriel³³. Différents types de seuils sont identifiés. Le **seuil de perception** correspond à la concentration la plus faible d'une substance chimique pouvant être détectée de manière fiable. Les odeurs sont en effet le plus souvent perçues à des concentrations très faibles, bien inférieures aux valeurs limites reconnues comme pouvant porter atteinte à la santé. Cependant à des concentrations plus faibles que celle nécessaire pour atteindre le seuil de perception, l'intensité de stimulation n'aura pas suffisamment d'effet sur le système sensoriel et ne pourra pas être perçue. Dans ces conditions, aucune qualité de goût ou d'odeur ne peut être détectée, on distingue simplement quelque chose de différents entre l'échantillon et le témoin. Le seuil de reconnaissance correspond à la concentration la plus faible pour laquelle une qualité d'odeur est discernée. Le seuil de différence correspond à la plus petite différence entre deux stimuli perçus et détectés par un sujet. Le seuil d'identification est le seuil pour lequel une odeur est perçue et identifiable, on peut la nommer et la décrire. Cela est dû à des concentrations supérieures à celle de la perception simple. Le dernier seuil est celui du phénomène d'adaptation, en effet, lorsqu'une exposition au stimulus est maintenue, une diminution de l'intensité sensorielle est perçue. Lors de la suppression de ce stimulus, la sensibilité sensorielle est récupérée.

L'« Odour Activity Value » (OAV, valeur d'activité d'odeur) permet de mesurer le potentiel aromatique de chaque composé et son importance pour l'odeur d'un aliment. L'OAV est calculée comme le rapport de concentration du composé dans un produit alimentaire et la concentration seuil de perception olfactive (OT) déterminée dans le même type de produit³⁴. Un OAV>1 indique que le composé est présent au-dessus de sa valeur seuil et est considéré comme contribuant individuellement à l'arôme de l'aliment²².

c. Facteurs influençant la perception

Influence de la matrice alimentaire

Une matrice alimentaire est composée de divers constituants (protéines, matières grasses, sucres, eau...) pouvant interagir avec les COV et induire une augmentation ou une diminution de leur transfert dans l'air en fonction de leurs paramètres physico-chimiques (polarité, hydrophobicité, nature chimique...)³⁵. Les autres constituants de l'aliment et leurs proportions vont impacter la libération de molécules odorantes et donc leur perception. La température est également un facteur influençant la perception de la saveur³⁶. Selon la température du produit, les odeurs perçues ne seront pas les mêmes car ce ne sont pas les mêmes composés qui seront libérés de la matrice et atteindront le système olfactif. Les arômes alimentaires ne sont pas composés d'odorants uniques mais sont plutôt des mélanges odorants complexes qui n'ont généralement pas la qualité ni l'odeur ni le goût des composés lorsqu'ils sont étudiés seuls³⁷. Des études ont montré qu'un sujet humain est capable de percevoir jusqu'à trois composés dans des mélanges de goûts³⁸ ou quatre dans le cas des odeurs³⁹. Différents phénomènes sont observés sur la perception sensorielle d'un mélange lorsque les composés aromatiques ont des concentrations très différentes : phénomènes suppressif (l'intensité perçue du mélange est inférieure à celle du composant le plus fort), synergique (l'intensité du mélange est supérieure à la réponse sensorielle d'un seul composant) et hypoaditif (l'intensité d'un composant du mélange est égale à celle du mélange).

Les facteurs humains influençant la perception sensorielle

La perception sensorielle d'un aliment ou d'un produit peut être différente d'un individu à l'autre. Une grande variété de facteurs individuels, dont l'âge, le sexe, l'origine, la culture, les conditions médicales, la consommation de substances tel que les drogues, le tabac ou l'alcool, et les circonstances spécifiques de l'exposition, entraînent une grande variabilité du seuil de perception et d'identification des odeurs par un individu donné⁴⁰. Des études se sont intéressées à l'influence de l'âge^{41,42}, du sexe⁴³⁴⁴ ⁴⁵, de l'expérience et de la culture⁴⁶ sur nos capacités olfactives. De plus, la perception sensorielle est une combinaison complexe de facteurs neurologiques et physiologiques. La génétique est également un facteur déterminant la sensibilité d'un sujet aux stimuli gustatifs et olfactifs. Chez certaines personnes, on observe des troubles du système olfactif ou gustatif⁴⁷ tel que l'anosmie⁴⁸ (perte partielle ou totale de l'odorat), l'agueusie (perte totale du goût), l'hypogueusie (perte partielle du goût) et la dysgueusie (perception anormale, altération du goût). Ces troubles peuvent être liés à la génétique mais également aux processus de vieillissement ou survenir suite à des accidents.

B. Cas particulier du thé et du fromage

Dans le cadre de cette thèse, nous avons étudié principalement deux matrices agroalimentaires qui sont le thé et le fromage. Pour ces deux matrices très différentes, la perception sensorielle est cependant primordiale pour le consommateur.

1. Le thé

Le thé est une boisson aromatique ancestrale obtenue par infusion des feuilles du théier qui ont été séchées et soigneusement transformées. Cette boisson est l'une des plus célèbres au monde, historiquement originaire de Chine, et principalement produite en Asie. Le thé est un véritable produit de la terre avec une authenticité marquée par son écosystème. Les variations d'arôme et de qualité de cette boisson sont largement déterminées par l'interaction d'événements métaboliques complexes qui se produisent dans les feuilles avant la récolte (culture) et après la récolte (traitement)⁴⁹.

Les arômes du thé et leurs biosynthèses sont variés et complexes. La compréhension de la formation de l'arôme du thé est d'un grand intérêt. Plusieurs voies de réaction ont lieu lors de la formation de l'arôme du thé correspondant à trois phases de la production du thé : la culture, la cueillette et la transformation. Au cours de ces 3 phases, on distingue donc les arômes primaires qui sont les caractéristiques de la plante (type de cultivar, conditions climatiques, terroir, cueillette) et les arômes secondaires que l'on peut également nommer « arômes de manufacture ». Ils sont issus de la transformation du thé et vont faire ressortir les notes aromatiques. Ces arômes secondaires sont l'acquis du thé et représentent la manière dont il a été travaillé.

Différentes variétés de thé existent et dépendent des éléments suivants : les jardins d'origine, le type de cueillette pratiquée et la période de récolte, les différents traitements que vont subir les feuilles durant le processus de fabrication et le savoir-faire⁵⁰. Dans cette partie nous allons nous intéresser à l'impact des conditions de culture et de transformation sur la formation des composés volatils aromatique dans le thé.

a. Impact des conditions de culture

La culture du thé débute par la sélection du théier. Le théier est l'espèce "*Camellia Sinensis*" et appartient à la famille des Théacées. Les trois variétés botaniques principales sont : *Camellia Sinensis* var. *sinensis*, *Camellia Sinensis* var. *assamica* et *Camellia Sinensis* var. *cambodiensis*. A cela s'ajoute des milliers de cultivars de par le monde. Le terme « cultivar » est un mot composé de « cultiver » et de « variété » soit un synonyme de « variété cultivée ». Un cultivar désigne une variété de plante ou d'arbre obtenue en culture, généralement par sélection, pour ses caractéristiques réputées uniques. Des instituts de recherche de différents pays étudient l'évolution des théiers et développent de nouveaux cultivars qui apporteront leur contribution dans le développement aromatique. L'objectif de ces nouveaux cultivars est de créer une espèce de théier plus résistante, moins sensible aux insectes ravageurs, et adaptée aux conditions climatiques locales et présentant un bon équilibre entre rendement et qualité.

La phase de culture et de développement du théier au champ implique des réactions de biosynthèse au sein des cellules vivantes de la feuille, et notamment l'activation de certains gènes liés à l'arôme du thé et des réactions entre le substrat et les enzymes. Le stress biotique (attaque d'insectes) ou abiotique (conditions climatiques et action anthropique) subi par la plante va augmenter les niveaux d'expression des gènes de synthèse d'arômes et entraîner une accumulation de composés aromatiques caractéristiques dans les feuilles de thé⁵¹.

Le terroir est un élément important lors de la culture du théier. Il désigne une aire géographique considérée comme homogène à travers les ressources et les pratiques agricoles. Le terroir comprend donc un ensemble de caractéristiques qui influencent l'écosystème du théier : le sol, le sous-sol, le climat, la topographie, l'exposition, l'altitude et l'action de l'homme. Chaque région a ses propres caractéristiques qui auront une incidence sur la qualité du thé, sa typicité et ses propriétés organoleptiques.

La date de la récolte et la modalité de la cueillette sont primordiales et ont une influence importante sur le thé. Ainsi la récolte varie selon la région de production, de l'altitude du jardin, le type de cultivar ou encore le stade de développement des bourgeons et des feuilles qui sont prélevés. Une fois la feuille cueillie, les réactions anaboliques ralentissent, tandis que les réactions cataboliques se poursuivent.

Guo et al⁵² ont étudié la composition en composés volatils et les profils olfactifs des feuilles de thé fraîches et des thés Oolong après traitement issus de trois cultivars différents. Les profils aromatiques ont été analysés par chromatographie gazeuse couplée à la spectrométrie de masse (GC-MS), chromatographie gazeuse couplée à la spectrométrie de mobilité ionique (GC-IMS) et évalués sensoriellement. L'analyse en GC-MS a permis l'identification de 177 composés organiques volatils : des alcools, des cétones, des aldéhydes, des alcènes, des esters, des composés hétérocycliques et des hydrocarbures aliphatiques et aromatiques. Les résultats ont indiqué une augmentation de l'abondance d'alcools dans les échantillons de feuilles fraîches et de composés hétérocycliques dans les échantillons de thés transformés. Les auteurs ont montré que, lors de la transformation, la composition en composés hétérocycliques variait d'un cultivar à l'autre. Des différences de profils volatils ont été montrées entre les cultivars. Une bonne discrimination des échantillons de thés a été observée selon le type de cultivar utilisé mais également entre les feuilles fraîches et les feuilles séchées et traitées pour chaque cultivar. Ces travaux montrent que le cultivar a un impact sur la composition en molécules volatiles de la feuille fraîche qui va, par conséquent, jouer un rôle crucial dans le développement de l'arôme dans les feuilles de thé séchées et transformées.

b. Impact des étapes de transformation

Une fois les feuilles récoltées, celles-ci vont subir diverses étapes de transformation afin d'obtenir le produit fini. Cette transformation va conditionner la couleur du thé et ses qualités sensorielles.

Les étapes de transformation diffèrent d'un thé à l'autre mais se regroupent en 7 étapes principales : le flétrissage, le roulage, l'oxydation enzymatique, la fixation, le séchage, la post fermentation et l'affinage. Chaque étape va engager des voies métaboliques et des activités enzymatiques. Ces réactions sont réalisées en contrôlant la température, la teneur en humidité, le degré de perturbation du tissu foliaire et le temps de réaction. Ces étapes de transformation vont donner la couleur au thé selon son degré d'oxydation ce qui les distinguera les uns des autres (Figure 5).

Le flétrissage est un processus de déshydratation ayant pour objectif de rendre la feuille de thé plus souple pour faciliter le roulage. Cette étape permet d'amorcer le processus d'oxydation puisque les feuilles sont mises au contact de l'oxygène. Le roulage ou brassage consiste à casser les cellules de la feuille de thé et d'activer les réactions d'oxydation. Cela permet également de faire ressortir les sucres et huiles essentielles. L'oxydation des feuilles est l'étape clé qui va révéler l'astringence (ou amertume) du thé, sa saveur et sa couleur. Les feuilles de thé sont mises au contact de l'oxygène de l'air dans une atmosphère chaude et humide pour favoriser l'« oxydation enzymatique ». La fixation (ou dessiccation) a pour but de stopper le processus d'oxydation. Pour cela, les feuilles (encore humides) sont chauffées à une température élevée, de manière « brutale », pour les sécher. Ce séchage intense va détruire les enzymes responsables de l'oxydation. L'étape de séchage va déshydrater les feuilles de thé en vue de sa conservation tout en laissant un taux d'humidité résiduel de 2 à 10%. Un léger pourcentage d'humidité est conservé pour préserver la note aromatique car de nombreux composés aromatiques sont hydrophiles. La post fermentation est réalisée uniquement pour certains thé ; il ne s'agit plus d'une réaction d'oxydation enzymatique mais d'une fermentation microbienne. La fermentation implique l'intervention de micro-organismes (levures, bactéries ou champignons) dans des conditions anaérobies c'est-à-dire dans un environnement pauvre en oxygène. Les feuilles de thé sont placées sous une bâche mouillée pour les priver d'oxygène et accélérer la prolifération de micro-organismes qui vont transformer ses nutriments pour lui donner ce goût si particulier. Ces actions fermentaires feront évoluer la couleur et la note aromatique du thé, c'est un processus contrôlé et maîtrisé par l'homme. L'affinage se réalise sur certaines couleurs de thé pour le faire mûrir, évoluer ou se bonifier. Une fois le procédé de transformation fini et avant d'être emballées, les feuilles de thé traitées sont tamisées pour éliminer les poussières et les tiges puis triées en fonction de leur taille.

L'ensemble des étapes de transformation va participer à la formation des composés volatils du thé. Ces composés sont formés par différentes réactions issues de mécanismes chimiques ou biologiques comme la dégradation des caroténoïdes, la voie des dérivés des lipides, la voie d'hydrolyse des glycosides, la réaction de Maillard...⁵³ Deux voies de dégradation des caroténoïdes existent : la dégradation oxydative enzymatique médiée par les dioxygénases et une dégradation oxydative non enzymatique (photo-oxydation, l'auto-oxydation et la dégradation thermique)⁵⁴. La dégradation oxydative enzymatique du β -carotène va générer la production de β -ionone (Figure 6). La dégradation oxydative non enzymatique par photo-oxydation du phytofluène (Figure 7) entraîne la formation du nérolidol (arôme fleuri) qui produira de l' α -farnésène (arôme fruité)⁵³. La présence et la concentration en composés issus de la dégradation des caroténoïdes sont étroitement liées aux étapes d'oxydation qui ont lieu lors du processus de transformation mais également de la composition en caroténoïdes des feuilles fraîches.

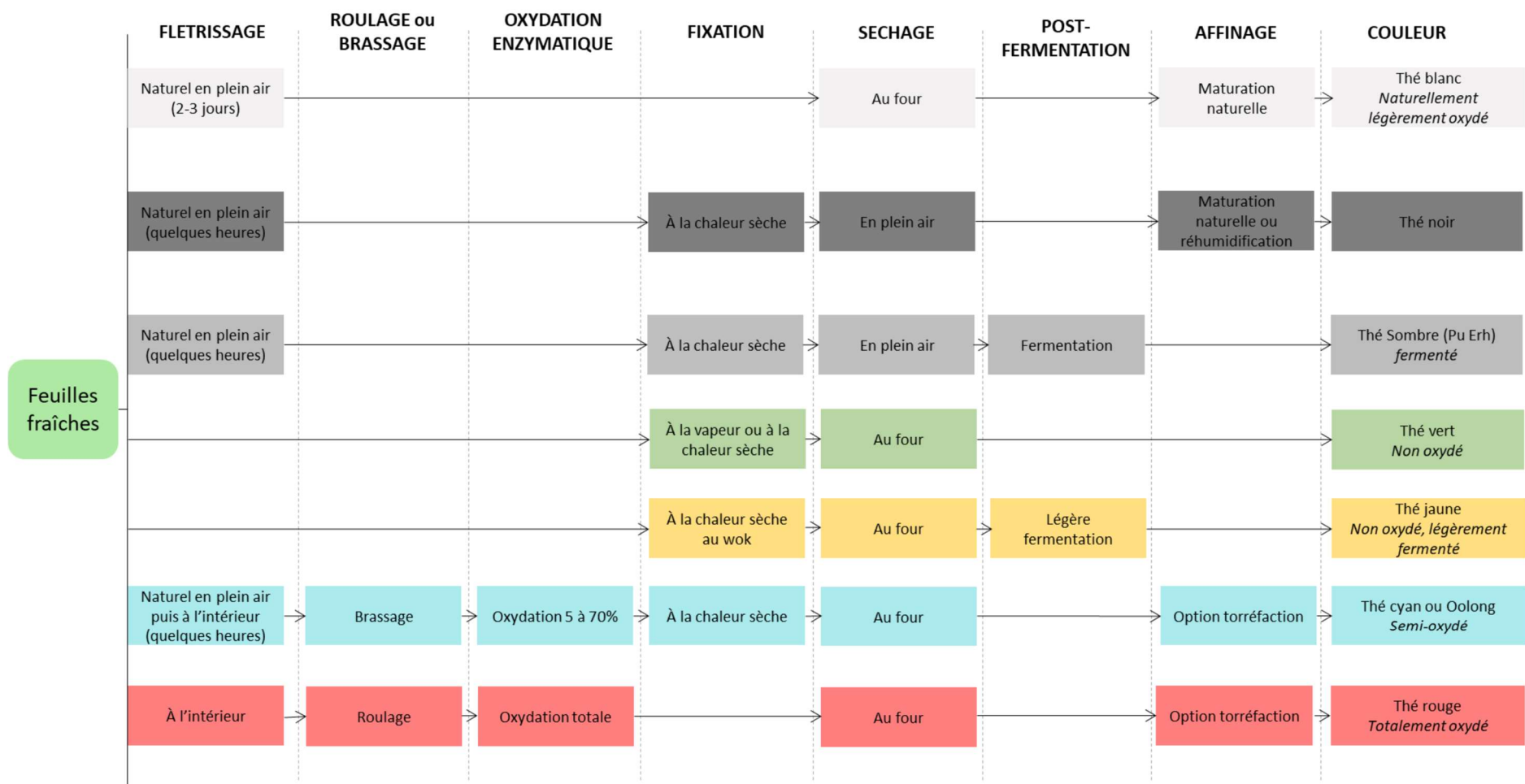


Figure 5 : Les différentes étapes de fabrication selon la couleur de thé d'après Lydia Gautier⁵⁵

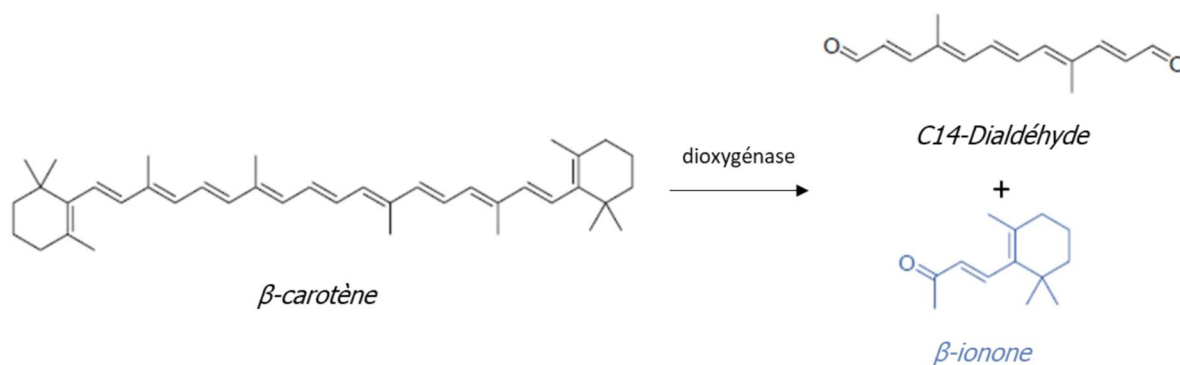


Figure 6 : Formation de la β -ionone par dégradation enzymatique du β -carotène

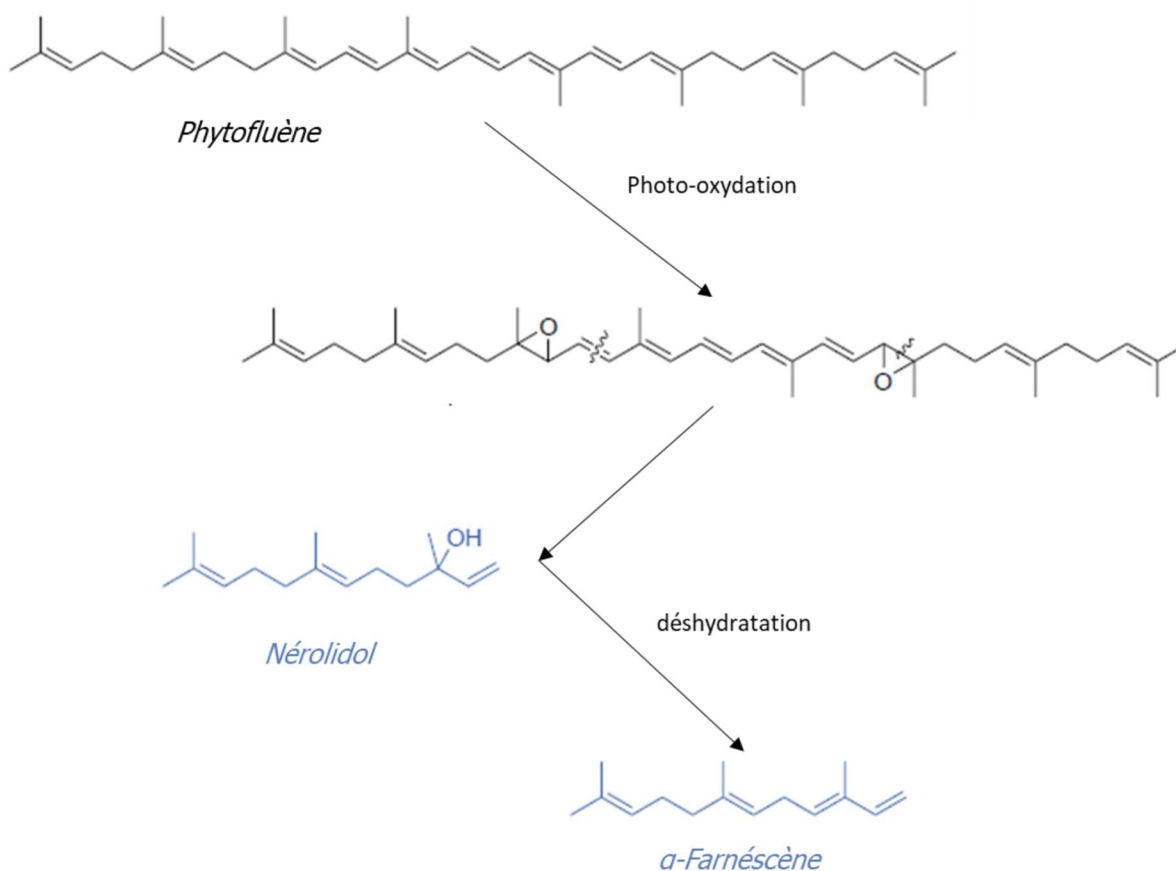


Figure 7 : Formation du nérolidol et de l' α -farnésène par photo-oxydation du phytofluène

La formation des arômes volatils à partir de l'oxydation des lipides du thé est généralement associée à deux voies principales : la réaction d'oxydation initiée par des radicaux libres et l'oxydation lipidique médiée par les lipoxygénases. Par exemple, l'oxydation de l'acide linoléique va former, par l'action de lipoxygénase (LOX) et d'hydroperoxyde lyase (HPL), un aldéhyde, le n-hexanal, qui pourra être réduit en alcool par l'alcool déshydrogénase (ADH) et former le n-hexanol (Figure 8). Les glycosides sont des composés non odorants présents dans les feuilles de thé. L'hydrolyse de ces composés va libérer des composés volatils tels que les alcools monoterpéniques comme le linalool et le géraniol à partir du pyrophosphate de géranyle (Figure 9) ou des alcools aromatiques (alcool benzylique et phényléthanol)⁵⁶. Cette hydrolyse est réalisée par l'action d'enzymes, les glycosidases, libérées des cellules des feuilles de thé lors du procédé de transformation. La concentration en glycosidases dans les feuilles de thé varie selon la saison de récolte⁵⁷.

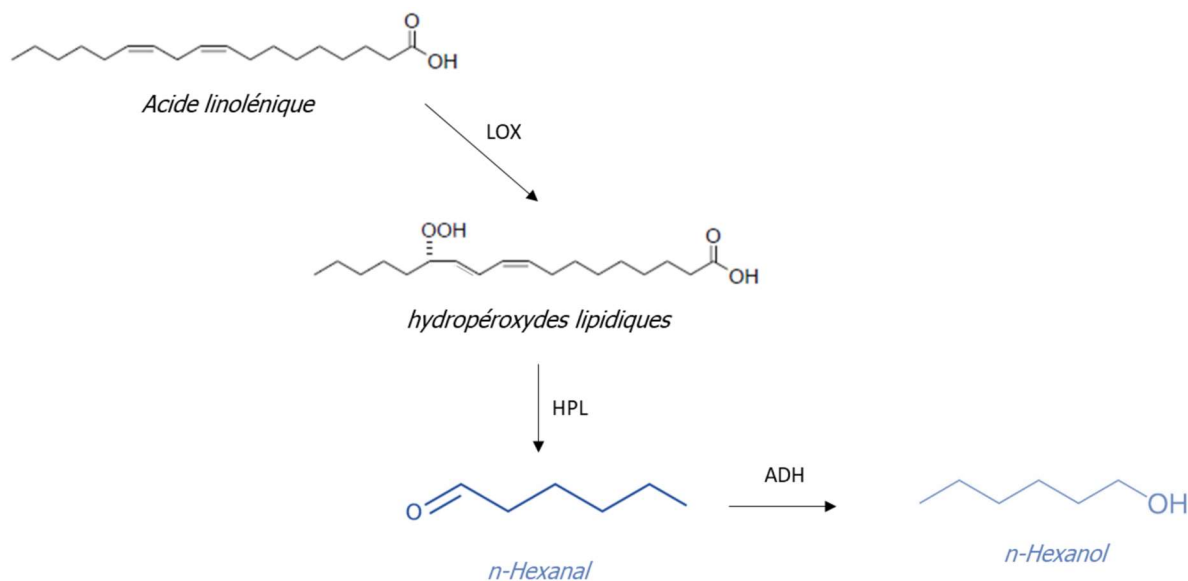


Figure 8 : Formation du n-hexanal et n-hexanol par dégradation de l'acide linoléinique

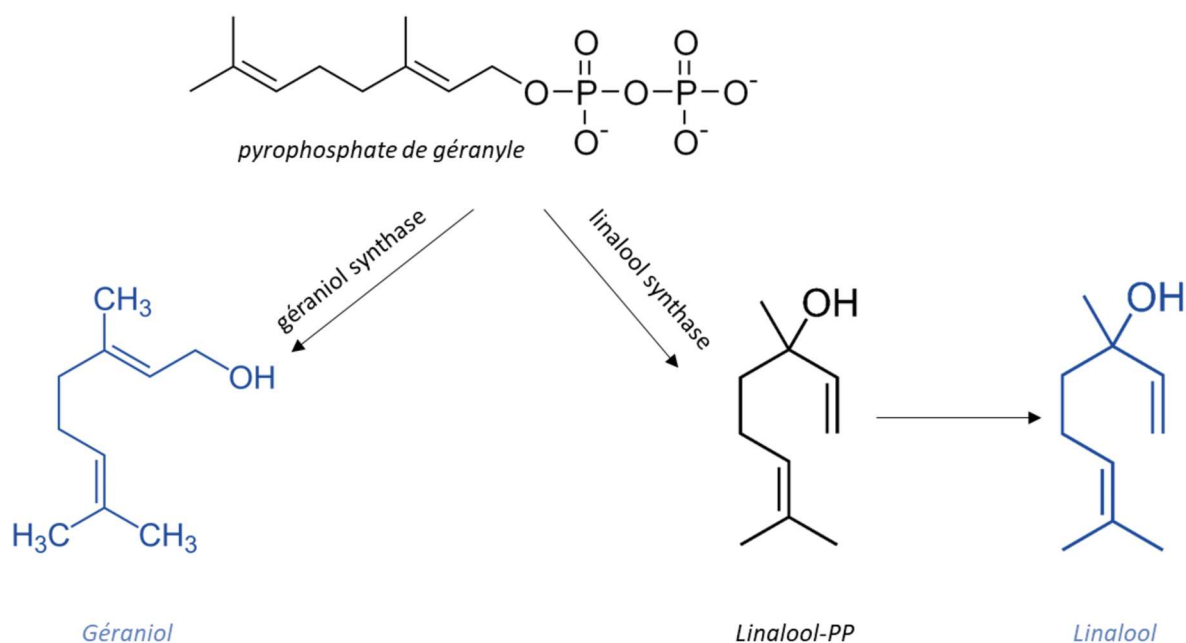


Figure 9 : Formation du Géraniole et du Linalool à partir du pyrophosphate de géranyle

Pendant le processus de transformation du thé, des réactions de Maillard (Figure 2) ont lieu et vont produire des composés hétérocycliques tels que le furane, le pyrrole et leurs dérivés. Un autre aspect de la réaction de Maillard ayant lieu pendant la transformation des feuilles de thé est la réaction de dégradation de Strecker associée à la dégradation des acides aminés⁵⁸ (Figure 2). La plupart des acides aminés ont leur aldéhyde de Strecker correspondant : par exemple, l'aldéhyde correspondant de la glycine est le formaldéhyde, celui de l'alanine est l'acétaldéhyde, et celui de la phénylalanine est le phénylacétaldéhyde⁵³. La méthionine est un acide aminé qui joue un rôle dans la formation de composés soufrés. Par la dégradation de Strecker, la méthionine va former son aldéhyde correspondant le méthional qui est le principal précurseur du méthanthiol. Ce composé est le précurseur direct de nombreux composés soufrés avec des seuils de perception relativement faibles comme le trisulfure de diméthyle ou le disulfure de diméthyle (7,6 ppb)⁵³ (Figure 10).

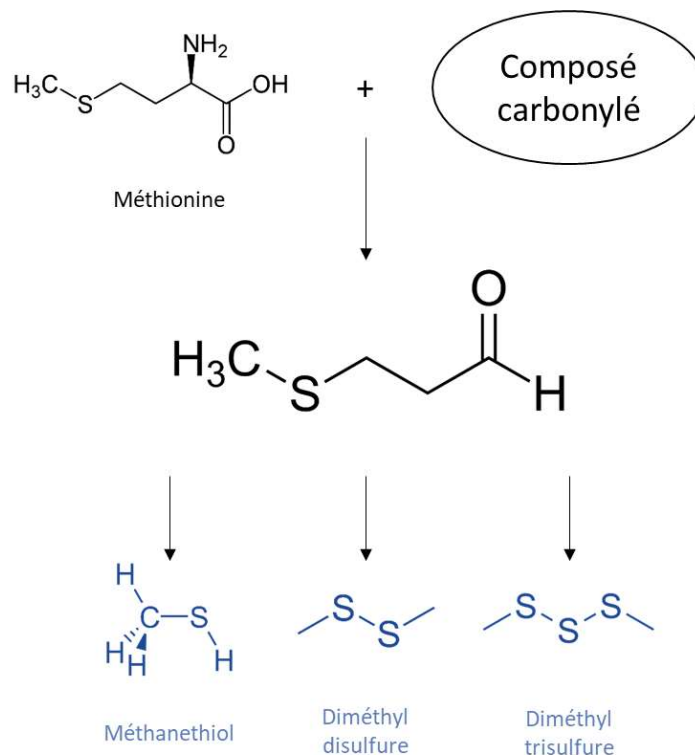


Figure 10 : Formation de composés soufrés par la dégradation de strecker de la Méthionine d'après Ho et al (2015)⁵³

Les différents processus de transformation impactent la formation, la nature, et la concentration des composés aromatiques. Le degré d'oxydation et de fermentation détermine la constitution et la concentration des principaux arômes. Feng et al⁵⁹ ont réalisé des travaux étudiant l'impact du procédé de transformation sur l'arôme du thé. Un thé de haute qualité est généralement fabriqué à partir d'un même cultivar, d'un même terroir et nécessite des étapes de traitement spécifiques. Pour étudier l'influence du procédé de transformation, ils ont fixé les facteurs de culture et ont utilisé les mêmes feuilles de thés fraîches pour produire les différents thés. Les profils de composés volatils contribuant à l'arôme ont été obtenus par une analyse en GC-MS. Les résultats observés ont montré un taux de formation rapide conduisant à une concentration élevée d'arômes dans les processus du thé noir et du thé blanc, dans lesquels les oxydations médiées par les enzymes sont prédominantes. Ils ont également montré que les réactions qui jouaient un rôle important dans la génération de l'arôme étaient les réactions de Maillard pour le thé vert et le thé jaune, l'oxydation des acides gras pour le thé noir et le thé oolong, et l'hydrolyse des glycosides ainsi que la dégradation des acides gras pour le thé blanc et le thé noir. Cette étude permet de confirmer l'hypothèse selon laquelle les procédés de transformation du thé favorisent certaines réactions de manière sélective et génèrent ainsi les caractères aromatiques typiques pour chaque catégorie de thé.

Le Tableau 1 présente une liste non exhaustive des principaux composés aromatiques volatils identifiés dans différents types de thé, leurs notes aromatiques et leur seuil de perception.

Composés	Note aromatiques	Seuil de perception ($\mu\text{g}\cdot\text{L}^{-1}$ dans l'eau)	Ref
Alcools			
(Z)-3-hexenol	Vert		60
1-penten-3-ol	Beurre, vert		61
Benzyl alcohol	goût de brûlé, faiblement aromatique		61
2-phenyl ethanol	miel	1000	60
1-hexanol	Vert		62
(Z)-3-hexen-1-ol	Vert et herbeux		60,62
1-octen-3-ol	Champignon		62
Aldéhydes			
Hexanal	Herbeux, vert		60
Pentanal	Piquant, malté, amande		62
(E,Z)-2,6-nonadienal	Arôme de concombre	0,03	60
(Z)-4-heptanal	Foin	0,06	60
(E)-2-hexenal	vert	190	61
(E,E)-2,4-decadienal	Gras	0,16	60
3-methylbutanal	malté	1,2	60
Phenyl-acetaldehyde	Miel	6,3	60
Methional	Pomme de terre	0,2	63
2-methylbutanal	malté		62
nonanal	gras, huileux		62
Benzaldehyde	parfumée, douce, amande		62
benzeneactetaldehyde	amande		62
Cétones			
β -Damascenone	fruité, semblable à la pomme	0,004	60
(Z)-1,5-octadien-3-one	Arôme du géranium		60
1-octen-3-one	Arôme de champignon		60
Cis-Jasmone	Floral, arôme de jasmin		61
2,3-butanedione	Beurre	10	60,64
1-penten-3-one	Piquant		61
2,3-Pentanedione	Beurre		64
(E,E)-3,5-Octadien-2-one	Fruité, foin, oxydé, vert		61,65
6-Methyl-5-hepten-2-one	Herbacée, piquante, huileuse, sucrée et fruitée		61,62
Esters			
Methyl salicylate	mentholé		61,62
Acides			
Hexanoic acid	Sueur, vert	890	60
Acetic acid	Acide		65
Butanoic acid	Sueur		

3-methylbutanoic acid	Sueur		60,65
2-methylbutanoic acid	Sueur, fromage		65
Pentanoic acid	Sueur		60
Terpènes			
β -ionone	Boisé, violette	0,2	60,61
α -ionone	Boisé, foin		61
Limonène	Citron		62
(E)-Linalool oxide I, furanoid	Doux, floral, vert, fruité, agrume		62,66
(Z)-Linalool oxide II, furanoid	Doux, floral, agrume, rafraîchissant		62,66
Linalool	Citron, floral		60,62,65,66
Safranal	Végétale		6
Nerol	Rosé, floral		6
(E)-Geraniol	rose	3,2	60,61
β -myrene	résineux		62
β -cyclocitral	vert doux, mentholé, fruité		62
α -terpineol	Huileux, anis, menthe		62
Furane / Furanone			
4-hydroxy-2,5-dimethyl-3(2H)-furanone	Caramel	30	60
2-ethyl-furan	Brûlé, sucré, café		61
2-pentyl-furan	Haricot, vert, beurre		62
Furfural	amande		61
Furfuryl alcohol	sucre brûlé		61
Dihydrofuran-2(3H)-one (γ -butyrolactone)	Caramel		65
Composés soufrés			
Dimethyl disulfide	Ail	7,6	63
Trimethylsulfide	Pourri	0,01	63

Tableau 1 : Quelques exemples de composés aromatiques identifiés dans le thé

2. Le fromage

Le fromage est un aliment produit à partir du lait caillé, égoutté puis plus ou moins affiné. A partir de lait de vache, de brebis, de chèvre ou d'autres mammifères (ex : bufflonne), la fabrication du fromage est apparue peu après la domestication des animaux, il y a environ 8 000 ans⁶⁷. Il existe plusieurs grandes familles de fromage qui dépendent du lait utilisé et du procédé de transformation (Figure 11). Ainsi plusieurs centaines de fromages différents sont produits dans le monde avec des différences de goût et de texture. Lors de la fabrication du fromage, de nombreuses transformations biochimiques ont lieu qui vont générer des modifications externes (croûte), internes (cœur) et organoleptiques. Ces transformations vont donner naissance aux arômes du fromage. La qualité et les caractéristiques sensorielles des fromages résultent d'une combinaison entre la composition de la matière première, les communautés microbiennes, les procédés de fabrication et les caractéristiques du terroir et des savoir-faire locaux.

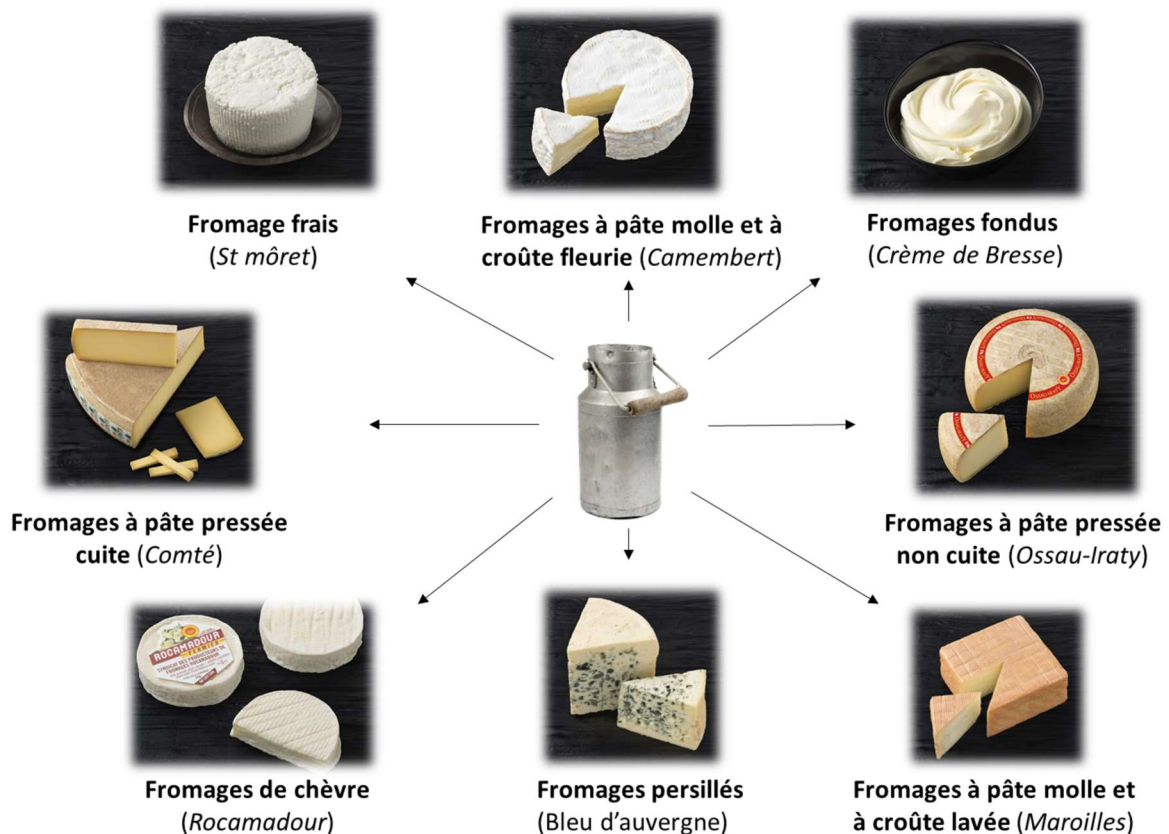


Figure 11 : Les grandes familles de fromages avec des exemples

a. Impact du lait et de l'alimentation du bétail

La qualité du fromage commence par le travail à la ferme auprès des animaux et notamment au travers de leur alimentation. En effet, cela influence la composition du lait qui varie également selon les saisons et les stades de lactation.

L'alimentation des animaux n'a pas d'impact direct sur la composition aromatique du fromage mais un effet indirect puisqu'elle influence la composition du lait, en effet, les laits d'estives se traduisent par plus de sesquiterpènes et terpènes provenant des pâtures. Selon l'alimentation, les caractéristiques trouvées dans le lait peuvent être transférées dans le fromage en rendant disponibles différents substrats pour les fonctions métaboliques des bactéries. Des études, réalisées majoritairement sur les vaches en pâture, ont montré l'influence de la composition botanique des pâturages ou encore de la complémentation alimentaire (selon la quantité, la qualité et la fréquence) sur la composition en acide gras et en terpènes⁶⁸⁻⁷⁰.

Outre l'alimentation, Park et al⁷¹ ont listé d'autres paramètres concernant les animaux qui ont un impact sur le lait comme leurs caractéristiques génétiques, leur état de santé et leurs caractéristiques physiologiques, ou encore le stade de lactation.

Les modifications des caractéristiques sensorielles des fromages dues à des facteurs impactant la composition du lait peuvent dépendre de différents mécanismes⁷² :

- Le transfert, sans modification, de molécules présentes dans le lait au fromage,
- Des modifications concernant la teneur, la structure et la composition de certains composés (protéines et matières grasses)
- Le transfert dans le fromage d'enzymes endogènes qui modifient la protéolyse et/ou la lipolyse pendant l'affinage
- Des modifications de l'écosystème microbien (composition, dynamique et activité).

b. Impact du procédé de transformation

La transformation du lait en fromage implique six grandes étapes : la préparation du lait, la coagulation, l'égouttage, le moulage, le salage et l'affinage (Figure 12)⁷³. La préparation du lait consiste à sélectionner le lait qui sera utilisé pour la fabrication selon le fromage souhaité. Le choix s'effectue selon trois critères : le type de lait (vache, chèvre, brebis...), la nature du lait (entier, écrémé, demi-écrémé) et le traitement thermique (cru, thermisé ou pasteurisé). Le lait cru ne subira pas de traitement thermique avant l'étape de coagulation, à l'inverse du lait thermisé ou du lait pasteurisé. Le premier subira une thermisation, c'est-à-dire un chauffage à 60-65°C pendant 15 à 20 secondes qui détruira un maximum de bactéries pathogènes mais pas toute la flore microbienne, et la pasteurisation consiste à chauffer le lait à 72°C pendant 15 à 20 secondes pour détruire les germes. La coagulation ou caillage consiste à extraire la matière fromageable : les protéines (caséine), les glucides (lactose), la matière grasse (lipides) et les minéraux (calcium, potassium...) en éliminant l'eau. La coagulation est réalisée par l'action d'enzyme coagulante comme la présure (animale ou végétale) ou par acidification du lait par les bactéries lactiques. La durée de cette étape est variable en fonction du type de fromage attendu ; à la fin de cette étape on obtient le caillé. L'égouttage a pour objectif de déshumidifier plus ou moins le caillé en fonction du type de fabrication et de la durée d'affinage souhaités. Plus un fromage sera humide, plus sa vie microbienne sera riche et active, moins il pourra se conserver. Le moulage consiste à placer le caillé dans des moules pour lui donner la forme souhaitée. Cette étape permet également de poursuivre l'égouttage soit de façon spontanée soit en pressant le fromage. Le

salage, dernière étape avant l'affinage, consiste à saler le fromage soit au sel sec, soit dans un bain de saumure. Le sel va jouer plusieurs rôles importants dont la formation de la croûte, exhausteur de goût et agent de conservation. L'affinage correspond à la digestion enzymatique des constituants du caillé. C'est l'étape ultime reposant sur l'action des micro-organismes du fromage et des transformations biochimiques pour développer ses arômes, sa saveur, sa texture et sa couleur. Cette étape se déroule dans des caves d'affinage en respectant différents paramètres tel que la température, l'hygrométrie, la durée, les soins en cave (lavage, salage, retournement) ou encore la ventilation qui sont dépendants du type de fromage attendu.

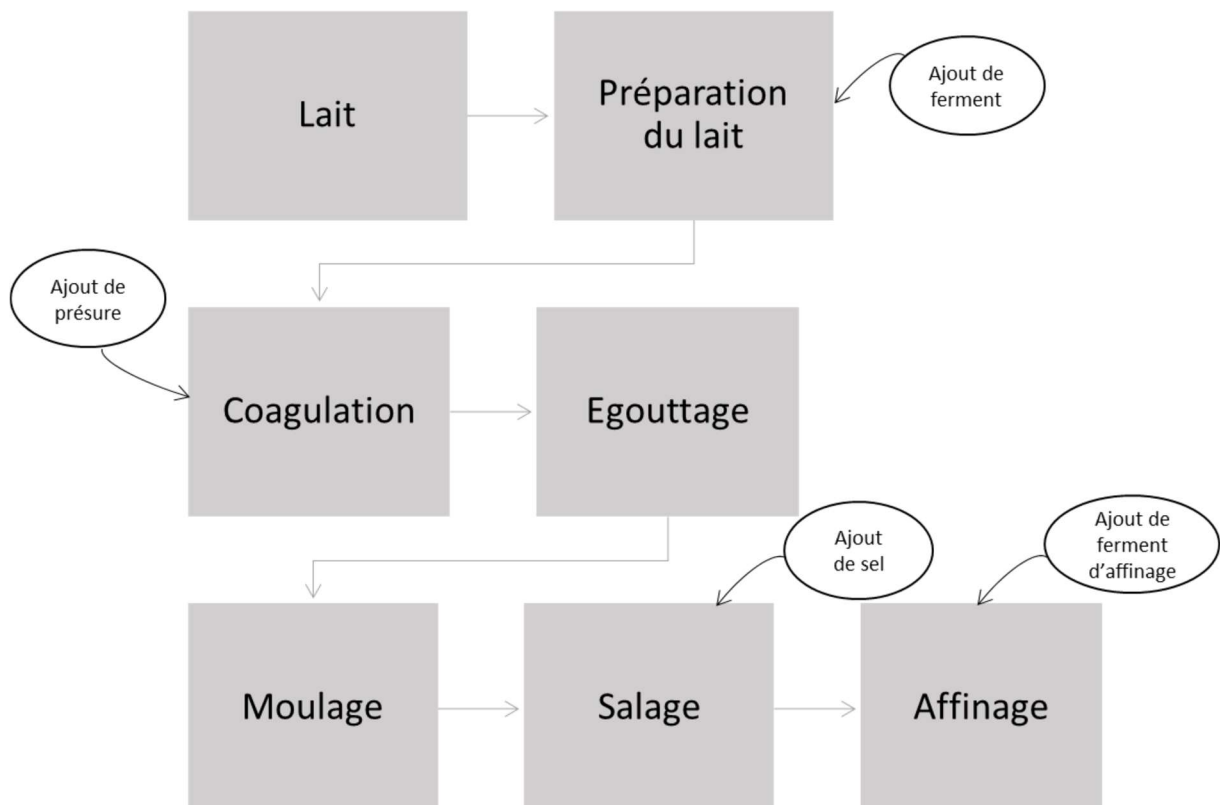


Figure 12 : les 6 grandes étapes de production du fromage

Les composés aromatiques sont majoritairement formés par l'activité des microorganismes à partir des différents composants du lait. Cette formation a lieu au cours de l'affinage qui est l'étape clé pour l'expression des arômes. Les composés aromatiques sont produits à partir de trois grands constituants du lait : le lactose, les lipides et les protéines (principalement les caséines). Le procédé de fabrication du fromage entraîne trois grandes transformations biochimiques⁷⁴ (Figure 13) :

- La glycolyse : hydrolyse du Lactose en lactate via les activités enzymatiques des ferments lactiques
- La protéolyse : hydrolyse des caséines en peptides (protéolyse primaire) sous l'action de la présure (chymosine) et à moindre mesure des enzymes naturellement présentes dans le lait (plasmines) qui sont à leur tour hydrolysées en acides aminés sous l'action des peptidases ou protéinases (protéolyse secondaire)
- La lipolyse : hydrolyse des lipides en acide gras par la lipoprotéine lipase.

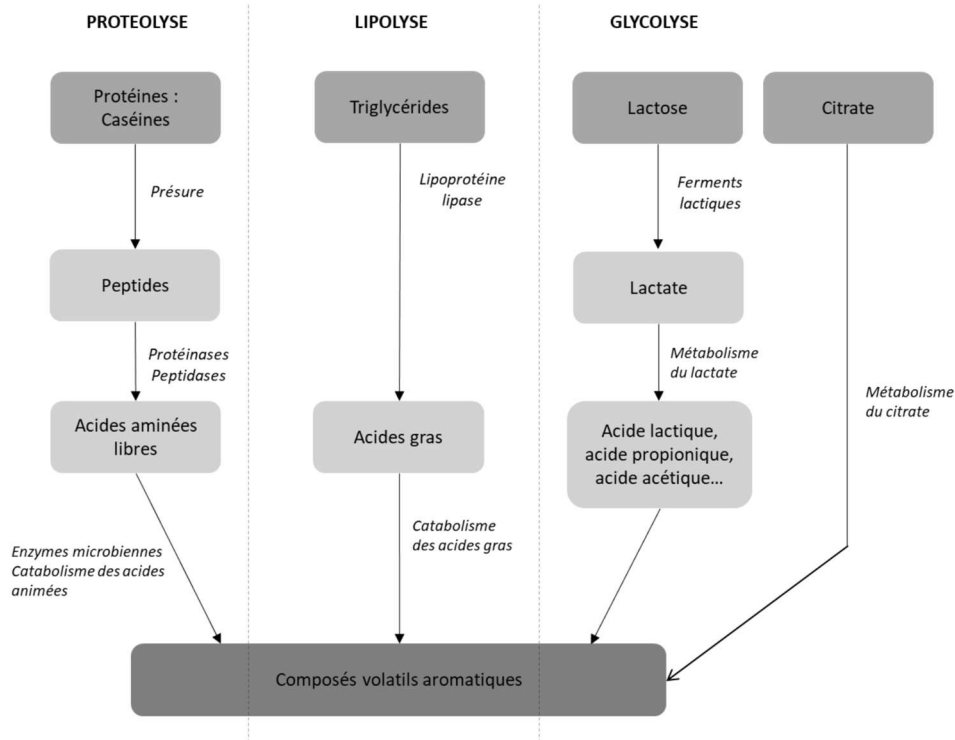


Figure 13 : Schéma résumé des trois voies biochimiques impliquées dans la formation des composés aromatiques du fromage d'après Andiç et al. (2015)⁷⁵

Les composés formés par ses trois voies biochimiques serviront de précurseurs pour la formation de composés volatils aromatiques en mettant en jeu 4 voies métaboliques⁷⁵⁻⁷⁷ :

- Le catabolisme des acides aminés (Figure 14),
- Le catabolisme des acides gras (Figure 15),
- Le métabolisme du citrate (Figure 16),
- Le catabolisme du lactate (Figure 17).

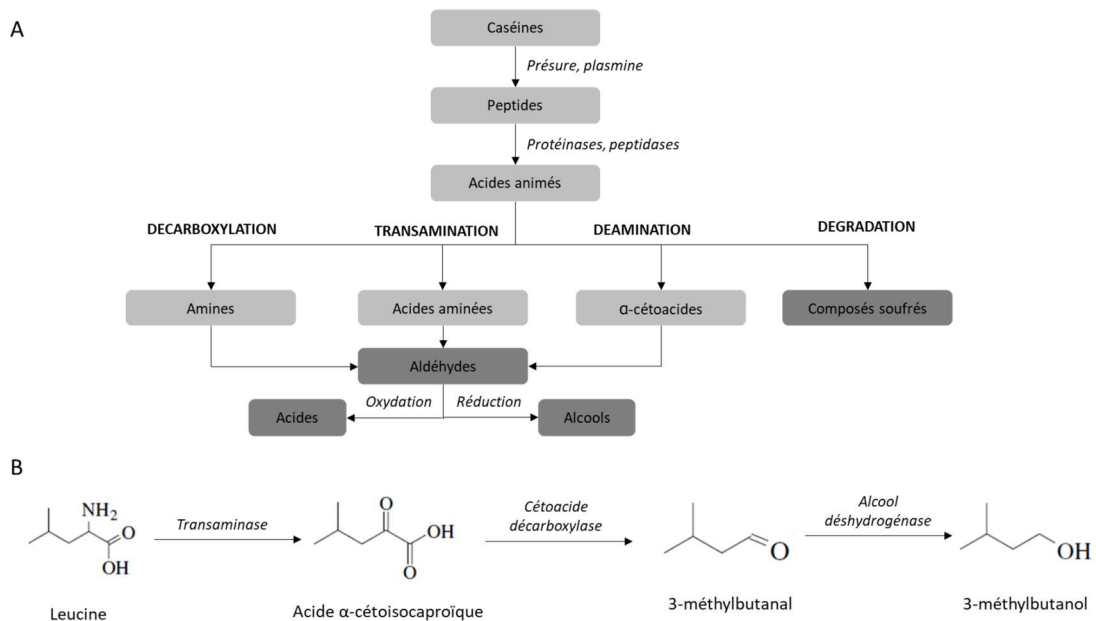


Figure 14 : (A) Schéma synthétique des différentes voies de formation des composés aromatiques par le catabolisme des acides aminés (B) Exemple du catabolisme de la leucine pour former du 3-méthylbutanal ou du 3-méthylbutanol d'après Smit et al. (2005)⁷⁸

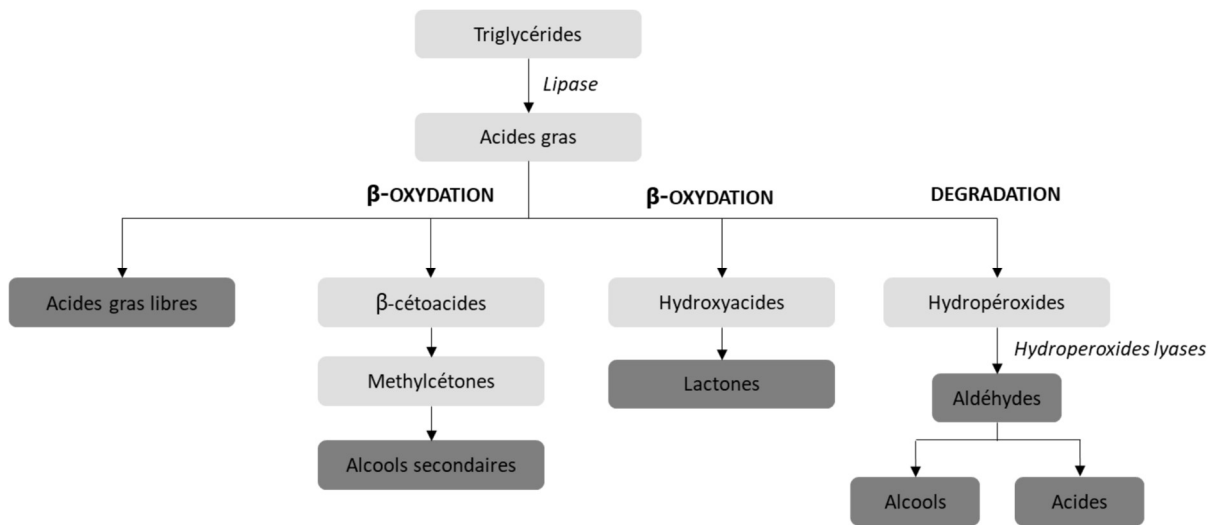


Figure 15 : Schéma synthétique des différentes voies de formation des composés aromatiques à partir des voies cataboliques des acides gras d'après Collins et al (2003)⁷⁹

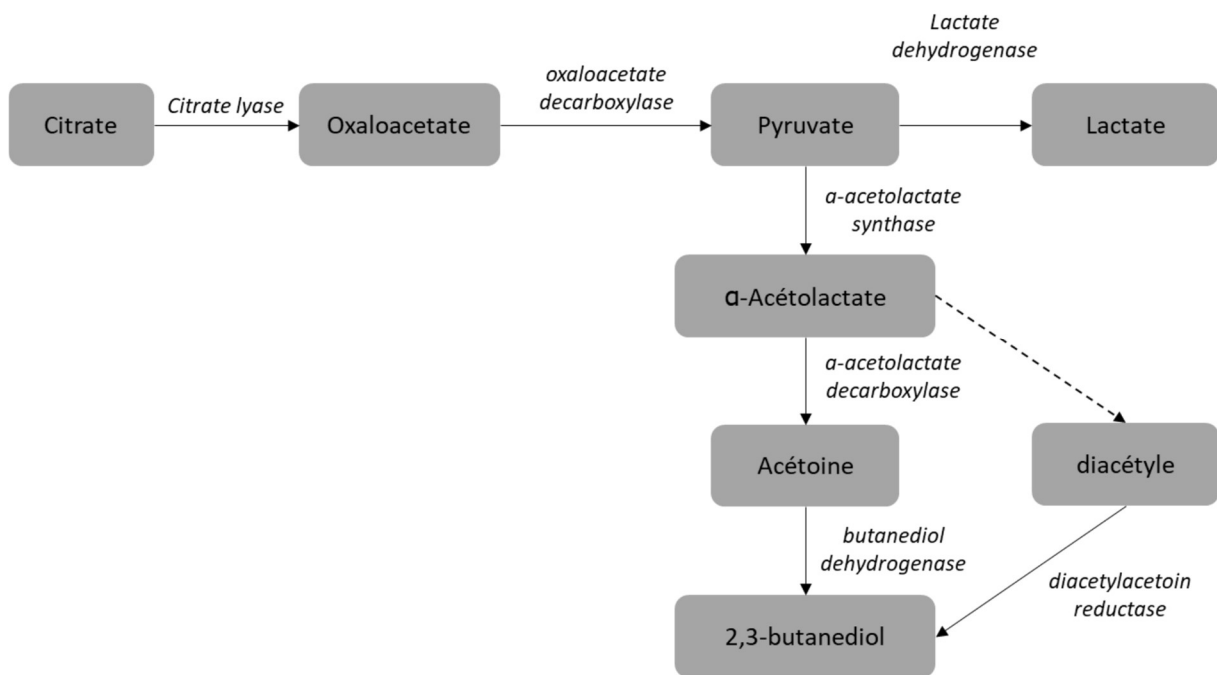


Figure 16 : Formation du 2,3-butanediol par la voie métabolique du citrate⁸⁰

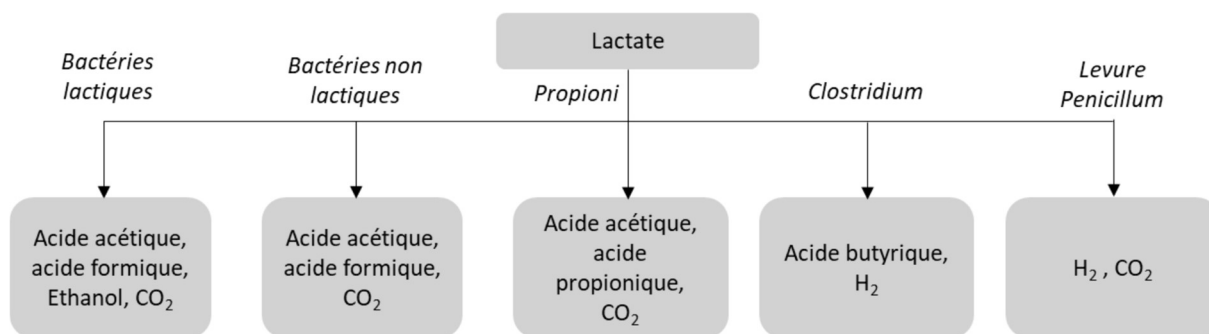


Figure 17 : Principales voies cataboliques du lactate d'après Andiç et al (2015)⁷⁵

De nombreuses études se sont intéressées à l'impact des étapes de transformation sur la composition en COV. Arques et al (2007)⁸¹ ont montré que le traitement du lait à haute pression avait un impact lors de la transformation mais pas sur le produit final. Ortigosa et al (2001)⁸² ont étudié l'impact de la pasteurisation du lait et l'utilisation d'une nouvelle culture indigène sur la formation des COV dans le fromage Roncal. Ils ont montré que la pasteurisation abaissait le niveau de certains composés volatils, en particulier les alcools, les aldéhydes et les cétones et que les fromages issus du lait pasteurisé présentaient des perceptions aromatiques plus faibles que pour ceux du lait cru. La culture indigène testée n'a exercé aucun effet significatif. Gomes et Torres (2015)⁸³ ont étudié l'altération des fromages au lait de brebis par cinq souches de *Clostridium* conduisant à un défaut de soufflage tardif. En effet, les spores ont été piégées dans la matrice de caséine pendant l'égouttage du lactosérum, une diminution du nombre de spores a été observée en raison de leur germination en cellules végétatives entraînant la voie de fermentation butyrique. Ces travaux ont montré des signes d'altération différents selon les stades de maturation et les espèces bactériennes testées. Ces fromages présentaient un profil volatil différent de celui du fromage témoin avec pour certains des niveaux plus élevés de 2-butanone, 2,3-butanedione et 2-butanol. Maria Izco et al (2000)⁸⁴ ont étudié l'impact de la protéolyse en ajoutant une préparation d'enzyme protéolytique dans le lait avant coagulation. Ils ont observé une augmentation du taux de libération des acides aminés et l'apparition de saveurs amères non-caractéristiques des fromages traditionnels non supplémentés en enzyme protéolytique. Rijnen et al (1999)⁸⁵ ont évalué l'effet de l'inactivation de l'araT (une aminotransférase aromatique) dans *Lactococcus lactis* sur le développement de l'arôme dans le fromage de type St Paulin. Leurs travaux ont montré que cette inactivation avait fortement empêché la formation de composés aromatiques floraux à partir d'acides aminés aromatiques, tandis qu'elle n'avait pas affecté la formation de composés aromatiques volatils à partir d'acides aminés à chaîne ramifiée et de méthionine. Cependant, ils n'ont pas relevé d'effet significatif sur l'aspect sensoriel, bien que l'odeur du fromage fabriqué avec le mutant ait eu tendance à être moins floral que celle du fromage fabriqué avec la souche de type sauvage. Delgado et al (2011)⁸⁶ se sont intéressés au profil volatil du fromage espagnol "Queso Ibores" au lait cru de chèvre pendant la maturation (jour 1, 30, 60 et 90). Les résultats ont montré que les composés les plus caractéristiques de l'arôme du fromage Ibores étaient les acides butanoïque, hexanoïque et octanoïque, certains alcools (2-butanol et 2-heptanol), les esters éthyliques des acides hexanoïque et butanoïque, certaines cétones méthyliques (2-butanone, 2-pentanone et 2-heptanone) et la δ -déalactone. Ils ont montré que le profil volatil pouvait être appliqué pour la caractérisation du fromage Ibores AOP et garantir le contrôle de qualité de ce fromage.

Des études se sont également intéressées aux conditions de stockage comme celles de Akpınar et al (2016) dont l'objectif était d'étudier l'impact de trois conditions d'emballage sous atmosphère modifiée sur des échantillons de fromage sepet traditionnel. Les composés volatils, la composition en

acides gras libres et les propriétés microbiologiques ont été analysés au cours de la période de stockage à 4°C. Ils ont montré que les conditions d'atmosphère modifiée avaient un impact sur la fraction aromatique et sur la composition en acide gras, cependant aucun effet significatif n'a été observé pour les propriétés microbiologiques. Ainsi, les fromages emballés dans des conditions modifiées différentes ont mieux protégé leurs spécialités.

Feng et al⁸⁷ ont étudié l'effet du dioxyde de carbone supercritique (SC-CO₂) avec variation de température et de pression pendant le stockage sur les micro-organismes du fromage mongol. Les résultats indiquent que le traitement par SC-CO₂ a permis de réduire de manière significative la génération de composés volatils (acides et esters), empêchant ainsi le développement de mauvaises odeurs pendant le stockage du fromage de Mongolie.

Le Tableau 2 présente une liste non exhaustive des principaux composés aromatiques volatils identifiés dans différents fromages, leurs notes aromatiques et leur seuil de perception.

Composés	Notes aromatiques	Seuil de perception (mg.m ⁻³ dans l'air)*	Ref
<i>Aldéhydes</i>			
(E,E)-2,4-Decadiénal	Frit	1,78.10 ⁻⁴	88
(E)-2-Nonénal	Ecurie, foin		89-91
(E)-2-Octénal	Frais, laiteux		89,92
(E,E)-2,4-Nonadiénal	Frit	2,51.10 ⁻⁴	88
(E,Z)-2,6-Nonadiénal	Concombre	6,03.10 ⁻⁵	89
(Z)-2-Nonénal	Stale, foin		89
(Z)-4-Heptenal	Rance		89
2,4-Dodécadiénal	Lacteo" (lactique)		88
2-Méthylbutanal	Chocolat noir, malté, légèrement caramélisé, noiseté		89,91,93
2-Méthylpropanal	Floral, malté	2,2.10 ⁻²	89-91
3-Méthylbutanal	Chocolat noir, malté, Vert, malté, pas mûr, cacao, malté, Herbacé	1,62.10 ⁻³	88,89,92,93
Acétaldéhyde	Éthéré, frais, vert, piquant	1	88,91,94
Heptanal	Vert, doux, savonneux	2,63.10 ⁻¹	90,95,96
Hexanal	Vert, herbe coupée		89,93
Méthional	Pomme de terre, Chou cuit, Pomme de terre bouillie, Sulfureux		89,94,97
Nonanal	Doux, floral, agrumes, herbeux	4,5.10 ⁻²	91,96
Octanal	Grasse, verte	7,7.10 ⁻³	90,93,96
trans-4,5-Epoxy-2-(E)-décénal	Métallique		88
Phénylacétaldéhyde	Rosé, styrène, fleuri, mielleux		89,96,97
<i>Alcools</i>			
1-Hexanol	Vert		395,98
2-Pentanol	Frais		
1-Octène-3-ol	Champignon		88,90
2-Phényléthanol	Rosy, floral	3,55.10 ⁻⁴	89,91,98
1-Pentanol	Alcoolisé, iodoforme, Fruité, Alcoolisé, vert, Balsamique	2,04	88,93,99
2-Heptanol	Terreux, huileux, sucré		88,96,98,100
2-Méthyl-1-propanol	Plastique, mauvais		88,98,101
2-Propanol	Fruité	7,943	86,88,101
3-Méthyl-2-buten-1-ol	Fromage, fruité, vert		88,102
Ethanol	Légère, éther, sèche, poussière	2	88,99,100
Propanol	Sucré (bonbon)		88,99,101
<i>Cétones</i>			
2,3-Butanedione	Beurré, crémeux, vanillé	4,9.10 ⁻³	89,96,99
1-Hexen-3-one	Bouteille d'eau en plastique		89
1-Octène-3-one	Champignon, champignon, terreux, fruité		88-90

2-Butanone	Vernis, sucré, fruité, caramel écossais		90,96,100
2-Heptanone	Fruité, épicé, cannelle	1,31	88,93,99,100
2-Hexanone	Floral, fruité		93,99,102
2-Nonanone	Fruité, moisi, fromage Gorgonzola, lait chaud, fromage fumé	1,69	88,93,99,100
2-Octanone	Fruité, Terreux	2,29.10 ⁻¹	88,96
2-Pentadécane	Humidité		88,102,103
2-Pentanone	Fruity, acetone, Sweet, fruity, cheesy, Orange peel	22,38	92,93,100
3-Hydroxy-2-butanone	Beurré, acidulé		89,99,100
3-Hydroxy-2-pentanone	Butyreux, herbacé, truffe		88,91,99
2-Tridécane	Chèvre		88,102
2-Undécane	Moisi, floral	2,00.10 ⁻⁴	86,88,96
3-Heptanone	Gongonzola, fruité, gras, épicé, herbacé, animaux, fromage bleu		88,90
8-Nonène-2-one	Fromage bleu		88,102
3-Octanone	Champignon, fruité	1,32.10 ⁻³	4
Acétone	Doux, fruité, pulpe de bois, foin		93,100
β-Damascénone	Compote de pommes		88,89
o-Aminoacétophénone	Tortilla de maïs, musqué		89
Acides gras volatils			
Acide méthylpropanoïque 2-	Doux, doux, pomme pourrie		96,98,100
Acide Methylbutanoïque 3-	Fruit pourri, doux, transpirant, rance, fécal, putride, fleuri		93,98,100
Acide acétique	Vinaigre, piquant, acide, Poivre, vert		89,99,100
Acide butanoïque	rance, tranchant, fromager, rance, sueur, aigre, putride		89,99,100
Acide décanoïque	Rance	8,5.10 ⁻²	89,90,98
Acide hexanoïque	Piquant, rance, fleuri		89,99,100
Acide propionique	Vinaigre, piquant, lait caillé		89,93,100
Acide phénylacétique	Rosé, mielleux		88
Acide octanoïque	Cire, savon, chèvre, moisi, rance, fruité	6,03.10 ⁻¹	89,99,100
Esters			
Butyrate de propyle	Fruité		88,101,104
3-Méthylbutanoate	Doux		86,88
Ethanoate de 3-méthylbutyle	Fruité, banane, bonbon		88,96,104
3-Octyle acétate	Floral		88
2-phényléthyle acétate	Floral, rose-like		88,91
2-Méthylbutyle hexanoate	Mentholé		88
Ethyle acetate	Comme un solvant, fruité, ananas,		90,90,98
Ethyle benzoate	Floral	6,03.10 ⁻⁴	88
Ethyle butanoate	Fruité, gomme à mâcher, Fruité, sucré, banane	1,74.10 ⁻¹	88,89,93,100
Ethyle hexanoate	Fruité, baie mûre, Fruité, pomme, banane	3,0.10 ⁻²	88,89,93,100
Ethyle octanoate	Fruité, banane, pomme, caramel	1,0.10 ⁻²	90,92,98
Ethyle pentanoate	Fruité		89,93,100
Ethyle-2-méthylbutanoate	Doux		88,93,96
Ethyl-2-méthylpropanoate	Fruit vert, non mûr		93,96
Ethyle-3-méthylbutanoate	Fruit vert + myrtille, fromage frais, fromagerie	4,47.10 ⁻¹	88,93,96
Isobutyle butanoate	Fruité	18,19	88
Pentyle acétate	Fruité	2,69.10 ⁻¹	88,96
Composés azotés			
2-Acétyle-1-pyrolline	Popcorn, grillé		89
2-Acétyle-2-thiazoline	Maïs soufflé, grillé		89
3-Isopropyl-2-méthoxypyrazin	Terreux, terreux		89
3-Méthylindole	Fécal, boules de naphtaline		89,94,97

Indole	Stable, jasmin		88
Composés soufrés			
Diméthylsulfure	Chou bouilli, chou-fleur, ail		93,98,100
Diméthylsulfure	Intense, sulfureux, chou		89-91
Diméthyltrisulfure	Sulfureux, chou, fécal		89,91,94
Alcanes			
Dodécane	Saleté		88,91,98
Terpènes			
Linalool	Orange, fleur		88
α -Pinène	Pin, vert		90,92
Lactones			
γ-Décalactone	Noix de coco		89,93,100
γ-Dodécalactone	Noix de coco, pêche, fruits frais, pêche, par, prune, noix de coco, beurre		88,89,91
γ-Nonalactone	Floral, pêche		89
δ -Octalactone	Floral, pêche, noix de coco		89,96,98
δ-Décalactone	Noix de coco, pêche		89,96,98
δ-Dodécalactone	Pêche, noix de coco, beurre, musqué, savonneux		91,98
δ-Nonalactone	Huile de noix de coco		89,91
Autres			
4-Méthylphénol	Vache, écurie, phénolique, moisi, stable	1,7.10 ⁻³	88,89,102
Toluène	Noix, amer, amande		90,92,95

Tableau 2 : Quelques exemples de composés aromatiques du fromage.
 * : les seuils de perception ont été déterminés d'après Devos et al. (1990)¹⁰⁵.

Les études décrites ci-dessus montrent que les profils volatils de produits agroalimentaires comme le thé ou le fromage sont impactés d'une part par la composition de la matière première et d'autre part par les procédés de transformation et de stockage. Ainsi les espèces végétales, animales et les conduites agricoles ont une grande influence sur les COV qui seront émis par l'aliment. Les activités enzymatiques et les modifications physicochimiques apportées lors de la transformation vont permettre d'obtenir le produit fini, sensoriellement conforme. Cependant, lorsque les équilibres des différentes chaînes métaboliques impliquées dans la transformation sont perturbés, cela peut se traduire par l'apparition de mauvaises odeurs. Ces profils volatils sont essentiels puisqu'ils participent grandement à l'appréciation de l'aliment par le consommateur.

C. Importance des COV dans le domaine agroalimentaire

L'analyse des COV responsables de l'arôme est un enjeu majeur pour l'industrie de l'agroalimentaire. L'arôme d'un aliment conditionnant son acceptabilité par le consommateur, il est important de connaître la nature exacte des composés qui le constituent et leur importance dans la modification du profil aromatique. Les COV ne sont pas seulement étudiés pour le côté aromatique mais également pour évaluer la sûreté alimentaire d'un produit, pour suivre des procédés de transformation ou de fabrication et pour tracer l'authenticité d'un produit. Dans un aliment, la composition en composés organiques volatils est caractéristique de sa qualité et de sa typicité, et apporte des informations essentielles quant à sa sûreté, sa maturation et son évolution au cours du temps.

1. Modification du profil aromatique - Arômes de synthèse

Le contrôle de la flaveur d'un produit alimentaire vise à satisfaire la demande du consommateur, c'est pour cela que les industries agro-alimentaires peuvent ajouter des arômes aux aliments pour modifier ou compléter un profil aromatique, aromatiser des produits initialement organoleptiquement neutres ou pour masquer des flaveurs désagréables.

On distingue deux types d'arômes alimentaires : les arômes naturels et les arômes de synthèse¹⁰⁶.

Les arômes naturels sont obtenus par extraction des produits naturels d'origine végétale ou animale selon des procédés physiques ou biotechnologiques de récupération¹⁰⁶⁻¹⁰⁸. Les arômes de synthèse comprennent :

- les arômes artificiels produits de manière chimique en laboratoire pour imiter le goût de l'arôme naturel dont sa structure moléculaire ne peut pas être retrouvée dans la nature
- les arômes de synthèse « identiques naturels »¹⁰⁶ synthétisés chimiquement mais identiques aux arômes naturels possédant exactement le même goût et la même odeur, il s'agit d'une copie pure et simple de l'arôme naturel.

On retrouve également les arômes de transformation et les arômes de fumée. Les arômes de transformation également appelés « arômes de Maillard »¹⁰⁹ sont obtenus par chauffage maîtrisé d'un mélange d'ingrédients réagissant entre eux pour former un type d'arômes. L'objectif étant de copier les réactions ayant lieu naturellement lors de la cuisson des aliments, afin d'en reproduire industriellement les arômes. Les arômes de fumée¹¹⁰ sont produits par un processus de combustion du bois appelé « pyrolyse » et confèrent une saveur « fumée » à l'aliment.

2. Sûreté alimentaire (pesticides, contamination, stockage, emballage...)

Les contaminants sont des substances qui se retrouvent dans des produits alimentaires et qui présentent un risque pour la santé animale ou humaine. Ils peuvent être de nature microbiologique ou chimique et proviennent de sources diverses. Ces contaminations vont entraîner des modifications dans le profil volatil des aliments pouvant modifier l'aspect aromatique ou non.

a. Pesticides

Les pesticides¹¹¹, qui comprennent principalement les pesticides organochlorés, les pesticides organophosphorés, les pesticides carbamates, les pesticides tels que des phénylurées et des pyréthroïdes, sont utilisés pour lutter contre les parasites, les champignons et les mauvaises herbes afin de limiter les pertes de récolte et accroître le rendement des cultures¹¹². Les pesticides ou autres produits agrochimiques peuvent pénétrer dans les tissus des fruits, des légumes, etc., où ils restent sous forme de résidus, entraînant la contamination des aliments. Largement utilisés dans l'agriculture, les pesticides peuvent également être transférés aux animaux de plusieurs façons (environnement, alimentaire...). Les résidus présents dans les aliments peuvent avoir des effets aigus et chroniques sur la santé humaine en raison de leur toxicité. Des méthodes analytiques fiables ont été développées pour déterminer les résidus de pesticides dans les aliments¹¹¹ principalement pour les quantifier afin de déterminer si leur présence est toxique pour les consommateurs. Ainsi, de nombreuses études se

sont focalisées sur le développement de méthodes pour la recherche et la quantification de pesticides dans divers produits agroalimentaires : les fruits et les végétaux¹¹³, le lait bovin¹¹⁴, le cacao¹¹⁵, les pommes^{116,117}, la laitue¹¹⁸, les jus de fruits¹¹⁹, le vin^{120,121}...

Cependant ces composés ainsi que leurs résidus sont des molécules de masse moléculaire élevée et sont rarement retrouvés lors de l'analyse des profils volatils des aliments.

b. Contamination

Les contaminations microbiologiques ou la présence de mycotoxines dans les aliments peut entraîner la synthèse de composés organiques volatils spécifiques modifiant la composition en COV de l'aliment et pouvant jouer un rôle sur sa note aromatique. Les COV produits par ces contaminants ne sont pas forcément nocifs mais la présence de l'agent microbien peut l'être. Ces composés peuvent donc être de bons marqueurs de contamination microbienne du produit. La recherche de contamination par des agents microbiens est réalisée par l'analyse microbiologique exigée par la réglementation pouvant être longue (3 à 4 jours). Ces dernières années, la recherche s'est intéressée à la détection de composés organiques volatils (COV) pouvant être associés à ces contaminations dans les aliments. Carraturo et al (2020)¹²² ont étudié l'évaluation qualitative et quantitative des COV produits par *Salmonella Typhimurium*, *Campylobacter jejuni* et *Staphylococcus aureus* dans différents types de viande crue (bœuf, porc, poulet). Ces travaux ont montré que l'identification de profils uniques de COV, marqueurs de contamination de la viande par des agents pathogènes spécifiques (par exemple, une nette augmentation de la quantité d'hexane lors de la contamination par *S.Typhimurium*), pourrait permettre de déterminer rapidement la contamination d'échantillons de viande spécifiques. D'autres travaux¹²³⁻¹²⁶ ont également démontré le potentiel de l'analyse des COV pour la détection d'une grande variété d'agents pathogènes et la possibilité d'application de cette approche à diverses matrices alimentaires.

c. Emballage

L'emballage est un élément indispensable du processus de protection des aliments. Plus de 30 plastiques différents sont utilisés comme matériaux d'emballage¹²⁷. Pour améliorer les performances de ces matériaux, de nombreux types d'additifs, tels que des antioxydants, des stabilisants, des lubrifiants et des agents antistatiques, ont été développés. Cependant, des constats récents ont montré que l'emballage représentait lui-même une source de contamination par la migration de substances le constituant vers les aliments¹²⁸. Les conditions d'emballage peuvent également accélérer ou réduire certaines voies métaboliques ou en produire de nouvelles. Kuuliala et al¹²⁹ ont également montré que différentes conditions d'emballage et de stockage affectent l'évolution de la flore microbienne du poisson et les COV générés dans l'espace de tête de l'emballage. D'autres études¹³⁰⁻¹³² ont montré l'influence des conditions de stockage sur la composition volatile des produits alimentaires en étudiant la formation et la libération des COV selon le type d'emballage, le temps et les conditions de conservation.

3. Suivi de procédés

Les composés organiques volatils peuvent être utilisés comme marqueurs pour le suivi des procédés de transformation alimentaire. Ils sont notamment utilisés pour suivre les facteurs influençant la production des arômes ou encore pour mettre en évidence des problèmes de contamination microbiologique ou chimique pouvant intervenir dans les différentes étapes de production. Par exemple, Huang et al¹³³ ont étudié l'impact de la température et du temps sur la torréfaction des graines de cacao, étape la plus importante dans le développement de la saveur du cacao, en suivant 60 composés volatils. Soukoulis et al¹³⁴ ont étudié la fermentation lactique, étape essentielle à la formation d'une grande variété de produits laitiers (yaourt, crème, fromages...). Ils ont suivi la production ou l'appauvrissement de certains COV tel que l'acétaldéhyde, le diacétyl, l'acétoïne et la 2-propanone et ont montré une bonne corrélation avec les taux d'acidification. Ils ont également suivi l'évolution des COV mineurs tels que l'acide acétique, la 2,3-pentanedione, l'éthanol, et les COV liés aux arômes indésirables tels que le sulfure de diméthyle et le furfural. Ces résultats démontrent la possibilité de suivre la fermentation lactique en mesurant l'un de ses sous-produits volatils. Tsevdou et al¹³⁵ ont également étudié la fermentation lactique en évaluant l'influence de paramètres technologiques (ajout de transluminale ou traitement thermique du lait) sur la cinétique de production des COV. Chen et al¹³⁶ ont examiné les changements de composition des composés organiques volatils (COV) dans le Liuyang Douchi, produit traditionnel chinois à base de soja fermenté, pendant la fermentation. Ils ont mis en évidence que la plupart des alcools et des acides diminuaient avec la fermentation, tandis que les esters, les cétones, les pyrazines et les phénols s'accumulaient. Ils ont identifié 49 COV marqueurs des différents stades de fermentation. Makhoul et al¹³⁷ ont suivi l'évolution des COV pendant la levée des pâtes à pain, et sur le pain cuit résultant, après inoculation avec différents starters industriels. Les résultats ont montré des différences entre les trois starters industriels sur la base de leur cinétique de production de COV pendant le levage donnant finalement des pains significativement différents. Yang et al¹³⁸ ont étudié le changement en composés volatils qui se produit pendant le séchage à l'air chaud du champignon *Flammulina velutipes*. Ils ont montré que le processus de séchage pouvait diminuer de manière significative la teneur relative en cétones et favorisait la génération d'alcools, d'acides et d'esters.

4. Traçabilité et authenticité des produits agroalimentaires (origine, fraude)

L'authenticité et la traçabilité alimentaire sont des enjeux importants permettant de rassurer les consommateurs sur les produits qu'ils consomment. La traçabilité désigne la capacité de retracer le cheminement d'une denrée alimentaire tout au long de sa chaîne de production et de distribution, soit depuis l'origine première du produit jusqu'à sa remise au consommateur final. L'authenticité s'attache à décrire les caractéristiques attendues du produit agroalimentaire, que ce soit sa provenance, la façon dont il a été produit et transformé mais également les qualités nutritionnelles, gustatives et olfactives du « véritable » produit. Les produits agroalimentaires de grande qualité bénéficient généralement d'une bonne image de marque et de ce fait sont recherchés par les consommateurs. La conséquence est qu'ils sont largement copiés par des acteurs peu scrupuleux, ce qui peut aboutir à des fraudes d'étiquetage et des adultérations consistant à ajouter des produits de moindre valeur à un produit conforme. C'est pour cela que les filières renforcent les contrôles au travers de labels déclaratifs et d'outils de certification (contrôle du cahier des charges de chaque producteur). Ces fraudes ont des conséquences économiques pour les producteurs, ainsi, l'authentification des produits est fondamentale non seulement pour empêcher la falsification des aliments, mais aussi pour garantir la

commercialisation des produits labélisés possédant une appellation d'origine protégée (AOP) ou d'une indication géographique protégée (IGP) par exemple¹³⁹. D'un point de vue sanitaire, la consommation de produits frauduleux ou altérés peut s'avérer dangereuse avec des effets à court ou à long terme¹⁴⁰.

Les systèmes de traçabilité permettent de répondre aux exigences des cahiers des charges définis par l'entreprise, par le syndicat de protection (AOP) ou par le client. Ils reposent majoritairement sur des documents administratifs ainsi que sur des contrôles de qualité. Prenons l'exemple de l'AOP Pomme du limousin : une traçabilité en amont c'est-à-dire aux champs est réalisée par les producteurs pour répondre aux obligations du cahier des charges ; une traçabilité interne vise à enregistrer les lots de pommes entrant dans les zones de stockage (code producteur associé au code de la parcelle) ou des analyses (teneur en sucre, fermeté et acidité) et un jury de conformité est effectué, puis une traçabilité aval correspondant à la distribution client est réalisée. Dans certains cas il s'avère que pour un lot, les analyses qualitatives (sucre, fermeté, acidité) soient correctes. En revanche, si le jury de conformité ne valide pas le lot, il sera alors déclassé. Les tests de conformité effectués par un jury ont pour objectif de déterminer l'authenticité du produit. Ils sont réalisés de manière inopinée mais ne sont pas appliqués à chaque lot. Ainsi, l'utilisation de méthodes physico-chimiques en lien avec les qualités organoleptiques des produits et répondant à diverses questions concernant la traçabilité permettrait de renforcer les systèmes de contrôle.

Les composés organiques volatils présents dans les aliments, responsable de l'odeur et de l'arôme, s'avèrent être de bons marqueurs pour répondre à ces enjeux. En effet, le profil en COV d'un aliment est affecté par l'origine géographique, les conditions de culture, les variations saisonnières, les procédés de transformation et de conservation et les différents stockages. L'étude de la similarité ou de la diversité des profils des COV de différents échantillons pourrait contribuer à garantir l'authenticité des aliments.

Plusieurs études ont démontré l'utilisation des COV et principalement des profils volatils pour déterminer l'authenticité des divers produits alimentaires. Les travaux de Perestrelo et al (2019)¹⁴¹ ont montré que la signature volatile du cidre permettait de différencier les cidres étudiés provenant de différentes régions géographiques de l'île de Madère. Les résultats fournissent des informations pertinentes liées à la signature volatile des cidres produits dans l'île de Madère, et pourraient ouvrir de nouvelles opportunités pour promouvoir le processus de fabrication du cidre. La discrimination géographique par les profils volatils a permis de reconnaître les caractéristiques uniques et distinctives de la qualité et de la typicité des produits originaires de certaines régions géographiques. Les COV comme marqueurs de typicité permettent de qualifier un produit pour lui attribuer un label afin de le mettre en valeur. Sousa et al (2020)¹⁴² ont également montré la discrimination du cidre selon leur origine de production par l'analyse des COV. D'autres études se sont intéressées à la discrimination des origines géographiques sur d'autres matrices alimentaires (huiles d'olives^{143,144}, fromage¹⁴⁵...). Les profils volatils ont également été appliqués pour identifier différentes origines botaniques de miel¹⁴⁶ ainsi que différentes variétés de miel¹⁴⁷, de graine de cacao¹⁴⁸, d'amande¹⁴⁹. L'authentification des huiles végétales est un des sujets les plus étudiés en raison de leur falsification par des huiles moins chères et de mauvaise qualité [59]. Les études de Ozcan-Scinir et al (2020)¹⁵⁰ et de Kharbach et al (2022)¹⁵¹ ont montré que la quantité et la composition des COV dans les huiles étaient liées à la fois à la qualité et au niveau de fraude. Sur la base des résultats obtenus, l'utilisation des COV pour déterminer la falsification de l'huile végétale peut être considérée comme une méthodologie de détection efficace.

D. Conclusion

Les composés organiques volatils dans les aliments sont impactés par la vie du produit et permettent de répondre à de nombreux enjeux. Comme le montre ce chapitre, de plus en plus d'études s'intéressent à l'analyse des COV pour répondre à des questions d'authenticité, de suivi de procédé, d'analyse des conditions de stockage et d'emballage ou encore pour rechercher des contaminations. Ces composés apportent donc de nombreuses informations et s'avèrent être de puissants marqueurs pour renforcer les systèmes de traçabilité. Pour cela, les sciences analytiques développent des méthodes d'analyses des COV permettant de répondre au mieux aux attentes des acteurs agroalimentaires mais également à leur contrainte. En effet, le challenge réside dans le développement d'outils pouvant réaliser des analyses rapides à hauts débits, si possible sans détruire l'échantillon et pouvant dans l'idéal être introduits sur des lignes de production.

II. Les méthodes analyse des COV

L'évaluation des odeurs et des arômes se fait majoritairement par l'analyse sensorielle¹⁵¹. Si ces méthodes semblent efficaces pour décrire les notes aromatiques évaluées ou déterminer l'acceptation du produit par les consommateurs, en revanche, elles ne sont pas adaptées à toutes les problématiques, comme par exemple celles relevant de la sécurité des aliments. L'analyse sensorielle nécessite de constituer des panels de testeurs qui pourront être formés, et devront se rendre disponibles. Cela peut être une limite. De plus, ces analyses restent assez subjectives et ne conduisent pas toujours à un consensus du panel. Un autre outil populaire pour évaluer rapidement le profil aromatique complet des produits alimentaires est le nez électronique (E-nose)¹⁵². Cette technique utilise différents types de capteurs dont les plus courants sont : les semi-conducteurs à oxydes métalliques, les polymères conducteurs, les capteurs piézoélectriques, les capteurs optiques ou calorimétriques^{153,154}. Néanmoins, les nez électroniques présentent certains inconvénients comme : l'incapacité à identifier individuellement les composés odorants ainsi que de prendre en compte les effets négatifs de l'humidité et de la température sur les stimuli sensoriels¹⁵⁵. Ces dernières années les sciences analytiques des aliments ont donc développé de nombreuses techniques pour l'analyse des COV comme par exemple, la spectrométrie de masse, couplée à la chromatographie gazeuse ou liquide¹⁵⁶. Des travaux plus récents ont montré l'intérêt des technologies de spectrométrie de masse à injection directe. Dans ce chapitre, deux techniques seront présentées : la technique de chromatographie gazeuse couplée à une double détection par spectrométrie de masse et ionisation de flamme (GC-MS-FID) et la technique de spectrométrie de masse à injection directe (Selected Ion Flow Tube-Mass Spectrometry, SIFT-MS).

A. Méthodes usuelles d'analyses des COV dans les matrices alimentaires

1. Analyse chromatographique

Pour analyser les composés organiques volatils dans les produits agroalimentaires, la chromatographie en phase gazeuse (GC) est aujourd'hui la méthode la plus largement utilisée¹⁵⁷. Parmi les nombreux détecteurs disponibles, le choix est guidé principalement par l'objectif analytique. Pour une analyse ciblée, des détecteurs spécifiques seront plutôt choisis comme, par exemple, un détecteur thermoionique (NPD) pour les composés azotés^{158,159} ou un détecteur à photométrie de flamme (FPD) pour les composés soufrés ou phosphorés¹⁶⁰⁻¹⁶². Cependant, dans le domaine de l'agroalimentaire, une détection par spectrométrie de masse est souvent privilégiée pour l'analyse des nombreuses molécules présentes² et leur identification structurale. Une détection par ionisation de flamme (FID)¹⁶³ est aussi utilisée pour sa robustesse, la linéarité de réponse et sa stabilité qui permettent une quantification fiable de molécules connues. Une double détection MS/FID est parfois mise en œuvre pour combiner en une seule analyse identification structurale (par MS) et quantification (par FID) des COV¹⁶⁴⁻¹⁶⁶. Dans le cas de l'analyse d'arômes, un détecteur olfactométrique peut aussi être couplé au GC (GC-O) en parallèle d'une détection par spectrométrie de masse ou FID pour identifier les COV contribuant à l'odeur^{52,167}.

Pour l'analyse chromatographique, le choix de la colonne est un paramètre clé qui dépend de la nature des composés à étudier. Les phases stationnaires sont soit apolaires composées de polymères siliconés dérivés du polydiméthylsiloxane (PDMS) soit polaires, à base de polyéthylène glycol (PEG). Diverses

colonnes capillaires polaires et non polaires sont disponibles. Dans le cas d'un mélange complexe, il est rare que tous les composés soient séparés à l'aide d'une seule phase stationnaire, ainsi des montages de colonne en série ou en tandem (GCxGC) peuvent être utilisés¹⁶⁸.

Les produits alimentaires sont des matrices complexes contenant des centaines voire des milliers de composés volatils, semi-volatils et non volatils. La qualité des analyses dépend fortement de la méthode de préparation des échantillons, dont les principales sont présentées ci-dessous.

2. Méthodes d'échantillonnage

La première étape et la plus cruciale, de toute procédure analytique est la préparation de l'échantillon¹⁶⁹. Cette étape est principalement utilisée pour purifier l'échantillon en éliminant les composés interférents, afin d'extraire et de préconcentrer les analytes cibles. Cette étape nécessite du temps et est sujette à des erreurs¹⁷⁰. Il existe de nombreuses méthodes de préparation et d'extraction des COV. Le choix de la méthode repose sur deux facteurs : la nature de la matrice à étudier et la finalité de l'analyse.

a. L'extraction par solvant

L'extraction par solvant est une opération de séparation d'un ou plusieurs composés se trouvant dans l'échantillon, solide ou liquide, grâce à l'utilisation d'un solvant dans lequel le ou les composés à extraire sont solubles. Parmi les techniques d'extraction par solvant, on retrouve l'extraction liquide-liquide (LLE) ou solide-liquide (SLE). Cette dernière est classiquement réalisée par Soxhlet ou plus récemment par ASE (Accelerated Solvent Extraction), qui, sous l'effet combiné de la pression et de la température, permet de réaliser des extractions plus rapidement.

L'extraction liquide-liquide (LLE) et l'extraction solide-liquide (SLE)

Ces méthodes consistent à mettre en contact l'échantillon liquide pour la LLE et solide pour la SLE avec un solvant. Les composés d'intérêt à extraire doivent avoir une plus forte affinité pour ce solvant que pour la matrice dans laquelle ils sont contenus. Ces méthodes peuvent être assistées par les ultrasons (UAE), la pression (PLE) ou les micro-ondes (MAE).

Dans l'agroalimentaire, la LLE est utilisée pour les matrices liquides tel que l'eau, le lait, le vin, pour analyser les hydrocarbures aromatiques (HAP), les arômes¹⁷¹⁻¹⁷³, les résidus de pesticides¹⁷⁴ ou d'antibiotiques¹⁷⁵, ou encore les contaminants^{176,177}. Ainsi, la norme ISO 15753 : 2006¹⁷⁸ indique que l'extraction des HAP contenus dans les corps gras d'origine animale ou végétale s'effectue avec un mélange de solvants polaires acétonitrile/acétone. Concernant la SLE, les techniques telles que le Soxhlet et plus récemment l'ASE, sont majoritairement utilisées.

Ces méthodes permettent d'extraire des composés de différentes polarités en fonction du solvant choisi, de préconcentrer les composés cibles et ainsi améliorer la sensibilité de détection. Cependant, les principaux inconvénients rencontrés sont la durée de préparation de l'échantillon, l'utilisation de grandes quantités de solvants toxiques et inflammables mais également la formation d'émulsions qui entraîne la perte partielle ou totale d'analytes.

Extraction par Soxhlet

L'extraction par Soxhlet est une méthode utilisant la solubilité des composés ciblés dans un solvant organique pour les extraire d'échantillons solides et l'insolubilité du reste de l'échantillon dans le solvant choisi. Le solvant d'extraction recircule continuellement à travers la matrice, par ébullition et condensation, les composés d'intérêt étant recueillis dans le solvant chaud.

Dans l'agroalimentaire, l'extraction par Soxhlet est principalement utilisée pour l'analyse des lipides¹⁷⁹⁻¹⁸¹ et des hydrocarbures aromatiques polycycliques (HAP)^{182,183}.

Les principaux avantages de cette méthode sont la grande disponibilité de solvants pouvant extraire des composés de différentes polarités et l'utilisation d'appareils à faible coût. Cependant, cette méthode utilise de grande quantité de solvant, nécessitant souvent des étapes d'évaporation/concentration après l'extraction, et des procédures longues. De plus, l'extraction des échantillons est généralement réalisée au point d'ébullition du solvant susceptible d'entraîner une décomposition des analytes thermolabiles. Enfin, il est important de connaître la nature du ou des composés à extraire afin de sélectionner le solvant approprié. Bien que l'extraction au Soxhlet soit simple, standard et robuste, l'opération manque d'automatisation, bien que plusieurs échantillons puissent être extraits en parallèle.

Extraction accélérée par solvant (ASE)

Cette technique utilise des températures et des pressions élevées avec des solvants liquides pour obtenir des extractions rapides et efficaces de composés présents dans des échantillons solides ou semi-solides. Pour obtenir une bonne efficacité, quantification et reproductibilité, divers paramètres tels que la température, le choix du solvant, la pression, les cycles d'extraction et le temps doivent être optimisés.

L'application de l'ASE dans les études de sécurité alimentaire¹⁸⁴ a été rapportée pour l'extraction de divers composés et contaminants tels que les hydrocarbures aromatiques polycycliques (HAP)¹⁸⁵⁻¹⁸⁸, les pesticides résiduels¹⁸⁹⁻¹⁹², ou encore les lipides¹⁹³⁻¹⁹⁵.

Comparée aux autres techniques d'extraction par solvant (Soxhlet par exemple), l'ASE est plus rapide avec une consommation de solvant plus faible ainsi que des températures et des pressions moins élevées peuvent être utilisées. Le milieu d'extraction est récupéré facilement sans avoir besoin d'une étape supplémentaire pour le séparer de la matrice. Ses principaux inconvénients sont les coûts d'investissement, plus élevés par rapport aux méthodes conventionnelles, la dégradation des matières thermosensibles due à l'utilisation de température élevée (entre 50 et 150°C).

b. Extraction sur phase solide (SPE)

L'extraction sur phase solide est une méthode basée sur le partage des composés à extraire entre une phase liquide, l'échantillon, et une phase solide ou stationnaire, l'adsorbant¹⁹⁶. Celle-ci repose sur la séparation des composés d'intérêt des autres composants du mélange. Le choix de l'adsorbant revêt une importance capitale qui va définir la sélectivité spécifique aux composés d'intérêts. Les paramètres physico-chimiques du (des) composé(s) à extraire vont donc conditionner la nature de l'adsorbant.

Dans l'agroalimentaire, la SPE est utilisée principalement pour la caractérisation des arômes¹⁹⁷⁻²⁰¹, la recherche de pesticides^{202,203} et d'HAP²⁰⁴⁻²⁰⁶.

La SPE a été développée comme une alternative à l'extraction liquide-liquide pour éliminer de nombreux inconvénients, tels que l'utilisation d'une grande quantité de solvant, la durée prolongée des opérations et un coût élevé. De plus, elle permet d'extraire simultanément une large gamme d'analytes pour une gamme de polarité. Cependant, cette méthode est difficilement automatisable, présente une faible sélectivité conditionnée par l'adsorbant et le solvant utilisé qui dépend de son affinité avec les composés cibles et une faible reproductibilité. Enfin, la SPE n'est applicable qu'à des échantillons liquides.

c. Micro extraction sur phase solide (SPME)

La micro extraction sur phase solide, inventée en 1989 par Janusz Pawliszyn (Université de Waterloo Ontario, Canada)²⁰⁷, est une technique d'échantillonnage passif basée sur la sorption des composés d'intérêt présents dans l'échantillon sur une fibre. Celle-ci est constituée d'un barreau de silice fondue d'environ 1 cm de long recouvert d'une phase stationnaire et rétractable dans une aiguille de protection en inox (Figure 18). Le processus d'extraction des composés sur la fibre est principalement réalisé de deux manières : extraction directe ou par immersion (DI-SPME) et par espace de tête (HS-SPME)²⁰⁸. La DI-SPME est spécifique aux échantillons liquides puisque la fibre est directement immergée dans l'échantillon. La HS-SPME s'applique à des échantillons solides, liquides ou gazeux, la fibre est exposée à la phase gazeuse au-dessus de l'échantillon liquide ou solide placé dans un flacon hermétiquement fermé. Les composés extraits sont ensuite désorbés de deux manières selon le type d'analyse chromatographique : par désorption thermique dans le port d'injection d'un chromatographe en phase gazeuse et par désorption liquide utilisant une petite quantité de solvant dans une chambre de désorption d'un chromatographe en phase liquide²⁰⁹.

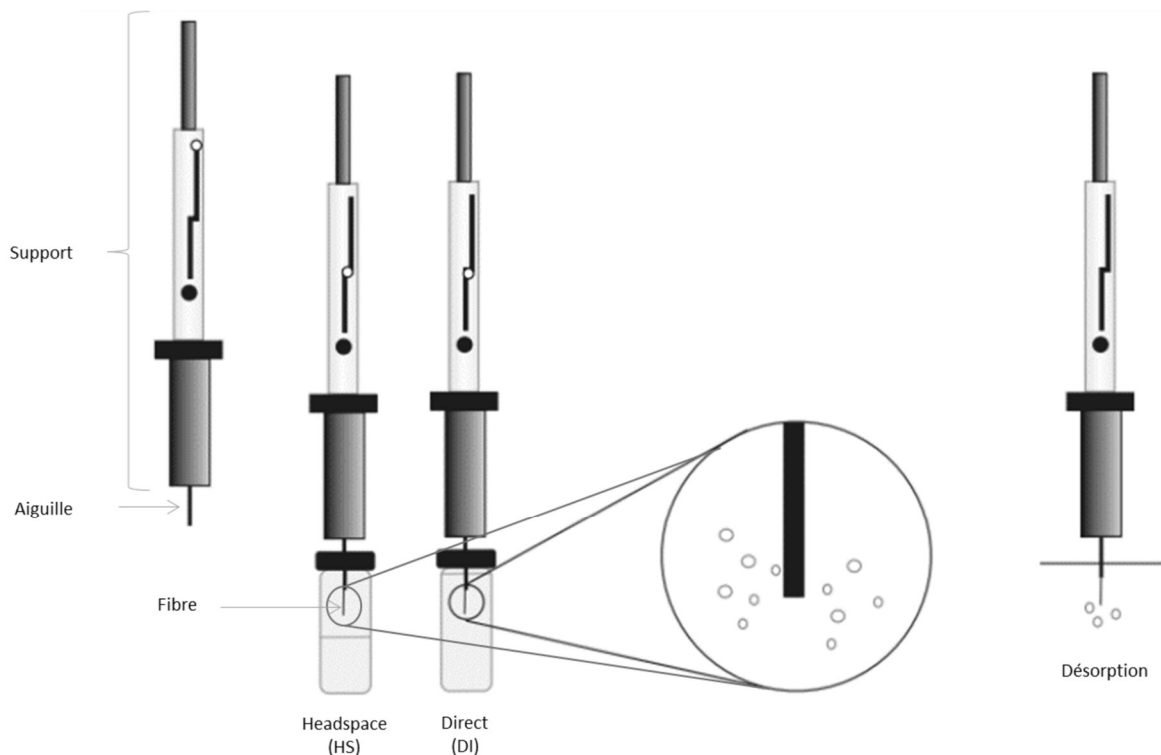


Figure 18 : Schéma de la micro-extraction en phase solide d'après Starowicz (2021)²

La SPME est une méthode très courante pour l'analyse des composés organiques volatils dans les aliments afin d'étudier les arômes, les mauvaises odeurs (off-flavors), les composés migrant à partir d'emballages alimentaires ou encore les composés toxiques volatils (Tableau 4).

La SPME présente de nombreux avantages par rapport aux méthodes d'extraction traditionnelles. En plus d'être une méthode rapide, simple et sans solvant, elle permet d'extraire et de pré-concentrer les composés en une seule étape en utilisant une faible quantité d'échantillon. Son utilisation pour des échantillons solides, liquides et gazeux en fait une méthode utilisée dans de nombreuses applications. De plus, la SPME est facilement automatisable permettant des analyses à haut débit.

Cette méthode présente, cependant, quelques inconvénients comme le nombre limité de phases stationnaires disponibles, la température de thermodésorption recommandée relativement haute (240-280 °C), l'instabilité et le gonflement dans les solvants organiques ainsi que la durée de vie limitée de la fibre²¹⁰. De plus, des interférences de composés non volatils et de poids moléculaire élevé peuvent affecter la reproductibilité, l'efficacité de l'extraction et la durée de vie du revêtement des fibres²¹¹.

Extraction sur barreau recouvert d'une phase adsorbante (Stir-Bar Sorptive Extraction, SBSE)

La SBSE est basée sur l'absorption et la désorption des analytes sur un revêtement polymérique enrobant un barreau magnétique placé dans l'échantillon (liquide). Le revêtement polymérique très largement utilisé en SBSE est le polyméthylsiloxane (PDMS).

Dans l'agroalimentaire, la SBSE est utilisée pour analyser les hydrocarbures aromatiques polycycliques (HAP)²¹²⁻²¹⁴, les pesticides et leurs résidus²¹⁵⁻²¹⁷, les COV^{218,219} et l'arôme des aliments^{220,221}.

Cette méthode présente les mêmes avantages que la SPME avec une quantité d'absorbant plus importante pour extraire davantage d'analytes puisque la couche de PDMS est beaucoup plus épaisse que celle des fibres SPME utilisant cette phase²²². L'inconvénient majeur de la SBSE concerne le choix limité de revêtements disponibles. De nouveaux revêtements à double phase utilisant différents adsorbants carbonés en plus du PDMS sont actuellement en cours de développement, ainsi que d'autres phases couvrant une large gamme de polarité et contenant des polymères comme le poly(méthacrylate de stéaryle-diméthacrylate d'éthylène) et le polypyrrole²²³. Cependant, la SBSE n'est pas encore aussi largement utilisée que la SPME en raison de la difficulté à parvenir à une automatisation complète.

d. Extraction par l'espace de tête (Headspace HS)

Les techniques d'espace de tête (ou « Headspace ») sont des méthodes d'extraction en phase gazeuse impliquant le partage des analytes entre une phase liquide ou solide et le ciel gazeux au-dessus. On distingue les extractions statiques et les extractions dynamiques²²⁴.

Extraction statique

L'espace de tête statique consiste à placer l'échantillon étudié dans un récipient hermétiquement clos qui peut être chauffé pour favoriser le transfert des analytes vers l'air situé au-dessus de la matrice (espace de tête). Cet espace de tête est prélevé et directement injecté dans un chromatographe en phase gazeuse pour l'analyse.

L'utilisation de l'espace de tête statique dans l'agroalimentaire concerne de nombreuses applications comme l'authentification²²⁵, le suivi de procédé^{226,227}, la discrimination d'espèce^{228,229}, la détermination en composés aromatiques²³⁰⁻²³² ou encore la recherche de marqueurs d'oxydation par exemple^{233,234}.

Cette technique est très fréquemment utilisée pour la séparation des composés volatils d'une matrice complexe puisqu'elle permet d'éviter les interférences dues aux composés non volatils qui ne sont pas transférés dans la phase gazeuse²³⁵. De plus, cette méthode est rapide, facile à mettre en œuvre, non destructive et ne nécessite pas de solvant. En revanche, cette méthode présente une sensibilité limitée et n'est pas très adaptée à l'extraction des composés peu volatils.

Extraction dynamique et la méthode de purge et de piégeage (purge and trap, P&T)

Les techniques d'espace de tête dynamique ont été développées dans le but de surmonter certaines limitations de l'échantillonnage statique, en particulier l'incapacité à détecter des substances présentes dans des échantillons à de faibles niveaux de concentration ou des substances caractérisées par un coefficient de partage défavorable entre la matrice et la phase gazeuse²³⁶.

La méthode « purge and trap »²³⁷ est une technique d'espace de tête dynamique qui implique la purge de gaz inerte à travers un échantillon liquide ou solide, suivie du piégeage des analytes volatils sur un adsorbant puis d'une désorption thermique et d'une analyse en GC. Le purge and Trap réduit les effets de matrice et augmente la sensibilité. Cette méthode d'échantillonnage a été largement utilisée pour l'analyse des COV dans les aliments²²⁴ afin d'étudier les composés responsables de l'arôme²³⁸⁻²⁴², les effets des procédés de transformation²⁴³⁻²⁴⁷, l'impact des variétés²⁴⁸ ou des sources botaniques²⁴⁹ sur le profil aromatique, les mauvaises odeurs¹⁴⁷ ou encore la libération des COV par les emballages alimentaires²⁵⁰.

e. Conclusion

La connaissance de chaque méthode d'échantillonnage ainsi que leurs avantages et inconvénients permettent à l'analyste de choisir la méthode adéquate en fonction de l'objectif de son étude mais également des composés à extraire et de la matrice.

Cette thèse a pour but de développer l'analyse d'empreintes volatiles c'est-à-dire d'étudier l'ensemble des COV émis par un aliment pour authentifier et tracer une origine ou un procédé de transformation sans destruction de l'échantillon. Pour cela, nous devons trouver une technique d'échantillonnage qui ne détruit pas l'échantillon, qui soit facilement applicable, rapide et sans solvant. L'objectif ultime est de sélectionner une méthode automatisable pour une analyse en routine aussi bien en laboratoire que sur une ligne de production.

Parmi les diverses méthodes citées, les méthodes d'extraction par solvant ou encore l'extraction en phase solide (SPE) ne correspondent pas puisqu'elles utilisent des solvants, sont relativement longues à mettre en œuvre et ne préservent pas l'échantillon. De plus, le Soxhlet et l'ASE ne s'utilisent que pour des matrices solides alors que la LLE et la SPE ne s'utilisent que pour des matrices liquides. Le SBSE n'utilise pas de solvant mais le choix limité de revêtements, l'automatisation difficile, ne permettent pas de choisir cette technique.

Notre choix se porte donc sur le couplage de l'extraction d'espace de tête statique avec la micro-extraction en phase solide (HS-SPME). Cette méthode est simple à mettre en œuvre, rapide,

automatisable et permet d'extraire et préconcentrer les composés volatils de l'échantillon sans le dégrader. Le principe de la HS-SPME est décrit dans le paragraphe suivant.

3. Le couplage espace de tête et micro-extraction en phase solide : HS-SPME

L'association de deux méthodes d'extraction telle que la technique d'espace de tête et la micro-extraction sur phase solide permet d'augmenter la sensibilité de détection des analytes recherchés et ainsi d'abaisser les limites de quantification et de détection. Ce couplage consiste à extraire sur la fibre SPME les composés présents dans l'espace de tête d'un produit alimentaire (Figure 19). Cette technique repose donc sur deux équilibres thermodynamiques : l'équilibre matrice/espace de tête et l'équilibre espace de tête/fibre.

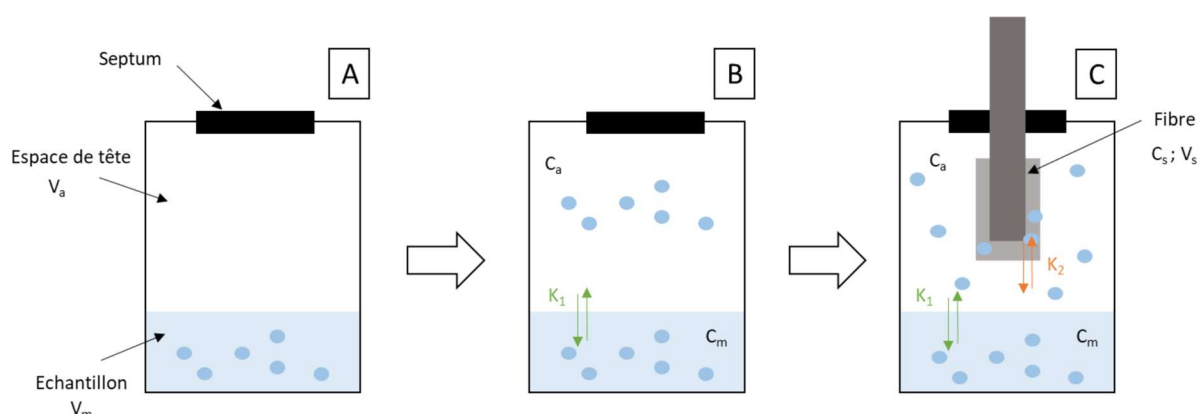


Figure 19 : Représentation schématique du principe de l'HS-SPME

a. Théorie et principe de l'HS-SPME

L'échantillonnage HS-SPME consiste dans un premier temps à placer l'échantillon dans un flacon hermétiquement clos et muni d'un septum (Figure 19-A). Ainsi commence la phase de transfert des composés dans l'espace de tête jusqu'à l'établissement d'un équilibre entre la concentration dans l'échantillon et la concentration dans la phase gazeuse (Figure 19-B). Cet équilibre matrice / espace de tête est régi par la constante de partage appelée K_1 (Équation 1) avec C_m concentration du composé dans la matrice et C_a , concentration dans l'espace de tête à l'équilibre.

$$K_1 = \frac{C_m}{C_a}$$

Équation 1

Dans un deuxième temps, la fibre SPME est introduite dans l'espace de tête du flacon (Figure 19-C). La SPME étant un échantillonneur passif, les COV diffusent alors de la phase gazeuse vers la surface de la fibre selon la 1^{ère} loi de Fick. En régime permanent, on obtient l'Équation 2, qui tient compte également des paramètres géométriques de la fibre²⁵¹ :

$$n = \frac{2\pi D_g L}{\ln(b + \frac{\delta}{b})} \int (C_a - C_s) dt$$

Équation 2

Avec n = quantité d'analyte collectée sur la surface de la fibre

t = temps de diffusion

D_g = coefficient de diffusion moléculaire de l'analyte

L = longueur de la fibre

b = rayon de la fibre

δ = épaisseur de la zone de diffusion

C_a = concentration de l'analyte dans l'air

C_s = concentration de l'analyte dans l'air à la surface de l'adsorbant

Au début du prélèvement, la cinétique d'adsorption est rapide par rapport à la cinétique de diffusion dans l'air. La concentration à la surface de l'adsorbant, C_s , est proche de zéro. On obtient alors l'Équation 3 suivante :

$$n = \frac{2\pi D_g L}{\ln(b + \frac{\delta}{b})} C_a t$$

Équation 3

Soit :

$$n = D C_a t$$

Équation 4

Avec $D = \frac{2\pi D_g L}{\ln(b + \frac{\delta}{b})}$

L'allure de la courbe $n = f(C_a \times t)$ est présentée Figure 20. La zone linéaire (zone 1) de pente D , correspond au domaine d'application de la 1^{ère} loi de Fick. Elle est généralement exploitée pour l'analyse quantitative des composés dans l'air dans des conditions de fonctionnement hors équilibre²⁵². Cependant, cela suppose que la concentration dans la phase gazeuse, C_a , reste constante durant le prélèvement. Cette hypothèse ne peut généralement pas être faite dans le cas d'un prélèvement en espace de tête en raison du faible volume gazeux (quelques mL) et donc de la faible quantité d'analyte en phase gazeuse.

La deuxième zone (zone 2), est une zone de transition où le débit d'échantillonnage diminue, c'est-à-dire que la quantité n qui s'adsorbe en fonction de la dose $C_a \times t$ diminue par rapport à la zone 1 en raison de la saturation progressive de l'adsorbant. La concentration à la surface de la fibre n'est plus négligeable. Ainsi, la quantification des composés à analyser selon la relation linéaire $n = f(C_a \times t)$ n'est plus possible.

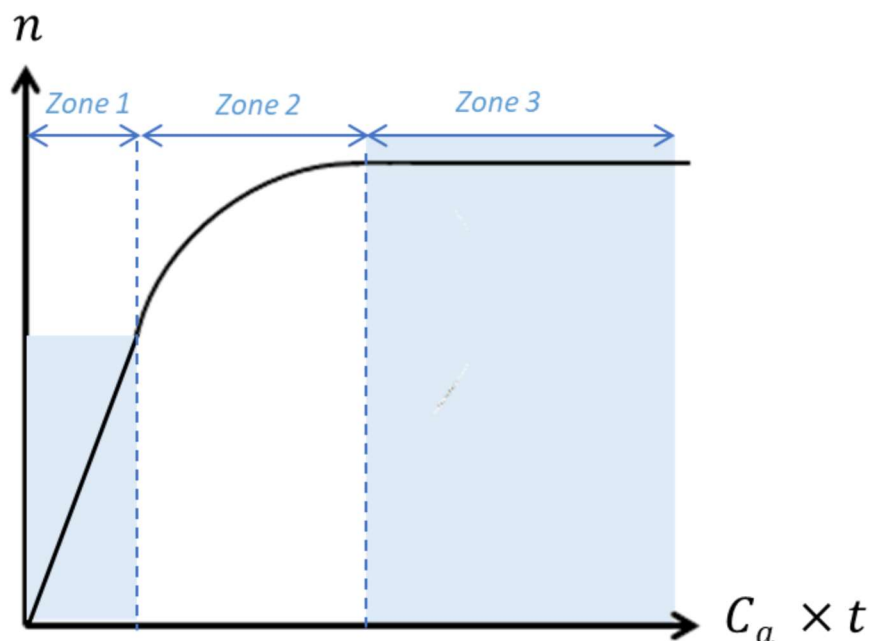


Figure 20 : Evolution de la quantité d'analyte n adsorbée en fonction du produit « $C_a \times t$ »

La dernière zone (zone 3) correspond à l'établissement d'un état d'équilibre entre la fibre et la phase gazeuse. L'extraction est alors optimale. La quantité adsorbée n est indépendante du temps et n'est plus fonction que de la concentration en phase gazeuse C_a et peut s'écrire selon l'Équation 5²⁵³:

$$n = \frac{K_2 V_s V_a C_a}{K_2 V_s + V_a}$$

Équation 5

Où K_2 est la constante de partage entre la fibre et l'air; V_s le volume de la fibre ; V_a le volume de la phase gazeuse; et C_a la concentration de l'analyte dans l'air²⁵⁴.

La quantification des COV dans l'air dans ces conditions d'équilibre est aussi possible²⁵⁵, mais souvent, comme précisé plus haut, une extraction hors équilibre cinétique (zone 1) est préférée pour limiter les compétitions d'adsorption sur la fibre.

Dans ces travaux de thèse, l'objectif n'est pas de chercher à quantifier précisément les composés présents dans la phase gazeuse au-dessus l'échantillon mais de comparer des profils volatils. Pour cela, on cherche à se placer dans les conditions permettant d'avoir un maximum de sensibilité afin de pouvoir détecter de manière la plus exhaustive possible tous les composés transférés en phase gazeuse depuis l'échantillon. Par conséquent, il a été choisi de se placer dans des conditions d'équilibre thermodynamique, matrice-espace de tête et gaz-fibre, permettant d'obtenir des résultats reproductibles avec un maximum de sensibilité. Cependant, en se plaçant dans ces conditions, des phénomènes de compétition d'adsorption sur la fibre peuvent se produire et conduire à ne pas détecter les COV ayant le moins d'affinité avec la fibre (généralement les composés les plus légers).

Une fois l'extraction sur la fibre réalisée, la fibre est rétractée dans son aiguille de protection pour être introduite dans l'injecteur du système d'analyse (GC) afin de désorber les composés à une température choisie en fonction des composés à désorber et de la nature de la fibre.

De nombreux paramètres peuvent agir sur les constantes de distribution (K_1 et K_2) comme la durée d'extraction, la température et l'agitation. En effet, ces paramètres peuvent favoriser le transfert vers la phase gazeuse des analytes. Dans le cas d'échantillons aqueux, des variations de pH et l'effet de sel peuvent également influencer l'équilibre matrice/espace de tête (K_1) et favoriser le transfert de certaines substances vers le ciel gazeux^{256,257}. Outre l'influence de la température et du temps d'extraction, la constante K_2 est impactée par la nature de la fibre qui sera discutée dans le prochain paragraphe.

b. Les différentes fibres

Différentes fibres sont disponibles avec des phases stationnaires de nature différente. On distingue deux grandes catégories : les fibres absorbantes et les fibres adsorbantes présentant chacune des modes de sorption différents^{209,258}.

Les fibres absorbantes sont principalement composées de deux matériaux : le polydiméthylsiloxane (PDMS) plus adapté au prélèvement des composés apolaires et le polyacrylate (PA), pour le prélèvement de composés polaires. Les fibres adsorbantes sont constituées de PDMS recouvert d'un solide poreux tel que le divinylbenzène (DVB), un polymère macroporeux, et un matériau carboné microporeux, le Carboxen (CAR). Le choix de la nature de la fibre s'effectue en fonction des composés à analyser (Tableau 3).

Fibre	Nature	Epaisseur de phase (μm)	Température de désorption ($^{\circ}\text{C}$)	Masse moléculaire des composés volatils conseillés pour ces fibres (g.mol^{-1})
PA	Absorbante	85	320	60-300 (composés semi-volatils polaires)
PDMS	Absorbante	100	220-280	60-275 (composés volatils)
	Absorbante	30	220-280	80-500 (composés semi-volatils apolaires)
	Absorbante	7	220-280	125-600 (composés apolaires lourds)
PDMS/DVB	Adsorbante	65	220-270	50-300 (composés volatils, amines, composés nitroaromatiques)
Carbowax (Cwax)/DVB	Adsorbante	65	220-265	40-275 (alcools, composés polaires)
PDMS/CAR	Adsorbante	75	240-320	30-255 (gaz, composés à faibles masse moléculaires)
PDMS/DBV/CAR	Adsorbante	50/30	230-270	40-275 (odeurs et saveurs)

Tableau 3 : Principales caractéristiques des phases absorbantes et adsorbantes des fibres SPME²⁵⁸

Dans certaines conditions, certains analytes ne peuvent pas être extraits ou détectés efficacement en raison de leurs propriétés physicochimiques ou de la composition de la matrice²⁵⁹. Ce qui est le cas,

par exemple, du formaldéhyde et d'autres composés carbonylés²⁶⁰. Dans ces cas, les composés doivent être dérivés pour améliorer leur extraction, leur désorption ou leur analyse chromatographique²⁵⁹. Il existe différentes méthodes de dérivation applicables à la SPME²⁰⁸ : (1) la dérivation directe consistant à ajouter l'agent de dérivation directement dans l'échantillon, (2) la dérivation sur la fibre SPME lors de l'extraction, (3) la dérivation après extraction (ou dérivation post-extraction) consistant à exposer la fibre SPME enrichie en composés à la solution d'agent de dérivation (4). La dérivation post-extraction a lieu dans le port d'injection d'un chromatographe en phase gazeuse par conversion thermique. Les techniques de dérivation sont principalement utilisées pour rechercher spécifiquement un composé ou une famille de composés difficilement détectables. En revanche, dans les analyses non ciblées, la dérivation est peu utilisée de par sa spécificité.

c. Paramètres à optimiser pour un échantillonnage par HS-SPME

De nombreux paramètres peuvent influencer les conditions de diffusion des composés de la matrice vers l'espace de tête puis de l'espace de tête vers la fibre. L'optimisation de ces paramètres est un étape clé dans le développement d'une méthode d'extraction par HS-SPME. On distingue les paramètres à optimiser pour l'espace de tête et ceux pour la fibre SPME :

- Les paramètres d'espace de tête à étudier sont : le volume d'échantillon, le volume du flacon, la température, le temps pour atteindre l'équilibre échantillon/phase gazeuse, l'agitation et l'effet de sel ou du pH (pour les matrices aqueuses et selon les composés)
- Les paramètres SPME sont : le temps d'extraction, la température, l'agitation, et l'un des plus importants, la nature de la fibre.

4. Application HS-SPME-GC-MS dans l'agroalimentaire

L'extraction par HS-SPME suivie d'une analyse par chromatographie gazeuse couplée à la spectrométrie de masse ou avec une détection par ionisation de flamme a été très largement utilisée pour l'analyse des COV dans les matrices alimentaires. Les travaux de recherche utilisant cette technologie sont regroupés en deux stratégies : l'analyse ciblée visant à détecter et quantifier spécifiquement un ou plusieurs COV et l'analyse non ciblée visant à étudier le profil volatil c'est-à-dire à détecter le maximum de composés. La principale différence entre les deux stratégies repose sur l'optimisation des conditions analytiques. Quelques exemples d'applications de ces stratégies sont présentés dans le Tableau 4 et discutés dans ce paragraphe.

a. L'analyse ciblée

L'analyse ciblée vise à suivre des analytes préalablement sélectionnés comme marqueurs. Pour cela, l'optimisation des paramètres analytiques est orientée et sélectionnée spécifiquement dans le but d'obtenir une détection sensible de ces analytes et de les quantifier. Les propriétés physico-chimiques des analytes recherchés et la complexité de la matrice sont les éléments principaux qui vont déterminer les paramètres d'échantillonnage.

Deux études ont été réalisées pour optimiser une méthode HS-SPME-GC-MS pour une détermination rapide du furane dans les aliments chauffés²⁶¹ et dans les aliments pour enfants²⁶². Différentes fibres ont été testées au cours de ces travaux et la fibre PDMS/CAR présente de meilleures performances. Après optimisation de cette méthode, Bianchi et al²⁶² ont déterminé une limite de détection (LOD =25,7 ng.kg⁻¹) et de quantification LOQ (41,7 ng.kg⁻¹) pour les échantillons de carotte, prouvant le potentiel de leur méthode pour déterminer le furane à des niveaux très bas.

Robertson et al²⁶³ ont évalué à l'aide d'une technique HS-SPME-GC-MS, la concentration correspondant au seuil de perception sensorielle de la géosmine, composé responsable de l'odeur et du goût terreux et moisi dans des produits aquacoles, et ont testé différents temps de lavage des poissons. Ils ont constaté que les temps de lavage nécessaires étaient directement liés à la concentration initiale en géosmine dans ces poissons et ont estimé le seuil sensoriel à 0,9 µgKg⁻¹. Jurado et al²⁶⁴ ont utilisé une méthode HS-SPME-GC-MS pour détecter et quantifier spécifiquement 18 composés permettant de différencier des échantillons de vins blancs d'appellations d'origine certifiées espagnoles. Chen et al²⁶⁵ ont développé une méthode HS-SPME-GC-MS pour mesurer le diacétyle dans les échantillons de beurre et émis dans l'air en optimisant les paramètres d'extraction HS-SPME. Pour la détection et la quantification du diacétyle dans les échantillons de beurre, ils ont choisi une fibre de polydiméthylsiloxane/divinyl benzène (PDMS/DVB) et ont déterminé une limite de détection de 0,0078 ppm. Panseri et al²⁶⁶ ont également étudié des échantillons de beurre et ont développé une méthode de détection et quantification de l'hexanal, un indicateur majeur de l'oxydation des lipides contribuant à une mauvaise odeur décrite comme « herbeuse ». L'extraction d'espace de tête a été réalisée en utilisant la fibre PDMS/CAR pendant 180 min à 4°C. Cette méthode a été appliquée pour suivre les phénomènes d'oxydation au cours du stockage. Lim et al²⁶⁷ ont développé une méthode HS-SPME-GC-MS en optimisant les paramètres (sélection de la fibre, température d'extraction, temps de chauffage et pH) pour détecter le 4-méthylimidazole (composé figurant dans la liste des cancérrogènes probables par l'état de Californie) dans les boissons gazeuses. Ils ont déterminé une limite de détection de 1,9 µg.L⁻¹ et une limite de quantification de 6,0 µg.L⁻¹. Abdulra'uf et al¹¹⁵ ont développé une méthode HS-SPME-GC-MS pour l'analyse des résidus de pesticides dans la poudre de cacao. Ils se sont principalement intéressés à l'optimisation des facteurs affectant la SPME, la fibre choisie est la PDMS. En utilisant les conditions optimisées, la linéarité s'étendait de 2,5 à 500 mg.kg⁻¹ grâce à un étalon interne et une gamme d'étalonnage.

Comme ces travaux le démontrent, l'analyse ciblée par HS-SPME-GC-MS permet de répondre à de nombreuses questions sur l'origine et la qualité de matrices agroalimentaires. Cependant, cela suppose de connaître et de choisir les molécules d'intérêt pour lesquelles la méthode doit être optimisée. Cette pré-orientation des molécules à analyser peut dans certains cas limiter les discriminations entre échantillons. Une approche non ciblée permet de visualiser la globalité des COV émis par le produit et de révéler de nouvelles informations en détectant autant de composés que possible en une seule étape.

b. Analyse non ciblée

Les analyses non ciblées partent du postulat que les molécules responsables du caractère étudié ne sont pas connues et donc considèrent l'ensemble des composés présents dans la matrice. Deux approches sont alors retrouvées dans les analyses non ciblées en fonction du traitement des données obtenues en HS-SPME-GC-MS avec une acquisition en mode full scan.

La première approche consiste dans un premier temps à identifier les composés clés par l'analyse puis à les quantifier. La méthode HS-SPME-GC-MS est alors généralement adaptée pour permettre l'extraction, la séparation et l'identification d'un maximum de familles de molécules chimiques. Les identifications sont réalisées par comparaison aux banques de données MS internes ou externes en s'appuyant sur un certain pourcentage de correspondance (match). Cette étape rejette donc les composés dont l'identification n'est pas certaine ou en dessous du pourcentage fixé, mais qui pourraient être importants pour répondre à la problématique posée. La seconde étape de quantification des composés sélectionnés consiste à retraiter les chromatogrammes en extrayant les signaux correspondants et dans certains cas à les convertir en concentrations. Cette approche est considérée comme non spécifique car, au départ, aucun composé cible n'est connu mais le retraitement effectué limite l'analyse aux composés sélectionnés. En revanche, cette approche reste très couramment utilisée pour l'analyse des COV de divers produits agroalimentaires notamment pour étudier les composés aromatiques volatils, l'impact du cultivar ou de la variété^{149,268-272}, des réactions (l'auto-oxydation, par exemple)²⁷³, l'impact des procédés de transformation²⁷⁴⁻²⁷⁷, la discrimination ou classification de produit^{145,147,148,278,279}, la recherche de mauvaises odeurs²⁸⁰ mais également pour l'impact des emballages^{281,282}.

La deuxième approche est basée sur la comparaison de profils de COV en traitant l'intégralité du signal comme une empreinte et en tenant compte des deux dimensions des données obtenues : celle correspondant aux chromatogrammes et celle correspondant aux spectres de masse. L'exploitation des données issues de ces analyses nécessite cependant l'utilisation d'outils de traitement spécifiques (§ III.B.) pour rechercher des motifs communs ou discriminants entre les empreintes des échantillons. Comme pour l'approche précédente, les paramètres d'extraction SPME et de séparation chromatographique seront choisis pour permettre la détection d'un maximum de molécules, de nature et de concentrations très différentes. Contrairement à la première approche d'analyse non ciblée, celle-ci ne va pas forcément jusqu'à la quantification précise des marqueurs discriminants mais s'arrête généralement à leur identification grâce à leurs spectres de masse et à la comparaison des intensités des signaux au sein des échantillons analysés. Cette approche est relativement moins utilisée car elle nécessite de nombreux développements pour faciliter son utilisation. Cependant, des études ont déjà démontré l'intérêt des empreintes volatiles obtenues en GC-MS pour des applications agroalimentaires afin d'évaluer des procédés de fermentation^{136,283,284}, de différencier des origines géographiques ou des variétés de cidres, de vins ou de riz^{141,285,286}, d'étudier l'impact de la matière première sur la gelée royale²⁸⁷ ou encore de rechercher des composés marqueurs de contamination de jus de fruits²⁸⁸.

5. Conclusion : forces et limites de l'HS-SPME-GC-MS

L'HS-SPME-GC-MS/FID est une technique facile à mettre en œuvre, peu coûteuse, sensible, reproductible, sélective (en fonction du type de fibre utilisé), facilement automatisable, tout en réduisant les interférences avec les composés non volatils des matrices. Cette méthode a cependant certaines limites. Une mise au point minutieuse des conditions opératoires est nécessaire selon l'objectif de l'analyse et pour des analyses en routine : choix de la fibre la plus appropriée, détermination du volume d'échantillon, de la température, du temps d'extraction et de désorption ; choix d'un étalon interne et/ou calibrations externes dans le cas d'analyses quantitatives. Le choix de la fibre est l'étape clé. Il est nécessaire d'identifier la nature de la fibre à plus fort potentiel c'est à dire celle susceptible de constituer un bon compromis entre les contraintes de mise en œuvre et les objectifs d'identification et de quantification des composés cibles dans la matrice. La sélection opérée

entre les différentes fibres repose sur plusieurs critères comme la nature de l'échantillon, le mode d'extraction (absorption ou adsorption), l'influence de l'humidité sur l'extraction, le temps d'extraction et la préconcentration directement liée à l'affinité des analytes pour la phase. La chromatographie en phase gazeuse permet de séparer les composés extraits pour les identifier et les quantifier. Cependant, elle apporte également quelques contraintes comme la sélectivité de la colonne qui va influencer la nature des composés qui seront séparés mais également la rampe de température qui doit permettre la séparation et l'analyse de tous les composés extraits. Un des atouts majeurs de cette méthode est le couplage avec la spectrométrie de masse qui permet d'identifier les composés présents.

En revanche, les méthodes HS-SPME-GC-MS sont généralement longues (30 minutes à 1 heure) ce qui peut parfois compliquer leur utilisation comme point de contrôle rapide tout au long de la chaîne de production agroalimentaire, qui est le défi actuel pour les acteurs de ce domaine. L'arrivée des nouvelles techniques de spectrométrie de masse à injection directe (DIMS) permettent d'analyser les constituants volatils d'une matrice gazeuse sans étapes de pré-concentration ni de séparation préalable. Parmi ces DIMS, la technologie SIFT-MS (Selected Ion Flow Tube Mass Spectrometry) sera présentée dans le paragraphe suivant.

Tableau 4 : Exemple d'applications de l'HS-SPME dans l'analyse des composés volatils de produits agroalimentaires

Matrice	Objectif	Analyse ciblée ou non ciblée	Fibre SPME	Colonne GC	Détection	Analytes cibles ou identifiés	LOD/LOQ	Ref	Années
Analyse ciblée									
Truite arc-en-ciel	Taux de dépuración	Ciblée	DVB/PDMS	ZB-5 : 5% Phenyl- 95% Diméthylpolysiloxane	MS	Géosmine		263	2005
Vin blanc espagnol	Différencier des marques de vin	Ciblée	CAR/PDMS	ZB-5ms : 5 % phényl-arylène, 95 % diméthylpolysiloxane SBL-5 : 5 % phényl-arylène, 95 % diméthylpolysiloxane	MS FID	2-méthyl-1-propanol, 2-phényléthyléthanol, 3-méthyl-1-butanol, 2- méthyl-1-butanol, acétate d'éthyle, acétate d'isoamyle, l'hexanoate d'éthyle, l'acétate d'hexyle, le sorbate d'éthyle, l'octanoate d'éthyle, le 2-phénéthyl octanoate d'éthyle, acétate de 2-phénéthyle, décanoate d'éthyle, octanoate d'isoamyle, 2-décénoate d'éthyle, laurate d'éthyle, myristate d'éthyle, et palmitate d'éthyle		264	2008
Beurre	Détection du diacétyl	Ciblée	PDMS/DVB	SPB-1 SULFUR : Poly(diméthyl siloxane)	MS-MS	Diacétyl	0,0078 ppm	265	2010
Beurre	Oxydation lipidique	Ciblée	CAR/PDMS	Rtx-Wax : polyethylene glycol (PEG)	MS	Hexanal	LOD : 1,02 ng.g ⁻¹ LOQ : 3,40 ng.g ⁻¹	266	2011
Boissons gazeuses (280)	Contaminant	Ciblée	DVB/PDMS	HP-InnoWax : polyethylene glycol (PEG)	MS	4-méthylimidazole	1,9 µg/L	267	2013
Fruits et légumes (287)	pesticides	Ciblée	PDMS	DB-5MS : 5 % phényl-arylène, 95 % diméthylpolysiloxane	MS	7 pesticides	0,35–8,33 µg/kg	115	2014
Caviar	Etudes des différents stades de maturation	Ciblée	DVB/CAR/PDMS	DB-5MS : 5 % phényl-arylène, 95 % diméthylpolysiloxane	MS	Nonanal, oct-1-en-3-ol, nonanoic acid, α-pinène, heptan-2-one, éthyl décanoate	LOD : 0,17 à 1,92 ng	289	2020
Analyse non ciblée avec identification et quantification des COV									
Saucisse	Profil aromatique	Non ciblée*	CAR/PDMS DVB/CAR/PDMS	DB-624 : 6% cyanopropyl/phenyl, 94% polydiméthylsiloxane,	MS	10 alcools, 17 aldéhydes, 7 cétones, 11 esters, 9 hydrocarbures, 9 acides, 2 furans, 1 composés soufrés 11 alcools, 14 aldéhydes, 9 cétones, 17 esters, 15 hydrocarbures, 11 acides, 2 composés soufrés, 2 furans		290	2004

Fromage	Profil aromatique	Non ciblée*	DVB/CAR/PDMS	DB-WAX : polyethylene glycol (PEG)	qMS EI-FID	2 Aldéhydes, 2 alcools, 2 acides, 2 esters, 2 cétones		291	2005
Jus d'orange (256)	Mauvaise odeur Espèces microbiennes	Non ciblée*	DVB/CAR/PDMS	DB-5 : 5% phényl, 95% méthylpolysiloxane	MS	Guaiacol, 2,6-Dichlorophenol, 2,6-Dibromophenol		280	2005
Abricots	Cultivar	Non ciblée*	DVB/PDMS	DB-WAX : polyethylene glycol (PEG)	MS O	acétate d'éthyle, acétate d'hexyle, limonène, b-cyclocitral, c-décalactone, 6-méthyl-5-hepten-2-one, linalol, b-ionone, menthone et (E)-hexen-2-al		268	2006
Huile de colza (259)	Auto-oxydation	Non ciblée*	CAR/DVB/PDMS	DB-5MS : 5 % phényl-arylène, 95 % diméthylpolysiloxane	MS	hexanal, 2,4-heptadiénal, 2-hepténal et 1-pentène-3-ol		273	2007
Légumes frais emballés (288)	Emballage, Détection COV produit par <i>Salmonella typhimurium</i>	Non ciblée*	CAR/PDMS	SPB-5 : poly(5% diphenyl/95% diméthylsiloxane)	MS	heptane, diméthyle sulfure, acide acétique, acétate d'éthyle, alcool méthylique, éthanol, 2-heptanol, 1-propanol, diméthyle disulfure, éthylbenzène, 1-pentanol, 3-octanone, 3-octanol et 1-heptène-3-ol.		281	2007
Bœuf	Emballage, vieillissement Détection COV associée à <i>Salmonella</i>	Non ciblée*		XTi-5 : 95% diméthyl/ 5% diphenyl polysiloxane	MS	2-méthylpropane, 2-méthylbutane, 2,3-butanedione, 2-butanone, acide acétique, 3-hydroxy-2-butanone, 2-oxopropanal, 3-méthyl-1-butanol, 2-méthyl-1-butanol, 3,7-diméthylnonane		282	2011
Raisins	Cultivars et procédé de transformation	Non ciblée*	DVB/PDMS	DB-35MS : (35%-phenyl)-méthylpolysiloxane	MS	18 esters, 15 composés carbonylés, 9 terpènes, 5 composés en C6		274	2011
Amande	Cultivar	Non ciblée*	CAR/PDMS	SPB-5 : poly(5% diphenyl/95% diméthylsiloxane)	MS	9 aldéhydes, 8 alcanes, 2 alcools, 2 acides, 1 hydrocarbure		149	2011
Huile d'amande	Authenticité du cultivar	Non ciblée*	DVB/CAR/PDMS	SPB-5 : poly(5% diphenyl/95% diméthylsiloxane)	MS	hexanal, (E)-2-heptenal, (E)-2-octenal, nonanal, (E)-2-nonenal, (E,E)-2,4-nonadienal, (E,E)-2,4-decadienal		269	2011
Lait, fromage et poudre de lactosérum	analyse des composés aromatiques	Non ciblée*	DVB/CAR/PDMS	DB-5 : 5% phényl, 95% méthylpolysiloxane	MS	5 aldéhydes, 6 acides, 3 cétones, 1 alcools, 1 terpènes, 1 composés soufrés, 1 esters		292	2013
Bœuf cuit	Profil aromatique	Non ciblée*	DVB/CAR/PDMS	VF-5 et ZB-5 : 5% Phényl-95% Diméthylpolysiloxane	MS FID	11 aldéhydes, 4 composés soufrés, 3 alcools, 2 cétones, 2 furanes, 1 terpène, 1 alcane		293	2013
Thé	Profil aromatique	Non ciblée*	DVB/CAR/PDMS	Rtx-5 : 95% diméthyl/ 5% diphenyl polysiloxane	MS	6 alcools aliphatiques, 4 alcools aromatiques, 6 aldéhydes aliphatiques, 2 aldéhydes aromatique, 6 cétones, 3 furans, 1 pyridine, 1 pyrazine, 1 furanone, 3 acides, 1 ester, 5 autres composés aromatiques		294	2013

Thé	Cultivar	Non ciblée*	PDMS/DVB	HP-InnoWax : polyethylene glycol (PEG)	MS	6 terpènes, 2 alcools terpéniques, 3 alcanes, 2 furans, 2 esters, 2 aldéhydes, 1 cétone, 1 alcool	270	2013
Vins de cerise	Discrimination des vins de cerise	Non ciblée*	DVB-CAR-PDMS	HP-InnoWax : polyethylene glycol (PEG)	MS	29 esters, 22 alcools, 8 acides, 3 ketones, 5 aldéhydes, and 8 miscellaneous compounds	295	2014
Boisson de café	COV émis d'un champignon	Non ciblée*	DVB/CAR	SUPELWAX : polyethylene glycol (PEG)	MS	9 alcools, 6 aldéhydes, 4 cétones, 5 autres composés	296	2014
Bœuf haché	Altération	Non ciblée*	DVB/CAR/PDMS	HP-5MS : 5 % phényl-arylène, 95 % diméthylpolysiloxane	MS	10 alcools, 10 aldéhydes, 8 cétones, 5 esters	297	2015
Liqueur chinoise	COV	Non ciblée*	DVB/CAR/PDMS - CAR/PDMS - DVB /PDMS - PA	CP-WAX : polyethylene glycol (PEG)	MS	33 esters, 10 alcools, 8 composés aromatiques et phénoliques, 6 hydrocarbures, 4 acides, 3 aldéhydes, 3 cétones, 1 pyrazine	298	2015
Fruit du Monstera deliciosa	Suivi de procédé de transformation	Non ciblée*	DVB/PDMS	BP-20 : polyethylene glycol (PEG)	qMS	49 esters, 12 composés carbonylés, 8 terpènes, 7 alcools, 4 lactones	299	2015
Cultivars de cerises douces	Cultivars	Non ciblée*	CAR/PDMS	DB-5MS : 5 % phényl-arylène, 95 % diméthylpolysiloxane	MS	8 alcools, 7 composés carbonylés, 3 autres composés	271	2015
Riz	Cultivar et vieillissement	Non ciblée*	DVB/CAR/PDMS	SE52 : 95% polydimethylsiloxane, 5% phenyl	MS	12 aldéhydes, 3 cétones, 7 alcools, 2 esters, 2 composés hétérocycliques	300	2015
Fromages Van Herby	Procédé de transformation Matière première	Non ciblée*	DVB/CAR/PDMS	HP-InnoWax : polyethylene glycol (PEG)	MS-FID	8 esters, 5 cétones, 5 aldéhydes, 9 acides, 6 alcools, 14 hydrocarbures et terpènes	276	2015
Oranges	traçabilité des oranges fraîches	Non ciblée*	CAR/PDMS/DVB	DB-5MS : 5 % phényl-arylène, 95 % diméthylpolysiloxane	MS	6 aldéhydes, 1 cétone, 4 alcools, 23 terpénoïdes, 10 esters, 1 furan	301	2016
Citri Reticulatae Pericarpium	Discrimination entre les variétés de Citri Reticulatae Pericarpium	Non ciblée*	DVB/CAR/PDMS	TG-5SILMS : 5 % diphényl / 95 % diméthyl polysiloxane	MS	Principalement des terpènes et des sesquiterpènes	302	2018
Miel	Discrimination de l'origine du miel	Non ciblée*	DVB/CAR/PDMS	5% phényl, 95% méthylpolysiloxane	MS	7 alcools, 14 aldéhydes, 12 cétones, 3 acides, 3 pyrazines, 1 éther, 1 ester, and 1 terpène	LOQ of the 8 identified markers: 0.5–320 2.5 µg/kg 147	2019
Produit à base de soja	Etude de la maturation	Non ciblée*	DVB/CAR/PDMS	DB-5 : 5% phényl, 95% méthylpolysiloxane	MS	Aldehydes, alcools, acides, esters	303	2019
Fève de cacao	Classification	Non ciblée*	DVB/CAR/PDMS	DB-WAXETR : polyethylene glycol (PEG)	qMS E-nose	Aldéhydes, cétones, composés soufrés, esters, hydrocarbures, dérivés de furan,	148	2019

						pyrazines, alcools, pyrroles, terpènes, acides, lactones			
Thé	Influence de la température de l'eau lors de l'infusion	Non ciblée*	DVB/CAR/PDMS	HP-5 : 5% phényl, 95% méthylpolysiloxane	MS	20 Alcools, 19 alcanes, 14 aldéhydes, 9 cétones, 9 esters, 7 alcènes, 2 phénols, 1 acide, 1 lactone, 1 naphthalène.	304	2019	
Huile de coco	Discrimination selon le profil volatile	Non ciblée*	PDMS/DVB	DB-WAX : polyethylene glycol (PEG)	MS	3 aldéhydes, 5 cétones, 5 alcools, 6 acides, and 9 esters	279	2020	
Fromages	Discrimination des fromages Pecorino typiques d'Italie	Non ciblée*	PDMS CAR/PDMS DVB/CAR/PDMS	VF-WAX : polyethylene glycol (PEG)	MS	7 acides, 9 composés carbonylés, 8 esters, 9 alcools, 3 éthers, 2 terpènes	145	2021	
Champignons	Comparaison des profils volatils	Non ciblée*	DVB/CAR/PDMS	DB-5 : 5% phényl, 95% méthylpolysiloxane	MS E-nose	Aldéhydes, cétones, esters, hydrocarbures, terpènes, composés aromatiques, phénols, alcools, acides, alcanes	305	2022	
Analyse non ciblée / Profil volatil									
Vin	Vérifier l'origine botanique	Non ciblée	PDMS	J&W DB 1701 : (14%-Cyanopropyl-phenyl)-methylpolysiloxane	MS	/	285	2014	
Moût de raisin fermenté	Suivi de fermentation	Non ciblée	DVB/CAR	J&W FFAP : Polyéthylène glycol modifié par de l'acide nitrotéréphtalique	MS	méthionol, 2-phényléthanol, acétate d'isoamyle, benzaldéhyde, 1-hexanol, acide isobutyrique, hexanoïque, propanol, 2-phényléthanol, alcool benzylique, acétate d'isoamyle, alcool isobutyrique, butanoate d'éthyle et octanoate d'éthyle	283	2014	
Grappes de raisin	Différents pathogènes, pourriture fongique, nuisibles COV	Non ciblée	PDMS/DVB/CAR	DB-WAXETR : polyethylene glycol (PEG)	MS	1,5-diméthylnaphtalène, benzaldéhyde, sesquiterpènes inconnus, acide 2-(4-hexyl-2,5-dioxo-2,5dihydrofuran-3-yl)acétique, m-crésol, 2-pentylfuran et en γ -nonalactone	288	2019	
Cidre	Différenciation géographique	Non Ciblée	DVB/CAR/PDMS	BP-20 : polyethylene glycol (PEG)	MS	41 esters, 21 alcools, 24 terpénoïdes, 13 acides, 3 composés carbonylés, 4 lactones and 1 phénol	141	2019	
Vin (Merlot Rouge)	Procédé de vinification (maturation et fermentation)	Non ciblée	DVB/CAR/PDMS	ZB-WAX : polyethylene glycol (PEG)	MS	acides phénoliques, esters, alcools et terpènes	284	2020	
Liuyang Douchi	Fermentation	Non ciblée	DVB/CAR/PDMS	CD-WAX : polyethylene glycol (PEG)	MS	hexanol, hexanal, acide propanoïque, 1-octen-3-ol, 3-octanone, acétate d'éthyle et 2,6-diméthylpyrazine	136	2021	

Gelée royale	Matière première de production	Non ciblée	DVB/CAR/PDMS	DB-5MS : 5 % phényl-arylène, 95 % diméthylpolysiloxane	MS	16 alcanes, 9 alcools, 9 cétones, 8 esters and lactones, 5 phénols, 2 acides, 3 autres composés		287	2021
Riz	Variété Origine géographique	Non ciblée	DVB/CAR/PDMS	DB-WAX : polyethylene glycol (PEG)	MS	1-hexanol, 2-phényléthanol, alcool benzylique, 2-tridécanone, indole, benzoate de méthyle, hexanal, palustrol, 2-butyl-2-octénal, 3,5-octadiène-2-one, 1,3,3-triméthyl-indoline-2-one, 2-méthylpropanoate de 2-éthyl-3-hydroxyhexyle, α-méthoxyphénylacétate de méthyle, 1,7-diméthyl-naphtalène, 2-tert-butyl-phénol et 2-méthylpropyl benzoate.		286	2022

B. Nouvelle technologie de spectrométrie de masse à injection directe : SIFT-MS (Selected Ion Flow Tube Mass Spectrometer)

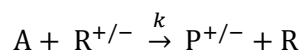
Des techniques de spectrométrie de masse à injection directe (Direct Injection Mass Spectrometry DIMS)³⁰⁶⁻³¹¹ telles que le spectromètre de masse à ionisation chimique à pression atmosphérique (APCI-MS)³¹², l'analyse directe en spectrométrie de masse en temps réel (DART-MS)³¹³, le spectromètre de masse à ionisation par électronébuliseur (ESI-MS)³¹⁴, le spectromètre de masse à réaction de transfert de protons (PTR-MS)³¹⁵ ou la spectrométrie de masse par flux d'ions sélectionnés (SIFT-MS)³¹⁶ permettent d'analyser les constituants volatils d'une matrice gazeuse sans séparation préalable. Ces technologies basent leur sélectivité sur des réactions d'ionisations chimiques douces entre un ou plusieurs ions et les COV présents dans l'air, conduisant à la formation d'ions produits, détectés et quantifiés par un détecteur de masse quadripolaire ou à temps de vol (Time of Flight, TOF). Ces technologies peuvent être utilisées pour la détection et la quantification ciblées de marqueurs³¹⁷ avec le mode Single Ion Monitoring (SIM) ou Multi Ion Monitoring (MIM) ou bien pour l'analyse non ciblée avec un mode de détection full scan, qui correspond à l'ensemble des signaux de tous les ions produits en fonction de leur rapport masse sur charge (m/z). En fonction du nombre d'ions précurseurs utilisés et du nombre de COV présents dans l'échantillon, ces technologies génèrent une grande quantité de signaux qui peuvent être parfois difficiles à interpréter. La méthode d'échantillonnage principalement utilisé dans l'application de ces technologies au secteur agroalimentaire est la technique de l'espace de tête statique (Tableau 11).

1. Présentation SIFT-MS

La technique du tube à flux d'ions sélectionnés a été mise au point par Adams et Smith¹⁰. Le principe du SIFT-MS, inspiré des travaux de Fergusson et al³²⁰ (Figure 21) est de faire réagir séquentiellement dans un tube d'écoulement des ions précurseurs produits par un plasma, avec les molécules volatiles de la matrice gazeuse et de détecter les ions produits par cette réaction³¹⁸. Grâce à la connaissance de la constante cinétique de la réaction entre l'ion précurseur et l'analyte, il est possible de calculer la concentration de l'analyte. Par la suite, de nombreux travaux ont été réalisés sur un instrument de type « SIFT-MS profile 3 » présentant la particularité de pouvoir changer de gaz d'appoint (par exemple : hélium ou azote) et de pouvoir injecter l'échantillon gazeux à différents niveaux du tube d'écoulement³²¹⁻³²⁷. Dans le cadre de cette thèse, un SIFT-MS Voice 200 Ultra, développé et commercialisé par la société SYFT (SYFT Technologies, Christchurch, New Zealand) a été utilisé. Ce modèle d'instrument a la particularité de disposer d'une source plasma « dual polarity » capable de produire à la fois des ions positifs et des ions négatifs.

Le SIFT-MS Voice 200 Ultra (Figure 21) possède une source plasma hyperfréquence qui génère les ions précurseurs à partir d'air et d'eau. Les espèces chargées produites par cette source sont sélectionnées par un filtre de masse quadripolaire encadré par deux réseaux de lentilles pour ne conserver qu'une seule espèce ionique d'un rapport masse/charge donné : l'« ion précurseur » R. Les ions précurseurs positifs (H_3O^+ , O_2^{*+} et NO^+) et négatifs (O^* , OH^- , O_2^{*-} , NO_2^- et NO_3^-) ainsi sélectionnés sont injectés séquentiellement dans le tube d'écoulement via un orifice Venturi où ils se déplacent par convection le long du tube de réaction. Parallèlement, les molécules de la matrice gazeuse sont introduites dans le tube d'écoulement par simple aspiration à un débit fixe d'environ 20 mL min^{-1} , via une ligne d'échantillonnage chauffée (High Precision Inlet, HPI). Enfin un gaz d'appoint (azote) vient compléter ce mélange dans le tube d'écoulement afin d'en stabiliser les conditions de réaction. Chaque analyte

A présent dans la matrice gazeuse rencontre les ions précurseurs dans le tube d'écoulement et, en fonction de ses propriétés chimiques, peut réagir pour former un ou plusieurs ions produits P (Équation 6).



Équation 6

Les « ions produits » nouvellement formés passent au travers d'un second orifice et d'un guide d'ion avant d'être dirigé vers un second quadripôle ou ils sont triés en fonction de leur rapport masse sur charge (m/z). Les ions produits triés traversent alors un réseau de lentilles pour atteindre le détecteur à multiplicateur d'électrons. Le SIFT-MS possède donc une résolution unitaire en masse.

Deux modes d'acquisition des données sont possibles : le mode full scan permettant d'obtenir le signal de chaque ion (m/z) produit et le mode SIM (Single Ion Monitoring) permettant l'acquisition du signal d'un seul ion.

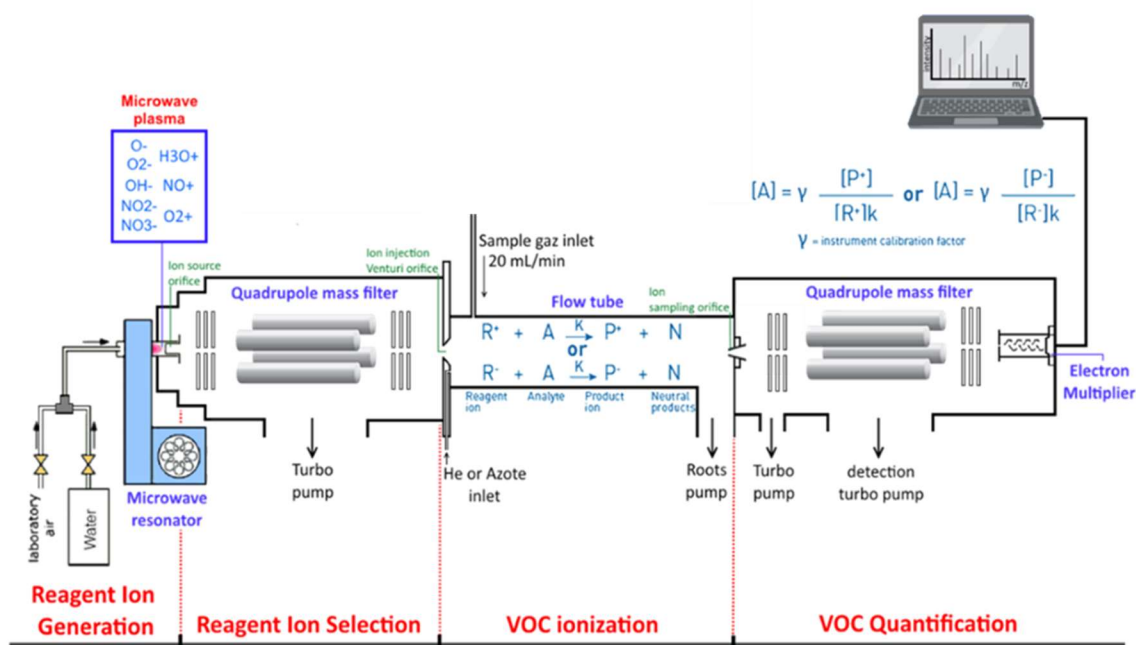


Figure 21 : Principe du SIFT-MS

2. Principe d'Ionisation et de réactivité chimique

L'ionisation chimique est la réaction entre une espèce ionique réactive (R⁺ ou R⁻) produite par un plasma et une molécule A impliquant des mécanismes différents dont le trait commun est de produire des cations ou anions en phase gazeuse. L'ionisation chimique est dite positive dans le cas d'espèces ioniques réactives positives (R⁺) produisant principalement des ions AH⁺ et négative dans le cas d'espèces ioniques réactives négatives (R⁻) produisant principalement des ions A⁻.

En opposition aux spectromètres de masse à impact électronique, l'ionisation chimique est dite douce car elle produit des ions par un processus chimique qui limite la fragmentation, favorise l'apparition de l'ion moléculaire et le préserve. La stabilité des ions va jouer un rôle crucial sur leur réactivité et donc leur capacité à se fragmenter.

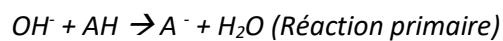
Les principales réactions retrouvées en ionisation chimique avec les 8 huit ions réactifs par SIFT-MS (Tableau 5) sont : le transfert de proton^{328,329} (Équation 7, Équation 8), le transfert d'électron (Équation 9), le transfert dissociatif de charge³²⁶ (Équation 10), l'abstraction d'hydrure³³⁰(~~Équation 11~~~~Équation 11~~), l'association(Équation 12), le détachement (Équation 13), l'élimination (Équation 14), le transfert d'hydrogène (Équation 15) et le détachement électronique associatif^{331,332}(Équation 16).

Process	Reaction	H ₃ O ⁺	O ₂ ^{•+}	NO ⁺	OH ⁻	O ₂ ^{•-}	O ^{•-}	NO ₂ ⁻	NO ₃ ⁻
Protonation	$RH^+ + A \rightarrow AH^+ + R$ Équation 7	✓							
Déprotonation	$R^- + AH \rightarrow A^- + RH$ Équation 8				✓	✓	✓	✓	✓
Transfert d'électron	$R^+ + A \rightarrow A^+ + R$ $R^- + A \rightarrow A^- + R$ Équation 9		✓	✓		✓			
Transfert dissociatif de charge/électron	$R^+ + A \rightarrow F_1^+ + F_2 + R$ Équation 10		✓	✓					
Abstraction d'hydrure	$R^+ + AH \rightarrow A^+ + HR$ Équation 11			✓					
Association	$R^{+/-} + A + M \rightarrow A.R^{+/-} + M$ Équation 12	✓		✓	✓	✓	✓		
Déplacement	$R^- + XA \rightarrow XR + A^-$ Équation 13				✓		✓		
Élimination	$R^- + XA \rightarrow X_1 + X_2R + A^-$ Équation 14				✓		✓		
Transfert d'hydrogène*	$R^- + AH \rightarrow A + RH^-$ Équation 15						✓		
Détachement électronique associatif*	$R^- + A \rightarrow RA + e^-$ Équation 16				✓	✓	✓		

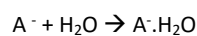
Tableau 5 : Réaction d'ionisation chimique possible pour les 8 ions précurseurs.

R = ions réactifs ; A = Analyte ; F = Fragment ; M = molécule neutre du gaz porteur (He ou N₂). * ce mécanisme de réaction peut se produire en SIFT-MS mais aucun ion issu de l'Analyte n'étant formé, la réaction ne peut pas être détectée.

Les ions produits issus de ces réactions sont de deux types fondamentalement différents : les ions primaires et les ions secondaires. Les ions primaires proviennent de la réaction initiale entre l'ion réactif et l'analyte (Équation 17). Les ions secondaires quant à eux, proviennent de la réaction entre les ions issus de la réaction précédente et d'autres composés présents en forte concentration dans l'échantillon comme l'eau (très fréquent) ou d'autres analytes présents (Équation 18). L'importance des produits secondaires dépendra donc de la matrice étudiée, de l'humidité et de la concentration en analytes. Pour confirmer la formation d'ions secondaires, une gamme de concentrations de l'analyte est généralement réalisée dans différentes conditions d'humidité. Ainsi, si le signal de l'analyte diminue avec l'humidité relative et qu'un ou plusieurs autres signaux augmentent avec l'humidité, nous avons confirmation de la formation d'ions secondaires. Il conviendra alors de les prendre en compte pour la quantification exacte de l'analyte (§ II.B.4).

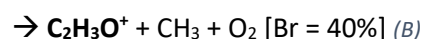
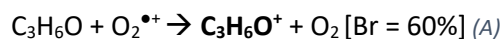


Équation 17 : exemple de réaction primaire



Équation 18 : exemple de réaction secondaire de l'ion produit A^- avec l'eau pour former le complexe $A^- \cdot H_2O$

Une réaction entre un ion et un analyte donné peut former plusieurs ions produits. Par exemple, considérant la réaction entre O_2^{*+} et l'acétone (C_3H_6O), deux produits sont formés : l'un par transfert de charge (Équation 19-A) et l'autre par transfert de charge dissociatif (Équation 19-B). Le rapport entre les deux produits est appelé : « Branching ratio » (Br). Ce branching ratio est le rapport entre la constante de vitesse d'un produit d'une réaction et la constante de vitesse de l'ensemble des réactions possibles. Il s'agit d'un paramètre important qui indique la probabilité qu'une réaction se produise mais également si les conditions du tube de réaction (température, pression et potentiel électrique) sont favorables ou non à la réaction. Lorsque l'on cherche à quantifier une molécule inconnue, il est important de connaître les branching ratio des réactions qui ont lieu avec l'ensemble des ions précurseurs afin de sélectionner pour la quantification, la réaction la plus favorable.



Équation 19

3. Influence de l'eau, de l'oxygène et du dioxyde de carbone

La technologie SIFT-MS a été principalement développée pour l'analyse des COV dans l'air. En raison de sa fréquence d'analyse élevée et de sa simplicité de mise en œuvre, l'analyse SIFT-MS est également très appliquée dans les domaines de la biologie³³³⁻³³⁵, de la santé³³⁶⁻³³⁹, de l'archéologie³⁴⁰, de l'agroalimentaire^{97,324,341} et de l'environnement³⁴²⁻³⁴⁴. Ces différentes matrices pouvant être riches en eau, en oxygène ou en dioxyde de carbone, il est donc important de connaître l'influence de ces composés sur la réaction ion précurseur-analyte dans le tube de réaction.

Dans plusieurs travaux³⁴⁵⁻³⁴⁸, l'influence de l'eau, de l'oxygène et du dioxyde de carbone, présents l'air sur la réaction entre des analytes et des ions précurseurs a été étudiée. Le Tableau 6 reprend les ions produits par réaction de l'eau, de l'oxygène et du dioxyde de carbone avec les huit ions précurseurs

positifs et négatifs dans le tube de réaction. Les constantes de réaction expérimentales de ces ions avec l'eau, l'oxygène et le dioxyde de carbone ont été déterminées par Ghislain et al.³⁴⁶ et sont résumées dans le Tableau 7. En présence de ces trois composés (H₂O, O₂ et CO₂), des réactions secondaires peuvent se produire dans le tube de réaction entre les ions produits primaires et ces composés pour conduire à la formation d'ions cluster composé-analyte.

Ions réactifs m/z	H ₂ O m/z		O ₂ m/z		CO ₂ m/z		
H ₃ O ⁺	19	H ₃ O ⁺ .H ₂ O H ₃ O ⁺ .(H ₂ O) ₂ H ₃ O ⁺ .(H ₂ O) ₃	37 55 73	NR		NR	
O ₂ ^{•+}	32	NR		NR		NR	
NO ⁺	30	NO ⁺ .H ₂ O	48	NR		NR	
NO ₂ ⁻	46	NR		NR		NR	
NO ₃ ⁻	62	NR		NR		NR	
O ₂ ^{•-}	32	O ₂ ^{•-} .(H ₂ O)	50	O ₃ ^{•-}	48	CO ₂ . O ₂ ^{•-}	76
OH ⁻	17	OH ⁻ .H ₂ O OH ⁻ .(H ₂ O) ₂	35 53	O ₂ ^{•-} O ₃ ^{•-}	32 48	CO ₂ . O ^{•-} HCO ₃ ⁻ CO ₂ . O ₂ ^{•-} HCO ₃ ⁻ .H ₂ O	60 61 76 79
O ^{•-}	16	OH ⁻ 17 OH ⁻ .H ₂ O OH ⁻ .(H ₂ O) ₂	17 35 53	O ₂ ^{•-} O ₃ ^{•-}	32 48	CO ₂ . O ^{•-} HCO ₃ ⁻ CO ₂ . O ₂ ^{•-} HCO ₃ ⁻ .H ₂ O	60 61 76 79

NR = "No reaction" Pas de réaction entre l'ions précurseurs et le composé étudié.

Tableau 6 : Ions produits lors des réactions entre les huit ions réactifs et l'eau, le dioxyde de carbone et l'oxygène. Les valeurs des rapports masse sur charge m/z des différents ions sont en italiques. D'après l'étude de Ghislain et al. (2020)³⁴⁶

	H ₃ O ⁺	O ₂ ^{•+}	NO ⁺	OH ⁻	O ^{•-}	O ₂ ^{•-}	NO ₂ ⁻	NO ₃ ⁻
H ₂ O	7,16.10 ⁻¹¹	/	6,08.10 ⁻¹³	1,06.10 ⁻¹²	1,32.10 ⁻¹²	7,73.10 ⁻¹³	/	/
O ₂	/	/	/	1,5.10 ⁻¹¹	ND	1,04.10 ⁻¹²	/	/
CO ₂	/	/	/	2,92.10 ⁻¹²	ND	2,63.10 ⁻¹³	/	/

ND = non déterminé dû à l'ion précurseur OH⁻ présent même sans ajout d'eau dans le tube d'écoulement

Tableau 7 : Constantes expérimentales de réaction pour chaque ion précurseur avec l'eau, l'oxygène et le dioxyde de carbone d'après l'étude de Ghislain et al. (2020)³⁴⁶

4. Quantification par SIFT-MS

La technologie SIFT-MS permet dans certaines conditions la quantification des composés en temps réel par la mesure de l'ion précurseur, de l'ion produit et par la connaissance de la constante cinétique d'ordre zéro apparent de la réaction. Ce calcul est basé sur l'hypothèse que la réaction a lieu en présence d'un large excès d'ion précurseur. L'ensemble des calculs permettant la quantification est présenté ci-dessous.

Dans un premier temps, la concentration d'un ion dans le tube de réaction est déterminée en multipliant le signal de l'ion par le facteur de calibration (Équation 20) :

$$[P^{+/-}] = P^{+/-} \times ICF_P \text{ ou } [R^{+/-}] = R^{+/-} \times ICF_R$$

Équation 20

$[P^{+/-}]$: concentration de l'ion produit dans le tube d'écoulement (coup.s⁻¹)

$[R^{+/-}]$: concentration de l'ion précurseur dans le tube d'écoulement (coup.s⁻¹)

$P^{+/-}$: signal de l'ion produit primaire, mesuré par l'instrument (Hz)

$R^{+/-}$: signal de l'ion précurseur, mesuré par l'instrument (Hz).

ICF_P : Facteur de calibration de l'instrument de l'ion produit, dépendant de l'instrument.

ICF_R : Facteur de calibration de l'instrument de l'ion précurseur, dépend de l'instrument.

Le facteur de calibration (ICF) est automatiquement déterminé par l'instrument lors de l'étape de validation. Cette étape consiste en une série de mesures des performances de l'instrument des plus élémentaires aux plus complexes ; vérification des quadripôles, du débit de gaz vecteur, des pressions, du débit d'échantillon, de la ligne de transfert et du détecteur. L'ICF est mis à jour quotidiennement par l'analyse d'une bouteille de gaz étalon contenant sept analytes dont les masses moléculaires couvrent la plage de masses de l'instrument. En comparant les concentrations obtenues pour les 7 analytes (Tableau 8) avec celles certifiées par le fournisseur du gaz de calibration, un facteur de correction (ICF) est calculé pour chaque m/z. Ce facteur permet de transférer des données de caractérisation des composés d'un instrument à l'autre.

Gaz étalons	MW (g mol ⁻¹)	Concentration (ppm)	Incertitude analytiques (RSD)	Ions précurseurs utilisés
1,2,3,4,tetrafluoro Benzene	150,07	2,100	± 5%	O ₂ ^{•+} , O ₂ ^{•-}
Benzene	78,11	2,110	± 5%	H ₃ O ⁺ , O ₂ ^{•+}
Ethylbenzene	106,17	2,040	± 5%	O ₂ ^{•+}
Isobutane	58,12	2,020	± 5%	NO ⁺
Octafluorotoluene	236,06	2,110	± 5%	O ₂ ^{•+} , OH ⁻
Hexafluorobenzène	186,06	2,100	± 5%	O ₂ ^{•+} , O ^{•-} , OH ⁻
Toluene	92,14	2,130	± 5%	H ₃ O ⁺ , NO ⁺

Tableau 8 : Composition du mélange de calibration du SIFT-MS

Après avoir mesuré les concentrations en ions et les paramètres thermodynamiques et cinétiques de réaction d'ionisation dans le tube de réaction (pression, température, débit et nature du gaz d'appoint), la présence en large excès d'ion précurseurs R et la connaissance de la constante cinétique k d'ordre zéro apparent permettent de quantifier directement les analytes dans le tube de réaction (Équation 21).

$$[A] = \frac{[P^{+/-}]}{[R^{+/-}]k \times tr}$$

Équation 21

$[A]$: Concentration de l'analyte dans le tube de réaction (molécule.cm⁻³)

tr : temps de réaction dans le tube d'écoulement, mesuré par l'instrument (s)

$[P^{+/-}]$: Concentration de l'ion produit dans le tube de réaction (coup.s⁻¹)

$[R^{+/-}]$: Concentration de l'ion réactif dans le tube de réaction (coup.s⁻¹)

k : constante de vitesse de réaction (cm³ molécule⁻¹ s⁻¹)

A partir de la concentration de l'analyte dans le tube de réaction et des paramètres instrumentaux, la concentration d'analyte dans un échantillon peut être déterminée selon l'Équation 22 :

$$C_{\text{échantillon}} = [A] \frac{k_B T_{FT} 10^{15} (\Phi_s + \Phi_c)}{P_{FT} \Phi_s}$$

Équation 22

[A] : Concentration de l'analyte dans le tube de réaction (ppb)

Φ_s : débits de l'échantillon ($m^3 \cdot s^{-1}$)

Φ_c : débits de gaz porteur ($m^3 \cdot s^{-1}$)

T_{FT} : température du tube de réaction (K)

P_{FT} : pression dans le tube de réaction (Pa)

k_B : constante de Boltzmann ($1,380649 \times 10^{-23} m^2 \cdot kg \cdot s^{-2} \cdot K^{-1}$)

L'utilisation du logiciel Labsyft fournit par le fabricant du SIFT-MS permet de quantifier directement un composé s'il est connu de la base de données, c'est-à-dire si les ions produits et les valeurs des constantes cinétiques et de branching ratio correspondantes sont connus. Le logiciel dispose ainsi d'une banque de données de constantes cinétiques k de références pour environ 400 molécules, déterminées par le constructeur ou issues de différents travaux dont les références sont accessibles. En revanche, cette banque de données fournit uniquement des informations pour l'ionisation positive c'est-à-dire pour les ions précurseurs H_3O^+ , $O_2^{+\bullet}$ et NO^+ et aucune donnée n'est actuellement disponible pour l'ionisation négative avec les 5 ions précurseurs (O^\bullet , OH^- , $O_2^{\bullet-}$, NO_2^- et NO_3^-). Il est à noter qu'avec ces données constructeurs, la précision des quantifications par SIFT-MS n'est généralement pas suffisante. Il convient donc de déterminer les constantes cinétiques dans les conditions de la mesure pour atteindre une précision de +/- 10%.

Pour quantifier des composés non répertoriés dans la banque de données Labsyft ou utiliser les ions précurseurs négatifs, les ions produits, les branching ratio et les constantes cinétiques doivent être déterminés expérimentalement. Pour cela, des calibrations externes sont réalisées dans des gammes de concentration d'analyte allant de quelques ppbV à quelques ppmV dans des conditions contrôlées (humidité, température...). Pour chaque point de la gamme de calibration, la concentration en analyte dans le flow tube sera calculée selon l'équation suivante (Équation 23):

$$[A] = C_{\text{analyte}} \frac{\Phi_s}{(\Phi_s + \Phi_c)} \frac{P_{FT}}{k_B T_{FT} 10^{15}}$$

Équation 23

Après avoir calculé la concentration en analyte présent dans le tube de réaction, la constante de réactivité k est définie par l'Équation 24 suivante :

$$k = \frac{P^{+/-} ICF_P}{[A] L_r R^{+/-} ICF_R}$$

Équation 24

A noter que deux analytes différents peuvent produire des ions de même ratio m/z avec le même ion précurseur (Tableau 9) ; on parle alors de conflit ou d'interférence qui complique l'élucidation de la composition de la matrice. Il est cependant possible de déterminer la concentration d'un de ces deux analytes par la détection d'un autre ion produit (non interféré) et de déduire les concentrations relatives. La présence des 8 huit ions précurseurs permet potentiellement de résoudre dans certains cas les problèmes des composés isomériques ou isobariques. Par exemple, l'acétone et le propanal, deux molécules de masse moléculaire $58,08 g \cdot mol^{-1}$ vont former des ions produits de même m/z avec

H_3O^+ 58 et O_2^{*+} 59. En revanche, il est possible de les différencier par leur réaction avec NO^+ ou O_2^{*+} (Tableau 10). Une étude récente de Baerenzung dit Baron et al³⁴⁹ portant sur la discrimination de grappes de raisin selon leur variété a montré le potentiel de l'ion O_2^{*+} pour discriminer des composés isobares en comparaison avec les ions H_3O^+ et NO^+ . De nombreuses études ont montré que NO^+ tend à produire des ions avec très peu voire pas de fragmentation tandis que l'ion H_3O^+ a tendance à protoner les molécules avec un niveau de fragmentation moyen. L'ion O_2^{*+} a quant à lui la capacité de fragmenter des molécules de mêmes masses moléculaires, notamment les terpènes comme le nérol, le géraniol et l'alpha-terpinéol pour former plusieurs ions produits dont certains peuvent être discriminants.

En revanche, il peut arriver que certains conflits ne puissent être résolus, principalement dans le cas de matrices complexes ou de la présence de molécules de nature chimique proche. Geeraerts et al³⁵⁰ et Langford et al³⁵¹ ont été confrontés à ce problème dans leurs études sur des matrices alimentaires pour la quantification du 2-méthyl-1-butanol et du 3-méthyl-1-butanol. Ces deux molécules produisent les mêmes ions produits et ne peuvent être distinguées. Les auteurs les ont donc regroupés dans une concentration globale de méthylbutanol.

Composé	MM (g/mol)	Formule brute	Réaction avec H_3O^+	Ions produit (m/z)
2-hexen-1-ol	100,16	$C_6H_{12}O$	$C_6H_{12}O + H_3O^+ \rightarrow C_6H_{12}O.H^+ + H_2O$	101
2-hexanone	100,18	$C_6H_{12}O$	$C_6H_{12}O + H_3O^+ \rightarrow C_6H_{12}O.H^+ + H_2O$	101
2,3-pentanedione	100,12	$C_5H_8O_2$	$C_5H_8O_2 + H_3O^+ \rightarrow C_5H_8O_2.H^+ + H_2O$	101

Tableau 9 : exemple de conflits d'ions pour la quantification de 2-Hexen-1-ol, de 2-Hexanone et de 2,3-pentanedione

Composé	MM (g/mol)	Formule brute	Ions précurseurs	Réaction	Ions produits (m/z)
Acétone	58,08	C_3H_6O	H_3O^+	$(CH_3)_2CO + H_3O^+ \rightarrow (CH_3)_2COH^+ + H_2O$	59
			O_2^{*+}	$(CH_3)_2CO + O_2^{*+} \rightarrow (CH_3)_2CO^+ + O_2$	58
				$\rightarrow CH_3CO^+ + CH_3 + O_2$	43
NO^+	$(CH_3)_2CO + NO^+ \rightarrow C_3H_6O.NO^+$	88			
Propanal	58,08	C_3H_6O	H_3O^+	$CH_3CH_2CHO + H_3O^+ \rightarrow CH_3CH_2CHOH^+ + H_2O$	59
			O_2^{*+}	$CH_3CH_2CHO + O_2^{*+} \rightarrow C_3H_6O^+ + O_2$	58
				$\rightarrow CH_3CH_2CO^+ + HO_2$	57
NO^+	$CH_3CH_2CHO + H_3O^+ \rightarrow CH_3CH_2CO^+ + HNO$	101			

Tableau 10 : Exemple de conflits d'ions pour des composés isomériques avec l'acétone et du propanal

5. Application à l'analyse des COV dans le domaine agroalimentaire

L'utilisation de la technologie SIFT-MS pour l'analyse des COV dans les matrices alimentaires est encore en développement. Dans ce paragraphe, les travaux de recherche utilisant la technologie SIFT-MS pour étudier les COV présents dans des matrices alimentaires sont présentés. A ce jour, 20 études ont démontré le potentiel de cette technologie (Tableau 11). Parmi ces récents développements, de même que pour les analyses SPME-GC-MS, deux stratégies sont appliquées : l'analyse ciblée et l'analyse non ciblée.

a. Analyse ciblée

Cette stratégie vise à quantifier des analytes préalablement sélectionnés pour suivre un procédé de transformation^{94,133,352}, discriminer ou comparer des produits^{317,351,353}. Harper et al³¹⁷ sont les premiers à avoir utilisé la technologie SIFT-MS pour analyser les composés soufrés émis par des fromages suisses afin de discriminer l'âge des fromages et étudier l'impact du ratio de l'acide propionique et de l'acide acétique sur la formation des composés soufrés lors de la fermentation. Pour cela, une méthode SIM a été établie pour suivre 18 composés spécifiques et les quantifier grâce à la base de données du logiciel LabSyft. Dans cette étude, les paramètres d'espace de tête ont été optimisés : temps d'équilibre, température et quantité de fromage. Leur étude a ainsi montré le potentiel du SIFT-MS pour mesurer les composés soufrés à fort impact dans les fromages suisses, à des concentrations de l'ordre du ppbV dans l'espace de tête avec une préparation d'échantillon simple. Ils mettent cependant en évidence la nécessité d'identifier et de quantifier des ions non conflits pour avoir des résultats fiables.

Huang et al¹³³ ont étudié le procédé de torréfaction du cacao en suivant 60 composés par prélèvement direct et en continu à la sortie du torréfacteur puis dans les graines torréfiées par espace de tête. Ils ont étudié l'effet du temps de torréfaction, de la température et du pH. Bowman et al³⁵² ont également étudié le procédé de torréfaction des graines de citrouilles en sélectionnant des composés spécifiques via la base de données LabSyft. Ils ont ainsi montré l'impact de la variété, de l'année de récolte, de la composition et la teneur en humidité sur le développement aromatique. Pour cela, des prélèvements à la sortie du torréfacteur et en espace de tête avec des graines de citrouille torréfiées ont été analysés. Ces deux études montrent le potentiel du SIFT-MS pour suivre en continu des composés au cours d'un procédé de transformation. Il a été mis en évidence que les concentrations des composés volatils des graines torréfiées variaient selon la variété de la graine et que l'augmentation du taux d'humidité dans les graines réduisait la formation de COV. Wood et al³⁵⁴ ont étudié la qualité des produits laitiers (lactose, concentrés de protéines de lactosérum et concentrés de protéines de lait) en suivant des composés aromatiques par SIFT-MS en utilisant différentes approches chimométriques pour analyser les données. Pour chaque matrice, les conditions d'échantillonnage en espace de tête ont été optimisées (température et temps d'incubation). Ils ont démontré le potentiel d'application en routine d'une méthode SIFT-MS en suivant des marqueurs spécifiques pour étudier la qualité de produits laitier en complément des analyses sensorielles réalisées par des panels. Langford et al³⁵¹, Taylor et al³⁵³ et Castada et al^{94,97,355} ont utilisé le SIFT-MS pour classer des échantillons de fromages selon leur origine de production, selon la formation d'ouvertures (« trous ») ou selon les ferments utilisés. Ces 4 études ont utilisé un échantillonnage en espace de tête et un mode d'acquisition SIM. L'étude menée par Oscar-Sinir et al¹⁵⁰ portait sur la détection d'adultération des huiles d'olive vierge par SIFT-MS. 72 composés présents dans l'espace de tête de diverses huiles d'olive ont été suivis pour détecter des altérations et permettre de discriminer les huiles altérées des huiles non altérées. Ces études ont montré la possibilité d'utiliser le SIFT-MS comme outil de contrôle qualité ou d'authenticité, de détection de défauts sur des produits finis ou au cours de la production et pour étudier des modifications de procédés. Kuulialia et al¹²⁹ ont utilisé la technologie SIFT-MS pour suivre en temps réel une vingtaine de composés dans l'espace de tête des emballages sous atmosphère contrôlée de morues. L'utilisation du SIFT-MS a permis aux auteurs de suivre l'évolution des COV générés dans les emballages sous atmosphère modifiée de manière rapide et sans préparation d'échantillon. Ces résultats représentaient en temps réel la détérioration de la qualité du poisson.

Chacune de ces études basées sur l'analyse ciblée par SIFT-MS a démontré l'intérêt de cette technologie dans l'analyse des COV alimentaires notamment pour sa simplicité et la rapidité de mesure

mais a également soulevé la problématique des ions conflits pouvant interférer dans la quantification de certains composés. Il est à noter que dans ces études, aucune calibration par ajouts dosés sur matrice n'a été effectuée pour s'assurer de la justesse des concentrations en COV mais que de simples comparaisons de concentrations relatives ont permis de tirer des conclusions intéressantes.

b. Analyse non ciblée

Une autre variation de l'approche par analyse ciblée est d'effectuer dans un premier temps une analyse non ciblée en full scan en balayant une large gamme de m/z d'ions produits puis dans un deuxième temps de traiter les données afin de convertir les signaux bruts obtenus pour un ou plusieurs ions en concentration de composés. C'est la stratégie adoptée par Van Kerrebroeck et al³⁵⁶ qui a permis d'étudier l'impact de la culture de départ, des ingrédients et de la qualité de la farine sur la formation des composés volatils de la mie de pain. Avec une approche similaire Geeraerts et al^{132,350} ont pu suivre la formation de composés volatils dans un emballage sous atmosphère modifiée du jambon cuit et des produits de volaille cuits. Ces études ont démontré le potentiel de l'analyse non ciblée suivie d'une quantification de marqueurs par SIFT-MS.

L'utilisation d'une démarche d'analyse totalement non ciblée est beaucoup plus récente et est de plus en plus étudiée depuis 2015 (Tableau 11). Elle est basée sur l'acquisition du profil volatil ou « empreinte volatile » d'un produit c'est-à-dire qu'elle tient compte de l'ensemble des COV émis par le produit. L'objectif n'est alors pas d'identifier chaque composé analysé, mais de comparer les profils ou « les empreintes » de l'ensemble des COV selon le procédé de transformation, la nature du produit, d'éventuelles altérations... Cette approche tient compte de l'ensemble des ions produits par la mesure SIFT-MS en mode full scan pour chaque ion précurseur étudié. La première étude traitant d'une analyse full scan menée par Carraspiso et al³⁵⁷ date de 2015. L'objectif de cette étude était d'examiner le potentiel de la technologie SIFT-MS à la fois pour quantifier les composés volatils émis par du jambon ibérique (mode SIM) et pour classer les jambons secs en fonction des conditions d'engraissement des animaux (mode full scan). Les acquisitions des mesures SIFT-MS en espace de tête ont été effectuées en mode full scan sur une gamme de m/z de 10 à 210 ainsi qu'en mode SIM. Les données obtenues pour chacun des trois ions précurseurs (H_3O^+ , $O_2^{\bullet+}$ et NO^+) provenant de l'analyse full scan ont été analysées indépendamment pour discriminer les 4 groupes de jambons ibériques. Ils ont ainsi démontré que l'empreinte obtenue par l'analyse full scan était plus performante pour classer les jambons que la méthode SIM de mesure de composés sélectionnés.

Bajoub et al³⁵⁸ ont été les premiers à n'utiliser que de l'analyse SIFT-MS en mode full scan pour assurer l'authentification d'huiles d'olive d'origine méditerranéenne. L'analyse a été effectuée en espace de tête sur une gamme de m/z de 10 à 200 pour les 3 ions précurseurs positifs (H_3O^+ , $O_2^{\bullet+}$, NO^+). Les spectres de masse ont été extraits à l'aide du logiciel LabSyft pour former trois jeux de données, un pour chaque ion précurseur. Leur étude a ainsi montré que l'utilisation d'analyses chimiométriques sur les jeux de données obtenues en mode full scan permettait de discriminer les différentes origines géographiques. Kharbach et al¹⁴⁴ ont également utilisé le SIFT-MS en mode full scan pour tracer l'origine géographique d'huiles d'argan marocaines sur une gamme de m/z de 15 à 250 en utilisant les 3 ions précurseurs positifs (H_3O^+ , $O_2^{\bullet+}$, NO^+). Pour classer les huiles d'argan issues de 5 origines géographiques à partir des données full scan, différentes analyses chimiométriques ont été testées. Ils ont montré que l'acquisition en mode full scan couplée à des analyses chimiométriques permettait de déterminer les origines géographiques des huiles. Dans une autre étude de Kharbach et al³⁵⁹, l'analyse non spécifique SIFT-MS couplée à des outils chimiométriques a permis d'authentifier des huiles d'argan

vierges, cette fois en fusionnant les données obtenues pour les trois ions précurseurs dans un même jeu de données. Ces études recommandent l'utilisation de la technologie SIFT-MS pour mesurer des empreintes à partir de la fraction volatile des huiles comme méthode fiable, précise, rapide et non destructive.

Vendel et al³⁶⁰ ont également utilisé la technologie SIFT-MS pour analyser les arômes de fraise afin de détecter de manière rapide, objective et reproductible des différences d'arômes et classer les fraises selon la variété, le stade de maturation et/ou le moment de la récolte. Pour cela, ils ont comparé une acquisition en mode SIM pour suivre 12 composés connus de l'arôme de fraise et une acquisition en mode full scan sur une gamme de m/z de 16 à 250 permettant l'obtention d'empreintes. Ils ont également comparé la mesure HS-SIFT-MS avec une mesure traditionnelle très fréquemment utilisée, la HS-SPME-GC-MS. De meilleurs résultats ont été obtenus par la méthode SIFT-MS full scan pour séparer les échantillons en fonction des différents stades de maturation. Les résultats obtenus par SPME-GC-MS et par SIFT-MS SIM montrent des chevauchements de groupes d'échantillons, indiquant qu'aucune discrimination nette n'est obtenue par ces approches. Les auteurs ont relevé des difficultés concernant des tentatives d'identification des composés par l'analyse SIFT-MS full scan, limitée par l'absence de techniques de séparation, par la possibilité que différents composés forment un même ion de même valeur m/z et par la base de données Labsyft moins étendue que la NIST (National Institute of Standards and Technology) utilisée pour l'identification en GC-MS. Leur étude montre que la technologie SIFT-MS peut être considérée comme une technique complémentaire utile pour étudier l'impact de certains paramètres tel que la variété, le stade de maturation et le moment de la récolte.

Baerenzung Dit Baron et al³⁴⁹ ont utilisé l'analyse SIFT-MS full scan pour discriminer rapidement différentes variétés de vigne en se basant sur l'empreinte volatile obtenue par l'analyse des grappes de raisin en espace de tête. Ils ont montré que le profil qui permettait la meilleure discrimination en utilisant une approche chimométrique était celui obtenu avec l'ion O_2^{*+} . Cet ion a une capacité à générer plus de fragments que les autres ions permettant ainsi de différencier des composés similaires, notamment les terpènes. L'originalité de leur approche a porté sur l'exploitation des ions produits entre 78 et 250 m/z afin de s'affranchir des ions conflits plus fréquents à faible ratio m/z.

Ces études n'ont utilisé que l'ionisation positive du SIFT-MS c'est-à-dire les ions H_3O^+ , O_2^{*+} , NO^+ . Actuellement, aucune étude concernant l'analyse de matrices alimentaires par ionisation négative n'a été publiée.

6. Conclusions : forces et limites de l'analyse des COV émis par des produits agroalimentaires par SIFT-MS

Pour l'ensemble des études utilisant la technologie SIFT-MS pour caractériser des matrices agroalimentaires, l'instrument est calibré quotidiennement par l'utilisation du mélange de gaz calibrant mais aucune étude ne réalise d'étalonnage des composés étudiés notamment en mode SIM. En raison de la complexité des matrices alimentaires, des étalonnages seraient nécessaires pour augmenter la précision des mesures quantitatives. De façon similaire à ce qui se pratique dans le cadre d'une analyse chromatographique, des étalonnages par ajouts dosés permettraient d'étudier l'influence de la matrice sur la réponse du composé en SIFT-MS. Un étalonnage externe des composés permet de vérifier la justesse de la constante de réaction et d'augmenter la résolution et la précision des quantifications. Cependant, cela représente une quantité de travail importante pour aboutir à des

mesures précises de concentrations en COV, qui finalement ne présentent pas beaucoup d'intérêt ; c'est davantage la comparaison des niveaux des signaux entre plusieurs produits qui importe.

Ces études montrent l'intérêt de la technologie SIFT-MS comme outil simple, rapide et non destructif de l'échantillon pour analyser les COV présents dans des matrices alimentaires. En revanche, cette méthode présente certaines limites. La présence d'eau dans certaines matrices peut générer des clusters avec les ions précurseurs, conduire à la formation d'ions secondaires et compliquer la détection de certains composés. Outre la présence d'eau, la complexité de certaines matrices contenant de très nombreuses molécules de familles chimiques diverses peut générer de nombreux conflits d'ions et empêcher les identifications et les mesures quantitatives. Enfin, la présence d'un composé majoritaire au sein d'un produit peut impacter l'analyse en consommant les ions précurseurs, normalement en excès, et en limitant les réactions avec les autres composés du produit. C'est par exemple le cas de l'éthanol que l'on retrouve en forte concentration dans des boissons alcoolisées mais également dans des produits fermentés comme le fromage. Un autre facteur limitant pour l'identification des composés est la base de données LabSyft qui est à ce jour beaucoup moins étendue que, par exemple, la base de données NIST pour GC-MS.

Une limite majeure pour l'analyse des COV de matrices alimentaires par SIFT-MS réside dans l'utilisation de méthodes SIM pour quantifier des marqueurs, en raison de la présence inhérente aux matrices complexes d'ions conflits inconnus. L'utilisation récente du mode full scan, méthode qualitative, s'avère performante pour comparer des profils volatils de matrices alimentaires complexes et ainsi discriminer des échantillons selon différents critères comme des altérations, contaminations ou autre... Cependant, après avoir mis en évidence les ions produits discriminants à l'aide d'outils statistiques adaptés, il peut être intéressant d'identifier les molécules responsables pour comprendre leur origine (voies enzymatiques impliquées par exemple) et ainsi améliorer un procédé de fabrication ou de transformation ou optimiser des conditions de stockage. L'identification de ces molécules clés est difficile sans un détecteur à haute résolution en masse. En effet, le détecteur quadripolaire du SIFT-MS Voice 200 ultra ne permet généralement pas d'élucider les conflits d'ions. Le chapitre 3 présente les différentes stratégies qui nous ont permis d'incriminer certaines molécules dans le classement d'échantillons de fromages ou de thés. Enfin, lors d'une analyse non spécifique full scan, la comparaison des profils est effectuée sur les signaux bruts et interroge sur la stabilité des signaux dans le temps et sur leur transposition ou fusion d'un instrument à l'autre pour permettre de réaliser des collections d'empreintes volatiles sur de longues périodes.

C. Conclusion

Les analyses non spécifiques des COV ont fait naître le terme « volatolome » pour décrire tous les composés volatils émis par un organisme, un écosystème ou un produit (tel qu'un aliment). En conséquence, l'analyse ciblant la détection, la caractérisation et la quantification de tous composés volatils est la volatolomique, sous-domaine de la métabolomique. L'acquisition des empreinte volatiles ou profils volatils génère des volumes importants de données et nécessite d'utiliser des méthodes chimométriques adaptées, déjà appliquées dans de nombreux domaines des sciences « omiques » (génomiques, protéomiques, métabolomiques) et des sciences alimentaires. Le concept de la « Big Science » est aujourd'hui en développement.

Tableau 11: Exemple d'applications de la technologie SIFT-MS dans l'analyse des composés volatils de produits agroalimentaires

Matrice	Objectif	Condition d'espace de tête	Mode d'acquisition	Ions réactifs	Etalon	Molécules	Ref	Année
Fromages	Développement de l'analyse des composés volatils soufrés par SIFT-MS dans des fromages suisses de différents âges	5g dans une bouteille de 500mL 60°C pendant 1h	SIM (18 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Acides, Composés soufrés, Alcool, Pyrazine et Furanéol.	317	2011
Cacao	Suivi des COV du cacao produits pendant la torréfaction et dans le produit fini	<i>Suivi torréfaction</i> : prélèvement direct et en continue à la sortie du torréfacteur <i>Fèves torréfié</i> : 50g dans un bocal hermétique 50°C pendant 90min	SIM (60 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Pyrazines, aldéhydes, cétones, alcools, acides, furanes, esters, composés azotés, phénols, hydrocarbures et composés soufrés.	133	2011
Fromages	Discrimination de fromages parmesans italiens et néo-zélandais	4g dans une bouteille de 1L 37°C pendant 1h	SIM (38 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Esters, aldéhydes, cétones, acides, pyrazines et composés soufrés.	351	2012
Citrouilles	Suivi des COV dans les graines de citrouille grillées	50g (graines non torréfié ou torréfié) + 100 mL d'eau : Broyé Bouillie introduite dans un flacon de 500mL. 50°C pendant 90 min	SIM (33 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Esters, aldéhydes, alcools, cétones, acides, pyrazines, furanes et composés soufrés.	352	2012
Fromages	Discrimination de fromage suisse provenant de 5 usines différentes	5g dans une bouteille de 1L 40°C pendant 1h	SIM (24 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Aldéhydes, acides, cétones, composés soufrés, esters, furanes, pyrazines	353	2013
Fromages	Analyse des COV dans des échantillons de fromage suisse divisés et non divisés	5g dans un flacon de 500 mL 40°C dans un bain marie pendant 1h	Full scan (m/z 15-200) SIM (30 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Alcools, aldéhydes, cétones, acides, esters, indole, composés soufrés, et furanes	97	2014
Fromages	Analyse des COV dans divers mélanges binaires	100 mL de solution dans une bouteille de 500mL 40°C dans un bain marie pendant 1h	SIM (30 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Alcools, aldéhydes, cétones, acides, esters, indole, composés soufrés, et furanes	94	2015
Jambons ibériques	Discrimination de jambons ibériques provenant de porcs élevés dans quatre conditions différentes	Headspace 15 g dans un sac de film à haute barrière 60min à 18-20°C	Full scan (10-210 m/z) SIM (39 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Aldéhydes, cétones, alcools, acides, composés soufrés, autres	357	2015
Morue	Etude du stockage sous atmosphère modifiée	Cinétique au cours du temps à deux température 4°C et 8°C	SIM (20 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Acides, Alcools, Aldéhydes, Esters, Composés soufrés, Amines.	129	2018

Mie de pain	Evaluer le profil aromatique de la mie de pain en utilisant différents levains liquides	0,2 g dans un flacon de 500mL 50°C pendant 20 min	Full scans 15 -200 m/z	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Aldehydes, alcools, acides, terpènes	361	2018
Huiles d'olive	Développer une méthode rapide et fiable pour la discrimination des huiles d'olive marquées d'une indication géographique par les empreintes volatiles	10g dans une bouteille de 100mL 1h à 40°C	Full scan (10-200 m/z)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	X	358	2018
Huile d'argan	Classification d'huile d'Argan marocaine d'origine certifié et comparaison avec le profil chimique	10g dans une bouteille de 100mL 30 min à température ambiante	Full Scan (15-250 m/z)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	X	144	2018
Volaille cuits	Suivi de la production de COV dans les produits de volaille cuits	Headspace Tedlar bag Suivi au cours du temps à 4-6°C, 7-9°C et 11-13°C.	Full scan (15-170 m/z)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Suivi de 36 ions correspondants 26 composés (alcools, aldéhydes, cétones, acides, composés soufrés et terpènes) Calculs des concentration	350	2019
Fraise	Etude du cultivar, de la maturation et du moment de la récolte.	1 fraise dans contenant plastique de 250 mL 3 min à température ambiante	Full scan 16-250 m/z SIM (12 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Aldéhydes, cétones, acides, esters et terpènes	360	2019
Fromages	Comprendre la variabilité de la saveur de fromage suisse, sans défauts de saveur perçus, en relation avec les COV et les attributs sensoriels	5g dans un flacon de 500mL 1h à 40°C	SIM (28 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	6 Aldéhydes, 7 composés soufrés, 3 alcools, 3 esters, 2 cétones, 3 acides, 2 furanes et 1 pyrazine	355	2019
Jambon cuit	Suivre la formation de substances volatiles dans le jambon cuit emballé sous atmosphère modifiée	Headspace Tedlar bag Suivi au cours du temps à 4-6°C, 7-9°C et 11-13°C.	Full scan 15 to 170 m/z	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Suivi de 35 ions correspondants à des alcools, aldéhydes, cétones, acides, composés soufrés et terpènes	132	2020
Huile d'olive	Détection d'adultération	5g dans un flacon de 500mL 30min à 30°C	SIM (72 composés)	H ₃ O ⁺ , NO ⁺	/	Alcools, aldéhydes, cétones, acides, esters, et terpènes	150	2020
Produit laitier	Qualité / Créer des modèles de seuil d'arômes identifiés	3g dans un vial de 20mL 70°C pendant 15 min	SIM (81 composés)	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	X	354	2021
Huile d'argan	Authentification de l'huile d'argan extra vierge par profil volatile SIFT-MS	10g d'huile dans une bouteille de 100mL 30 min à Température ambiante	Full scan 15-250 m/z	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	X	359	2022
Raisins	Discrimination des variétés de vigne sur la base des empreintes volatiles des raisins	Headspace 6 heure à RT puis 40 min dans un bain marie à 40°C	Full scan 15 to 250 m/z	H ₃ O ⁺ , O ₂ ^{•+} , NO ⁺	/	Terpène et sesquiterpène	349	2022

III. Chimiométrie

La chimiométrie a pour objectif d'aider l'analyste dans le traitement des données et dans l'interprétation des résultats. Cette discipline utilise un ensemble d'outils mathématiques et des méthodes statistiques pour analyser des données chimiques multivariées de manière optimale. La chimiométrie, apparue dans la littérature des années 70, a été proposée par Svante Wold (Umea University, Suède) qui la décrit comme l'art d'extraire des informations chimiquement pertinentes à partir de données chimiques mesurées et de les représenter³⁶². Ces dernières décennies ont vu naître une autre révolution : la croissance rapide de la puissance de calcul, permettant de traiter les données de nouveaux instruments rapides et automatisés, générant de grands volumes de données. La chimiométrie a ainsi permis à la chimie analytique de passer dans l'ère numérique des big data. A leur tour les techniques analytiques ont enrichi la chimiométrie par l'apport de données toujours plus complexes, nécessitant le développement de nouveaux algorithmes : on parle alors de Machine Learning³⁶³.

La métabolomique offre une description approfondie des événements biochimiques qui régissent un système par la mesure de l'ensemble des métabolites qui le caractérisent en s'appuyant sur des techniques analytiques de pointe³⁶⁴. C'est également le cas de la volatolomique, sous-domaine de la métabolomique, qui s'attache à étudier uniquement les composés organiques volatils (COV) présents ou émis par système. Cette approche très prometteuse permet l'analyse non invasive d'un écosystème ou d'organismes. Les développements technologiques récents permettent aujourd'hui de mesurer simultanément en une seule analyse des milliers de métabolites volatils présents dans un échantillon et génèrent des données complexes et massives, qui posent des problèmes particuliers d'analyse de données. La chimiométrie apporte des réponses efficaces en proposant des outils spécifiques pour extraire des informations pertinentes au sein de ces vastes ensembles de données³⁶⁵.

Les deux composantes essentielles de la chimiométrie sont le prétraitement des données utilisé pour obtenir des données lisibles et exploitables à partir des données brutes et les traitements statistiques pour interpréter et visualiser les résultats.

A. Prétraitement des données de volatolomique

La performance d'une analyse chimiométrique repose sur la qualité et la pertinence des données. Malheureusement, la présence de facteurs délétères (bruit de fond, valeurs manquantes, données incohérentes...) peut fortement influencer la qualité et l'intérêt des résultats. Des données de faible qualité conduiront à des informations de faible qualité³⁶⁶. Ainsi, le prétraitement des données est une étape majeure et essentielle dont le but principal est d'obtenir des ensembles de données finales qui peuvent être considérés comme cohérents et utilisables pour les algorithmes d'exploration de données ultérieures (traitement statistique)³⁶⁷.

Le prétraitement est un outil puissant dont l'objectif est de préparer les données à leur analyse et de maximiser la chance de détection des composés d'intérêt³⁶⁸. On distingue quatre grandes étapes : (i) la collecte des données, (ii) la correction de dérive analytique, (iii) le nettoyage des données, (iv) la transformation des données (normalisation notamment) et dans certains cas la réduction des données pour obtenir une représentation réduite de l'ensemble de données en ne conservant que les plus pertinentes et en produisant la même qualité de résultats analytiques. Après les étapes de

prétraitement des données, l'ensemble final obtenu peut être considéré comme une source fiable et appropriée pour appliquer des algorithmes d'exploration de données (traitements statistiques)³⁶⁹.

Dans ce chapitre, les prétraitements effectués sur les données obtenues par analyse chromatographique (GC-MS, GC-FID) et SIFT-MS seront présentés pour des analyses ciblées ou non ciblées.

1. Prétraitement des données GC-MS et GC-FID

Comme nous l'avons vu précédemment, l'analyse des arômes par chromatographie en phase gazeuse (GC) couplée à un détecteur par spectrométrie de masse (MS) ou associée à un détecteur par ionisation de flamme (FID) est généralement utilisée. Ces analyses génèrent de grandes quantités de données multidimensionnelles comme :

- Le temps de rétention, caractéristique du composé élué en fonction de la température et de la colonne chromatographique
- L'aire du pic chromatographique, proportionnelle à la quantité de composé injectée, pouvant être obtenue pour un signal FID ou MS
- Pour une détection en MS, le spectre de masse du composé, représentant l'intensité relative des fragments de molécule en fonction de leur rapport masse sur charge (m/z)

La manipulation d'ensembles de ces données est complexe et peut nécessiter des outils de traitement informatique.

a. Analyse ciblée

Les techniques de GC-MS ou de GC-FID sont très couramment utilisées pour mener des analyses ciblées permettant de quantifier des composés spécifiques dans des échantillons. Le temps de rétention permet d'identifier un composé par rapport à un étalon analysé dans les mêmes conditions. Dans le cas de l'utilisation d'un détecteur de spectrométrie de masse, la structure moléculaire du composé peut être déterminée ou confirmée et les éventuelles co-élutions de pics peuvent être identifiées. L'identification se fait généralement par comparaison du spectre obtenu expérimentalement avec des spectres de références regroupés dans des bases de données. Celles-ci peuvent être commerciales, comme la bibliothèque NIST (National Institute of Standards and Technology) qui contient plusieurs milliers de spectres, ou internes au laboratoire. La quantification du composé est réalisée par un étalonnage externe ou par ajouts dosés.

Dans cette thèse, un chromatographe en phase gazeuse couplé à une double détection par spectrométrie de masse et ionisation de flamme (GC-MS-FID) a été utilisé. Le détecteur FID étant plus stable et ayant une gamme de linéarité plus étendue que le détecteur MS, il est préféré pour la quantification. Les interprétations des données obtenues en GC-MS-FID pour identifier et quantifier spécifiquement des composés sont réalisées la plupart du temps en utilisant les logiciels des fabricants d'équipements.

Beltran et al¹⁴⁹, après avoir analysé par HS-SPME-GC-MS leur échantillon d'huile d'amande, ont identifié les composés recherchés par comparaison de leurs temps de rétention avec ceux de composés étalons et en comparant leur spectre de masse par rapport aux spectres de la librairie NIST.

Lorsque les étalons n'étaient pas disponibles, les composés volatils ont été identifiés provisoirement en utilisant uniquement les spectres de masse : seuls les composés dont le spectre avait une similarité de plus de 90 % avec le spectre proposé par la bibliothèque du NIST ont été considérés. Ils ont ensuite mesuré les réponses chromatographiques (aire de pic) de chaque composé détecté. Cette approche a été appliquée dans de nombreuses études^{149,268-276,278,280-282,290-293,296-298,300,301,305,370} pour traiter les résultats d'analyse ciblée par HS-SPME-GC-MS d'échantillons de champignons.

b. Analyse non ciblée

Pour les analyses non spécifiques consistant à acquérir des profils de composés volatils, l'ensemble du chromatogramme et, pour une détection par spectrométrie de masse, toutes les valeurs m/z des spectres de masse sont importantes. Cependant, l'analyse de toutes ces données présente quelques difficultés. Certains pics chromatographiques peuvent être difficiles à distinguer du bruit de fond, il peut aussi y avoir des co-élutions avec comme conséquence un spectre de masse correspondant au mélange de plusieurs composés. Une autre difficulté concerne la fusion de plusieurs chromatogrammes, issus de divers échantillons, obtenus à des périodes différentes avec de possibles décalages de temps de rétention¹⁰. Ces difficultés rendent le prétraitement des données complexe mais crucial. En général, les outils de prétraitement des données brutes comprennent quatre modules de base : le filtrage du bruit de fond et la correction de la ligne de base qui permet de distinguer les "vrais" pics (correspondant à un composé analysé) du bruit dans un chromatogramme, la détection et la déconvolution spectrale de composés co-élus, l'alignement qui permet de tenir compte des dérives du temps de rétention et la normalisation³⁷¹.

Pour exécuter ces différentes étapes, plusieurs algorithmes ont été développés pour traiter les données MS brutes acquises par spectrométrie de masse couplée à des techniques de chromatographie liquide (LC-MS) ou gazeuse (GC-MS). La disponibilité de logiciels informatiques permettant un prétraitement fiable des données, un débit rapide et la production d'informations fiables est essentielle pour la qualité des résultats analytiques et leur interprétation.

Les logiciels de traitement des données se présentent sous un format propriétaire ou libre. De nombreuses solutions « open-source » existent avec différents degrés de sophistication ; certains logiciels prennent en charge l'ensemble du traitement des données, tandis que d'autres programmes sont spécialisés dans des tâches spécifiques telles que la détection de caractéristiques, l'analyse statistique ou l'identification de composés. Des études^{372,373} comparant différents logiciels libres ont déjà été réalisées. Les critères d'évaluation étaient : les fonctionnalités des différents logiciels, les compatibilités des données d'entrée utilisées, le mode d'intégration des pics chromatographiques, la disponibilité d'outils statistiques, les modes d'identification des composés et la dépendance vis à vis d'autres logiciels (Tableau 12).

Nom du logiciel	Alignement des chromatogrammes	Déconvolution des spectres de masses	Formats du fichier de données d'entrée	Sortie/Type de quantification	Statistiques	Identification des composés
AMDIS	Oui	Non	netCDF	Valeurs m/z et aire de pic	Oui	Librairie utilisateur ou en ligne
SpectConnect	Non	Oui	ADMIS.elu	Aire de pics ou à l'abondance relative	Test de student seulement	Librairie utilisateur
MeltDB	Oui	Non	NetCDF, mzXML, mzDATA	Aire de pic	Oui	Librairie utilisateur ou en ligne
XCMS	Oui	Non	NetCDF, mzXML, mzDATA, Agilent.d	Aire de pic	Test t de Welch seulement	Librairie utilisateur ou en ligne
Metabolite detector	Oui	Oui	NetCDF	Aire de pic	ACP seulement	Librairie utilisateur importée
MetAlign	Non	Non	NetCDF, Agilent .d mzData, mzXML, Masslynx, Xcalibur	Aire de pic	Non	Librairie utilisateur
MzMine	Non	Oui	NetCDF, mzML, mzXML, mzData, Xcalibur	Aire de pic	Oui	Librairie utilisateur ou en ligne
TagFinder	Oui	Non	NetCDF MetAlign output	Groupes corrélés en intensité	Non	Librairie utilisateur
Gavin	Oui	Non	NetCDF, AMDIS .elu	Aire de pic	Non	Librairie utilisateur

Tableau 12 : Description de quelques logiciels libres d'accès pour l'exploitation non ciblée d'analyses GC/LC-MS

L'étude de Morales (2017)³⁷³ a montré que la déconvolution des pics effectués par le logiciel AMDIS couplé au logiciel Gavin pour effectuer l'alignement des pics s'est avérée être le moyen le plus utile et le plus convivial pour étudier les changements de profils volatils dans une boisson à la fraise. De plus, les résultats obtenus étaient similaires à ceux obtenus par intégration manuelle.

Il existe également des logiciels propriétaires principalement développés par les constructeurs d'instruments tels que : MarkerLynx (Waters) (spécialisé pour la LC-MS), MassHunter (Agilent), TraceFinder (Thermo), ChromCompare (Markes international).

Les logiciels libres et propriétaires présentent chacun leurs avantages et leurs inconvénients. Les logiciels libres ont la particularité d'être génériques et non spécifiques, c'est-à-dire qu'ils peuvent traiter des données provenant de plusieurs instruments. Cela s'avère être un avantage pour développer une approche commune à plusieurs instruments, cependant, cela risque de les rendre potentiellement moins performants que les logiciels propriétaires spécifiques. Les logiciels propriétaires sont souvent coûteux mais simples d'utilisation avec un support technique, à contrario des logiciels libres qui sont gratuits mais plus complexes à utiliser, en raison des nombreux paramètres à régler sans assistance ou sans formation³⁷⁴.

Conclusion sur le choix du logiciel

Malgré les divers avantages des logiciels libres, leur complexité d'utilisation nous a orientés vers le choix d'un logiciel propriétaire. Ainsi, pour traiter les données obtenues par l'analyse non ciblée HS-SPME-GC-MS, le logiciel choisi est le logiciel ChromCompare + (Markes International)³⁷⁵. A l'inverse des autres logiciels propriétaires fournisseurs, celui-ci permet de traiter des données obtenues à partir d'instruments de plusieurs marques.

2. Prétraitement des données SIFT-MS

Une analyse non ciblée réalisée en mode full scan par SIFT-MS peut générer de grands volumes de données. Cette technologie dispose de 8 ions précurseurs (H_3O^+ , $O_2^{•+}$, NO^+ , $O^{•+}$, OH^+ , $O_2^{•+}$, NO_2^+ , NO_3^+) pouvant réagir avec les analytes présents dans les échantillons pour générer des ions produits. La détection de ces ions produits se fait en balayant une gamme de rapport m/z allant de 15 à 400 amu pour chaque réaction avec les ions précurseurs. 385 données peuvent donc être générées pour chaque ion précurseur lors d'une analyse et, en tenant compte de tous les ions précurseurs, atteindre 3 080 données. Le logiciel « LabSyft » est fourni par le fabricant de l'instrument SIFT-MS pour traiter les données de façon conventionnelle non ciblée.

a. Analyse ciblée

La majorité des études réalisées avec le SIFT-MS concerne des analyses spécifiques c'est-à-dire la détection et la quantification en temps réel de certains composés cibles que ce soit dans le domaine alimentaire, environnemental ou la santé. Les expérimentateurs utilisent le logiciel LabSyft en mode d'acquisition SIM (Single ion monitoring) ou MIM (Multiple ion monitoring). Pour cela, ils créent une méthode de quantification où les composés à quantifier présents dans la base de données du logiciel sont sélectionnés. Cette base de données renseigne les ions produits par les analytes avec les différents ions précurseurs et donne les constantes de réactivité associées. Lors de la création d'une méthode, le logiciel met en évidence les composés présentant des ions conflits c'est-à-dire des ions identiques produits par deux ou plusieurs composés lors de la réaction avec un ion précurseur. Le prétraitement très largement utilisé dans ce cas consiste à retirer les ions produits conflictuels afin d'assurer une bonne quantification des composés. Il arrive cependant que certains composés ne puissent pas être identifiés séparément car ils produisent les mêmes ions produits avec de mêmes valeurs de m/z. Certains auteurs décident soit de ne pas reporter ces composés³⁷⁶, soit de rapporter leurs concentrations ensembles^{377,378}. Geeraerts *et al*³⁵⁰ ont regroupé le 2-méthyl-1-butanol et le 3-méthyl-1-butanol, appelés par la suite méthylbutanol, ainsi que le 2-méthylbutanal et le 3-méthylbutanal, appelés par la suite méthylbutanal. Les concentrations de ces composés volatils ont été calculées à partir des données cinétiques des isomères 2-méthyl. Wood *et al*³⁵⁴ ont simplement soustrait le bruit de fond afin d'obtenir un spectre de masse plus propre et obtenir de meilleurs résultats (intensité en coups par seconde). D'autre part, au moment de la réalisation du calcul de concentration d'un analyte, si plusieurs ions produits sont mesurés, l'algorithme LabSyft va exclure une mesure qui serait supérieure aux autres, supposant ainsi que le signal de cet ion est impacté par un autre composé non listé dans la méthode.

b. Analyse non ciblée

Les analyses non spécifiques sont basées sur des mesures en mode full scan détectant tous les ions produits. La complexité des spectres obtenus nécessite des outils chimiométriques pour extraire l'information et parfois des systèmes de prétraitement des données. Le logiciel Labsyft fournit un tableau au format CSV reprenant l'abondance de chaque ion produit par analyse.

Les études portant sur l'analyse des profils volatils par SIFT-MS utilisent différentes stratégies de traitement des données. Carraspiro *et al*³⁵⁷, les premiers à avoir démontré le potentiel du SIFT-MS en mode full scan pour classer les jambons secs en fonction des conditions d'engraissement des animaux, ont utilisé les données fournies par l'instrument sans étape de prétraitement avant d'effectuer des analyses statistiques. Bajoub *et al*³⁵⁸ ont prétraité les données full scan obtenues pour chaque ion précurseur en réalisant une méthode d'« auto-scaling » (c'est-à-dire qu'ils ont normalisé les colonnes de la matrice de données traitées pour qu'elles aient une moyenne nulle et une variance unitaire) sur les trois matrices de données correspondant à chaque ion précurseur avant l'analyse chimiométrique. Kharbach *et al*^{144,359}, dans leurs deux études, n'ont pas prétraité les données full scan obtenues et les ont utilisées telles quelles. Ils ont étudié les données full scan de chaque ion précurseur séparément puis ensemble. Vendel *et al*³⁶⁰ ont regroupé les données obtenues pour les 3 ions précurseurs étudiés et ont traité les données en excluant les valeurs des ions précurseurs et les m/z présentant de faibles abondances afin de présélectionner les ions pertinents. Baerensung Dit Baron *et al*³⁴⁹, dont le but était de discriminer le volatolome des grappes de raisin, ont judicieusement nettoyé les jeux de données obtenus pour chaque ion précurseur en excluant les valeurs de m/z inférieures à 100 et toutes les valeurs de m/z dont les abondances étaient inférieures au bruit de fond. Leur étude se focalisant sur la recherche des terpènes, ce prétraitement leur a permis de réduire le bruit de fond et d'éliminer la plupart des ions de fragmentation générés par les terpènes et qui sont similaires entre eux ou avec d'autres composés.

En se basant sur les études antérieures à ces travaux de thèse, notre choix a été de tester différents prétraitements des données brutes, notamment pour s'affranchir des variations de performance de l'instrument dont les niveaux d'ions précurseurs ont beaucoup évolué au cours de ces dernières années. Nous avons également choisi de fusionner les signaux obtenus avec l'ensemble des ions précurseurs afin de profiter de l'augmentation de la sélectivité apportée par l'ionisation à double polarité.

B. Traitements statistiques des données

Après le prétraitement des données et l'obtention d'un tableau de données corrigées et nettoyées, ces données peuvent alors être explorées. De nombreuses méthodes statistiques existent et se distinguent en plusieurs groupes caractérisés par le nombre de variables et le nombre d'échantillons pris en compte dans l'analyse : des méthodes univariées, bivariées aux méthodes multivariées. Les méthodes univariées analysent une variable sans tenir compte des autres et des interactions existantes entre elles. Les méthodes bivariées ont pour but de mettre en relation deux informations différentes afin de vérifier si elles sont ou non liées. L'analyse multivariée regroupe les méthodes statistiques qui s'attachent à l'observation et au traitement simultané de plusieurs variables statistiques³⁷⁹. Au sein de ces trois groupes, on trouve des méthodes dites « exploratoires » et des méthodes dites « confirmatoires ». De plus, au-delà de 2 variables, on peut avoir sur chacune des rôles différents : des

variables à expliquer ou des variables explicatives. Cette différenciation conduit à parler de méthodes supervisées ou non-supervisées que nous décrirons plus loin.

Les méthodes exploratoires n'utilisent que les données et s'attachent à fournir des éléments graphiques pour décrire les données sans variables cibles à expliquer³⁸⁰. Parmi ces méthodes, on retrouve l'analyse en composantes principales (ACP) qui est l'une des méthodes d'analyse de données multivariées les plus fréquemment utilisées³⁸¹ dans l'exploration de liens linéaires (ou corrélations) entre plusieurs variables. Elle permet d'étudier des ensembles de données multidimensionnelles avec des variables quantitatives ayant en général toutes le même rôle, c'est une méthode non supervisée. Le principe de l'ACP consiste à transformer des variables corrélées (liées entre elles) en nouvelles variables décorrélées appelées composantes principales. Par construction on peut également réduire le nombre de variables pertinentes dans l'analyse en sélectionnant un nombre approprié de composantes principales. L'ACP a donc pour objectif d'extraire les informations pertinentes d'un tableau de données et de réduire la dimensionnalité des données en conservant l'information présente³⁸¹. Elle permet de visualiser la répartition des échantillons dans un espace à une, deux ou trois dimensions (les dimensions sont les composantes principales). L'ACP permet également de voir la structure du jeu de données selon un facteur qualitatif des échantillons étudiés, par exemple leur origine, ou selon un facteur analytique tel que l'ordre d'analyse des échantillons qui est introduit en variable illustrative. Dans les méthodes exploratoires, on retrouve également la classification non-supervisée (clustering) qui vise à classer des échantillons les uns par rapport aux autres. C'est le cas de la classification ascendante hiérarchique (CAH) qui assemble ou dissocie successivement les ensembles d'échantillons et de la classification par la méthode des nuées dynamiques (k-means)³⁸².

Les méthodes confirmatoires sont des méthodes basées sur des tests statistiques et des modèles (elles sont alors supervisées ou explicatives). Certaines de ces dernières méthodes visent à expliquer la répartition ou le classement des individus (échantillons) selon une variable qualitative à expliquer (dites explicatives) en tenant compte de l'ensemble des autres variables du jeu de données³⁸⁰. Dans cette approche, l'algorithme effectue des prédictions itératives sur les données afin de discriminer les groupes étiquetés³⁸³. Pour ces analyses, les variables descriptives des échantillons sont renseignées et participent à la construction du modèle statistique. De plus, les méthodes explicatives peuvent être utilisées dans un but prédictif et servent alors au développement de modèle d'apprentissage. Les différents algorithmes des méthodes supervisées génèrent ainsi une fonction qui associe les entrées aux sorties souhaitées. Ces méthodes mettent en évidence les données qui expliquent les différences entre des groupes (ou classes). Certaines méthodes confirmatoires peuvent également être exploratoires c'est le cas de la PLS-DA (Partial Least Square Discriminant Analysis) qui fournit des éléments graphiques décrivant l'organisation des données en plus du modèle statistique.

Le choix des méthodes d'analyses se fait en fonction de l'objectif recherché (explorer, décrire ou expliquer) mais également du type de variables disponibles (quantitatives ou qualitatives), de leur nombre et de leur rôle. Le choix des tests statistiques dépend de la distribution de probabilité de la ou des variables : des tests paramétriques seront utilisés lorsque la distribution suit une loi statistique (généralement la loi normale) alors que des tests non paramétriques seront choisis pour les populations ne suivant pas une loi statistique. Un paramètre important pour la sélection des méthodes statistiques à appliquer est la complexité et la taille du jeu de données étudié.

Dans le domaine de l'agroalimentaire et pour l'interprétation des données obtenues par l'analyse des COV, diverses méthodes statistiques sont utilisées. Cette diversité de méthode est liée aux nombreux logiciels statistiques disponibles et dépend également des affinités des analystes pour les méthodes statistiques disponibles. Le [Tableau 13](#) regroupe les principales méthodes utilisées pour l'analyse ciblée ou non ciblée des COV dans les produits alimentaires.

1. Traitements statistiques pour l'analyse ciblée des COV

L'analyse ciblée des COV s'attache à mesurer précisément et spécifiquement quelques COV (de l'ordre de la dizaine). Les tableaux de données générés par ces analyses sont assez simples à explorer avec un nombre d'individus supérieurs au nombre de variables. D'un point de vue statistique, il est souhaitable que le nombre d'individus soit nettement supérieur au nombre de variables utilisées pour décrire ces ensembles³⁸⁴. En théorie, le nombre d'individus doit augmenter de façon exponentielle avec le nombre de variables si l'on veut faire des inférences sur les données à savoir utiliser des méthodes confirmatoires. La plupart des analyses ciblées respectent les critères théoriques statistiques présents derrière chaque méthode. Le choix des méthodes statistiques utilisées (univariées, bivariées ou multivariées, paramétriques ou non, exploratoires ou explicatives) pour interpréter des données de COV cibles s'effectuera en fonction de l'objectif recherché, du nombre de variables à étudier et de la dispersion des variables. De plus, le choix dépendra également du logiciel statistique utilisé et de l'affinité des analystes pour certaines méthodes.

2. Traitements statistiques pour l'analyse non ciblée des COV

Le développement des approches non ciblées d'analyse des COV et le développement des techniques de spectrométrie de masse à introduction directe ont confronté les analystes à l'obtention de données massives (méga-données, ou en anglais « big data »). En effet, dans ces approches on ne s'attache plus à quantifier spécifiquement certains composés mais à traiter l'intégralité des informations obtenues par l'analyse des COV, sans choix a priori des informations à traiter. Ces méthodes génèrent des milliers voire des millions de variables. De plus, la rapidité des nouvelles technologies permet des analyses d'échantillons à haut débit. Les big data sont donc des données massives en termes de volume, d'intensité et de complexité qui dépassent la capacité de gestion et d'analyse des outils logiciels standards³⁸⁶. Mais ces masses de données non structurées n'apportent une valeur ajoutée que si elles sont traitées et analysées.

Ces données issues d'analyses non ciblées des COV sont comparables aux données obtenues par les méthodes omiques (génomique, protéomique et métabolomique) où le nombre de variables est largement supérieur au nombre d'individus³⁸⁴. Par conséquent, la réduction de la dimensionnalité de ces données avant de procéder aux tests statistiques multiples est fortement recommandée. Les analyses statistiques multivariées le permettent, elles ont la propriété de réduire la dimension des données en utilisant des variables synthétiques (composantes), qui sont définies comme des combinaisons linéaires de toutes les variables initiales³⁸⁰.

Le défi apporté par les méthodes omiques ou de « big data » est l'adaptation des méthodes statistiques classiques aux nouveaux types de données produites. En effet, deux obstacles informatiques existent : 1) les données peuvent être trop volumineuses pour être stockées dans la mémoire d'un ordinateur ; 2) la tâche de calcul peut prendre trop de temps pour atteindre les résultats.

Ces obstacles peuvent être surmontés grâce au développement de nouvelles méthodologies statistiques et/ou informatiques³⁸⁶. Pour traiter les données big data obtenues par les sciences omiques, de nouveaux logiciels comme IOAT^{387,388} et Perseus³⁸⁹ mais également de nouveaux packages implémentés dans des logiciels usuels comme le package Mixomics^{390,391} dans R et le package / module Scikit-learn^{392,393} dans Python ont été créés.

Méthodes	Nom	Principe	Supervisée ou non	Ref
Exploratoires	ACP	Permet de visualisation des données dans un espace multivarié, par compression des données, en réduisant la dimensionnalité et en supprimant le bruit	Non supervisée	141,145,147, 148,270,272, 277– 279,287,305, 349,355,357 –360,370
	CAH	Méthode qui assemble ou dissocie successivement les ensembles d'échantillons, elle permet de grouper des individus (échantillons) dans des classes (clusters) en fonction de similarités ou de différences entre les uns et les autres	Non supervisée	149,270,272, 278,287,349, 361
Confirmatoires	ANOVA , MANOVA	Permet de tester la présence de différences significatives ou non entre des moyennes de deux populations ou plus	Supervisée	133,141,147, 150,277,287, 291,305,349, 357,361,385
	Méthodes de régression (MLR, PCR, PLS)	L'analyse de régression calcule la relation estimée entre une variable dépendante et une ou plusieurs variables explicatives.	Supervisée	94,97,97,144 ,150,270,355 ,360,385
	Linear discriminant e analysis (LDA)	Permet d'expliquer et de prédire l'appartenance d'un individu à une classe (groupe) prédéfinie à partir de ses caractéristiques mesurées	Supervisée	145,149,270
	Sélection de variables (iPLS ; VIP ; UVE-PLS ...)	Identifie les variables les plus discriminantes et élimine les variables non pertinentes	Supervisée	144,147,149, 287
Confirmatoires et exploratoires	SIMCA	Méthode statistique de classification supervisée des données	Supervisée	97,144,147,1 50,351,353,3 59
	PLS-DA	Méthode de classification et de discrimination. Elle permet de savoir si les informations quantitatives caractérisent suffisamment précisément les échantillons pour les affecter à la classe à laquelle ils appartiennent et également de classer un nouvel échantillon de classe inconnue a priori	Supervisée	141,144,145, 147,270,272, 279,287,301, 358,359

Tableau 13 : Principales méthodes statistiques utilisées pour interpréter des données obtenues par l'analyse des COV dans des études agroalimentaires.

ACP : Analyse en composante principale ; CAH : Classification ascendante hiérarchique ; ANOVA : Analysis of Variance ; MANOVA : Multivariate ; MLR : Multiple Linear Regression ; PCR : Principal Component Regression ; PLS : Partial Least Square ; iPLS : interval Partial Least Square ; VIP : Variable Important Projection ; UVE-PLS : Uniformative Variable Elimination-Partial Least Square ; SIMCA : Soft independent modelling of class analogies ; PLS-DA : Partial Least Square-Discriminante Analysis.

C. Conclusion

Pour traiter les empreintes volatiles obtenues par l'analyse non ciblée des COV dans des produits alimentaires par des méthodes HS-SPME-GC-MS-FID et HS-SIFT-MS, notre choix s'est porté sur l'utilisation du package MixOmics, spécialisé pour l'analyse des données omiques, du logiciel open source R³⁹¹. En effet, MixOmics est un package R convivial dédié à l'exploration, l'exploitation, l'intégration et la visualisation de grands ensembles de données³⁹⁰. Il offre des fonctionnalités attrayantes telles que (i) des visualisations perspicaces avec réduction de la dimension, (ii) l'identification de signatures moléculaires et (iii) une utilisation améliorée avec des appels communs à toutes les méthodes de visualisation et d'évaluation des performances.

IV. Conclusion générale

Cette analyse bibliographique a permis de choisir les méthodes qui seront utilisées pour répondre aux objectifs de la thèse. Pour développer l'acquisition des empreintes volatiles par SIFT-MS et GC-MS/FID, l'échantillonnage des COV par espace de tête a été choisi pour les deux méthodes avec une pré-concentration SPME pour l'analyse GC-MS/FID. Afin d'avoir des résultats répétables et représentant le maximum de composés volatils, nous avons choisi d'effectuer les mesures une fois les conditions d'équilibre thermodynamique atteintes pour chaque méthode d'échantillonnage. Le mode d'acquisition des données choisi pour les deux instruments est le mode full scan permettant de détecter un maximum de composés sans à priori sur les molécules recherchées. Pour le traitement des données, différents systèmes seront étudiés pour chaque méthode : le retraitement des données full scans obtenus en analyse ciblée, traditionnellement utilisée, et le traitement de l'intégralité du signal c'est-à-dire une analyse non ciblée. Pour les données GC-MS/FID, un logiciel disponible sur le marché sera utilisé, en revanche, pour le traitement des données SIFT-MS, un algorithme sera développé en raison de l'absence de logiciel existant. Cet algorithme a pour objectif de rendre l'extraction et le traitement des données faciles et automatisables. Enfin, pour interpréter les données, nous avons choisi d'utiliser des méthodes statistiques supervisées et non-supervisées dédiées aux méthodes omiques, pour lesquelles le nombre des variables est plus grand que le nombre d'échantillons étudiés. L'ensemble de ce développement est expliqué dans le chapitre suivant.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. Lytou, A. E., Panagou, E. Z. & Nychas, G.-J. E. Volatilomics for food quality and authentication. *Curr. Opin. Food Sci.* S2214799319300724 (2019) doi:10.1016/j.cofs.2019.10.003.
2. Starowicz, M. Analysis of Volatiles in Food Products. *Separations* **8**, 157 (2021).
3. Tylewicz, U., Inchingolo, R. & Rodriguez-Estrada, M. T. Food Aroma Compounds. in *Nutraceutical and Functional Food Components* 297–334 (Elsevier, 2017). doi:10.1016/B978-0-12-805257-0.00009-0.
4. Siegmund, B. Biogenesis of aroma compounds. in *Flavour Development, Analysis and Perception in Food and Beverages* 127–149 (Elsevier, 2015). doi:10.1016/B978-1-78242-103-0.00007-2.
5. Cadwallader, K. R. Aromas. in *Encyclopedia of Food Chemistry* (eds. Melton, L., Shahidi, F. & Varelis, P.) 22–29 (Academic Press, 2019). doi:10.1016/B978-0-08-100596-5.21623-5.
6. Takeo, T. & Tsushida, T. Changes in lipoxygenase activity in relation to lipid degradation in plucked tea shoots. *Phytochemistry* **19**, 2521–2522 (1980).
7. Pérez, A. G. & Sanz, C. Formation of fruit flavour. in *Fruit and Vegetable Flavour* 41–70 (Elsevier, 2008). doi:10.1533/9781845694296.2.41.
8. Ames, J. M. THE MAILLARD REACTION. *Biochem. Food Proteins* 99–153 (1992) doi:https://doi-org.inc.bib.cnrs.fr/10.1007/978-1-4684-9895-0_4.
9. Nursten, H. E. Recent developments in studies of the maillard reaction. *Food Chem.* **6**, 263–277 (1981).
10. Jousse, F., Jongen, T., Agterof, W., Russell, S. & Braat, P. Simplified Kinetic Scheme of Flavor Formation by the Maillard Reaction. *J. Food Sci.* **67**, 2534–2542 (2002).
11. Molimard, P. & Spinnler, H. E. Review: Compounds Involved in the Flavor of Surface Mold-Ripened Cheeses: Origins and Properties. *J. Dairy Sci.* **79**, 169–184 (1996).
12. Labuda, I. Flavor Compounds. in *Encyclopedia of Microbiology* 305–320 (Elsevier, 2009). doi:10.1016/B978-012373944-5.00148-6.
13. Maga, J. A. & Katz, I. Furans in foods. *C R C Crit. Rev. Food Sci. Nutr.* **11**, 355–400 (1979).
14. Reineccius, G. Off-Flavors in Foods. *Crit. Rev. Food Sci. Nutr.* **29**, 381–402 (1991).
15. Ridgway, K. 4 - Analysis of taints and off-flavours. in *Flavour Development, Analysis and Perception in Food and Beverages* (eds. Parker, J. K., Elmore, J. S. & Methven, L.) 63–81 (Woodhead Publishing, 2015). doi:10.1016/B978-1-78242-103-0.00004-7.
16. Pegg, R. B. & Shahidi, F. Off Flavors and Rancidity in Foods. in *Handbook of Meat, Poultry and Seafood Quality* 127–139 (John Wiley & Sons, Ltd, 2012). doi:10.1002/9781118352434.ch9.
17. Bonneau, M. Compounds responsible for boar taint, with special emphasis on androstenone: A review. *Livest. Prod. Sci.* **9**, 687–705 (1982).
18. Simpson, R. F. & Sefton, M. A. Origin and fate of 2,4,6-trichloroanisole in cork bark and wine corks. *Aust. J. Grape Wine Res.* **13**, 106–116 (2007).
19. Peña-Neira, A. *et al.* Presence of cork-taint responsible compounds in wines and their cork stoppers. *Eur. Food Res. Technol.* **211**, 257–261 (2000).
20. Houle, S. *et al.* Geosmin causes off-flavour in arctic charr in recirculating aquaculture systems. *Aquac. Res.* **42**, 360–365 (2011).
21. Tucker, C. S. Off-Flavor Problems in Aquaculture. *Rev. Fish. Sci.* **8**, 45–88 (2000).
22. Rocha, S. M., Costa, C. P. & Martins, C. Aroma Clouds of Foods: A Step Forward to Unveil Food Aroma Complexity Using GC × GC. *Front. Chem.* **10**, (2022).
23. DeMaria, S. & Ngai, J. The cell biology of smell. *J. Cell Biol.* **191**, 443–452 (2010).

24. Marques, C., Correia, E., Dinis, L.-T. & Vilela, A. An Overview of Sensory Characterization Techniques: From Classical Descriptive Analysis to the Emergence of Novel Profiling Methods. *Foods* **11**, 255 (2022).
25. ISO/TC 34/SC 12. *Analyse sensorielle — Évaluation (détermination et vérification) de la durée de conservation des produits alimentaires*. (2015).
26. ISO/TC 34/SC 12. *Analyse sensorielle — Méthodes pour évaluer les modifications de la flaveur des aliments causées par l’emballage*. (2003).
27. Stevenson, R. J. Perceptual learning with odors: implications for psychological accounts of odor quality perception. *Psychon. Bull. Rev.* **8**, 708–712 (2001).
28. Jönsson, F. U., Tchekhova, A., Lönnér, P. & Olsson, M. J. A Metamemory Perspective on Odor Naming and Identification. *Chem. Senses* **30**, 353–365 (2005).
29. AFNOR NF X43-103. *Qualité de l’air. Mesurage de l’odeur d’une atmosphère gazeuse*. (1996).
30. Étiévant, P. X. Pathologies et troubles de l’odorat : quels outils en pratique ? *Nutr. Clin. Métabolisme* **22**, 113–120 (2008).
31. Ramalho, O. Vers une métrologie olfactive de la qualité de l’air intérieur. Correspondances entre les données de l’analyse sensorielle, de l’analyse chimique et d’un “nez électronique”. 194.
32. Medjkoune, M. Vers une approche non orientée pour l’évaluation de la qualité des odeurs. (Université de Montpellier, 2018).
33. Bi, J. & Ennis, D. M. SENSORY THRESHOLDS: CONCEPTS AND METHODS. *J. Sens. Stud.* **13**, 133–148 (1998).
34. Parker, J. K. Introduction to aroma compounds in foods. in *Flavour Development, Analysis and Perception in Food and Beverages* 3–30 (Elsevier, 2015). doi:10.1016/B978-1-78242-103-0.00001-1.
35. Guichard, E. Interactions between flavor compounds and food ingredients and their influence on flavor perception. *Food Rev. Int.* **18**, 49–70 (2002).
36. Haese, G., Humeau, P., De Oliveira, F., Le Callet, P. & Le Cloirec, P. Tastes and Odors of Water—Quantifying Objective Analyses: A Review. *Crit. Rev. Environ. Sci. Technol.* **44**, 2455–2501 (2014).
37. Frijters, J. E. R. Psychophysical Models for Mixtures of Tastants and Mixtures of Odorants. *Ann. N. Y. Acad. Sci.* **510**, 67–78 (1987).
38. Marshall, K., Laing, D., Jinks, A., Effendy, J. & Hutchinson, I. Perception of temporal order and the identification of components in taste mixtures. *Physiol. Behav.* **83**, 673–681 (2005).
39. Livermore, A. & Laing, D. G. The influence of chemical complexity on the perception of multicomponent odor mixtures. *Percept. Psychophys.* **60**, 650–661 (1998).
40. Greenberg, M. I., Curtis, J. A. & Vearrier, D. The perception of odor is not a surrogate marker for chemical exposure: a review of factors influencing human odor perception. *Clin. Toxicol.* **51**, 70–76 (2013).
41. Koskinen, S. & Tuorila, H. Performance on an odor detection and identification test as a predictor of ortho- and retronasal odor intensity ratings in the young and elderly. *Food Qual. Prefer.* **16**, 383–392 (2005).
42. Doty, R. L. *et al.* Smell identification ability: changes with age. *Science* **226**, 1441–1443 (1984).
43. Doty, R. L. & Cameron, E. L. Sex differences and reproductive hormone influences on human odor perception. *Physiol. Behav.* **97**, 213–228 (2009).
44. Dalton, P., Doolittle, N. & Breslin, P. A. S. Gender-specific induction of enhanced sensitivity to odors. *Nat. Neurosci.* **5**, 199–200 (2002).

45. Kuga, M., Ikeda, M. & Suzuki, K. Gustatory Changes Associated With the Menstrual Cycle. *Physiol. Behav.* **66**, 317–322 (1999).
46. Ferdenzi, C. *et al.* Individual Differences in Verbal and Non-Verbal Affective Responses to Smells: Influence of Odor Label Across Cultures. *Chem. Senses* bjw098 (2016) doi:10.1093/chemse/bjw098.
47. Mannella, J., Bobowski, N. & Liem, D. G. Taste and Smell. in *Swaiman's Pediatric Neurology* 58–64 (2017).
48. Labows, J. N. & Wysocki, C. J. Individual Differences in Odor Perception. *Perfum. Flavorist* **9**, 5 (1984).
49. Agarwal, B. FACTORS AFFECTING QUALITY OF TEA DURING PROCESSING. *Sri Lanka J. Tea Sci.* **9** (1989).
50. Zheng, X.-Q., Li, Q.-S., Xiang, L.-P. & Liang, Y.-R. Recent Advances in Volatiles of Teas. *Molecules* **21**, 338 (2016).
51. Zeng, L., Watanabe, N. & Yang, Z. Understanding the biosyntheses and stress response mechanisms of aroma compounds in tea (*Camellia sinensis*) to safely and effectively improve tea aroma. *Crit. Rev. Food Sci. Nutr.* **59**, 2321–2334 (2019).
52. Guo, X., Schwab, W., Ho, C.-T., Song, C. & Wan, X. Characterization of the aroma profiles of oolong tea made from three tea cultivars by both GC–MS and GC-IMS. *Food Chem.* **376**, 131933 (2022).
53. Ho, C.-T., Zheng, X. & Li, S. Tea aroma formation. *Food Sci. Hum. Wellness* **4**, 9–27 (2015).
54. Sanderson, G. W. & Graham, H. N. On the Formation of Black Tea Aroma. **10**.
55. Gautier, L. *Portraits de Thés. Voyage dans 40 pays producteurs.* (2018).
56. Takeo, T. Production of linalol and geraniol by hydrolytic breakdown of bound forms in disrupted tea shoots. *Phytochemistry* **20**, 2145–2147 (1981).
57. Takeo, T. Variation in amounts of linalol and geraniol produced in tea shoots by mechanical injury. *Phytochemistry* **20**, 2149–2151 (1981).
58. Saijō, R. & Takeo, T. The Formation of Aldehydes from Amino Acids by Tea Leaves Extracts. *Agric. Biol. Chem.* **34**, 227–233 (1970).
59. Feng, Z. *et al.* Tea aroma formation from six model manufacturing processes. *Food Chem.* **285**, 347–354 (2019).
60. Schuh, C. & Schieberle, P. Characterization of the Key Aroma Compounds in the Beverage Prepared from Darjeeling Black Tea: Quantitative Differences between Tea Leaves and Infusion. *J. Agric. Food Chem.* **54**, 916–924 (2006).
61. Wang, K. *et al.* Comparison of catechins and volatile compounds among different types of tea using high performance liquid chromatograph and gas chromatograph mass spectrometer: Catechins and volatile compounds of tea. *Int. J. Food Sci. Technol.* **46**, 1406–1412 (2011).
62. Qin, Z. *et al.* Evaluation of Chinese tea by the electronic nose and gas chromatography–mass spectrometry: Correlation with sensory properties and classification according to grade level. *Food Res. Int.* **53**, 864–874 (2013).
63. Zhu, J. *et al.* Comparison of Aroma-Active Volatiles in Oolong Tea Infusions Using GC–Olfactometry, GC–FPD, and GC–MS. *J. Agric. Food Chem.* **63**, 7499–7510 (2015).
64. Kumazawa, K. & Masuda, H. Identification of potent odorants in different green tea varieties using flavor dilution technique. *J. Agric. Food Chem.* **50**, 5660–5663 (2002).
65. Cheng, Y., Huynh-Ba, T., Blank, I. & Robert, F. Temporal changes in aroma release of Longjing tea infusion: interaction of volatile and nonvolatile tea components and formation of 2-butyl-2-octenal upon aging. *J. Agric. Food Chem.* **56**, 2160–2169 (2008).

66. Lv, H.-P. *et al.* Aroma characterisation of Pu-erh tea using headspace-solid phase microextraction combined with GC/MS and GC-olfactometry. *Food Chem.* **130**, 1074–1081 (2012).
67. Fox, P. F. Cheese: An Overview. in *Cheese: Chemistry, Physics and Microbiology* (ed. Fox, P. F.) 1–36 (Springer US, 1993). doi:10.1007/978-1-4615-2650-6_1.
68. Noni, I. D. & Battelli, G. Terpenes and fatty acid profiles of milk fat and “Bitto” cheese as affected by transhumance of cows on different mountain pastures. *Food Chem.* **109**, 299–309 (2008).
69. Revello Chion, A. *et al.* Variation of fatty acid and terpene profiles in mountain milk and “Toma piemontese” cheese as affected by diet composition in different seasons. *Food Chem.* **121**, 393–399 (2010).
70. Aprea, E. *et al.* Effects of grazing cow diet on volatile compounds as well as physicochemical and sensory characteristics of 12-month-ripened Montasio cheese. *J. Dairy Sci.* **99**, 6180–6190 (2016).
71. W Park, Y., Jeanjulien, C. & Siddique, A. Factors Affecting Sensory Quality of Goat Milk Cheeses: A Review. *Adv. Dairy Res.* **05**, (2017).
72. Coulon, J.-B., Delacroix-Buchet, A., Martin, B. & Pirisi, A. Relationships between ruminant management and sensory characteristics of cheeses: a review. *Le Lait* **84**, 221–241 (2004).
73. Nacef, M. M. Maroilles fermiers et industriels : quelles sont les différences ? : Une approche pluridisciplinaire allant du consommateur aux caractérisations sensorielles, physico-chimiques et microbiologiques. (Université de Lille (2018-2021), 2018).
74. Murtaza, M. A., Ur-Rehman, S., Anjum, F. M., Huma, N. & Hafiz, I. Cheddar Cheese Ripening and Flavor Characterization: A Review. *Crit. Rev. Food Sci. Nutr.* **54**, 1309–1321 (2014).
75. Andiç, S., Tunçtürk, Y. & Boran, G. Changes in Volatile Compounds of Cheese. in *Processing and Impact on Active Components in Food* 231–239 (Elsevier, 2015). doi:10.1016/B978-0-12-404699-3.00028-7.
76. McSweeney, P. L. H. & Sousa, M. J. Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. *Le Lait* **80**, 293–324 (2000).
77. Marilley, L. & Casey, M. G. Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. *Int. J. Food Microbiol.* **90**, 139–159 (2004).
78. Smit, G., Smit, B. A. & Engels, W. J. M. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* **29**, 591–610 (2005).
79. Collins, Y. F., McSweeney, P. L. H. & Wilkinson, M. G. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *Int. Dairy J.* **13**, 841–866 (2003).
80. Zuljan, F. A. *et al.* Lactic acid bacteria decarboxylation reactions in cheese. *Int. Dairy J.* **62**, 53–62 (2016).
81. Arqués, J. L., Garde, S., Fernández-García, E., Gaya, P. & Nuñez, M. Volatile Compounds, Odor, and Aroma of La Serena Cheese High-Pressure Treated at Two Different Stages of Ripening. *J. Dairy Sci.* **90**, 3627–3639 (2007).
82. Ortigosa, M., Torre, P. & Izco, J. M. Effect of Pasteurization of Ewe’s Milk and Use of a Native Starter Culture on the Volatile Components and Sensory Characteristics of Roncal Cheese. *J. Dairy Sci.* **84**, 1320–1330 (2001).
83. Gómez-Torres, N., Garde, S., Peirotén, Á. & Ávila, M. Impact of *Clostridium* spp. on cheese characteristics: Microbiology, color, formation of volatile compounds and off-flavors. *Food Control* **56**, 186–194 (2015).
84. Izco, J. M., Irigoyen, A., Torre, P. & Barcina, Y. Effect of the activity levels of the added proteolytic enzyme mixture on free amino acids in ripening Ossau-Iraty cheese. *J. Chromatogr. A* **881**, 69–79 (2000).

85. Rijnen, L. *et al.* Inactivation of lactococcal aromatic aminotransferase prevents the formation of floral aroma compounds from aromatic amino acids in semi-hard cheese. *Int. Dairy J.* **9**, 877–885 (1999).
86. Delgado, F. J., González-Crespo, J., Cava, R. & Ramírez, R. Formation of the aroma of a raw goat milk cheese during maturation analysed by SPME–GC–MS. *Food Chem.* **129**, 1156–1163 (2011).
87. Feng, J., Zheng, Y., Zhang, X., Zhou, R. & Ma, M. Effect of supercritical carbon dioxide on bacterial community, volatile profiles and quality changes during storage of Mongolian cheese. *Food Control* **143**, 109225 (2023).
88. Curioni, P. M. G. & Bosset, J. O. Key odorants in various cheese types as determined by gas chromatography-olfactometry. *Int. Dairy J.* **12**, 959–984 (2002).
89. Avsar, Y. K. *et al.* Characterization of Nutty Flavor in Cheddar Cheese. *J. Dairy Sci.* **87**, 1999–2010 (2004).
90. Bergamaschi, M. & Bittante, G. From milk to cheese: Evolution of flavor fingerprint of milk, cream, curd, whey, ricotta, scotta, and ripened cheese obtained during summer Alpine pasture. *J. Dairy Sci.* **101**, 3918–3934 (2018).
91. Bontinis, Th. G., Mallatou, H., Pappa, E. C., Massouras, Th. & Alichanidis, E. Study of proteolysis, lipolysis and volatile profile of a traditional Greek goat cheese (Xinotyri) during ripening. *Small Rumin. Res.* **105**, 193–201 (2012).
92. Bezerra, T. K. A. *et al.* Optimization of the HS-SPME-GC/MS technique for the analysis of volatile compounds in caprine Coalho cheese using response surface methodology. *Food Sci. Technol.* **36**, 103–110 (2016).
93. Bergamaschi, M. *et al.* Proton transfer reaction time-of-flight mass spectrometry: A high-throughput and innovative method to study the influence of dairy system and cow characteristics on the volatile compound fingerprint of cheeses. *J. Dairy Sci.* **98**, 8414–8427 (2015).
94. Castada, H. Z., Wick, C., Harper, W. J. & Barringer, S. Headspace quantification of pure and aqueous solutions of binary mixtures of key volatile organic compounds in Swiss cheeses using selected ion flow tube mass spectrometry: Headspace quantification of VOCs using SIFT-MS. *Rapid Commun. Mass Spectrom.* **29**, 81–90 (2015).
95. Bezerra, T. K. A. *et al.* Volatile profile in goat coalho cheese supplemented with probiotic lactic acid bacteria. *LWT - Food Sci. Technol.* **76**, 209–215 (2017).
96. Boltar, I., Majhenič, A. Č., Jarni, K., Jug, T. & Kralj, M. B. Research of volatile compounds in cheese affected by different technological parameters. *J Food Nutr Res* **58**, 10 (2019).
97. Castada, H. Z., Wick, C., Taylor, K. & Harper, W. J. Analysis of Selected Volatile Organic Compounds in Split and Nonsplit Swiss Cheese Samples Using Selected-Ion Flow Tube Mass Spectrometry (SIFT-MS): Analysis of selected volatile organic.... *J. Food Sci.* **79**, C489–C498 (2014).
98. Bertolino, M., Dolci, P., Giordano, M., Rolle, L. & Zeppa, G. Evolution of chemico-physical characteristics during manufacture and ripening of Castelmagno PDO cheese in wintertime. *Food Chem.* **129**, 1001–1011 (2011).
99. Attaie, R. Quantification of volatile compounds in goat milk Jack cheese using static headspace gas chromatography. *J. Dairy Sci.* **92**, 2435–2443 (2009).
100. Battelli, G. *et al.* Modifications of the volatile and nonvolatile metabolome of goat cheese due to adjunct of non-starter lactic acid bacteria. *LWT* **116**, 108576 (2019).
101. Bosset, J. O. & Gauch, R. Comparison of the volatile flavour compounds of six european 'AOC' cheeses by using a new dynamic headspace GC-MS method. *Int. Dairy J.* **3**, 359–377 (1993).

102. Di Cagno, R. *et al.* Comparison of the microbiological, compositional, biochemical, volatile profile and sensory characteristics of three Italian PDO ewes' milk cheeses. *Int. Dairy J.* **13**, 961–972 (2003).
103. Dirinck, P. & De Winne, A. Flavour characterisation and classification of cheeses by gas chromatographic–mass spectrometric profiling. *J. Chromatogr. A* **847**, 203–208 (1999).
104. Delgado, F. J., González-Crespo, J., Cava, R., García-Parra, J. & Ramírez, R. Characterisation by SPME–GC–MS of the volatile profile of a Spanish soft cheese P.D.O. Torta del Casar during ripening. *Food Chem.* **118**, 182–189 (2010).
105. *Standardized Human Olfactory Thresholds.* (Oxford University Press, 1990).
106. Saqqa, G. S. R. A. What to Know about Food Flavor? A Review. *Jordan J. Agric. Sci.* **18**, 1–15 (2022).
107. Sales, A., Paulino, B. N., Pastore, G. M. & Bicas, J. L. Biogenesis of aroma compounds. *Curr. Opin. Food Sci.* **19**, 77–84 (2018).
108. Paulino, B. N. *et al.* Recent advances in the microbial and enzymatic production of aroma compounds. *Curr. Opin. Food Sci.* **37**, 98–106 (2021).
109. Starowicz, M. & Zieliński, H. How Maillard Reaction Influences Sensorial Properties (Color, Flavor and Texture) of Food Products? *Food Rev. Int.* **35**, 707–725 (2019).
110. Maga, J. A. The flavor chemistry of wood smoke. *Food Rev. Int.* **3**, 139–183 (1987).
111. LeDoux, M. Analytical methods applied to the determination of pesticide residues in foods of animal origin. A review of the past two decades. *J. Chromatogr. A* **1218**, 1021–1036 (2011).
112. Xu, C.-H. Applications of solid-phase microextraction in food analysis. *Trends Anal. Chem.* **18** (2016).
113. Abdulra'uf, L. B. & Tan, G. H. Multivariate study of parameters in the determination of pesticide residues in apple by headspace solid phase microextraction coupled to gas chromatography–mass spectrometry using experimental factorial design. *Food Chem.* **141**, 4344–4348 (2013).
114. Merib, J., Nardini, G. & Carasek, E. Use of Doehlert design in the optimization of extraction conditions in the determination of organochlorine pesticides in bovine milk samples by HS-SPME. *Anal. Methods* **6**, 3254–3260 (2014).
115. Abdulra'uf, L. B. & Tan, G. H. Chemometric study and optimization of headspace solid-phase microextraction parameters for the determination of multiclass pesticide residues in processed cocoa from Nigeria using gas chromatography/mass spectrometry. *J. AOAC Int.* **97**, 1007–1011 (2014).
116. Hercegová, A., Dömötöröová, M., Hrouzková, S. & Matisová, E. Study on pesticide residues in apples, apple-based baby food, and their behaviour during processing using fast GC–MS multiresidue analysis. *Int. J. Environ. Anal. Chem.* **87**, 957–969 (2007).
117. Abdulra'uf, L. B. & Tan, G. H. Multivariate study of parameters in the determination of pesticide residues in apple by headspace solid phase microextraction coupled to gas chromatography–mass spectrometry using experimental factorial design. *Food Chem.* **141**, 4344–4348 (2013).
118. Melo, A., Aguiar, A., Mansilha, C., Pinho, O. & Ferreira, I. M. P. L. V. O. Optimisation of a solid-phase microextraction/HPLC/Diode Array method for multiple pesticide screening in lettuce. *Food Chem.* **130**, 1090–1097 (2012).
119. Cortés Aguado, S., Sánchez-Morito, N., Garrido Frenich, A., Martínez Vidal, J. L. & Arrebola, F. J. Screening Method for the Determination at Parts Per Trillion Levels of Pesticide Residues in Vegetables Combining Solid-Phase Microextraction and Gas Chromatography-Tandem Mass Spectrometry. *Anal. Lett.* **40**, 2886–2914 (2007).

120. Martins, J., Esteves, C., Simões, T., Correia, M. & Delerue-Matos, C. Determination of 24 Pesticide Residues in Fortified Wines by Solid-Phase Microextraction and Gas Chromatography–Tandem Mass Spectrometry. *J. Agric. Food Chem.* **59**, 6847–6855 (2011).
121. Ravelo-Pérez, L. M., Hernández-Borges, J., Borges-Miquel, T. M. & Rodríguez-Delgado, M. Á. Solid-phase microextraction and sample stacking micellar electrokinetic chromatography for the analysis of pesticide residues in red wines. *Food Chem.* **111**, 764–770 (2008).
122. Carraturo, F. *et al.* Metabolomic profiling of food matrices: Preliminary identification of potential markers of microbial contamination. *J. Food Sci.* **85**, 3467–3477 (2020).
123. Xu, Y., Cheung, W., Winder, C. L. & Goodacre, R. VOC-based metabolic profiling for food spoilage detection with the application to detecting *Salmonella typhimurium*-contaminated pork. *Anal. Bioanal. Chem.* **397**, 2439–2449 (2010).
124. Parlapani, F. F., Mallouchos, A., Haroutounian, S. A. & Boziaris, I. S. Microbiological spoilage and investigation of volatile profile during storage of sea bream fillets under various conditions. *Int. J. Food Microbiol.* **189**, 153–163 (2014).
125. Arroyo-Manzanares, N. *et al.* Head-space gas chromatography coupled to mass spectrometry for the assessment of the contamination of mayonnaise by yeasts. *Food Chem.* **289**, 461–467 (2019).
126. Spadafora, N. D. *et al.* Detection of *Listeria monocytogenes* in cut melon fruit using analysis of volatile organic compounds. *Food Microbiol.* **54**, 52–59 (2016).
127. Lau, O.-W. & Wong, S.-K. Contamination in food from packaging material. *J. Chromatogr. A* **882**, 255–270 (2000).
128. Arvanitoyannis, I. S. & Bosnea, L. Migration of Substances from Food Packaging Materials to Foods. *Crit. Rev. Food Sci. Nutr.* **44**, 63–76 (2004).
129. Kuuliala, L. *et al.* Microbiological, chemical and sensory spoilage analysis of raw Atlantic cod (*Gadus morhua*) stored under modified atmospheres. *Food Microbiol.* **70**, 232–244 (2018).
130. Bhadury, D., Nolvachai, Y., Marriott, P. J., Tanner, J. & Tuck, K. L. Detection of Volatiles from Raw Beef Meat from Different Packaging Systems Using Solid-Phase Microextraction GC–Accurate Mass Spectrometry. *Foods* **10**, 2018 (2021).
131. Lozano, P. R., Miracle, E. R., Krause, A. J., Drake, M. & Cadwallader, K. R. Effect of Cold Storage and Packaging Material on the Major Aroma Components of Sweet Cream Butter. *J. Agric. Food Chem.* **55**, 7840–7846 (2007).
132. Geeraerts, W., Borremans, W., De Vuyst, L., Leroy, F. & Van Kerrebroeck, S. The application of selected ion flow tube-mass spectrometry to follow volatile formation in modified-atmosphere-packaged cooked ham. *Food Res. Int.* **123**, 601–611 (2019).
133. Huang, Y. & Barringer, S. A. Monitoring of Cocoa Volatiles Produced during Roasting by Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS). *J. Food Sci.* **76**, C279–C286 (2011).
134. Soukoulis, C. *et al.* Proton transfer reaction time-of-flight mass spectrometry monitoring of the evolution of volatile compounds during lactic acid fermentation of milk. *Rapid Commun. Mass Spectrom.* **24**, 2127–2134 (2010).
135. Tsevdou, M. *et al.* Monitoring the effect of high pressure and transglutaminase treatment of milk on the evolution of flavour compounds during lactic acid fermentation using PTR-ToF-MS. *Food Chem.* **138**, 2159–2167 (2013).
136. Chen, Y. *et al.* Characteristic fingerprints and volatile flavor compound variations in Liuyang Douchi during fermentation via HS-GC-IMS and HS-SPME-GC-MS. *Food Chem.* **361**, 130055 (2021).

137. Makhoul, S. *et al.* Proton-transfer-reaction mass spectrometry for the study of the production of volatile compounds by bakery yeast starters. *J. Mass Spectrom.* **49**, 850–859 (2014).
138. Yang, W. *et al.* Effect of hot air drying on volatile compounds of *Flammulina velutipes* detected by HS-SPME–GC–MS and electronic nose. *Food Chem.* **196**, 860–866 (2016).
139. Castro-Puyana, M. & Herrero, M. Metabolomics approaches based on mass spectrometry for food safety, quality and traceability. *TrAC Trends Anal. Chem.* **52**, 74–87 (2013).
140. Sharma, S. & Kaushal, P. Food fraud: A menace to public health. *World J. Adv. Res. Rev.* **12**, 466–475 (2021).
141. Perestrelo, R. Untargeted fingerprinting of cider volatiles from different geographical regions by HS-SPME/GC-MS. *Microchem. J.* **9** (2019).
142. Sousa, A. *et al.* Geographical differentiation of apple ciders based on volatile fingerprint. *Food Res. Int.* **137**, 109550 (2020).
143. Bajoub, A. *et al.* A metabolic fingerprinting approach based on selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics: A reliable tool for Mediterranean origin-labeled olive oils authentication. *Food Res. Int.* **106**, 233–242 (2018).
144. Kharbach, M. *et al.* Selected-ion flow-tube mass-spectrometry (SIFT-MS) fingerprinting versus chemical profiling for geographic traceability of Moroccan Argan oils. *Food Chem.* **263**, 8–17 (2018).
145. Donato, F. D. HS-SPME/GC–MS volatile fraction determination and chemometrics for the discrimination of typical Italian Pecorino cheeses. *Microchem. J.* **11** (2021).
146. Machado, A. M., Antunes, M., Miguel, M. G., Vilas-Boas, M. & Figueiredo, A. C. Volatile Profile of Portuguese Monofloral Honeys: Significance in Botanical Origin Determination. *Molecules* **26**, 4970 (2021).
147. Wang, X. *et al.* Untargeted and Targeted Discrimination of Honey Collected by *Apis cerana* and *Apis mellifera* Based on Volatiles Using HS-GC-IMS and HS-SPME-GC–MS. *J. Agric. Food Chem.* **67**, 12144–12152 (2019).
148. Barbosa-Pereira, L., Rojo-Poveda, O., Ferrocino, I., Giordano, M. & Zeppa, G. Assessment of volatile fingerprint by HS-SPME/GC-qMS and E-nose for the classification of cocoa bean shells using chemometrics. *Food Res. Int.* **123**, 684–696 (2019).
149. Beltrán Sanahuja, A., Ramos Santonja, M., Grané Teruel, N., Martín Carratalá, M. L. & Garrigós Selva, M. C. Classification of Almond Cultivars Using Oil Volatile Compound Determination by HS-SPME–GC–MS. *J. Am. Oil Chem. Soc.* **88**, 329–336 (2011).
150. Ozcan-Sinir, G. Detection of adulteration in extra virgin olive oil by selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics. *Food Control* **118**, 107433 (2020).
151. Lewkowska, P., Dymerski, T. & Namiesnik, J. Use of Sensory Analysis Methods to Evaluate the Odor of Food and Outside Air. *Crit. Rev. Environ. Sci. Technol.* **45**, 2208–2244 (2015).
152. Roy, M. & Yadav, B. K. Electronic nose for detection of food adulteration: a review. *J. Food Sci. Technol.* **59**, 846–858 (2022).
153. Sanaeifar, A., ZakiDizaji, H., Jafari, A. & Guardia, M. de la. Early detection of contamination and defect in foodstuffs by electronic nose: A review. *TrAC Trends Anal. Chem.* **97**, 257–271 (2017).
154. Bhattacharyya, N. & Bandhopadhyay, R. Electronic Nose and Electronic Tongue. in *Nondestructive Evaluation of Food Quality: Theory and Practice* (ed. Jha, S. N.) 73–100 (Springer, 2010). doi:10.1007/978-3-642-15796-7_4.

155. Mohd Ali, M., Hashim, N., Abd Aziz, S. & Lasekan, O. Principles and recent advances in electronic nose for quality inspection of agricultural and food products. *Trends Food Sci. Technol.* **99**, 1–10 (2020).
156. Luykx, D. M. A. M. & van Ruth, S. M. An overview of analytical methods for determining the geographical origin of food products. *Food Chem.* **107**, 897–911 (2008).
157. Lehotay, S. J. & Hajslova, J. Application of gas chromatography in food analysis. *Trac-Trends Anal. Chem.* **21**, 686–697 (2002).
158. Georgakopoulos, P., Mylona, A., Athanasopoulos, P., Drosinos, E. H. & Skandamis, P. N. Evaluation of cost-effective methods in the pesticide residue analysis of non-fatty baby foods. *Food Chem.* **115**, 1164–1169 (2009).
159. Vaquero-Fernandez, L. *et al.* Development of a liquid-liquid extraction method for the determination of pyrimethanil, metalaxyl, dichlofluanid and penconazol during the fermentative process of must by GC-NPD. *J. Sci. Food Agric.* **89**, 750–757 (2009).
160. Ye, D.-Q. *et al.* Evolutions of volatile sulfur compounds of Cabernet Sauvignon wines during aging in different oak barrels. *Food Chem.* **202**, 236–246 (2016).
161. Moreira, N., de Pinho, P. G., Santos, C. & Vasconcelos, I. Volatile sulphur compounds composition of monovarietal white wines. *Food Chem.* **123**, 1198–1203 (2010).
162. Li, C. *et al.* Cysteine residues are responsible for the sulfurous off-flavor formed in heated whey protein solutions. *Food Chem. Mol. Sci.* **5**, 100120 (2022).
163. Paolini, M., Tonidandel, L. & Larcher, R. Development, validation and application of a fast GC-FID method for the analysis of volatile compounds in spirit drinks and wine. *Food Control* **136**, 108873 (2022).
164. Anastasaki, E. *et al.* Geographical differentiation of saffron by GC-MS/FID and chemometrics. *Eur. Food Res. Technol.* **229**, 899–905 (2009).
165. Garcia-Vico, L. *et al.* Volatile Compound Profiling by HS-SPME/GC-MS-FID of a Core Olive Cultivar Collection as a Tool for Aroma Improvement of Virgin Olive Oil. *Molecules* **22**, 141 (2017).
166. Charry-Parra, G., DeJesus-Echevarria, M. & Perez, F. J. Beer Volatile Analysis: Optimization of HS/SPME Coupled to GC/MS/FID. *J. Food Sci.* **76**, C205–C211 (2011).
167. Angeloni, S. *et al.* Characterization of Odor-Active Compounds, Polyphenols, and Fatty Acids in Coffee Silverskin. *Molecules* **25**, 2993 (2020).
168. Jirovetz, L., Smith, D. & Buchbauer, G. Aroma compound analysis of *Eruca sativa* (Brassicaceae) SPME headspace leaf samples using GC, GC-MS, and olfactometry. *J. Agric. Food Chem.* **50**, 4643–4646 (2002).
169. Dugheri, S. *et al.* Advanced Solid-Phase Microextraction Techniques and Related Automation: A Review of Commercially Available Technologies. *J. Anal. Methods Chem.* **2022**, 1–15 (2022).
170. Dugheri, S. *et al.* Solid phase microextraction techniques used for gas chromatography: a review. *Acta Chromatogr.* **32**, 1–9 (2020).
171. Hernanz, D., Gallo, V., Recamales, A. F., Melendez-Martinez, A. J. & Heredia, F. J. Comparison of the effectiveness of solid-phase and ultrasound-mediated liquid-liquid extractions to determine the volatile compounds of wine. *Talanta* **76**, 929–935 (2008).
172. Ivanova, V. *et al.* Validation of a Method for Analysis of Aroma Compounds in Red Wine using Liquid-Liquid Extraction and GC-MS. *Food Anal. Methods* **5**, 1427–1434 (2012).
173. Zhang, Y., Fraatz, M. A., Horlamus, F., Quitmann, H. & Zorn, H. Identification of Potent Odorants in a Novel Nonalcoholic Beverage Produced by Fermentation of Wort with Shiitake (*Lentinula edodes*). *J. Agric. Food Chem.* **62**, 4195–4203 (2014).

174. Farajzadeh, M. A., Djozan, D., Mogaddam, M. R. A. & Bamorowat, M. Extraction and preconcentration technique for triazole pesticides from cow milk using dispersive liquid-liquid microextraction followed by GC-FID and GC-MS determinations. *J. Sep. Sci.* **34**, 1309–1316 (2011).
175. Lakew, A., Megersa, N. & Chandravanshi, B. S. Quantitative extraction of multiclass antibiotic residues in cow milk utilizing salting-out assisted liquid-liquid extraction followed by liquid chromatographic determination. *Int. J. Environ. Anal. Chem.* doi:10.1080/03067319.2022.2044029.
176. Hamed, A. M., Arroyo-Manzanares, N., Garcia-Campana, A. M. & Gamiz-Gracia, L. Determination of Fusarium toxins in functional vegetable milks applying salting-out-assisted liquid-liquid extraction combined with ultra-high-performance liquid chromatography tandem mass spectrometry. *Food Addit. Contam. Part -Chem. Anal. Control Expo. Risk Assess.* **34**, 2033–2041 (2017).
177. Zapf, A., Heyer, R. & Stan, H. Rapid Micro Liquid-Liquid-Extraction Method for Trace Analysis of Organic Contaminants in Drinking-Water. *J. Chromatogr. A* **694**, 453–461 (1995).
178. ISO/TC 34/SC 11. *Corps gras d'origines animale et végétale — Détermination des hydrocarbures aromatiques polycycliques.* (2006).
179. Virot, M., Tomao, V., Colnagui, G., Visinoni, F. & Chemat, F. New microwave-integrated Soxhlet extraction an advantageous tool for the extraction of lipids from food products. *J. Chromatogr. A* **1174**, 138–144 (2007).
180. Ahmad, U. K., Ujang, Z., Woon, C. H., Indran, S. & Mian, M. N. Development of extraction procedures for the analysis of polycyclic aromatic hydrocarbons and organochlorine pesticides in municipal sewage sludge. *Water Sci. Technol.* **50**, 137–144 (2004).
181. Priego-Capote, F., Ruiz-Jimenez, J. & Luque De Castro, M. D. Identification and quantification of trans fatty acids in bakery products by gas chromatography-mass spectrometry after focused microwave Soxhlet extraction. *Food Chem.* **100**, 859–867 (2007).
182. Smith, K. E. C., Northcott, G. L. & Jones, K. C. Influence of the extraction methodology on the analysis of polycyclic aromatic hydrocarbons in pasture vegetation. *J. Chromatogr. A* **1116**, 20–30 (2006).
183. Helaleh, M. I. H., Al-Omair, A., Nisar, A. & Gevao, B. Validation of various extraction techniques for the quantitative analysis of polycyclic aromatic hydrocarbons in sewage sludges using gas chromatography-ion trap mass spectrometry. *J. Chromatogr. A* **1083**, 153–160 (2005).
184. Sun, H., Ge, X., Lv, Y. & Wang, A. Application of accelerated solvent extraction in the analysis of organic contaminants, bioactive and nutritional compounds in food and feed. *J. Chromatogr. A* **1237**, 1–23 (2012).
185. Wang, G., Lee, A. S., Lewis, M., Kamath, B. & Archer, R. K. Accelerated Solvent Extraction and Gas Chromatography/Mass Spectrometry for Determination of Polycyclic Aromatic Hydrocarbons in Smoked Food Samples. *J. Agric. Food Chem.* **47**, 1062–1066 (1999).
186. Jira, W. A GC/MS method for the determination of carcinogenic polycyclic aromatic hydrocarbons (PAH) in smoked meat products and liquid smokes. *Eur. Food Res. Technol.* **218**, 208–212 (2004).
187. Yusà, V., Pardo, O., Martí, P. & Pastor, A. Application of accelerated solvent extraction followed by gel performance chromatography and high-performance liquid chromatography for the determination of polycyclic aromatic hydrocarbons in mussel tissue. *Food Addit. Contam.* **22**, 482–489 (2005).

188. Ziegenhals, K., Jira, W. & Speer, K. Polycyclic aromatic hydrocarbons (PAH) in various types of tea. *Eur. Food Res. Technol.* **228**, 83–91 (2008).
189. Garrido Frenich, A., Martínez Vidal, J. L., Cruz Sicilia, A. D., González Rodríguez, M. J. & Plaza Bolaños, P. Multiresidue analysis of organochlorine and organophosphorus pesticides in muscle of chicken, pork and lamb by gas chromatography–triple quadrupole mass spectrometry. *Anal. Chim. Acta* **558**, 42–52 (2006).
190. Adou, K. Multiresidue Method for the Analysis of Pesticide Residues in Fruits and Vegetables by Accelerated Solvent Extraction and Capillary Gas Chromatography | Journal of Agricultural and Food Chemistry.
191. Suchan, P., Pulkrabová, J., Hajšlová, J. & Kocourek, V. Pressurized liquid extraction in determination of polychlorinated biphenyls and organochlorine pesticides in fish samples. *Anal. Chim. Acta* **520**, 193–200 (2004).
192. Wu, G. *et al.* Analysis of multi-pesticide residues in the foods of animal origin by GC–MS coupled with accelerated solvent extraction and gel permeation chromatography cleanup. *Food Chem.* **126**, 646–654 (2011).
193. Ullah, S. M. R., Murphy, B., Dorich, B., Richter, B. & Srinivasan, K. Fat Extraction from Acid- and Base-Hydrolyzed Food Samples Using Accelerated Solvent Extraction. *J. Agric. Food Chem.* **59**, 2169–2174 (2011).
194. Toschi, T. G., Bendini, A., Ricci, A. & Lercker, G. Pressurized solvent extraction of total lipids in poultry meat. *Food Chem.* **83**, 551–555 (2003).
195. Dodds, E. D., McCoy, M. R., Geldenhuys, A., Rea, L. D. & Kennish, J. M. Microscale recovery of total lipids from fish tissue by accelerated solvent extraction. *J. Am. Oil Chem. Soc.* **81**, 835–840 (2004).
196. Buszewski, B. & Szultka, M. Past, Present, and Future of Solid Phase Extraction: A Review. *Crit. Rev. Anal. Chem.* **42**, 198–213 (2012).
197. Zhao, P., Qian, Y., He, F., Li, H. & Qian, M. Comparative Characterization of Aroma Compounds in Merlot Wine by LiChrolut-EN-Based Aroma Extract Dilution Analysis and Odor Activity Value. *Chemosens. Percept.* **10**, 149–160 (2017).
198. Uckun, O. & Selli, S. Characterization of key aroma compounds in a representative aromatic extracts from citrus and astragalus honeys based on aroma extract dilution analyses. *J. Food Meas. Charact.* **11**, 512–522 (2017).
199. Sanchez-Palomo, E., Trujillo, M., Garcia Ruiz, A. & Gonzalez Vinas, M. A. Aroma profile of malbec red wines from La Mancha region: Chemical and sensory characterization. *Food Res. Int.* **100**, 201–208 (2017).
200. Chen, X. *et al.* Aroma characterization of Hanzhong black tea (*Camellia sinensis*) using solid phase extraction coupled with gas chromatography–mass spectrometry and olfactometry and sensory analysis. *Food Chem.* **274**, 130–136 (2019).
201. Mangas, J. J., Gonzalez, M. P., Rodriguez, R. & Blanco, D. Solid-phase extraction and determination of trace aroma and flavour components in cider by GC-MS. *Chromatographia* **42**, 101–105 (1996).
202. Sun, Q., Zhang, M., Yuan, D., Chen, Z. & Gu, C. Determination of nine emerging pesticides at trace level in aqueous samples using fully automated on-line solid phase extraction coupled with liquid chromatography-mass spectrometry. *Int. J. Environ. Anal. Chem.* **93**, 970–983 (2013).

203. Kawasaki, M., Inoue, T., Fukuhara, K. & Uchiyama, S. Study on CC/MS (SIM) for determination of carbamate and organonitrogen pesticides in foods with simple clean-up by SPE method. *J. Food Hyg. Soc. Jpn.* **40**, 382–390 (1999).
204. Warzecha, L., Strozyk, M., Janoszka, B., Blaszczyk, U. & Bodzek, D. Determination of polycyclic aromatic hydrocarbons, azaarenes, and aminoazaarenes in meat samples by solid-phase extraction (SPE) and HPLC. *Acta Chromatogr.* **12**, 104–121 (2002).
205. Stuppner, S. *et al.* An automated preloaded pipet tip SPE method for the accurate quantification of carcinogenic polycyclic aromatic hydrocarbons from tea. *Anal. Methods* **12**, 1827–1833 (2020).
206. Purcaro, G., Moret, S. & Conte, L. S. Rapid SPE-HPLC determination of the 16 European priority polycyclic aromatic hydrocarbons in olive oils. *J. Sep. Sci.* **31**, 3936–3944 (2008).
207. Arthur, C. L. & Pawliszyn, Janusz. Solid phase microextraction with thermal desorption using fused silica optical fibers. *Anal. Chem.* **62**, 2145–2148 (1990).
208. Merkle, S., Kleeberg, K. K. & Fritsche, J. Recent Developments and Applications of Solid Phase Microextraction (SPME) in Food and Environmental Analysis—A Review. *Chromatography* **2**, 293–381 (2015).
209. Kataoka, H., Lord, H. L. & Pawliszyn, J. Applications of solid-phase microextraction in food analysis. *J. Chromatogr. A* **880**, 35–62 (2000).
210. Nerín, C., Salafranca, J., Aznar, M. & Batlle, R. Critical review on recent developments in solventless techniques for extraction of analytes. *Anal. Bioanal. Chem.* **393**, 809–833 (2009).
211. Larroque, V., Desauziers, V. & Mocho, P. Comparison of two solid-phase microextraction methods for the quantitative analysis of VOCs in indoor air. *Anal. Bioanal. Chem.* **386**, 1457–1464 (2006).
212. Itoh, N., Tao, H. & Ibusuki, T. Optimization of aqueous acetylation for determination of hydroxy polycyclic aromatic hydrocarbons in water by stir bar sorptive extraction and thermal desorption–gas chromatography–mass spectrometry. *Anal. Chim. Acta* **535**, 243–250 (2005).
213. Kolahgar, B., Hoffmann, A. & Heiden, A. C. Application of stir bar sorptive extraction to the determination of polycyclic aromatic hydrocarbons in aqueous samples. *J. Chromatogr. A* **963**, 225–230 (2002).
214. Prieto, A. *et al.* Stir-bar sorptive extraction: A view on method optimisation, novel applications, limitations and potential solutions. *J. Chromatogr. A* **1217**, 2642–2666 (2010).
215. Nakamura, S. & Daishima, S. Simultaneous determination of 64 pesticides in river water by stir bar sorptive extraction and thermal desorption–gas chromatography–mass spectrometry. *Anal. Bioanal. Chem.* **382**, 99–107 (2005).
216. Guerrero, E. D., Castro Mejías, R., Marín, R. N. & Barroso, C. G. Optimization of stir bar sorptive extraction applied to the determination of pesticides in vinegars. *J. Chromatogr. A* **1165**, 144–150 (2007).
217. Liu, W., Hu, Y., Zhao, J., Xu, Y. & Guan, Y. Determination of organophosphorus pesticides in cucumber and potato by stir bar sorptive extraction. *J. Chromatogr. A* **1095**, 1–7 (2005).
218. Guerrero, E. D., Marín, R. N., Mejías, R. C. & Barroso, C. G. Optimisation of stir bar sorptive extraction applied to the determination of volatile compounds in vinegars. *J. Chromatogr. A* **1104**, 47–53 (2006).
219. Corbi, E., Benanou, D., Cantet, J. & Tabet, J. C. Identification of new odorous compounds in Swedish water: mixed haloanisoles and phenolic precursors. *Water Sci. Technol.* **55**, 185–193 (2007).

220. Alves, R. F., Nascimento, A. M. D. & Nogueira, J. M. F. Characterization of the aroma profile of Madeira wine by sorptive extraction techniques. *Anal. Chim. Acta* **546**, 11–21 (2005).
221. Zalacain, A., Marín, J., Alonso, G. L. & Salinas, M. R. Analysis of wine primary aroma compounds by stir bar sorptive extraction. *Talanta* **71**, 1610–1615 (2007).
222. Sanchez-Rojas, F., Bosch-Ojeda, C. & Manuel Cano-Pavon, J. A Review of Stir Bar Sorptive Extraction. *Chromatographia* **69**, S79–S94 (2009).
223. Kataoka, H. Recent developments and applications of microextraction techniques in drug analysis. *Anal. Bioanal. Chem.* **396**, 339–364 (2010).
224. Rouseff, R. & Cadwallader, K. Headspace Techniques in Foods, Fragrances and Flavors. in *Headspace Analysis of Foods and Flavors: Theory and Practice* (eds. Rouseff, R. L. & Cadwallader, K. R.) 1–8 (Springer US, 2001). doi:10.1007/978-1-4615-1247-9_1.
225. Gorska-Horczyzak, E., Zalewska, M. & Wierzbicka, A. Chromatographic fingerprint application possibilities in food authentication. *Eur. Food Res. Technol.* **248**, 1163–1177 (2022).
226. Medina, I., Satue-Gracia, M. T. & Frankel, E. N. Static headspace gas chromatographic analyses to determine oxidation of fish muscle lipids during thermal processing. *J. Am. Oil Chem. Soc.* **76**, 231–236 (1999).
227. Shen, M. *et al.* Simultaneous determination of furan and 2-alkylfurans in heat-processed foods by automated static headspace gas chromatography-mass spectrometry. *Lwt-Food Sci. Technol.* **72**, 44–54 (2016).
228. Chen, Y.-J., Lin, C.-Y., Cheng, S.-S. & Chang, S.-T. Rapid Discrimination and Feature Extraction of Three *Chamaecyparis* Species by Static-HS/GC-MS. *J. Agric. Food Chem.* **63**, 810–820 (2015).
229. Morais da Silva, P. de L., de Lima, L. S., Caetano, I. K. & Torres, Y. R. Comparative analysis of the volatile composition of honeys from Brazilian stingless bees by static headspace GC-MS. *Food Res. Int.* **102**, 536–543 (2017).
230. Zhu, W., Benkwitz, F., Sarmadi, B. & Kilmartin, P. A. Validation Study on the Simultaneous Quantitation of Multiple Wine Aroma Compounds with Static Headspace-Gas Chromatography-Ion Mobility Spectrometry. *J. Agric. Food Chem.* **69**, 15020–15035 (2021).
231. Zhao, J. & Boatright, W. L. Static headspace analysis of odorants in commercial rice proteins. *Food Chem.* **221**, 345–350 (2017).
232. Ganguly, S., Mahanty, A., Mitra, T., Raman, R. K. & Mohanty, B. P. Volatile compounds in hilsa (*Tenualosa ilisha*, Hamilton) as detected by static headspace gas chromatography and mass spectrometry. *J. Food Process. Preserv.* **41**, e13212 (2017).
233. Sharma, A. *et al.* Determination of Hexanal Using Static Headspace GC-FID Method and Its Correlation with Oxidative Rancidity in Edible Oils. *Food Anal. Methods* doi:10.1007/s12161-022-02320-4.
234. Vickackaite, V., Pipiraite, G., Poskus, V., Jurkute, I. & Bugelyte, B. Static headspace gas chromatographic determination of hexanal as a marker of lipid oxidation in fat-rich food. *Chemija* **31**, 95–100 (2020).
235. Ettre, L. S. Headspace-gas chromatography - An ideal technique for sampling volatiles present in non-volatile matrices. in *Headspace Analysis of Foods and Flavors: Theory and Practice* (eds. Rouseff, R. L. & Cadwallader, K. R.) vol. 488 9–32 (Kluwer Academic/Plenum Publ, 2001).
236. Wojnowski, W., Majchrzak, T., Dymerski, T., Gebicki, J. & Namiesnik, J. Dynamic Headspace Sampling as an Initial Step for Sample Preparation in Chromatographic Analysis. *J. Aoac Int.* **100**, 1599–1606 (2017).

237. Harries, M. E. & Bruno, T. J. Headspace Analysis: Purge and Trap ☆. in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* B978012409547214404X (Elsevier, 2018). doi:10.1016/B978-0-12-409547-2.14404-X.
238. Ruiz-Bevia, F., Font, A., Garcia, A. N., Blasco, P. & Ruiz, J. J. Quantitative analysis of the volatile aroma components of pepino fruit by purge-and-trap and gas chromatography. *J. Sci. Food Agric.* **82**, 1182–1188 (2002).
239. Qian, M. & Reineccius, G. Potent aroma compounds in Parmigiano Reggiano cheese studied using a dynamic headspace (purge-trap) method. *Flavour Fragr. J.* **18**, 252–259 (2003).
240. Fredes, A. *et al.* Quantification of prominent volatile compounds responsible for muskmelon and watermelon aroma by purge and trap extraction followed by gas chromatography-mass spectrometry determination. *Food Chem.* **190**, 689–700 (2016).
241. Topi, D., Amanpour, A., Kelebek, H. & Selli, S. Screening of aroma profiles in Albanian cvs. Kalinjot and Bardhi Tirana olive oils using purge and trap extraction technique. *Riv. Ital. Delle Sostanze Grasse* **96**, 101–108 (2019).
242. Sarhir, S. T., Amanpour, A., Bouseta, A. & Selli, S. Fingerprint of aroma-active compounds and odor activity values in a traditional Moroccan fermented butter ‘Smen’ using GC-MS-Olfactometry. *J. Food Compos. Anal.* **96**, 103761 (2021).
243. Hwang, I.-S. & Kim, M. K. Influence of Processing Conditions on the Flavor Profiles of Mulberry (*Morus alba*Linn) Fruits Using Instrumental Flavor Analysis and Descriptive Sensory Analysis. *Foods* **9**, 581 (2020).
244. Huang, X.-H. *et al.* Flavor formation in different production steps during the processing of cold-smoked Spanish mackerel. *Food Chem.* **286**, 241–249 (2019).
245. Sabia, E., Gaulty, M., Napolitano, F., Cifuni, G. F. & Claps, S. The effect of different dietary treatments on volatile organic compounds and aromatic characteristics of buffalo Mozzarella cheese. *Int. J. Dairy Technol.* **73**, 594–603 (2020).
246. Qin, L. *et al.* Changes in Aroma Profile of Shiitake Mushroom (*Lentinus edodes*) during Different Stages of Hot Air Drying. *Foods* **9**, 444 (2020).
247. Keser, D. *et al.* Characterization of aroma and phenolic composition of carrot (*Daucus carota* ‘Nantes’) powders obtained from intermittent microwave drying using GC-MS and LC-MS/MS. *Food Bioprod. Process.* **119**, 350–359 (2020).
248. Hakala, M. A., Lapvetelainen, A. T. & Kallio, H. P. Volatile compounds of selected strawberry varieties analyzed by purge-and-trap headspace GC-MS. *J. Agric. Food Chem.* **50**, 1133–1142 (2002).
249. Cristina Soria, A., Martinez-Castro, I. & Sanz, J. Some aspects of dynamic headspace analysis of volatile components in honey. *Food Res. Int.* **41**, 838–848 (2008).
250. Nerin, C., Acosta, D. & Rubio, C. Potential migration release of volatile compounds from plastic containers destined for food use in microwave ovens. *Food Addit. Contam.* **19**, 594–601 (2002).
251. Tuduri, L., Desauziers, V. & Fanlo, J. L. A simple calibration procedure for volatile organic compounds sampling in air with adsorptive solid-phase microextraction fibres. *The Analyst* **128**, 1028 (2003).
252. Pawliszyn, J. Quantitative aspects of SPME. in *Applications of Solid Phase Microextraction* 3–21 (1999). doi:10.1039/9781847550149-00003.
253. Koziel, J. A. *et al.* Evaluation of Sample Recovery of Malodorous Livestock Gases from Air Sampling Bags, Solid-Phase Microextraction Fibers, Tenax TA Sorbent Tubes, and Sampling Canisters. *J. Air Waste Manag. Assoc.* **55**, 1147–1157 (2005).

254. Lord, H. & Pawliszyn, J. Evolution of solid-phase microextraction technology. *J. Chromatogr. A* **885**, 153–193 (2000).
255. Larroque, V., Desauziers, V. & Mocho, P. Development of a solid phase microextraction (SPME) method for the sampling of VOC traces in indoor air. *J. Environ. Monit.* **8**, 106–111 (2006).
256. Cavalcante, R. M., de Andrade, M. V. F., Marins, R. V. & Oliveira, L. D. M. Development of a headspace-gas chromatography (HS-GC-PID-FID) method for the determination of VOCs in environmental aqueous matrices: Optimization, verification and elimination of matrix effect and VOC distribution on the Fortaleza Coast, Brazil. *Microchem. J.* **96**, 337–343 (2010).
257. Safarova, V. I. *et al.* Gas chromatography–mass spectrometry with headspace for the analysis of volatile organic compounds in waste water. *J. Chromatogr. B* **800**, 325–330 (2004).
258. Wardencki, W., Michulec, M. & Curyło, J. A review of theoretical and practical aspects of solid-phase microextraction in food analysis. *Int. J. Food Sci. Technol.* **39**, 703–717 (2004).
259. Stashenko, E. E. & Martínez, J. R. Derivatization and solid-phase microextraction. *TrAC Trends Anal. Chem.* **23**, 553–561 (2004).
260. Bourdin, D. & Desauziers, V. Development of SPME on-fiber derivatization for the sampling of formaldehyde and other carbonyl compounds in indoor air. *Anal. Bioanal. Chem.* **406**, 317–328 (2014).
261. Goldmann, T., Périsset, A., Scanlan, F. & Stadler, R. H. Rapid determination of furan in heated foodstuffs by isotope dilution solid phase micro-extraction-gas chromatography – mass spectrometry (SPME-GC-MS). *The Analyst* **130**, 878 (2005).
262. Bianchi, F., Careri, M., Mangia, A. & Musci, M. Development and validation of a solid phase micro-extraction–gas chromatography–mass spectrometry method for the determination of furan in baby-food. *J. Chromatogr. A* **1102**, 268–272 (2006).
263. Robertson, R. F., Jauncey, K., Beveridge, M. C. M. & Lawton, L. A. Depuration rates and the sensory threshold concentration of geosmin responsible for earthy-musty taint in rainbow trout, *Onchorhynchus mykiss*. *Aquaculture* **245**, 89–99 (2005).
264. Jurado, J. M. *et al.* Differentiation of certified brands of origins of Spanish white wines by HS-SPME-GC and chemometrics. *Anal. Bioanal. Chem.* **390**, 961–970 (2008).
265. Chen, Y., Shirey, R. E. & Sidisky, L. M. Determination of Diacetyl in Butter and Air Samples by SPME Coupled with GC–MS. *Chromatographia* **72**, 999–1004 (2010).
266. Panseri, S., Soncin, S., Chiesa, L. M. & Biondi, P. A. A headspace solid-phase microextraction gas-chromatographic mass-spectrometric method (HS-SPME-GC/MS) to quantify hexanal in butter during storage as marker of lipid oxidation. *Food Chem.* **127**, 886–889 (2011).
267. Lim, H.-H. & Shin, H.-S. Simple Determination of 4-Methylimidazole in Soft Drinks by Headspace SPME GC–MS. *Chromatographia* **76**, 97–101 (2013).
268. Guillot, S. *et al.* Aroma characterization of various apricot varieties using headspace-solid phase microextraction combined with gas chromatography-mass spectrometry and gas chromatography-olfactometry. *Food Chem.* **96**, 147–155 (2006).
269. Beltrán, A., Ramos, M., Grané, N., Martín, M. L. & Garrigós, M. C. Monitoring the oxidation of almond oils by HS-SPME–GC–MS and ATR-FTIR: Application of volatile compounds determination to cultivar authenticity. *Food Chem.* **126**, 603–609 (2011).
270. Lin, J. *et al.* Discrimination of oolong tea (*Camellia sinensis*) varieties based on feature extraction and selection from aromatic profiles analysed by HS-SPME/ GC–MS. *Food Chem.* **7** (2013).

271. Vavoura, M. V., Badeka, A. V., Kontakos, S. & Kontominas, M. G. Characterization of Four Popular Sweet Cherry Cultivars Grown in Greece by Volatile Compound and Physicochemical Data Analysis and Sensory Evaluation. *Molecules* **20**, 1922–1940 (2015).
272. Zheng, Y., Zeng, X., Peng, W., Wu, Z. & Su, W. Study on the Discrimination between Citri Reticulatae Pericarpium Varieties Based on HS-SPME-GC-MS Combined with Multivariate Statistical Analyses. *Molecules* **23**, 1235 (2018).
273. Jeleń, H. H., Mildner-Szkudlarz, S., Jasińska, I. & Wąsowicz, E. A Headspace–SPME–MS Method for Monitoring Rapeseed Oil Autoxidation. *J. Am. Oil Chem. Soc.* **84**, 509–517 (2007).
274. Yang, C., Wang, Y., Wu, B., Fang, J. & Li, S. Volatile compounds evolution of three table grapes with different flavour during and after maturation. *Food Chem.* **4**, 823–830 (2011).
275. Spínola, V., Perestrelo, R., Câmara, J. S. & Castilho, P. C. Establishment of Monstera deliciosa fruit volatile metabolomic profile at different ripening stages using solid-phase microextraction combined with gas chromatography-mass spectrometry. *Food Res. Int.* **67**, 409–417 (2015).
276. Ocak, E., Javidipour, I. & Tunçturk, Y. Volatile compounds of Van Herby cheeses produced with raw and pasteurized milks from different species. *J. Food Sci. Technol.* **52**, 4315–4323 (2015).
277. Fan, X. *et al.* Characterization of Volatile Compounds by SPME-GC-MS During the Ripening of Kedong Sufu, a Typical Chinese Traditional Bacteria-Fermented Soybean Product. *J. Food Sci.* **84**, 2441–2448 (2019).
278. Xiao, Z. *et al.* Discrimination of Cherry Wines Based on Their Sensory Properties and Aromatic Fingerprinting using HS-SPME-GC-MS and Multivariate Analysis: Discrimination of cherry wine... *J. Food Sci.* **79**, C284–C294 (2014).
279. Chang, M. *et al.* Characteristic volatiles fingerprints and profiles determination in different grades of coconut oil by HS-GC-IMS and HS-SPME-GC-MS. *Int. J. Food Sci. Technol.* **55**, 3670–3679 (2020).
280. Gocmen, D., Elston, A., Williams, T., Parish, M. & Rouseff, R. L. Identification of medicinal off-flavours generated by Alicyclobacillus species in orange juice using GC-olfactometry and GC-MS. *Letts. Appl. Microbiol.* **40**, 172–177 (2005).
281. Siripatrawan, U. & Harte, B. R. Solid phase microextraction/gas chromatography/mass spectrometry integrated with chemometrics for detection of Salmonella typhimurium contamination in a packaged fresh vegetable. *Anal. Chim. Acta* **581**, 63–70 (2007).
282. Bhattacharjee, P. *et al.* A comparative qualitative study of the profile of volatile organic compounds associated with Salmonella contamination of packaged aged and fresh beef by HS-SPME/GC-MS. *J. Food Sci. Technol.* **48**, 1–13 (2011).
283. Silva Ferreira, A. C. *et al.* Monitoring Alcoholic Fermentation: An Untargeted Approach. *J. Agric. Food Chem.* **62**, 6784–6793 (2014).
284. Arcena, M. R., Kebede, B., Leong, S. Y., Silcock, P. & Oey, I. Feasibility of using integrated fingerprinting, profiling and chemometrics approach to understand (bio) chemical changes throughout commercial red winemaking: A case study on Merlot. *Food Res. Int.* **127**, 108767 (2020).
285. Springer, A. E. *et al.* Validated Modeling for German White Wine Varietal Authentication Based on Headspace Solid-Phase Microextraction Online Coupled with Gas Chromatography Mass Spectrometry Fingerprinting. *J. Agric. Food Chem.* **62**, 6844–6851 (2014).
286. Zhao, Q. *et al.* A comparative HS-SPME/GC-MS-based metabolomics approach for discriminating selected japonica rice varieties from different regions of China in raw and cooked form. *Food Chem.* **385**, 132701 (2022).

287. Qi, D., Ma, C., Wang, W., Zhang, L. & Li, J. Gas Chromatography-Mass Spectrometry Analysis as a Tool to Reveal Differences Between the Volatile Compound Profile of Royal Jelly Produced from Tea and Pagoda Trees. *Food Anal. Methods* **14**, 616–630 (2021).
288. Schueuermann, C. *et al.* A GC–MS untargeted metabolomics approach for the classification of chemical differences in grape juices based on fungal pathogen. *Food Chem.* **270**, 375–384 (2019).
289. Lopez, A. *et al.* Volatile Organic Compounds Profile in White Sturgeon (*Acipenser transmontanus*) Caviar at Different Stages of Ripening by Multiple Headspace Solid Phase Microextraction. *15* (2020).
290. Marco, A., Navarro, J. L. & Flores, M. Volatile compounds of dry-fermented sausages as affected by solid-phase microextraction (SPME). *Food Chem.* **9** (2004).
291. Mallia, S., Fernández-García, E. & Olivier Bosset, J. Comparison of purge and trap and solid phase microextraction techniques for studying the volatile aroma compounds of three European PDO hard cheeses. *Int. Dairy J.* **15**, 741–758 (2005).
292. Tunick, M. H., Iandola, S. K. & Van Hekken, D. L. Comparison of SPME Methods for Determining Volatile Compounds in Milk, Cheese, and Whey Powder. *Foods* **2**, 534–543 (2013).
293. Ma, Q. L., Hamid, N., Bekhit, A. E. D., Robertson, J. & Law, T. F. Optimization of headspace solid phase microextraction (HS-SPME) for gas chromatography mass spectrometry (GC–MS) analysis of aroma compounds in cooked beef using response surface methodology. *Microchem. J.* **111**, 16–24 (2013).
294. Lee, J., Chambers, D., Chambers, E., Adhikari, K. & Yoon, Y. Volatile Aroma Compounds in Various Brewed Green Teas. *Molecules* **18**, 10024–10041 (2013).
295. Xiao, Z. *et al.* Discrimination of Cherry Wines Based on Their Sensory Properties and Aromatic Fingerprinting using HS-SPME-GC-MS and Multivariate Analysis: Discrimination of cherry wine.... *J. Food Sci.* **79**, C284–C294 (2014).
296. Iamanaka, B. T. *et al.* Potential of volatile compounds produced by fungi to influence sensory quality of coffee beverage. *Food Res. Int. Ott. Ont* **64**, 166–170 (2014).
297. Argyri, A. A., Mallouchos, A., Panagou, E. Z. & Nychas, G.-J. E. The dynamics of the HS/SPME-GC/MS as a tool to assess the spoilage of minced beef stored under different packaging and temperature conditions. *Int. J. Food Microbiol.* **193**, 51–58 (2015).
298. Du, L., He, T., Li, W., Wang, R. & Xiao, D. Analysis of volatile compounds in Chinese Laobaigan liquor using headspace solid-phase microextraction coupled with GC-MS. *Anal. Methods* **7**, 1906–1913 (2015).
299. Spínola, V., Perestrelo, R., Câmara, J. S. & Castilho, P. C. Establishment of *Monstera deliciosa* fruit volatile metabolomic profile at different ripening stages using solid-phase microextraction combined with gas chromatography-mass spectrometry. *Food Res. Int.* **67**, 409–417 (2015).
300. Griglione, A. *et al.* High-quality Italian rice cultivars: Chemical indices of ageing and aroma quality. *Food Chem.* **172**, 305–313 (2015).
301. Cuevas, F. J., Moreno-Rojas, J. M. & Ruiz-Moreno, M. J. Assessing a traceability technique in fresh oranges (*Citrus sinensis* L. Osbeck) with an HS-SPME-GC-MS method. Towards a volatile characterisation of organic oranges. *Food Chem.* **221**, 1930–1938 (2017).
302. Zheng, Y., Zeng, X., Peng, W., Wu, Z. & Su, W. Study on the Discrimination between *Citri Reticulatae* Pericarpium Varieties Based on HS-SPME-GC-MS Combined with Multivariate Statistical Analyses. *Molecules* **23**, 1235 (2018).

303. Fan, X. *et al.* Characterization of Volatile Compounds by SPME-GC-MS During the Ripening of Kedong Sufu, a Typical Chinese Traditional Bacteria-Fermented Soybean Product. *J. Food Sci.* **84**, 2441–2448 (2019).
304. Wang, Z. *et al.* Effects of Different Steeping Temperatures on the Leaching of Aroma Components in Black Tea by SPME–GC–MS Coupled with Chemometric Method. *J. AOAC Int.* **102**, 1834–1844 (2019).
305. Guo, Q. Comparative analysis of volatile profiles in four pine-mushrooms using HS-SPME/GC-MS and E-nose. *Food Control* **10** (2022).
306. Biasioli, F., Yeretdzian, C., Märk, T. D., Dewulf, J. & Van Langenhove, H. Direct-injection mass spectrometry adds the time dimension to (B)VOC analysis. *TrAC Trends Anal. Chem.* **30**, 1003–1017 (2011).
307. Deuscher, Z. *et al.* Volatile compounds profiling by using proton transfer reaction-time of flight-mass spectrometry (PTR-ToF-MS). The case study of dark chocolates organoleptic differences. *J. Mass Spectrom.* **54**, 92–119 (2019).
308. Merfort, I. Review of the analytical techniques for sesquiterpenes and sesquiterpene lactones. *J. Chromatogr. A* **967**, 115–130 (2002).
309. Beneito-Cambra, M. *et al.* Direct analysis of olive oil and other vegetable oils by mass spectrometry: A review. *TrAC Trends Anal. Chem.* **132**, 116046 (2020).
310. Dryahina, K., Som, S., Smith, D. & Španěl, P. Reagent and analyte ion hydrates in secondary electrospray ionization mass spectrometry (SESI-MS), their equilibrium distributions and dehydration in an ion transfer capillary: Modelling and experiments. *Rapid Commun. Mass Spectrom.* **35**, (2021).
311. Jiang, Y. *et al.* Microbial diversity and volatile profile of traditional fermented yak milk. *J. Dairy Sci.* S0022030219309427 (2019) doi:10.3168/jds.2019-16753.
312. French, J. B., Thomson, B. A., Davidson, W. R., Reid, N. M. & Buckley, J. A. Atmospheric Pressure Chemical Ionization Mass Spectrometry. in *Mass Spectrometry in Environmental Sciences* (eds. Karasek, F. W., Hutzinger, O. & Safe, S.) 101–121 (Springer US, 1985). doi:10.1007/978-1-4613-2361-7_6.
313. Guo, T. *et al.* Applications of DART-MS for food quality and safety assurance in food supply chain. *Mass Spectrom. Rev.* **36**, 161–187 (2017).
314. Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J. & Udseth, H. R. New developments in biochemical mass spectrometry: electrospray ionization. *ACS Publications* <https://pubs.acs.org/doi/pdf/10.1021/ac00208a002> (2002) doi:10.1021/ac00208a002.
315. Lindinger, W. & Jordan, A. Proton-transfer-reaction mass spectrometry (PTR–MS): on-line monitoring of volatile organic compounds at pptv levels. *Chem. Soc. Rev.* **27**, 347–375 (1998).
316. Smith, D. & Španěl, P. Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. *Mass Spectrom. Rev.* **24**, 661–700 (2005).
317. Harper, W. J., Kocaoglu-Vurma, N. A., Wick, C., Elekes, K. & Langford, V. Analysis of Volatile Sulfur Compounds in Swiss Cheese Using Selected Ion Flow Tube Mass Spectrometry (SIFT-MS). in *Volatile Sulfur Compounds in Food* (eds. Qian, M. C., Fan, X. & Mahattanatawee, K.) vol. 1068 153–181 (American Chemical Society, 2011).
318. Adams, N. G. & Smith, D. The Selected Ion Flow Tube (SIFT); a technique for studying ion-neutral reactions. *Int. J. Mass Spectrom. Ion Phys.* 349–359 (1976).
319. Adams, N. G., Smith, D. & Grief, D. Reactions of HnCO⁺ ions with molecules at 300K. *Int. J. Mass Spectrom. Ion Phys.* **26**, 405–415 (1978).

320. Ferguson, E. E., Fehsenfeld, F. C. & Schmeltekopf, A. L. Flowing Afterglow Measurements of Ion-Neutral Reactions. in *Advances in Atomic and Molecular Physics* vol. 5 1–56 (Elsevier, 1969).
321. Smith, D. & Spanel, P. Ions in the terrestrial atmosphere and in interstellar clouds. *Mass Spectrom. Rev.* **14**, 255–278 (1995).
322. Smith, D. & Spanel, P. The Novel Selected-ion Flow Tube Approach to Trace Gas Analysis of Air and Breath. *Rapid Commun. Mass Spectrom.* **10**, 1183–1996 (1996).
323. Španěl, P., Rolfe, P., Rajan, B. & Smith, D. The selected ion flow tube (SIFT)—A novel technique for biological monitoring. *Ann. Occup. Hyg.* **40**, 615–626 (1996).
324. Spanel, P. & Smith, D. Selected ion flow tube - mass spectrometry: detection and real-time monitoring of flavours released by food products. *Rapid Commun Mass Spectrom* **12** (1999).
325. Španěl, P., Ji, Y. & Smith, D. SIFT studies of the reactions of H₃O⁺, NO⁺ and O₂⁺ with a series of aldehydes and ketones. *Int. J. Mass Spectrom. Ion Process.* **165–166**, 25–37 (1997).
326. Spanel, P. & Smith, D. SIFT studies of the reactions of H₃ O⁺, NO ⁺ and with a series of alcohols .pdf. *Int. J. Mass Spectrom. Ion Process.* **167/168**, 375–388 (1997).
327. Španěl, P. & Smith, D. SIFT studies of the reactions of H₃O⁺, NO⁺ and O₂⁺ with a series of volatile carboxylic acids and esters. *Int. J. Mass Spectrom. Ion Process.* **172**, 137–147 (1998).
328. Sekimoto, K., Sakai, M. & Takayama, M. Specific Interaction Between Negative Atmospheric Ions and Organic Compounds in Atmospheric Pressure Corona Discharge Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **23**, 1109–1119 (2012).
329. Dahlingi, P. *et al.* Ionization by proton bastraction in negative ion field desorption mass spectrometry. *Int. J. Mass Spectrom. Ion Phys.* 301–304 (1983).
330. Green, R. J., Qian, J., Kim, H.-T. & Anderson, S. L. Hydride abstraction by NO⁺ from ethanol: Effects of collision energy and ion rotational state. *J. Chem. Phys.* **113**, 3002–3010 (2000).
331. Fehsenfeld, F. C., Albritton, D. L., Burt, J. A. & Schiff, H. I. Associative-detachment reactions of O⁻ and O₂⁻ by O₂(1Δg). 3.
332. Hassan, S. Z. *et al.* Associative detachment in anion-atom reactions involving a dipole-bound electron. *Nat. Commun.* **13**, 818 (2022).
333. Gibson, A., Malek, L., Dekker, R. F. H. & Ross, B. Detecting volatile compounds from Kraft lignin degradation in the headspace of microbial cultures by selected ion flow tube mass spectrometry (SIFT-MS). *J. Microbiol. Methods* **112**, 40–45 (2015).
334. Scotter, J. M., Langford, V. S., Wilson, P. F., McEwan, M. J. & Chambers, S. T. Real-time detection of common microbial volatile organic compounds from medically important fungi by Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS). *J. Microbiol. Methods* **63**, 127–134 (2005).
335. Wang, T., Smith, D. & Španěl, P. Selected ion flow tube, SIFT, studies of the reactions of H₃O⁺, NO⁺ and O₂⁺ with compounds released by *Pseudomonas* and related bacteria. *Int. J. Mass Spectrom.* **233**, 245–251 (2004).
336. Lourenço, C. *et al.* Monitoring type 2 diabetes from volatile faecal metabolome in Cushing’s syndrome and single Afmid mouse models via a longitudinal study. *Sci. Rep.* **9**, 18779 (2019).
337. Navaneethan, U. *et al.* Volatile organic compounds in bile for early diagnosis of cholangiocarcinoma in patients with primary sclerosing cholangitis: a pilot study. *Gastrointest. Endosc.* **81**, 943-949.e1 (2015).
338. Storer, M., Curry, K., Squire, M., Kingham, S. & Epton, M. Breath testing and personal exposure—SIFT-MS detection of breath acetonitrile for exposure monitoring. *J. Breath Res.* **9**, 036006 (2015).

339. Catala, A. *et al.* Prediction and detection of human epileptic seizures based on SIFT-MS chemometric data. *Sci. Rep.* **10**, 18365 (2020).
340. La Nasa, J. *et al.* SIFT-ing archaeological artifacts: Selected ion flow tube-mass spectrometry as a new tool in archaeometry. *Talanta* **207**, 120323 (2020).
341. Biasioli, F. PTR-MS monitoring of VOCs and BVOCs in food science and technology. *Trends Anal. Chem.* **30**, 10 (2011).
342. Prince, B. J., Milligan, D. B. & McEwan, M. J. Application of selected ion flow tube mass spectrometry to real-time atmospheric monitoring.pdf. *Rapid Commun. Mass Spectrom.* **24**, 1763–1769 (2010).
343. Spanel, P. & Cocker, J. VALIDATION OF THE SIFT TECHNIQUE FOR TRACE GAS ANALYSIS OF BREATH USING THE SYRINGE INJECTION TECHNIQUE. *Ann Occup Hyg* **41**, 373–382 (1997).
344. Storer, M., Salmond, J., Dirks, K. N., Kingham, S. & Epton, M. Mobile selected ion flow tube mass spectrometry (SIFT-MS) devices and their use for pollution exposure monitoring in breath and ambient air—pilot study. *J. Breath Res.* **8**, 037106 (2014).
345. Spanel, P. & Smith, D. Influence of water vapour on selected ion flow tube mass spectrometric analyses of trace gases in humid air and breath. *Rapid Commun. Mass Spectrom.* **14**, 1898–1906 (2000).
346. Ghislain, M. *et al.* High frequency air monitoring by selected ion flow tube-mass spectrometry (SIFT-MS): Influence of the matrix for simultaneous analysis of VOCs, CO₂, ozone and water. *Microchem. J.* **153**, 104435 (2020).
347. Dhooghe, F. *et al.* A selected ion flow tube study of the reactions of H₃O⁺, NO⁺ and O₂⁺ with a series of sesquiterpenes. *Int. J. Mass Spectrom.* **272**, 137–148 (2008).
348. Dryahina, K., Poláček, M. & Španěl, P. A selected ion flow tube, SIFT, study of the ion chemistry of H₃O⁺, NO⁺ and O₂⁺ ions with several nitroalkanes in the presence of water vapour. *Int. J. Mass Spectrom.* **239**, 57–65 (2004).
349. Baerenzung dit Baron, T., Yobrégat, O., Jacques, A., Simon, V. & Geffroy, O. A novel approach to discriminate the volatilome of *Vitis vinifera* berries by Selected Ion Flow Tube Mass Spectrometry analysis and chemometrics. *Food Res. Int.* 111434 (2022) doi:10.1016/j.foodres.2022.111434.
350. Geeraerts, W., De Vuyst, L., Leroy, F. & Van Kerrebroeck, S. Monitoring of volatile production in cooked poultry products using selected ion flow tube-mass spectrometry. *Food Res. Int.* **119**, 196–206 (2019).
351. Langford, V. S. *et al.* Headspace Analysis of Italian and New Zealand Parmesan Cheeses. *J. Food Sci.* **77**, C719–C726 (2012).
352. Bowman, T. & Barringer, S. Analysis of Factors Affecting Volatile Compound Formation in Roasted Pumpkin Seeds with Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS) and Sensory Analysis. *J. Food Sci.* **77**, C51–C60 (2012).
353. Taylor, K., Wick, C., Castada, H., Kent, K. & Harper, W. J. Discrimination of Swiss Cheese from 5 Different Factories by High Impact Volatile Organic Compound Profiles Determined by Odor Activity Value Using Selected Ion Flow Tube Mass Spectrometry and Odor Threshold: Swiss cheese factory VOC variability.... *J. Food Sci.* **78**, C1509–C1515 (2013).
354. Wood, J. E. *et al.* Dairy product quality using screening of aroma compounds by selected ion flow tube–mass spectrometry: A chemometric approach. *Int. Dairy J.* **121**, 105107 (2021).

355. Castada, H., Hanas, K. & Barringer, S. Swiss Cheese Flavor Variability Based on Correlations of Volatile Flavor Compounds, Descriptive Sensory Attributes, and Consumer Preference. *Foods* **8**, 78 (2019).
356. Van Deun, K. *et al.* Weighted sparse principal component analysis. *Chemom. Intell. Lab. Syst.* **195**, 103875 (2019).
357. Carrapiso, A. I. *et al.* SIFT-MS analysis of Iberian hams from pigs reared under different conditions. *Meat Sci.* **104**, 8–13 (2015).
358. Bajoub, A. *et al.* A metabolic fingerprinting approach based on selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics: A reliable tool for Mediterranean origin-labeled olive oils authentication. *Food Res. Int.* **106**, 233–242 (2018).
359. Kharbach, M. *et al.* Authentication of extra virgin Argan oil by selected-ion flow-tube mass-spectrometry fingerprinting and chemometrics. *Food Chem.* **383**, 132565 (2022).
360. Vendel, I., Hertog, M. & Nicolăi, B. Fast analysis of strawberry aroma using SIFT-MS: A new technique in postharvest research. *Postharvest Biol. Technol.* **152**, 127–138 (2019).
361. Van Kerrebroeck, S., Comasio, A., Harth, H. & De Vuyst, L. Impact of starter culture, ingredients, and flour type on sourdough bread volatiles as monitored by selected ion flow tube-mass spectrometry. *Food Res. Int.* **106**, 254–262 (2018).
362. Wold, S. Chemometrics; what do we mean with it, and what do we want from it? *Chemom. Intell. Lab. Syst.* **30**, 109–115 (1995).
363. Lavine, B. & Workman, J. Chemometrics. *Anal. Chem.* **80**, 4519–4531 (2008).
364. Boccard, J. & Rudaz, S. Harnessing the complexity of metabolomic data with chemometrics. *J. Chemom.* **28**, 1–9 (2014).
365. Yi, L. *et al.* Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Anal. Chim. Acta* **914**, 17–34 (2016).
366. Pyle, D. *Data Preparation for Data Mining.* (1999).
367. García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M. & Herrera, F. Big data preprocessing: methods and prospects. *Big Data Anal.* **1**, 9 (2016).
368. Zeng, Z.-D., Hugel, H. M. & Marriott, P. J. Chemometrics in comprehensive multidimensional separations. *Anal. Bioanal. Chem.* **401**, 2373–2386 (2011).
369. Alexandropoulos, S.-A. N., Kotsiantis, S. B. & Vrahatis, M. N. Data preprocessing in predictive data mining. *Knowl. Eng. Rev.* **34**, e1 (2019).
370. Lee, J., Chambers, D., Chambers, E., Adhikari, K. & Yoon, Y. Volatile Aroma Compounds in Various Brewed Green Teas. *Molecules* **18**, 10024–10041 (2013).
371. Want, E. & Masson, P. Processing and Analysis of GC/LC-MS-Based Metabolomics Data. in *Metabolic Profiling* (ed. Metz, T. O.) vol. 708 277–298 (Humana Press, 2011).
372. Niu, W., Knight, E., Xia, Q. & McGarvey, B. D. Comparative evaluation of eight software programs for alignment of gas chromatography–mass spectrometry chromatograms in metabolomics experiments. *J. Chromatogr. A* **1374**, 199–206 (2014).
373. Morales, M. L., Callejón, R. M., Ordóñez, J. L., Troncoso, A. M. & García-Parrilla, M. C. Comparative assessment of software for non-targeted data analysis in the study of volatile fingerprint changes during storage of a strawberry beverage. *J. Chromatogr. A* **1522**, 70–77 (2017).
374. Delaporte, G. Développement d'une approche non-ciblée par empreinte pour caractériser la qualité sanitaire chimique de matrices agro-alimentaires complexes. (Université Paris-Saclay, 2018).

375. Spadafora, N. D., Mascrez, S., McGregor, L. & Purcaro, G. Exploring multiple-cumulative trapping solid-phase microextraction coupled to gas chromatography–mass spectrometry for quality and authenticity assessment of olive oil. *Food Chem.* **383**, 132438 (2022).
376. Huang, Y. & Barringer, S. A. Monitoring of Cocoa Volatiles Produced during Roasting by Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS). *J. Food Sci.* **76**, C279–C286 (2011).
377. Geeraerts, W., Borremans, W., De Vuyst, L., Leroy, F. & Van Kerrebroeck, S. The application of selected ion flow tube-mass spectrometry to follow volatile formation in modified-atmosphere-packaged cooked ham. *Food Res. Int.* **123**, 601–611 (2019).
378. Langford, V. S. *et al.* Headspace Analysis of Italian and New Zealand Parmesan Cheeses. *J. Food Sci.* **77**, C719–C726 (2012).
379. Dillon, W. R. & Goldstein, A. *Multivariate analysis : methods and applications.* (Wiley, 1984).
380. Saccenti, E., Hoefsloot, H. C. J., Smilde, A. K., Westerhuis, J. A. & Hendriks, M. M. W. B. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **10**, 361–374 (2014).
381. Lavine, B. K. Pattern Recognition. *Crit. Rev. Anal. Chem.* **36**, 153–161 (2006).
382. Sinaga, K. P. & Yang, M.-S. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **8**, 80716–80727 (2020).
383. Guerra, L. *et al.* Comparison between supervised and unsupervised classifications of neuronal cell types: a case study. *Dev. Neurobiol.* **71**, 71–82 (2011).
384. Cunningham, P. Dimension Reduction. in *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval* (eds. Cord, M. & Cunningham, P.) 91–112 (Springer, 2008). doi:10.1007/978-3-540-75171-7_4.
385. Lopez, A. *et al.* Volatile Organic Compounds Profile in White Sturgeon (*Acipenser transmontanus*) Caviar at Different Stages of Ripening by Multiple Headspace Solid Phase Microextraction. 15 (2020).
386. Wang, C., Chen, M.-H., Schifano, E., Wu, J. & Yan, J. *A Survey of Statistical Methods and Computing for Big Data.* (Cornell University, 2015).
387. Wu, L., Liu, F. & Cai, H. IOAT: an interactive tool for statistical analysis of omics data and clinical data. *BMC Bioinformatics* **22**, 326 (2021).
388. WISunshine. *IOAT-software.* (2021).
389. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
390. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLOS Comput. Biol.* **13**, e1005752 (2017).
391. mixOmics | Omics Data Integration Project. <http://mixomics.org/>.
392. scikit-learn: machine learning in Python — scikit-learn 1.1.2 documentation. <https://scikit-learn.org/stable/>.
393. Riquelme, G., Zabalegui, N., Marchi, P., Jones, C. M. & Monge, M. E. A Python-Based Pipeline for Preprocessing LC–MS Data for Untargeted Metabolomics Workflows. *Metabolites* **10**, 416 (2020).

CHAPITRE 2 : Développements analytiques

Le sujet de cette thèse porte sur les empreintes volatiles de produits agroalimentaires comme outil pour tracer l'authenticité et suivre des procédés. Pour y parvenir, l'obtention d'échantillons parfaitement tracés et identifiés était nécessaire. Ainsi, deux matrices agroalimentaires fournies par des partenaires intéressés par cette approche ont été étudiées : des fromages Ossau Iraty et des thés. En se basant sur l'étude bibliographique précédente (Cf. Chapitre 1), les techniques d'échantillonnage pour l'analyse des empreintes volatiles des échantillons sont l'espace de tête pour l'analyse par SIFT-MS (HS-SIFT-MS) et l'espace de tête couplé à la micro extraction sur phase solide pour les analyses par GC-MS-FID (HS-SPME-GC-MS-FID). Dans ce chapitre seront présentés les différents développements analytiques réalisés avec tout d'abord la préparation des échantillons, puis les conditions d'espace de tête. Dans un troisième temps seront présentées les mises au point d'extraction par SPME et les conditions d'analyse en GC-MS-FID. Enfin les travaux permettant la mesure SIFT-MS, l'extraction, la compilation et le traitement des données pour les deux méthodes seront présentés.

I. Description des échantillons étudiés

A. Thé

Le partenaire du projet EMVOL pour cette matrice est Lydia Gauthier, sommelière en thé à Paris et qui a la particularité de sélectionner elle-même les thés qu'elle vend. Elle parcourt donc différents pays afin de rencontrer et nouer des partenariats avec des producteurs de thés. Outre la qualité supérieure des produits ainsi sélectionnés, ces échantillons sont extrêmement intéressants du point de vue analytique parce qu'ils sont parfaitement tracés. En effet, dans la filière du thé, il est assez courant de mélanger différents lots de thés. Dans le cadre de cette étude nous avons analysé 26 échantillons de thés de 7 pays, provenant de 11 régions et pour lesquels nous disposons d'informations essentielles comme la région de production, le cultivar utilisé et le mode de transformation.

B. Fromage Ossau-Iraty

Dans le cadre d'une collaboration avec les partenaires du projet BioNaChol (financé par la région nouvelle aquitaine) dont l'objectif est d'étudier l'effet des plantes à métabolite secondaire **bio**actif dans l'alimentation des **ch**èvres et **ovin** laitiers sur la qualité des laits et des fromages en **Nouvelle Aquitaine**, nous avons étudié tout d'abord 9 fromages Ossau-Iraty, provenant de 2 producteurs différents. Le protocole défini par les partenaires était le suivant : pendant 2 semaines, les animaux sont nourris avec une ration alimentaire habituelle appelée « contrôle ». Au bout de ces deux semaines, un fromage est élaboré à partir du lait collecté dans chaque troupeau. Pendant les 2 semaines suivantes, le régime alimentaire est modifié par ajout de Sainfoin pour le producteur E1 ou de chicorée pour le producteur E2. Au bout d'une nouvelle période de 2 semaines, le lait est collecté

et transformé en fromage. Enfin, les animaux reçoivent de nouveau une ration alimentaire « contrôle » pendant 2 semaines supplémentaires au bout desquelles un dernier fromage contrôle est élaboré. Les Tableau 14 et Tableau 15 donnent la composition des rations alimentaires « contrôle » et « modifié » pour chacun des producteurs. Après transformation, les 3 fromages de chaque production sont affinés par un des partenaires selon un protocole identique. Ce protocole a été reproduit sur 2 années permettant ainsi d'obtenir 6 fromages pour chaque producteur. Ces fromages ont été testés par un jury de dégustation et 150 g de chaque fromage a été reçu pour les analyses d'empreintes volatiles.

	Contrôle	Modifié
Foin	8,6%	8,6%
Regain	34,3%	34,3%
Foin de luzerne	29,7%	18,2%
Sainfoin sec	0,0%	11,6%
Maïs en grains	10,9%	10,9%
Concentré azoté	16,5%	16,5%

Tableau 14 : Protocole d'alimentation du producteur E1

	Contrôle	Modifié
Pâturage	29,3%	29,3%
Regain	34,1%	0,0%
Chicorée	0,0%	34,1%
Foin	19,5%	19,5%
Concentré azoté	17,1%	17,1%

Tableau 15 : Protocole d'alimentation du producteur E2

II. Conditions d'échantillonnage

Pour chaque matrice étudiée, les conditions de préparation des différents échantillons ont été déterminées. Il s'agit de :

- la mise en œuvre de l'échantillon en fonction de sa nature : par exemple, le découpage du fromage ou l'infusion du thé,
- la masse d'échantillon analysée,
- le type de flacon d'espace de tête
- les conditions d'espace de tête (température et de temps) pour les analyses SIFT-MS
- les conditions d'espace de tête-SPME (HS-SPME) (température et de temps) pour les analyses en GC-MS/FID.

A. Mise en œuvre de l'échantillon

1. Thé

Pour la préparation des échantillons de thés, la première question posée est : faut-il réaliser des infusions de thé pour se rapprocher des analyses sensorielles ou analyser le thé sec ? En effet, dans la littérature, les deux cas sont retrouvés. Nous avons donc commencé par étudier les composés volatils émis par le thé sec, par le thé infusé et par l'infusât, c'est-à-dire la boisson. Pour cette étude préalable, un thé vert de Ceylan de la marque Ethiquable provenant du Sri Lanka, disponible en grande distribution (Figure 22) a été utilisé.



Figure 22 : Thé vert de Ceylan de la marque Ethiquable utilisé pour le développement analytique

Le thé sec a été utilisé tel quel, le thé humide et l'infusât sont obtenus après infusion du thé sec dans de l'eau chaude à 85°C pendant 3-4 minutes, selon les recommandations de notre experte en thé. 1g de thé sec, 1g de thé humide et 2mL d'infusât ont ensuite été placés dans trois vials d'espace de tête de 10 mL hermétiquement fermés par un bouchon à vis équipé d'un septum (Figure 23).

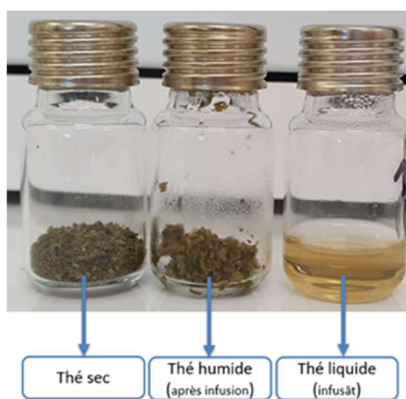


Figure 23 : Thé sec, humide et infusât

Les échantillons ont ensuite été analysés par HS-SPME-GC-MS/FID. Les conditions d'échantillonnage HS-SPME optimales n'ayant pas encore été déterminées, les conditions suivantes ont été fixées : incubation des vials pendant 4 min à 60°C, extraction sur fibre PDMS/DVB/CAR pendant 5 min à 60°C. La fibre a été désorbée à 250°C dans l'injecteur du GC-MS/FID pour analyse des COV extraits (voir condition analytique GC-MS/FID § III.A.). Cette expérience a été répétée 4 fois dans les mêmes conditions pour chaque préparation. Les chromatogrammes obtenus en spectrométrie de masse pour chaque préparation sont présentés sur la Figure 24 et montrent que c'est avec le thé sec que l'on

détecte le plus de COV avec la meilleure sensibilité. De plus, le chromatogramme obtenu avec le thé sec représente l'association des chromatogrammes obtenus avec le thé humide et l'infusât. L'objectif de ce travail étant d'obtenir une empreinte volatile qui représente de manière la plus exhaustive possible le produit, le mode de préparation le plus adapté pour répondre à cet objectif est le thé sec.

De plus, une des limitations de la technologie SIFT-MS est la présence d'eau dans les échantillons qui pourrait compromettre les résultats. En effet, il a été démontré que l'eau présente dans des échantillons peut réagir avec les ions précurseurs positifs notamment avec H_3O^+ et former des ions hydroniums hydratés plus ou moins réactifs^{1,2}. Ces ions hydratés peuvent être considérés à la fois comme de nouveaux ions réactifs et comme des ions produits et donc entrainer des erreurs de quantification pour certains métabolites. De plus, la teneur en eau des échantillons peut influencer les réactions en phase gazeuse impliquées lors de la réaction entre un ion précurseur et un analyte.

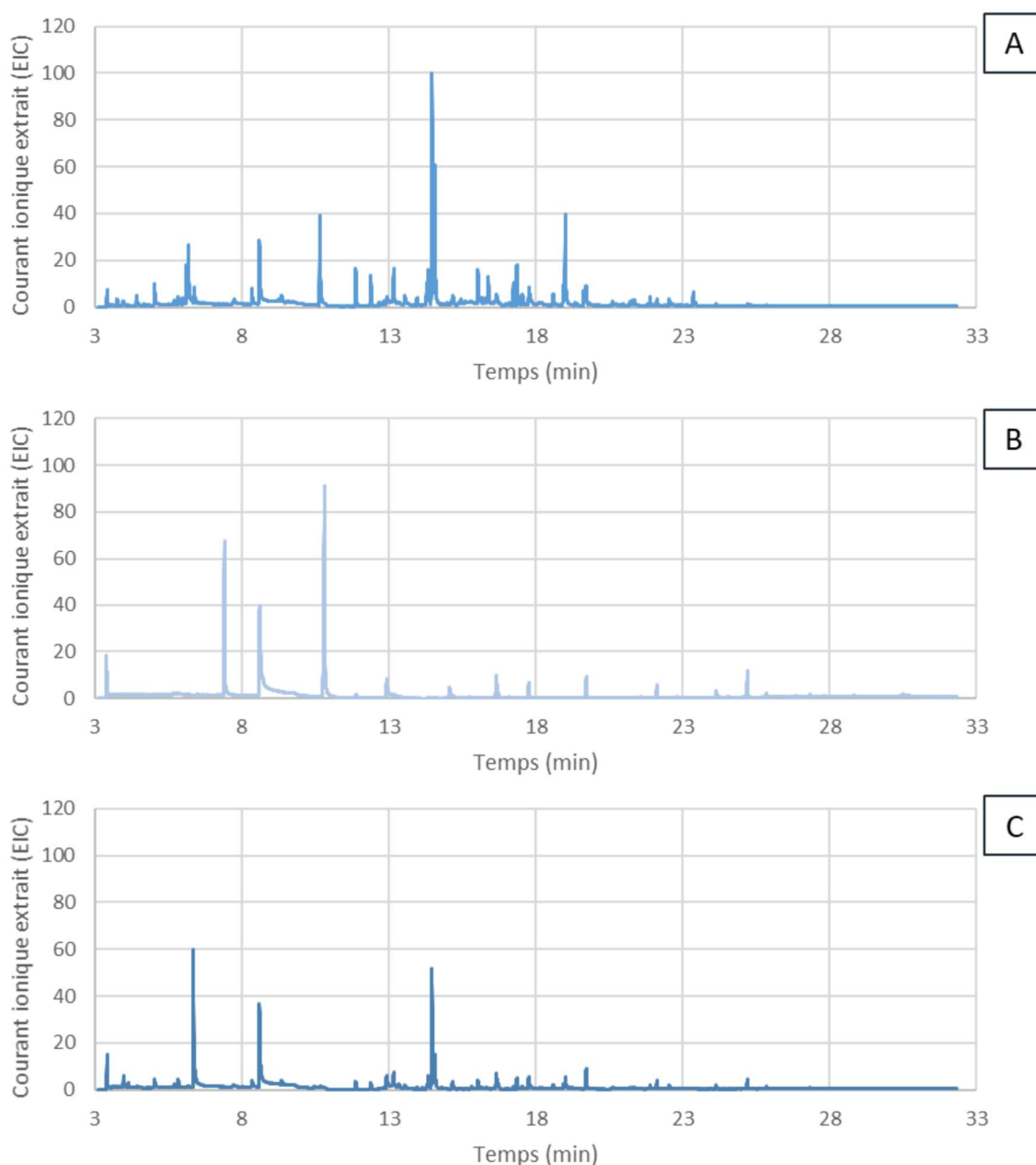


Figure 24 : Chromatogrammes obtenus pour le thé sec (A), humide (B) et l'infusât (C) (acquisition en mode scan 10-250 u.m.a et tracés en mode extraction d'ions (EIC) de 35 à 250 u.m.a)

2. Fromage

Pour les fromages, aucune mise en œuvre spécifique n'a été effectuée, ces aliments étant consommés tels quels. Le fromage est tranché en conservant la croûte et le cœur puis découpé en dés de 1 cm (Figure 25), avant d'être introduit dans les flacons d'espace de tête.

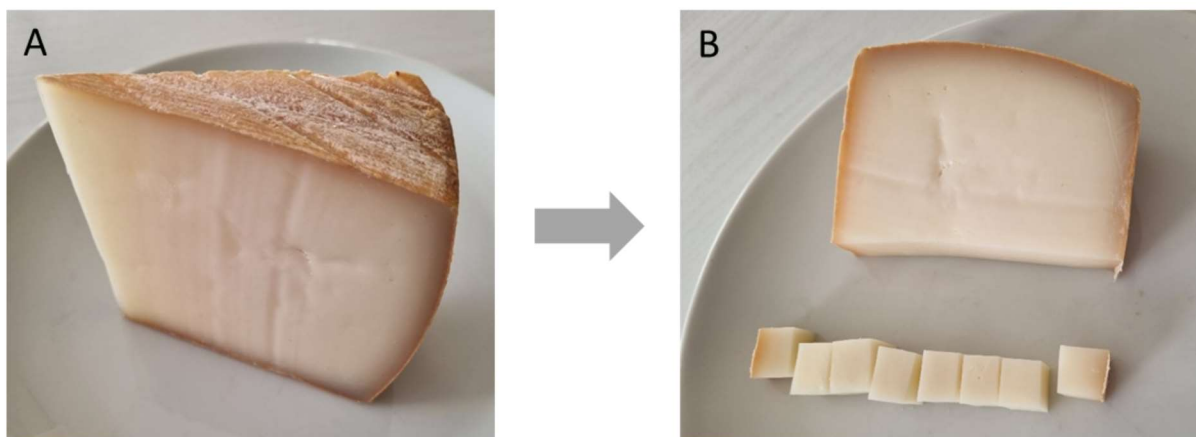


Figure 25 : Photo d'un morceau de fromage reçu (A) et de la découpe (B)

B. Conditions d'espace de tête pour l'analyse SIFT-MS

L'objectif de ces travaux étant de développer l'acquisition et le traitement des empreintes volatiles, l'atteinte de l'équilibre « espace de tête-matrice » est important pour avoir des données comparables d'un échantillon à l'autre. Ainsi, divers paramètres sont à optimiser comme : le volume du flacon, la masse de produit, la température d'extraction et le temps.

Cependant, les contraintes de prélèvement de l'espace de tête par le SIFT-MS ont restreint le choix du flacon et de son volume. En effet, une analyse SIFT-MS full scan positif et full scan négatif prend environ 3 minutes avec un débit d'aspiration fixe de $20 \text{ mL}\cdot\text{min}^{-1}$. Ces conditions non modifiables nous ont contraints à utiliser des flacons de 1L de façon à ce que le volume prélevé pour l'analyse reste inférieur à 10%, pourcentage considéré comme négligeable par rapport au volume total. Pour ces analyses SIFT-MS, il est également crucial que la pression dans le flow tube ne chute pas au cours de la mesure sans quoi le débit d'échantillon évoluera au cours de l'analyse. Le flacon d'espace de tête (Figure 26) est donc muni d'un bouchon étanche à 3 voies : la première est équipée d'un septum permettant d'injecter un étalon interne ou d'exposer une fibre SPME, la seconde d'un tube perfluoroalkoxy (PFA) 0,6mm qui sera relié au SIFT-MS ou à un GC-FID équipé d'une vanne à gaz et la dernière, munie d'un tube PFA 0,6mm de diamètre qui sera relié à un sac de prélèvement en Tedlar de 1L (Supelco, Bellefonte, PA, États-Unis). Le sac Tedlar rempli d'air zéro sec fourni par un générateur d'air Zéro F-DGS (91055 Evry, France) a pour but de compenser le volume d'air prélevé par l'analyse SIFT-MS (prélèvement de $20\text{mL}\cdot\text{min}^{-1}$ pendant 3 min) et éviter une dépression dans le flacon.

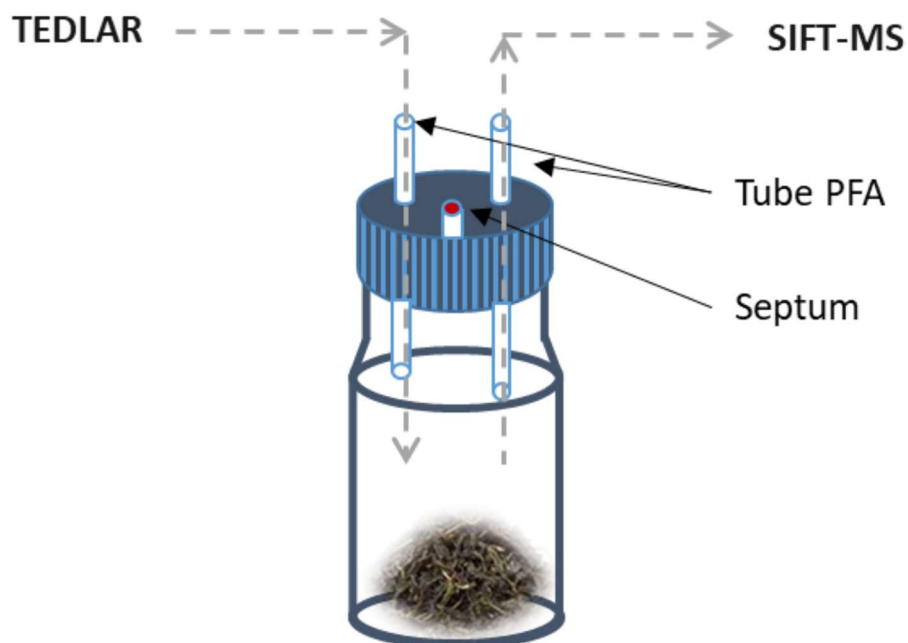


Figure 26 : Schéma du flacon d'espace de tête utilisé pour l'analyse SIFT-MS

1. Optimisation des paramètres pour le thé

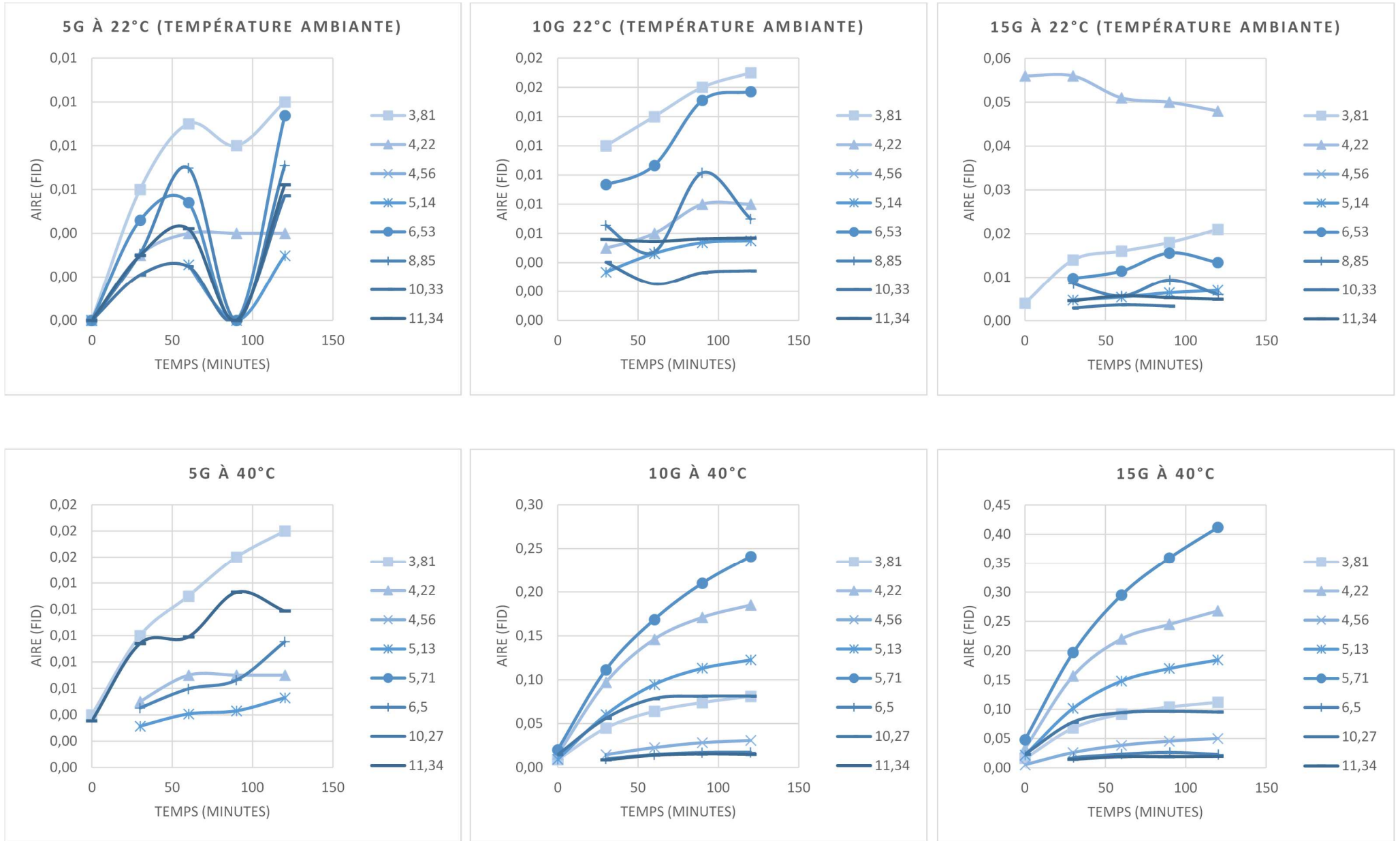
Le volume du flacon étant fixé, les trois paramètres d'espace de tête pour l'analyse du thé sec par SIFT-MS restent à déterminer : la quantité de thé à introduire, la température d'extraction et le temps nécessaire pour atteindre une concentration en COV stable dans la phase gazeuse (équilibre)

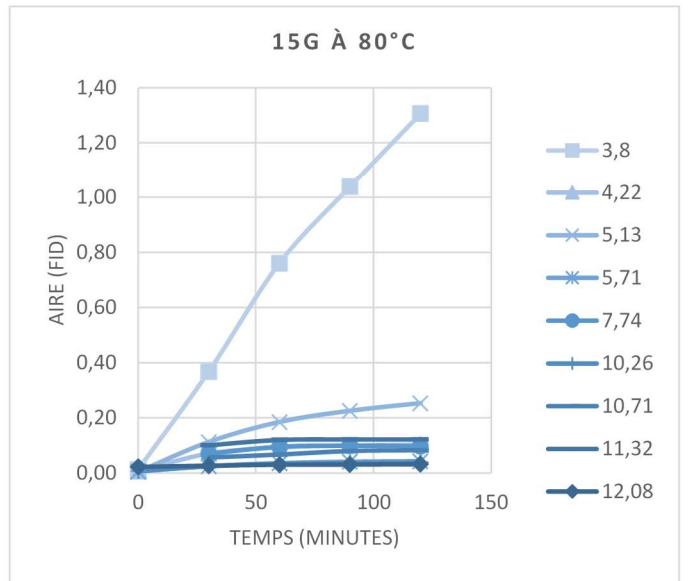
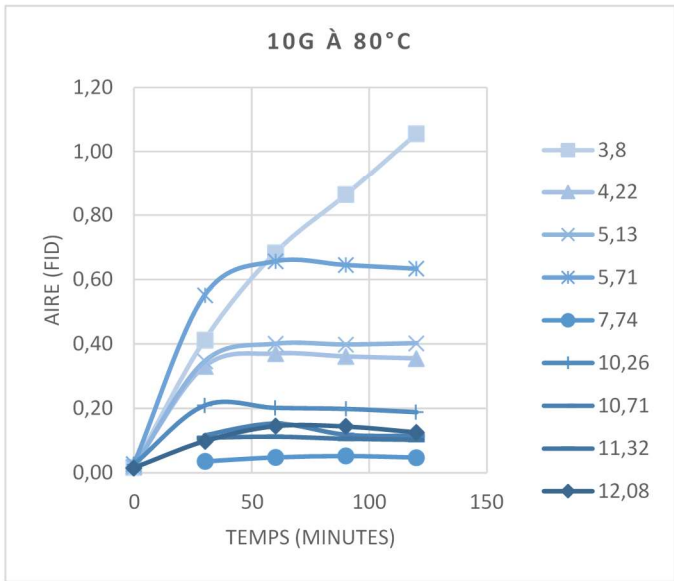
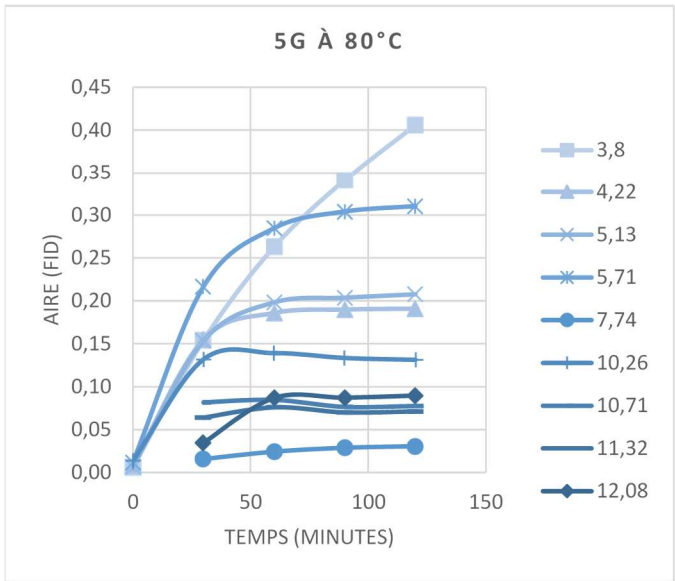
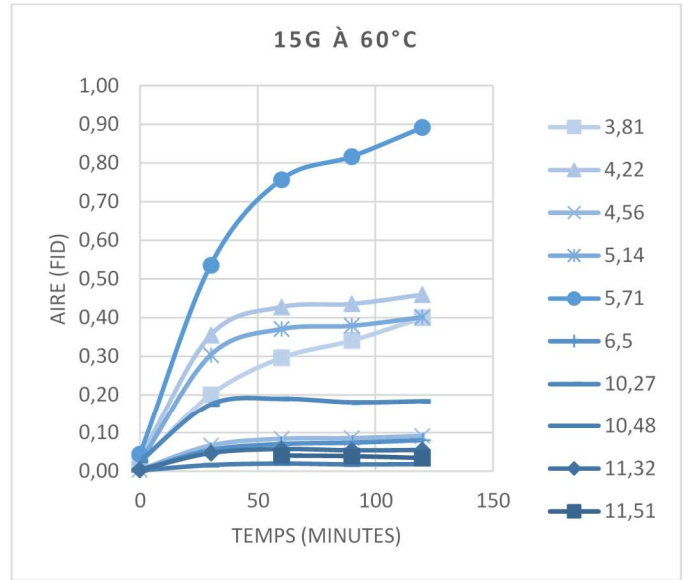
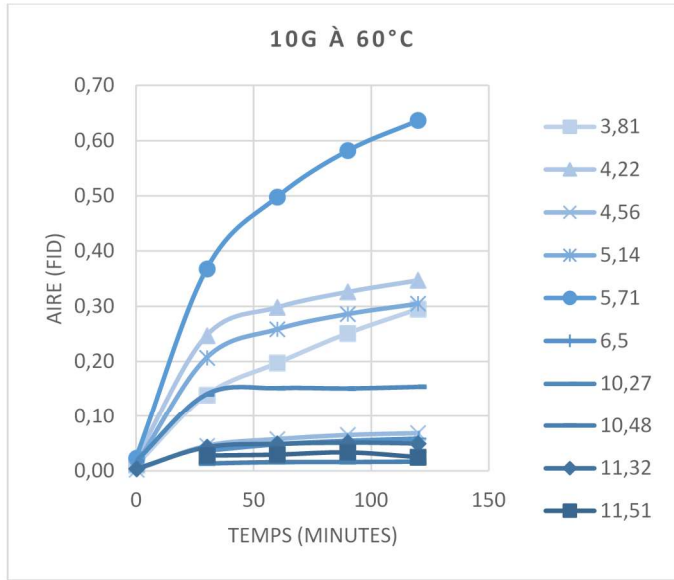
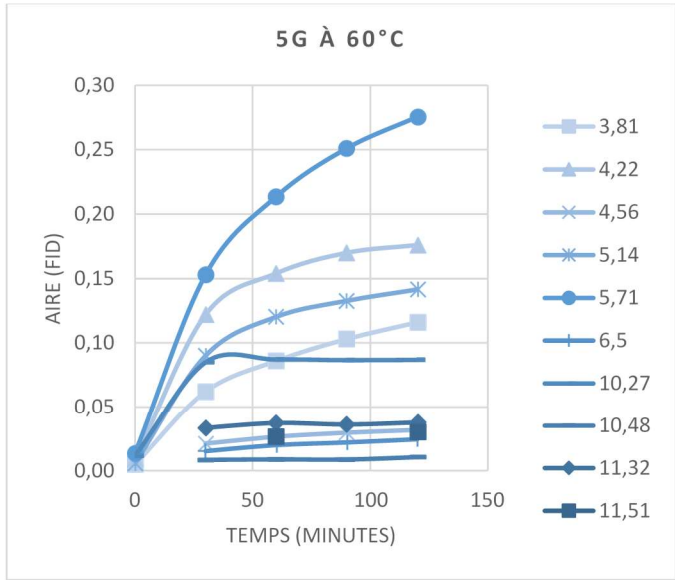
Pour optimiser ces paramètres, dans un premier temps, un chromatographe en phase gazeuse équipé d'un détecteur FID et d'une vanne à gaz automatique a été utilisé pour sa simplicité de mise en œuvre et la fiabilité des mesures en termes de stabilité de la réponse et de répétabilité. Par la suite, les conditions déterminées ont été vérifiées par analyse SIFT-MS.

a. Etude préalable du transfert échantillon-air par GC-FID équipé d'une vanne à gaz

Des cinétiques ont été réalisées en faisant varier la masse d'échantillon (5g – 15g) à différentes températures (température ambiante 22°C, 40°C, 60°C et 80°C). Ces expériences sont réalisées avec un GC-FID Trace 1300 (Thermo Fisher Scientific, Waltham, Massachusetts, États-Unis). Les conditions analytiques sont les suivantes : la boucle de la vanne à gaz est de 500 μL et est chauffée à 80°C. Le débit de la pompe est de 15 $\text{mL}\cdot\text{min}^{-1}$. L'injection de l'espace de tête est réalisée en mode splitless à 200°C dans l'injecteur du GC. La colonne utilisée est constituée d'une phase stationnaire apolaire de 100 % diméthylpolysiloxane (PDMS) (Rtx-1, 30 m \times 250 μm , épaisseur de film de 0,25 μm , Fisher Scientific, Hampton, New Hampshire, États-Unis). L'hélium a été utilisé comme gaz vecteur (débit constant de 1 mL/min). La température du four a été initialement maintenue à 40°C pendant 2 min, puis augmentée de 40°C à 220°C (20°C $\cdot\text{min}^{-1}$) et enfin maintenue à 220°C pendant 2 min. Les paramètres du FID sont : une température de 250°C, un débit d'air de 350 $\text{mL}\cdot\text{min}^{-1}$, un débit d'hydrogène de 35 $\text{mL}\cdot\text{min}^{-1}$ et un débit de make-up (hélium) de 40 $\text{mL}\cdot\text{min}^{-1}$.

Figure 27 : Cinétiques de transfert des COV du thé vers l'espace de tête à différentes températures et avec différentes quantités de thé.





Les cinétiques obtenues selon la quantité de matière et la température d'incubation sont présentées Figure 27. Ces résultats montrent que, très logiquement, les aires des pics augmentent avec la masse d'échantillon et avec la température. A température ambiante, les aires de pics correspondant aux composés étudiés sont relativement faibles, surtout lorsque la masse d'échantillon est faible. Pour 5g, une chute brutale des aires est observée à 90 minutes, due à la faible intensité du signal ne permettant pas d'intégrer les aires sous le pic : ce phénomène à ce temps précis de la cinétique ne s'explique que par un dysfonctionnement ponctuel de la vanne à gaz (pas ou peu de gaz prélevé). L'augmentation de la température à 60°C et 80°C permet de détecter plus de COV. Une augmentation du signal du composé au temps de rétention 3,8 min est observée pour une température de 80°C (60 fois plus grand qu'à température ambiante), que l'on attribue à la formation de ce composé, surement dû à la dégradation d'autres molécules. Pour éviter la dégradation de certains composés, la température d'incubation choisie pour atteindre l'équilibre en espace de tête est 60°C. A 60°C, pour 5g d'échantillon, l'équilibre n'est pas atteint après 2h. Pour des masses de 10g et 15g, l'équilibre est atteint après 60 minutes d'incubation. Pour des raisons de quantités d'échantillons disponibles, nous choisirons une masse de 10g. Ainsi, les conditions choisies pour l'échantillonnage en espace de tête sont une température de 60°C, une quantité de thé sec de 10g et un temps de 1 heure. Ces conditions sont un compromis pour obtenir des conditions d'équilibre, et donc un maximum de sensibilité, sans altération du produit.

b. Validation HS-SIFT-MS

A partir des expériences précédentes réalisées avec un GC-FID équipé d'une vanne à gaz, les conditions expérimentales ont été vérifiées avec le SIFT-MS. Pour cela, 10g de thé sec sont introduits dans un flacon de 1L puis incubés à 60°C. Des cinétiques ont été réalisées en full scan c'est-à-dire en balayant une gamme de m/z de 15 à 250 amu pour chacun des huit ions précurseurs (H_3O^+ , O_2^{*+} , NO^+ , O^+ , OH^- , O_2^{*-} , NO_2^- et NO_3^-). Les mesures ont été effectuées sur le même flacon pour les temps : 0, 2, 4, 6, 8, 10, 15, 20, 25, 60, 90 et 120 minutes. Des ions correspondants à quelques molécules préalablement identifiées sur les chromatogrammes (obtenus pour le thé sec) ont été sélectionnés et suivis au cours du temps (Tableau 16). Les graphiques (Figure 28) représentent le suivi des intensités des ions correspondant aux composés en fonction du temps. Les cinétiques confirment les résultats obtenus avec le GC-FID équipé de la vanne à gaz, avec un équilibre atteint au bout d'une heure d'incubation à 60°C. Ces conditions sont nécessaires pour obtenir des empreintes volatiles stables et comparables d'un échantillon à l'autre.

Ion Précurseurs	m/z	Composé
H3O+	101+	Hexanal
H3O+	87+	Pentanal
H3O+	99+	2-Hexenal, (E)- ; Cyclohexanone
H3O+	89+	Ethyl acetate
H3O+	107+	Benzaldehyde
H3O+	103+	Pentanoic acid
H3O+	117+	Hexanoic acid
H3O+	135+	M-Cymene
H3O+	137+	Camphene ; 3-Carene ; D-Limonene ; Eucalyptol ; Gamma Terpinene
H3O+	139+	Isophorone
H3O+	151+	Carvone
H3O+	205+	Caryophyllene
H3O+	115+	2-Heptanone ; Heptanal
H3O+	155+	l-Menthone
NO+	99+	Hexanal
NO+	85+	1-penten-3-ol ; Pentanal ; 2-Penten-1-ol, (Z)-
NO+	97+	2-Hexenal, (E)-
NO+	118+	Ethyl acetate
NO+	105+	Benzaldehyde
NO+	132+	Pentanoic acid
NO+	146+	Hexanoic acid
NO+	134+	M-Cymene
NO+	136+	Camphene ; 3-Carene ; D-Limonene ; Eucalyptol ; Gamma Terpinene
NO+	168+	Isophorone
NO+	180+	Carvone
NO+	204+	Caryophyllene
NO+	113+	Heptanal
NO+	144+	2-Heptanone
NO+	152+	(+)-2-Bornanone
NO+	154+	Eucalyptol ; l-Menthone
NO+	128+	Cyclohexanone

Tableau 16 : Liste des ions produits suivis et leur composé associé

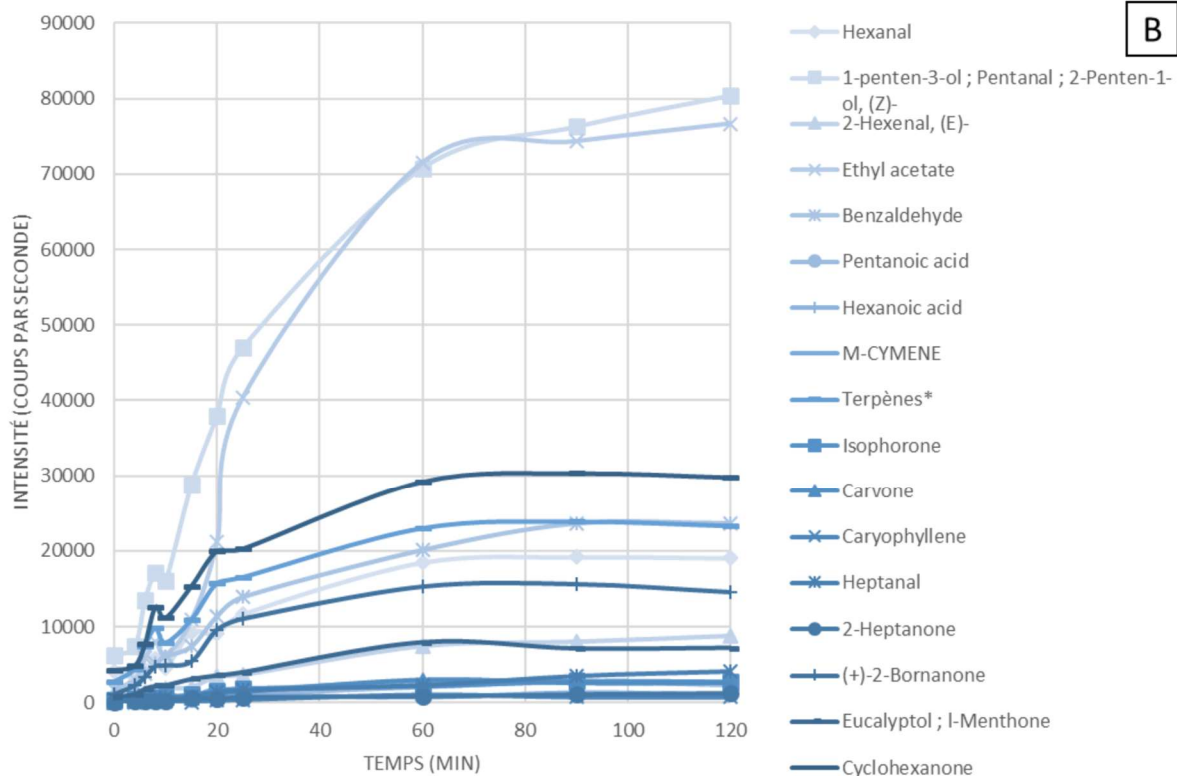
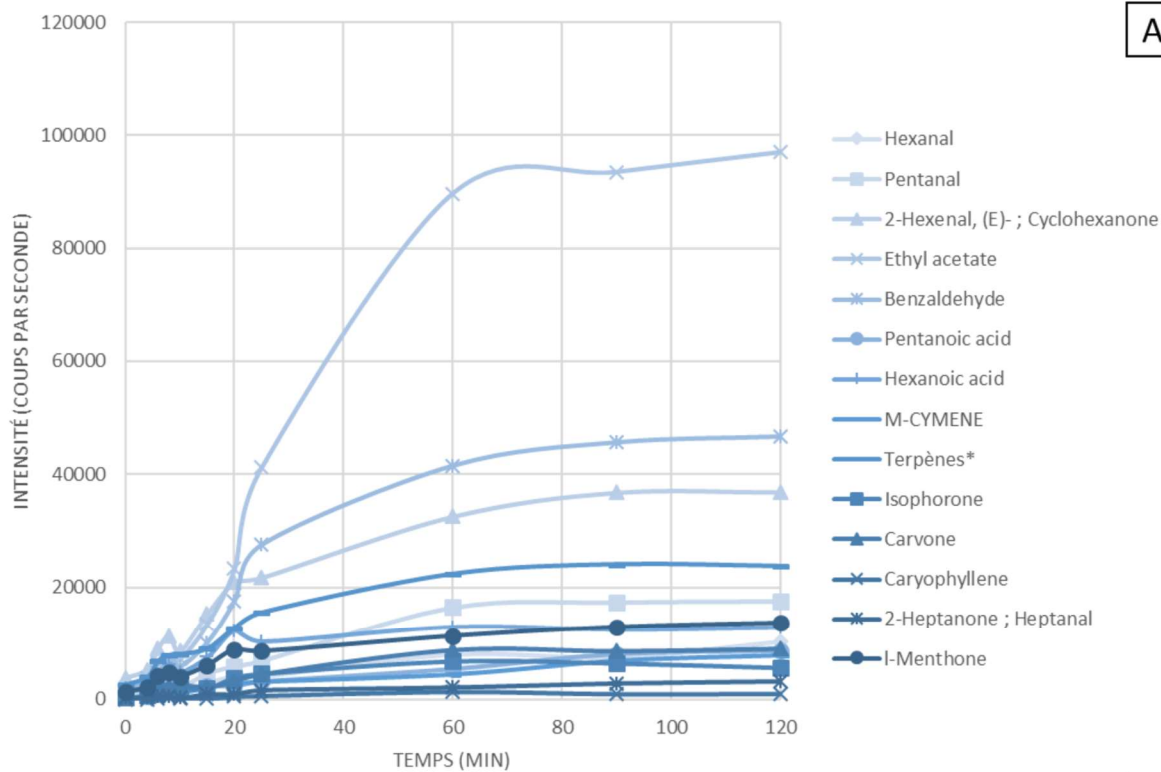


Figure 28 : Cinétique d'émission des COV par SIFT-MS, suivi avec l'ion H_3O^+ (A) et NO^+ (B)
 * Terpènes = Camphène ; 3-Carène ; D-Limonène ; Eucalyptol ; Gamma Terpinène

Les conditions d'échantillonnage par HS pour l'acquisition des empreintes volatiles par SIFT-MS seront : une température de 60°C, un temps d'équilibre de 1h pour une quantité de thé sec de 10g dans un flacon de 1L.

2. Optimisation des paramètres pour le fromage

Les mêmes expériences que précédemment ont été réalisées pour l'optimisation des paramètres d'espace de tête pour l'analyse du fromage par SIFT-MS. Les paramètres étudiés étaient : la quantité (5, 10, 15g), la température (ambiante et 40°C) et le temps d'extraction. Des cinétiques ont été réalisées par vanne à gaz GC-FID en faisant varier la quantité d'échantillon et la température. Les résultats (non présentés) ont montré que les meilleures conditions étaient une température de 22°C (température ambiante) permettant de ne pas dégrader le produit, et un temps d'incubation de 2h afin d'atteindre l'équilibre matrice/espace de tête pour une quantité de 15g et permettant d'avoir à la fois, le cœur du fromage et sa croûte (Annexe 1). Ces conditions ont été vérifiées et confirmées en réalisant une cinétique SIFT-MS (résultats non présentés).

C. Conditions HS-SPME pour l'analyse GC-MS/FID

Pour l'analyse GC-MS/FID, l'extraction des composés volatils des matrices alimentaires est réalisée par la méthode d'échantillonnage d'espace de tête suivie d'une préconcentration sur fibre SPME où deux échanges ont lieu, celui entre la matrice et l'espace de tête et celui entre l'espace de tête et la fibre. (Cf. Chapitre 1. II.A.3.)

1. Optimisation des paramètres HS-SPME pour le thé

Ce développement analytique a pour but de chercher à optimiser les paramètres d'extraction HS-SPME pour obtenir une meilleure sensibilité sur un maximum de COV et obtenir le profil le plus complet. Pour une matrice donnée, il est nécessaire de :

1. Sélectionner le volume du flacon et la quantité de thé
2. Sélectionner la fibre SPME la mieux adaptée,
3. Déterminer la température d'incubation et d'extraction,
4. Déterminer le temps d'incubation permettant d'atteindre un équilibre entre la matrice et l'espace de tête.
5. Déterminer le temps d'extraction de la fibre dans l'espace de tête permettant d'obtenir l'équilibre phase gazeuse/fibre.

a. Volume et quantité

Pour l'analyse SIFT-MS, une masse de 10g a été choisie pour un volume de flacon de 1L pour des raisons de contraintes analytiques. En revanche, pour l'analyse HS-SPME-GC-MS-FID, l'automatisation de l'extraction et donc l'utilisation du passeur automatique CombiPal ((Combi-Pal, CTC Analytics, Zwingen, Suisse) impose d'utiliser des flacons de 10mL. Afin de conserver le même ratio masse/volume que celui appliqué avec les flacons d'1L, il faudrait donc utiliser une masse de 0,1g de thé. Après une étude préliminaire, nous avons constaté que les signaux obtenus en HS-SPME-GC-MS-FID sont trop faibles avec cette quantité. Nous avons donc choisi d'augmenter la quantité de thé à 1g

pour avoir une détection suffisante des COV émis. A l'inverse, nous n'avons pas pu augmenter le ratio masse d'échantillon / volume de flacon pour les analyses SIFT-MS car nous ne disposons pas de quantités suffisantes des différents échantillons de thé. Notre choix est donc de ne pas avoir le même ratio volume / masse d'échantillon entre les deux méthodes : 10g seront utilisés pour l'analyse HS-SIFT-MS et 1g pour l'analyse HS-SPME-GC-MS/FID.

L'optimisation des paramètres HS-SPME pour une analyse par GC-MS/FID est présentée dans les paragraphes suivants.

b. Choix fibre

Différentes fibres ont été étudiées : Polyacrylate (PA) ; Polydiméthylsiloxane /divinylbenzène (PDMS/DVB) ; Polydiméthylsiloxane/ Carboxen (PDMS/CAR) ; Polydiméthylsiloxane/divinylbenzène /Carboxen (PDMS/DVB/CAR).

Pour étudier les capacités de chaque fibre, les autres paramètres influençant l'extraction des COV ont été fixés. Ainsi, 1g de thé sec a été déposé dans des flacons d'espace de tête en verre de 10 mL équipés de bouchons à vis muni d'un septum (Supelco, Sigma-Aldrich, Saint-Louis, Missouri, États-Unis) puis incubés à une température de 60°C pendant 4 minutes. La fibre SPME est alors introduite dans l'espace de tête du flacon pendant 5 minutes, toujours à 60°C, puis est désorbée dans l'injecteur du GC en mode splitless pendant 5 minutes à une température de 250°C pour être analysé.

Les résultats de cette étude sur les fibres SPME sont présentés dans le [Tableau 17](#) qui résume les différents composés détectés par chacune des fibres étudiées avec une indication de la détection ou non du composé. Ces résultats (Figure 29) montrent que le nombre et les aires de pics des composés extraits par les différentes fibres sont très différents. La fibre PA, adaptée à l'extraction des composés polaires³⁻⁵, est celle qui permet d'extraire le moins de composés (seulement 2 composés sur 23). La fibre PDMS/CAR permet d'extraire plus de composés que la fibre PA (10 composés/23) mais moins que les fibres PDMS/DVB et PDMS/DVB/CAR (22 composés/23). La fibre PDMS/CAR détecte moins de terpènes et de composés apolaires. Les fibres PDMS/DVB et PDMS/DVB/CAR sont celles qui permettent d'extraire le plus grand nombre et la plus grande diversité de composés chimiques avec des performances proches. Ces fibres permettent donc d'obtenir un maximum d'informations sur la composition en COV de la matrice en fournissant une empreinte plus riche que les 2 autres fibres testées. En comparant ces résultats expérimentaux à la littérature, ces deux fibres (PDMS/DVB et PDMS/DVB/CAR) sont couramment utilisées pour l'analyse non spécifique des COV dans le thé. Cependant la fibre PDMS/DVB/CAR reste la plus utilisée, que ce soit pour le thé⁶⁻⁸ ou pour d'autres matrices. Ainsi, notre choix se porte sur l'utilisation de cette fibre à trois phases.

Composé	CAS	Tr (min)	Famille	Masse moléculaire (g/mol)	Point d'ébullition (°C)	Pression de vapeur à 25°C	Aires de pic			
							PDMS-DVB-CAR	PDMS-DVB	PDMS-CAR	PA
1-propanol	71-23-8	4,4	Alcool	60,1	97	2,8 kPa	173 329,69	129 921,26	891 969,67	
Éthyle acétate	141-78-6	5,02	Ester	88,11	77	12,4 kPa	5 2362,16	46 775,40		
1-penten-3-ol	616-25-1	6,115	Alcool	86,13	114	1,49 kPa	191 539,19	288 588,47	110 007,93	
Pentanal	110-62-3	6,377	Aldéhyde	86,134	103	4,13 kPa	6 0682,615	44 845,32	31 938,26	
Unknown 1		7,685					8 007,2925	151 16,34		
2-Penten-1-ol, (Z)-	1576-95-0	7,727	Alcool	86,14	138	2,6 hPa	12 413,44	37 458,60		
Hexanal	66-25-1	8,34	Aldéhyde	100,16	129	1,46 kPa	57 728,935	74 370,40	21 235,99	
Unknown 2		8,585					28 221,295	28 906,01		
Cyclohexanone	108-94-1	10,646	Cétones	98,15	155,6	6,6 hPa	372 805,88	378 470,93	203 508,99	
3-carene	13466-78-9	11,866	Terpène	136,24	168	4 hPa	121 123,90	110 542,02	106 534,15	
Camphène	79-92-5	12,38	Terpène	136,24	159	2,6 hPa	93 749,96	85 498,55	125 187,43	
b-pinène	18172-67-3	13,161	Terpène	136,23	166	2,6 hPa	103 669,92	92 103,62	24 602,31	
1-hexanol, 2-ethyl	104-76-7	14,289	Alcool	130,23	184,7	18 Pa	48 656,865	66 799,07		
Cymène	535-77-3	14,332	Terpène	134,21	177	2 hPa	29 403,48	19 279,39		
D-limonène	5989-27-5	14,46	Terpène	136,24	176	2,63 hPa	537 802,04	509 566,78	41821,24	
Eucalyptol	470-82-6	14,56	Terpène	154,25	176-177	2,53 hPa	225 801,71	221 846,24	188 424,52	
Linalool	78-70-6	16,03	Terpène	154,25	198	21 Pa	94 772,47	84 035,73		
Isomenthone	491-07-6	17,36	Cétones	154,25	207	34,13 Pa	51 416,05	58 381,13		
Linalool, formate	115-99-1	18,949		182,26	200	4,53 Pa	48 847,53	46 636,88		
D-carvone	2244-16-8	19,008	Terpène	150,22	230	20,66 hPa	173 955,91	169 029,80		27 759,73
4-isopropoxybutanol	31600-69-8	21,362	Alcool	132,2	?	?				252 866,37
Caryophyllène	87-44-5	21,897	Terpène	204,36	254 - 257	1,73 Pa	15 282,808	15 102,37		
Unknown 3		23,357					22 181,93	23 306,15		

Tableau 17 : Identification et détection des composés extraits avec les différentes fibres étudiées

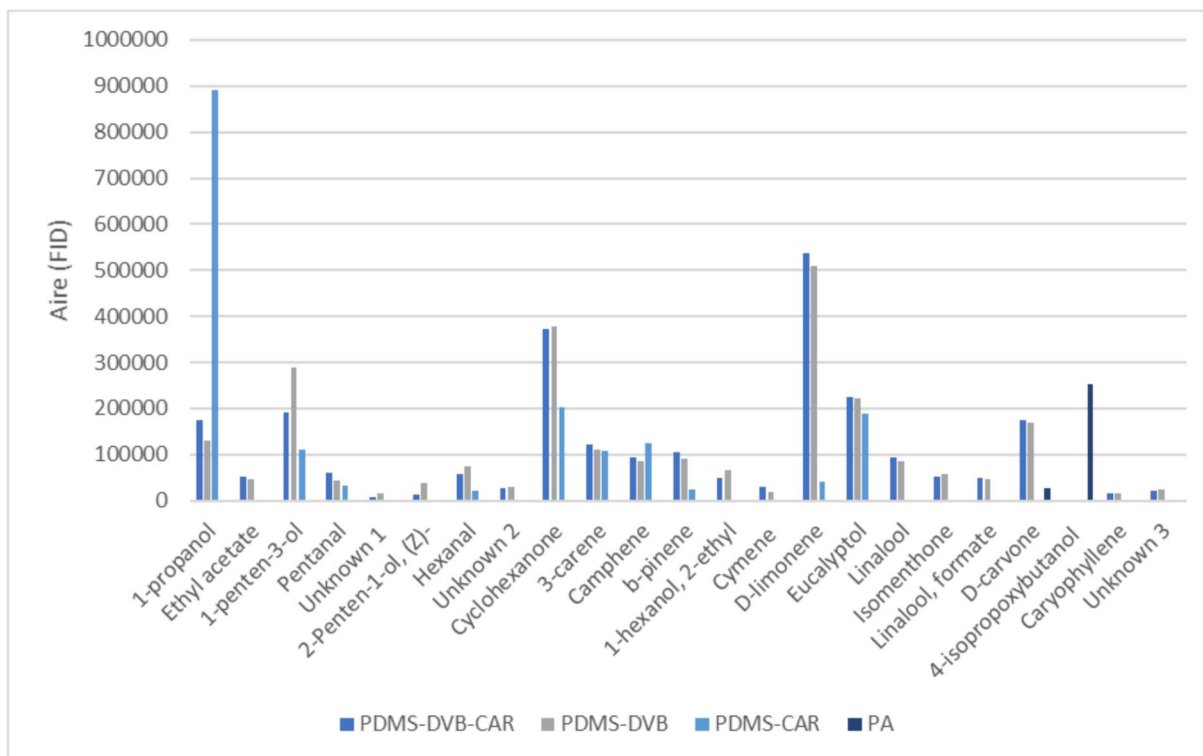


Figure 29 : Aires des pics chromatographiques obtenues à partir du signal FID de quelques composés extraits du thé avec différentes fibres

c. Température d'incubation et d'extraction

La température d'extraction influence le coefficient de partage défini comme le rapport des concentrations du composé d'une part entre la phase gazeuse de l'espace de tête et l'échantillon, et d'autre part le coefficient de partage entre le revêtement de la fibre SPME et la phase gazeuse de l'espace de tête. En effet, l'augmentation de la température d'extraction de l'échantillon va déplacer les équilibres vers la phase gazeuse en augmentant la quantité de COV extraits de l'échantillon mais également limiter l'adsorption de ceux-ci sur la fibre SPME⁹.

Pour déterminer la température optimale d'incubation et d'extraction SPME des COV, les temps d'incubation et d'extraction ont été arbitrairement fixés à 4 et 5 minutes respectivement. Une quantité d'échantillon de 1g a été introduite dans un flacon de 10mL. La fibre PDMS/DVB/CAR sélectionnée précédemment a été utilisée. Trois températures différentes ont été testées : 40, 60 et 80°C. Chaque essai a été réalisé en triplicat. On observe que la température peut avoir des effets différents sur l'extraction de l'analyte en fonction de ses propriétés physico-chimiques (nature chimique, température d'ébullition, volatilité, masse molaire...). Cet effet est illustré dans la Figure 30. L'augmentation de la température fait diminuer l'abondance des composés très volatils comme le 1-propanol pour lequel le signal diminue de 10% et 34% pour des températures de 60°C et 80°C respectivement par rapport au signal à 40°C. Cela peut s'expliquer par des phénomènes de compétition d'adsorption sur la fibre, les composés les plus lourds conduisant à la désorption des plus légers^{10,11}. En effet l'augmentation de température fait augmenter l'abondance des composés moins volatils comme le caryophyllène dont le signal à 40°C et à 60°C représente 14% et 49% respectivement du signal obtenu à 80°C. Même si, à ce stade du développement de la méthode, il n'est pas certain que

les équilibres matrice/air et air/fibre soient atteints, la température de 60°C a été choisie comme le meilleur compromis puisqu'elle permet la détection de composés légers et très volatils (alcools et aldéhydes légers) mais également des composés plus lourds et moins volatils (tels que les terpènes).

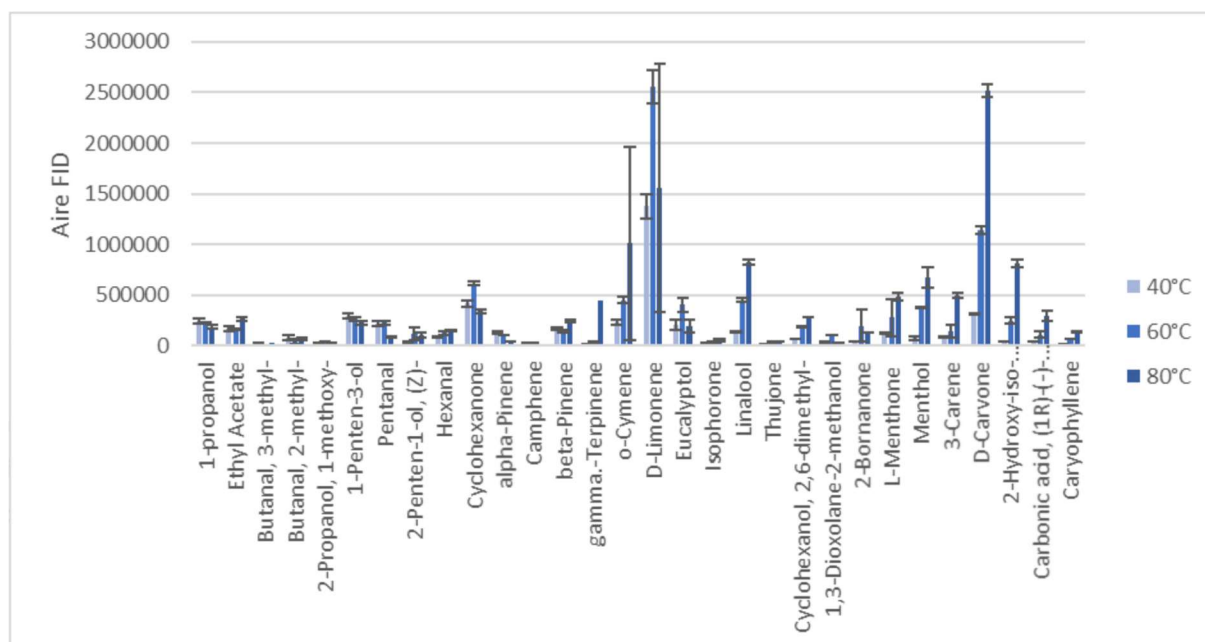


Figure 30 : Effet de la température sur l'extraction des COV présents dans le thé par HS-SPME-GC-FID avec la fibre PDMS/DVB/CAR (aires de pics obtenues en FID)

d. Optimisation du temps d'incubation

Le temps d'incubation optimal correspond au temps nécessaire pour atteindre un équilibre entre la concentration en composé dans la matrice et la concentration dans l'espace de tête. L'optimisation du temps d'incubation est importante pour obtenir une concentration maximale et stable dans l'air au-dessus de l'échantillon. Pour la détermination du temps d'incubation, la température d'incubation et d'extraction a été fixée à 60°C, le temps d'extraction SPME a été fixé à 5 minutes en utilisant la fibre PDMS/DVB/CAR. Sept temps d'incubation ont été testés : 2, 4, 6, 8, 10, 15 et 20 min. Les résultats sont présentés dans la Figure 31 et montrent que l'intensité de certains composés tels que l' α -terpinyl acétate, le benzofuranone, l'hexanal est relativement stable tout au long de la cinétique, alors que pour d'autres composés tels que le cyclohexanone, le 1-penten-3-ol, l'éthyl acétate, l'intensité augmente au début de la cinétique et se stabilise à partir de 10 minutes. Ces résultats indiquent un équilibre entre la concentration des composés dans l'espace de tête et dans la matrice. Cependant, pour la caféine, l'équilibre n'est toujours pas atteint après 15 min. Au vu des résultats, un temps d'incubation de 12 min a été choisi parce qu'il permet de travailler dans une zone d'équilibre pour tous les analytes sauf la caféine qui est un composé relativement lourd et peu volatil (masse molaire : 194,19 g.mol⁻¹ et température d'ébullition : 178°C).

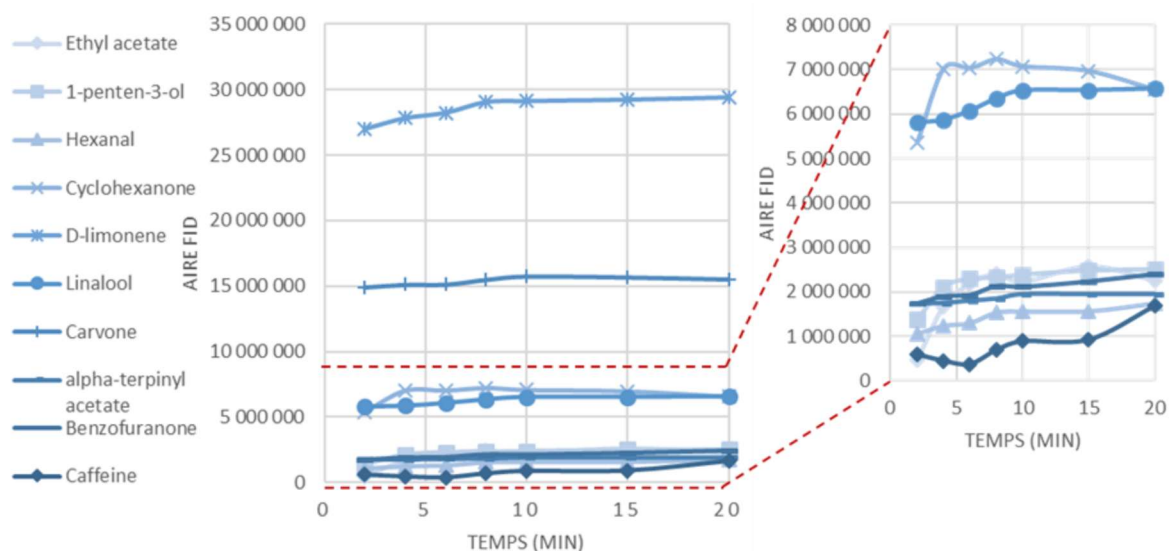


Figure 31 : Cinétique d'émission des COV (aires de pics obtenues en FID) présents dans le thé en faisant varier le temps d'incubation

e. Optimisation du temps d'extraction

Ce temps correspond au temps nécessaire pour atteindre l'équilibre entre la concentration en composé dans l'espace de tête et sur la fibre SPME. Pour la détermination du temps d'extraction sur la fibre PDMS/DVB/CAR, les autres paramètres ont ainsi été fixés : la température d'incubation et d'extraction a été fixée à 60°C avec un temps d'incubation fixé à 12 minutes, et 1g de thé dans un flacon scellé de 10 mL. Les résultats sont présentés dans la Figure 32. Ces résultats montrent une augmentation dans l'espace de tête de la quantité de certains composés tels que le limonène, la cyclohexanone, le 1-penten-3-ol, l'hexanal, l' α -terpinyl acétate et l'éthyle acétate puis d'une stabilisation du signal à partir de 10 minutes indiquant un équilibre pour ces composés entre la matrice et l'espace de tête. En revanche, pour la carvone, le linalool, la benzofuranone et la caféine, composés peu volatils dont la masse molaire est supérieure à 150 g.mol⁻¹ et la température d'ébullition supérieure à 170°C, l'équilibre n'est pas atteint même après 20 min d'extraction. Cependant, un temps d'extraction plus long ne serait pas favorable à l'extraction des composés plus volatils comme le montrent les résultats pour l'éthyle acétate ou le 1-penten-3-ol puisqu'une diminution du signal est observée entre 15 et 20 minutes, en raison très probablement de compétitions d'adsorption sur la fibre. Ainsi, un temps d'extraction de 10 minutes a été choisi comme compromis pour l'extraction des COV par SPME avec une fibre PDMS/DVB/CAR.

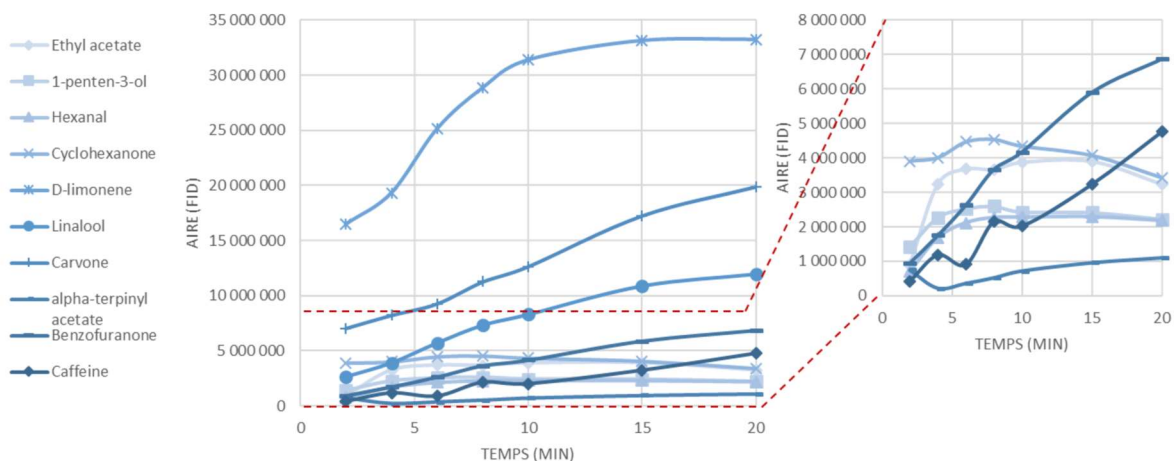


Figure 32 : Cinétique d'extraction des COV (aires de pics obtenues en FID) présents dans le thé par HS-SPME-GC-MS/FID avec une fibre PDMS/DVB/CAR, un temps d'incubation de 12 min et une température de 60°C

f. Conclusion

Pour obtenir un profil de COV le plus complet et le plus exhaustif possible d'échantillons de thés, les paramètres HS-SPME déterminés sont : une quantité de 1g placée dans un flacon de 10mL scellé, un temps d'incubation de 12 minutes à une température de 60°C, l'exposition d'une fibre PDMS/DVB/CAR à l'espace de tête pendant 10 minutes, toujours à 60°C.

2. Optimisation des paramètres HS-SPME pour le fromage

Le développement analytique des conditions HS-SPME pour l'analyse d'échantillon de fromage a consisté à étudier les mêmes paramètres que pour les échantillons de thé soit : la taille du flacon, la masse, la température, le temps d'incubation et d'extraction ainsi que le choix de la fibre. Les mêmes expériences ont été réalisées (résultats non présentés). Concernant la taille du flacon et la quantité de matrice, à l'inverse du thé, le volume de 1L choisi pour l'analyse SIFT-MS a été conservé afin d'assurer une bonne représentativité de l'échantillon de fromage (incluant le cœur et la croûte) et obtenir des résultats reproductibles d'une prise d'essai à l'autre. La quantité de 15g déterminée lors de l'optimisation des paramètres d'espace de tête pour l'analyse SIFT-MS est choisie ainsi que le temps d'incubation de 2h à température ambiante.

Les flacons de 1L sont équipés d'un septum permettant d'introduire la fibre SPME. Ainsi, le développement des conditions SPME a été réalisé directement sur ces flacons.

Les paramètres optimisés permettant l'obtention d'un profil en COV complet et exhaustif sont : un temps d'incubation de 2h à température ambiante (22°C) suivi d'une extraction sur fibre PDMS/DVB/CAR à température ambiante avec un temps d'extraction sur fibre de 15 minutes (Annexe 2).

D. Conclusion

Les protocoles d'échantillonnage développés pour l'analyse des COV par HS-SIFT-MS et HS-SPME-GC-MS/FID sont présentés dans la Figure 33 pour les échantillons de thés et dans la Figure 34 pour les échantillons de fromage.

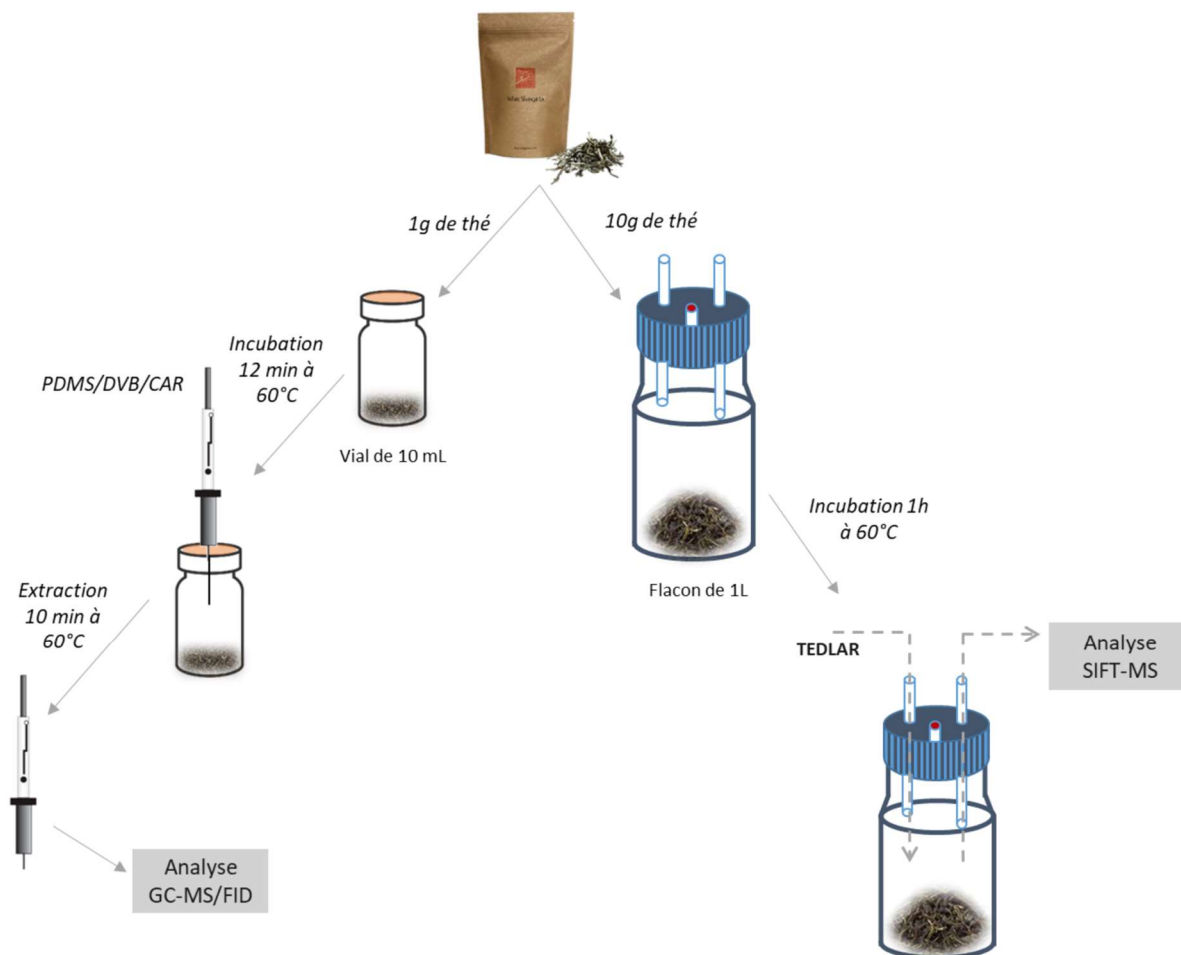


Figure 33: Schéma du protocole d'échantillonnage du thé pour l'analyse des COV par HS-SIFT-MS et HS-SPME-GC-MS/FID

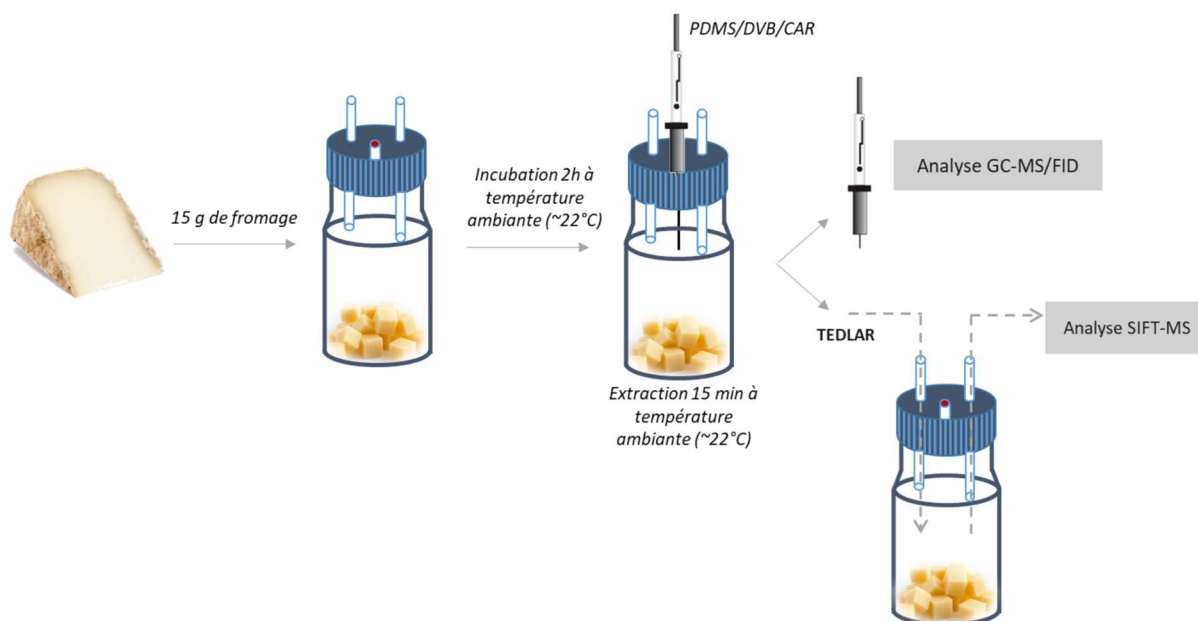


Figure 34 : Schéma du protocole d'échantillonnage du fromage pour l'analyse des COV par HS-SIFT-MS et HS-SPME-GC-MS/FID

III. Conditions analytiques

A. GC-MS/FID

Les composés volatils émis par les différents échantillons étudiés ont été analysés avec un chromatographe en phase gazeuse (GC) modèle Agilent 7890B (Agilent Technologies, Palo Alto, CA, USA) couplé à une double détection par spectrométrie de masse (modèle 5977B, Agilent Technologies, Palo Alto, CA, USA) et ionisation de flamme. La colonne utilisée est une colonne composée d'une phase stationnaire apolaire de méthylpolysiloxane 5%-Phényl- (ELITE-5MS, 60 m × 250 µm, épaisseur de film de 0,25 µm, Perkin-Elmer, Waltham, MA, États-Unis)¹². Ces colonnes plutôt généralistes sont couramment utilisées pour l'analyse non ciblée de composés dans une large gamme de propriétés physico-chimiques. Elles offrent un faible ressuage et une forte inertie même à des températures élevées. La désorption des fibres a été réalisée en mode splitless à 250°C pendant 6 minutes dans l'injecteur du GC équipé d'un liner spécifique de diamètre interne 0,8 mm. L'hélium a été utilisé comme gaz vecteur (débit constant de 1 mL/min). La température du four a été initialement maintenue à 40°C pendant 4 min, puis augmentée de 40°C à 90°C (15°C/min), puis de 90°C à 250°C (10°C/min) et enfin maintenue à 250°C pendant 5 min¹². Les températures de la ligne de transfert et de la source d'ions du spectromètre de masse ont été réglées à 250°C. En spectrométrie de masse, l'ionisation a été réalisée par impact électronique à 70 eV avec une acquisition en mode scan sur une gamme de masses de 15 à 250 u.m.a. La température du détecteur FID est réglée à 250°C. Le débit d'hydrogène dans le détecteur est de 30 mL.min⁻¹ et le débit d'air est de 400 mL.min⁻¹. L'hélium est utilisé comme gaz d'appoint avec un débit de 30 mL.min⁻¹.

B. SIFT-MS

Un SIFT-MS Voice 200 Ultra (SYFT Technologies, Christchurch, Nouvelle-Zélande) équipé d'une double source produisant des ions précurseurs d'ionisation douce positive et négative (H_3O^+ , $\text{O}_2^{\bullet+}$, NO^+ , $\text{O}^{\bullet-}$, OH^- , $\text{O}_2^{\bullet-}$, NO_2^- et NO_3^-) est utilisé. Cet instrument utilise l'azote comme gaz vecteur, l'échantillon est introduit par une ligne d'échantillonnage à température de 110 °C et à débit contrôlé de 20 mL.min⁻¹ (High Performance Inlet HPI®). L'instrument est calibré quotidiennement avec un gaz standard (mélanges de gaz Scott™, Air Liquid USA) contenant des composés à 2.0 ppmV dans l'azote (Air Liquide, Alphagaz 2) (Tableau 18).

Pour chaque échantillon, une expérience à blanc avec une bouteille vide est réalisée avant chaque analyse en triplicat. Chaque expérience à blanc et chaque échantillon est analysé en utilisant deux méthodes d'analyse full scan consécutives balayant une gamme de m/z de 15 à 250 u.m.a, l'une pour les ions précurseurs positifs (H_3O^+ , $\text{O}_2^{\bullet+}$ et NO^+) et l'autre pour les ions négatifs ($\text{O}^{\bullet-}$, OH^- , $\text{O}_2^{\bullet-}$, NO_2^- et NO_3^-). La vitesse de balayage du quadrupole est de 20 ms par masse et le nombre de répétitions de scan par ions précurseurs est de 2. Les fichiers de données brutes de balayage complet contenant les intensités des ions produits dans la gamme 15-250 u.m.a sont collectés pour une étape de compilation et de prétraitement.

Composés	Concentration (Mole %)	Incertitudes analytiques	Ions précurseurs utilisés
1,2,3,4,tetrafluoro Benzene	2,100	± 5%	$\text{O}_2^{\bullet+}$, $\text{O}_2^{\bullet-}$
Benzene	2,110	± 5%	H_3O^+ , $\text{O}_2^{\bullet+}$
Ethylbenzene	2,040	± 5%	$\text{O}_2^{\bullet+}$
Isobutane	2,020	± 5%	NO^+
Octafluorotoluene	2,110	± 5%	$\text{O}_2^{\bullet+}$, OH^-
Hexafluorobenzène	2,100	± 5%	$\text{O}_2^{\bullet+}$, $\text{O}^{\bullet-}$, OH^-
Toluène	2,130	± 5%	H_3O^+ , NO^+

Tableau 18 : Composition du mélange de gaz d'étalonnage

C. Etalonnage

Les méthodes de quantification s'appuient généralement sur le principe d'étalonnage consistant à utiliser un étalon pour déterminer le rapport entre la valeur affichée par l'instrument de mesure et la valeur attendue. L'étalonnage permet également de surveiller des variations instrumentales qui peuvent apparaître tout au long d'une campagne de mesure ou encore des effets matrices afin de corriger le signal. On retrouve trois types de méthodes d'étalonnage externe, interne et par ajouts dosés.

Ces travaux ne sont pas basés sur la quantification de marqueurs spécifiques mais sur la comparaison de mesure d'empreintes volatiles. Le questionnement sur l'ajout d'un étalon interne dans nos échantillons s'est donc posé pour contrôler la stabilité de l'instrument, la fiabilité des résultats et donc permettre de comparer les empreintes volatiles acquises. Le choix de l'étalon interne doit correspondre à la fois aux mesures HS-SPME-GC-MS/FID et aux mesures HS-SIFT-MS c'est-à-dire qu'il doit être facilement extrait par la fibre sélectionnée et analysé par la colonne chromatographique utilisée et réagir avec des ions précurseurs pour l'analyse SIFT-MS. Selon ces critères, l'étalon interne

choisi est l'hexafluorobenzène (C_6F_6) de masse moléculaire $186,06 \text{ g.mol}^{-1}$ possédant une température d'ébullition de 82°C et une pression de vapeur de $84,5 \text{ mm Hg}$

1. SIFT-MS

Dans un premier temps, un étalonnage externe de l'hexafluorobenzène a été réalisé. Deux solutions mères de C_6F_6 , notées « S1 » et « S2 » ont été réalisées avec, respectivement, $1\mu\text{L}$ et $3\mu\text{L}$ de C_6F_6 dans une bouteille de 1L remplie d'air zéro. La concentration en C_6F_6 dans les solutions « S1 » et « S2 » est de 207 ppmV et 623 ppmV , respectivement. A partir de ces deux solutions, une gamme d'étalonnage a été établie allant de 10 à 1557 ppbV dans de l'air zéro en utilisant des seringues à gaz (Tableau 19).

Concentration C_6F_6 (ppb)	10	52	104	208	518	623	1558
Volume S1 (μL)	50	250	500	1000	2500	/	/
Volume S2 (μL)	/	/	/	/	/	1000	2500
Volume flacon (L)	1	1	1	1	1	1	1

Tableau 19 : Gamme d'étalonnage C_6F_6

A partir de cette gamme, la constante de réaction du composé C_6F_6 envers l'ion précurseur O_2^{+*} a été recalculée dans les conditions de la mesure pour atteindre une meilleure précision ($\pm 10\%$) (Figure 35). La constante de réaction corrigée du C_6F_6 par la gamme est de $2,08 \cdot 10^{-9} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$.

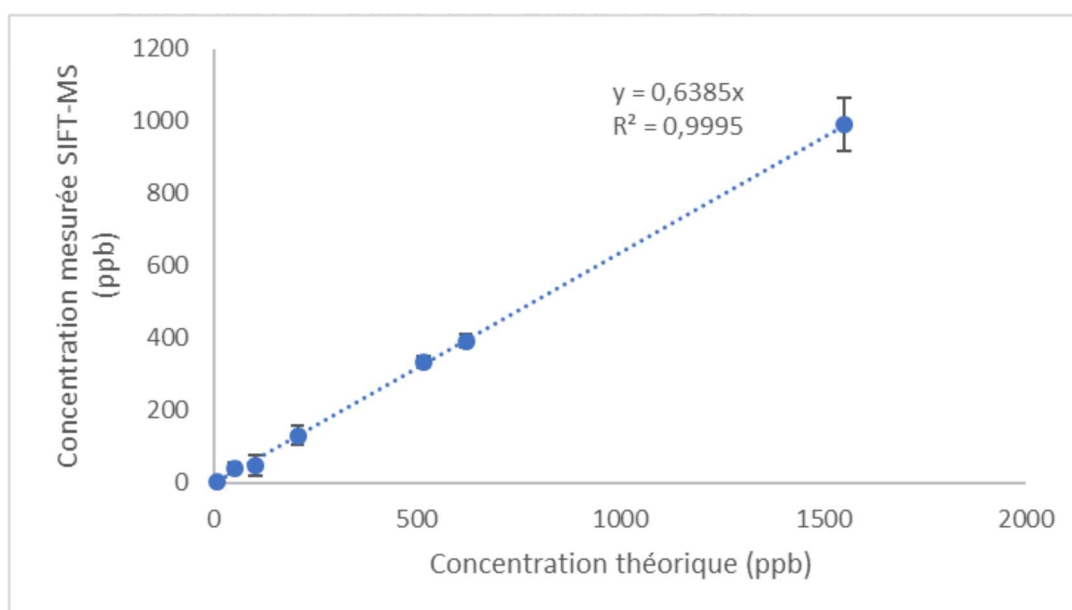


Figure 35 : Concentration expérimentale de C_6F_6 en fonction de la concentration théorique

Pour assurer la répétabilité de la quantification de C_6F_6 , 12 échantillons de concentrations connues ont été réalisés et analysés dans les mêmes conditions. Pour cela, une solution mère est réalisée avec $1\mu\text{L}$ de C_6F_6 introduit dans un flacon de 1L (207 ppm). Pour chaque répéta, $2,5 \text{ mL}$ du ciel gazeux de la solution mère sont prélevés pour être introduits dans un nouveau flacon de 1L. Les flacons sont ensuite analysés par SIFT-MS en mode SIM pour quantifier spécifiquement l'hexafluorobenzène. Les 10

répétas ont été effectués sur deux solutions mères, de 1 à 5 sur la première et de 5 à 10 sur la deuxième.

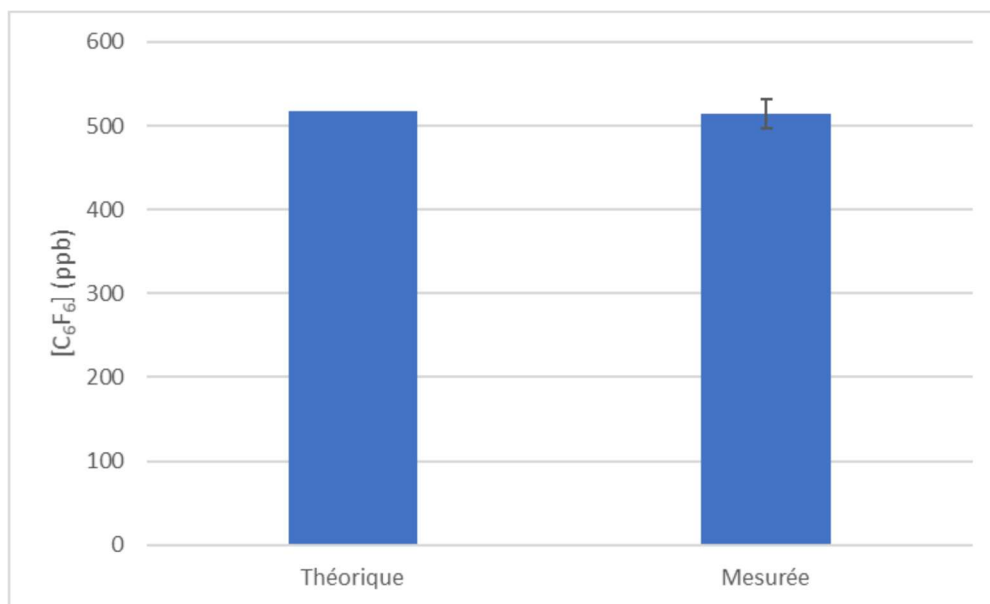


Figure 36 : Histogramme représentant la concentration théorique attendue et la moyenne des concentration mesurées en C_6F_6

Les résultats présentés en Figure 36 montrent une bonne répétabilité de la mesure pour la quantification de l'hexafluorobenzène avec une moyenne de 514,6 ppb pour les 10 répétas et un RSD de 3%.

Cependant, si la quantification de l'hexafluorobenzène seul s'est avérée être satisfaisante, lorsque celui-ci a été introduit comme étalon interne dans un échantillon et mesuré de façon non spécifique pour acquérir les empreintes volatiles, les résultats se sont montrés aberrants, sûrement en raison de la présence d'ions conflits. En effet, l'étalon interne C_6F_6 à une concentration connue (518 ppb) a été introduit à l'état gazeux dans l'espace de tête du flacon où se situe un échantillon de fromage avant analyse et sa concentration a été déterminée. La Figure 37 montre les valeurs de concentration déterminées dans 6 fromages analysés en triplicats. La concentration de l'étalon interne varie de 4924 à 9597 ppb. Ces valeurs sont environ 10 à 20 fois plus élevées que la concentration attendue (518 ppb). Cette surévaluation peut s'expliquer par la présence d'une ou plusieurs molécules au sein de la matrice qui produiraient le même ion produit que l'hexafluorobenzène c'est-à-dire l'ion O_2^{+} 186. Ces résultats montrent qu'il s'avère délicat de trouver un étalon interne dont la volatilité soit similaire à celle des molécules étudiées, qui réagisse avec les ions précurseurs du SIFT-MS mais qui n'induit pas de conflit avec les molécules de la matrice étudiée.

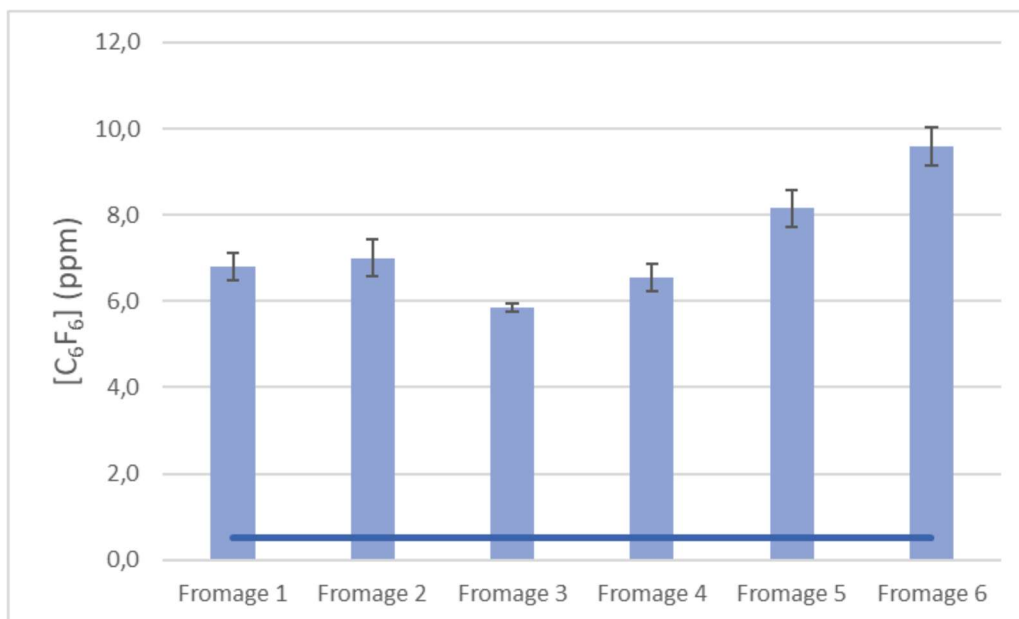


Figure 37 : Histogramme représentant la concentration de l'étalon interne C₆F₆ introduit dans l'espace de tête des flacons contenant des échantillons de fromages. La barre représente la concentration injectée dans les flacons.

Ainsi, l'utilisation d'un étalon interne pour le suivi de la stabilité de l'instrument a donc été abandonnée au profit de la mise en place des systèmes de contrôle présentés un peu plus loin (§ IV.B.2.a).

2. GC-MS/FID

L'utilisation de l'étalon interne C₆F₆ pour l'analyse HS-SPME-GC-MS/FID de matrices alimentaires a également été étudiée. Dans un premier temps, des mesures de l'étalon interne seul à une concentration connue ont été effectuées. Pour cela, une solution mère d'hexafluorobenzène à 207 ppmV a été préparée puis une dilution à une concentration attendue de 518 ppbV. L'espace de tête de ce flacon a été analysé par HS-SPME-GC-MS/FID dans les conditions suivantes : fibre PDMS/DBV/CAR exposée dans l'espace de tête du flacon pendant 15 minutes à température ambiante. Ces conditions correspondent à celles mises en place pour la matrice fromage puisque l'intérêt est d'étudier si l'étalon interne peut être utilisé dans ces conditions, similaires à celles développées pour l'analyse par SIFT-MS. Plusieurs réplicas ont été effectués pour suivre la stabilité du signal correspondant à l'étalon. Dans un deuxième temps, l'étalon interne a été ajouté dans l'espace de tête d'un flacon contenant un échantillon de fromage afin d'étudier la réponse de l'hexafluorobenzène en présence des composés émis par la matrice. Les chromatogrammes en signal FID obtenus pour ces deux expériences sont présentés en Figure 38. Le pic correspondant à C₆F₆ se situe à un temps de rétention de 4,99 minutes. Une bonne détection de C₆F₆ seul est observée, en revanche, lorsqu'il est introduit dans l'espace de tête du fromage, il se retrouve co-élué avec d'autres composés extraits de la matrice. En effet, le fromage est une matrice riche en composés volatils et notamment en acides dont les pics sont mal résolus sur une colonne apolaire comme celle utilisée. Les aires des pics de C₆F₆ seul ou dans la matrice sont similaires, indiquant que la même quantité de C₆F₆ a été extraite par la fibre (pas de compétition d'adsorption avec des composés volatils émis par le fromage). Cependant, une co-élution de C₆F₆ est observée avec l'acide acétique rendant impossible la quantification du composé (Figure 38-B). Ces résultats montrent que l'étalon interne C₆F₆ ne s'avère pas être non plus

un choix judicieux pour les analyses HS-SPME-GC-MS/FID. La richesse et la complexité du profil chromatographique obtenu pour le fromage complique le choix d'un étalon interne. En effet, dans l'approche des empreintes volatiles, trouver des étalons qui conviennent à plusieurs matrices alimentaires dont la nature des COV est très variable est un véritable challenge.

Pour l'analyse en HS-SPME-GC-MS/FID, nous avons donc aussi abandonné l'utilisation d'un étalon interne. Il a été choisi de suivre le signal FID des COV émis par les matrices. En effet, la réponse FID est plus stable dans le temps que celle d'un spectromètre de masse (MS) et possède une gamme de linéarité de réponse plus étendue.

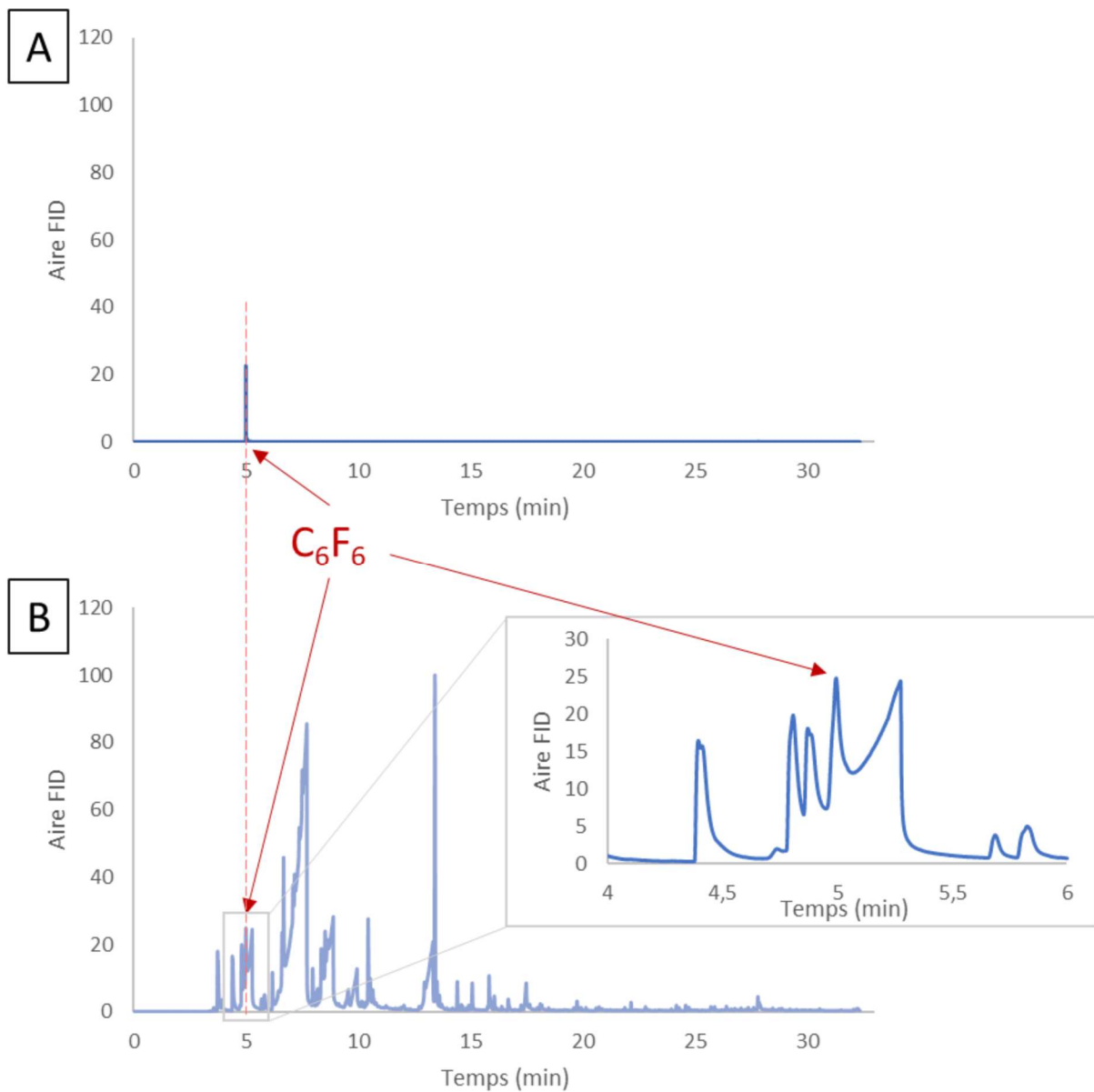


Figure 38 : Chromatogrammes (signal FID) obtenus pour l'analyse HS-SPME-GC-MS/FID de l'hexafluorobenzène seul (A) et dans l'espace de tête d'un échantillon de fromage (B)

IV. Traitement des signaux

L'approche de l'empreinte volatile vise à traiter l'intégralité des signaux obtenus par les techniques analytiques étudiées. L'objectif de ce paragraphe est de présenter les outils mathématiques et informatiques utilisés et développés pour extraire, prétraiter et visualiser les données obtenues. Enfin, les tests statistiques utilisés pour comparer les empreintes volatiles obtenues seront présentés.

A. Données HS-SPME-GC-MS/FID

1. Analyse ciblée

Le traitement des données obtenues par HS-SPME-GC-MS-FID est composé de deux étapes. La première consiste à identifier les molécules présentes sur le chromatogramme via les spectres de masse. Pour chaque composé, les spectres sont comparés à ceux présents dans la banque de données NIST (National Institute of Standards and Technology, v2.2-2014). Les composés présentant un pourcentage de correspondance supérieur à 70% ont été sélectionnés. La deuxième étape consiste à utiliser les aires des pics obtenus en FID comme données quantitatives. Le jeu de données obtenu comprend donc les molécules identifiées et leurs aires de pic en FID.

2. Analyse non ciblée

Pour le traitement des profils de composés volatils obtenus par HS-SPME-GC-MS, le logiciel Chromcompare + (SepSolve Analytical, Royaume-Uni) a été utilisé. La méthode créée permet de lire les fichiers « Mass Hunter » du spectromètre de masse, d'aligner l'ensemble des chromatogrammes d'une banque d'échantillons selon le temps de rétention et le pic moléculaire du composé, et de réaliser l'intégration des chromatogrammes pour obtenir des « tiles ». Ces « tiles » correspondent à des portions de chromatogramme de 5 secondes (plus ou moins 20%) associées à des « features » c'est-à-dire à l'ion moléculaire identifié dans cette plage de temps sur le spectre de masse (obtenu en mode scan sur une gamme de masses de 15 à 250 amu). En se basant sur l'étude de Spadafora et al (2022)¹³ et sur les conseils du fournisseur, un nettoyage des données est réalisé en effectuant un filtrage initial en choisissant un descripteur selon lequel on souhaite trier les échantillons (origine géographique, couleur ...). Le filtrage initial a été appliqué avec une intensité absolue minimale de 1000 et une fréquence minimale du descripteur de 100%. Une découverte de caractéristiques (appelée « Features discovery ») est appliquée en sélectionnant le descripteur étudié et le nombre souhaité de features. Les features correspondent aux variables (temps de rétention associé à un pic moléculaire) qui vont discriminer le paramètre étudié. Plus ce nombre sera faible, plus la discrimination sera forte, les variables communes seront éliminées et seules les variables discriminantes seront retenues.

B. Données HS-SIFT-MS

1. Analyse ciblée

La quantification des composés par SIFT-MS est réalisée en utilisant le logiciel du fournisseur : LabSyft. Il s'agit d'une base de données contenant des paramètres et des données de références pour détecter et quantifier des composés cibles. Pour quantifier des composés, une méthode SIM (Single Ion Monitoring) est créée en sélectionnant les composés d'intérêt, qui doivent être présents dans la base de données Syft ou dans celle de l'utilisateur. En revanche, des précautions doivent être prises pour évaluer chaque composé, comme la présence d'ions conflits, c'est-à-dire deux composés dont les rapports m/z peuvent être communs et ainsi interférer pour l'identification. Le logiciel permet de supprimer les masses inappropriées (conflituelles) ou identifiées comme potentiels ions conflits lors de la création de la méthode, ainsi ces ions peuvent être supprimés de l'analyse. En revanche, aucune information sur les ions pouvant être « conflits » avec d'autres composés de la matrice que ceux listés dans la méthode n'est donnée. L'analyse en mode SIM va mesurer le courant ionique c'est-à-dire le nombre d'ions (coups) par secondes atteignant le détecteur pour des masses d'ions produits sélectionnées. Le logiciel va par la suite transformer les coups par seconde en concentration par le biais des calculs décrits dans le chapitre 1 § II.B.4.

Par rapport aux analyses en full scan, les analyses SIM ciblent des composés spécifiques pour une analyse plus sensible. Les résultats quantitatifs obtenus via le mode SIM permettent d'évaluer en temps réel les changements de concentration des composés.

2. Analyse non ciblée

Pour rappel, l'analyse SIFT-MS en mode full scan fait intervenir 8 ions précurseurs (H_3O^+ , NO^+ , O_2^{*+} , O^+ , OH^- , O_2^{*-} , NO_2^- , NO_3^-) qui vont interagir avec les analytes émis par la matrice étudiée pour former des ions produits. La détection des ions produits par réaction avec chaque ion précurseur est réalisée en balayant une gamme de ratios m/z allant de 15 à 250 uma. L'analyse SIFT-MS dans son intégralité génère donc 1880 ions produits. Pour rappel, chaque échantillon est analysé en triplicats avec une mesure de blancs, ainsi, nous obtenons donc 6 fichiers contenant 1880 données à exploiter. Ces données brutes sous format XML sont relativement complexes à étudier sans logiciel d'extraction de données.

a. Développement algorithmique

Pour faciliter l'exploitation des données, un algorithme d'extraction nommé « **pysiftms** » a été développé (Annexe 3). Il s'agit d'un algorithme programmé sous Python fournissant un ensemble d'outils pour lire et analyser les données brutes d'une série d'expériences SIFT-MS en mode full scan. Cet algorithme permettant le prétraitement des données consiste à lire les données présentes dans les fichiers XML des blancs et des échantillons puis à extraire les informations nécessaires tels que l'ion précurseur, le rapport m/z de l'ion produit, la période, le datum correspondant aux coups par seconde et l'ICF, facteur de correction de l'instrument. L'algorithme soustrait le bruit de fond (signal du flacon vide) pour chaque réplica, puis fait la moyenne des triplicats pour chaque échantillon afin d'obtenir

une seule valeur de signal pour chaque ion. Ensuite, l'algorithme détermine la fréquence selon le calcul suivant :

$$F = \frac{Datum}{Period} \times 10^6$$

Avec $F =$ Fréquence (Hz)

$Datum =$ nombre moyen d'ions atteignant le détecteur (coups)

$Period =$ temps de mesure (μ s)

Par la suite, le calcul de correction de la fréquence (§ IV.B.2.c) est appliqué. Ainsi, le jeu de données final est organisé en matrice notée X avec n lignes correspondant aux échantillons analysés et p colonnes correspondant aux intensités (valeurs moyennes des fréquences corrigées des triplicats pour chaque échantillon) des ions du produit.

b. Contrôle

Pour assurer la fiabilité, la qualité et la reproductibilité des résultats obtenus, des points de contrôle ont été étudiés de même que le contrôle des blancs et la correction du signal.

Les blancs correspondent aux mesures effectuées sur les flacons d'espace de tête vides avant introduction de l'échantillon. Les flacons sont lavés au lave-vaisselle, mis à sécher à l'étuve et sont balayés par de l'air zéro toute la nuit (12h environ) précédant la mesure afin d'éliminer toute contamination. De plus, ces mesures de blancs permettent de vérifier l'état de l'instrument et l'absence de contaminants dans le SIFT-MS. Afin de s'assurer de l'absence de contamination, une carte de contrôle a été mise en place ainsi qu'une détection de valeur atypique. La carte de contrôle a pour but d'examiner l'état de l'instrument et des flacons d'espace de tête en fonction des jours de mesure. Pour cela, une moyenne des blancs réalisés sur une journée est effectuée et est comparée à un seuil de tolérance défini. Dans notre cas, le seuil de tolérance correspond à la moyenne de l'ensemble des blancs réalisés au cours de cette thèse à laquelle on ajoute trois fois l'écart type. Les graphiques (Figure 39) montrent deux exemples de résultats obtenus pour les cartes de contrôle des moyennes. Pour l'ion H_3O^+ 97, les moyennes de chaque jour de mesure sont inférieures à la limite déterminée. En revanche pour l'ion H_3O^+ 123, la moyenne des blancs du 9 Juin 2020 est supérieure au seuil déterminé. Cette valeur haute peut être due à une contamination de l'instrument ou des flacons utilisés.

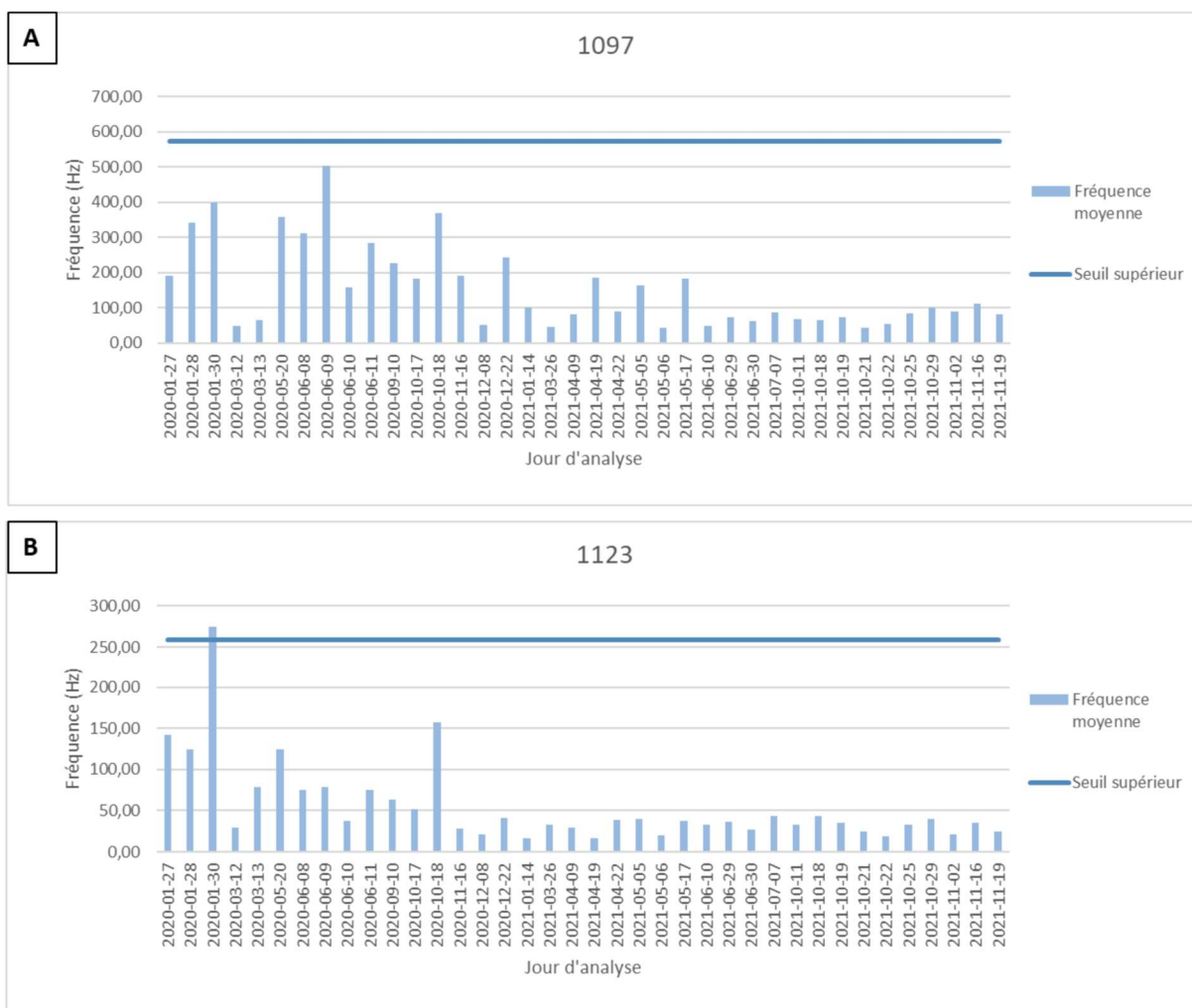


Figure 39 : Graphique des cartes de contrôle pour l'ion H_3O^+ 97 (A) et H_3O^+ 123 (B)

Pour répondre à cela, une détection de valeur aberrante est utilisée. Celle-ci consiste à détecter une valeur (point de données) très éloignée de la moyenne. Contrairement à la carte de contrôle, ce test évalue la dispersion de chaque mesure par rapport à une moyenne et à une limite supérieure déterminée par le calcul suivant : $Q_3 + 1,5(Q_3 - Q_1)$ où Q_3 correspond au 3^{ème} quartile du jeu de données c'est-à-dire la valeur pour laquelle 75% des valeurs sont inférieures et Q_1 le 1^{er} quartile du jeu de données c'est-à-dire la valeur pour laquelle 25% des valeurs sont inférieures. Les graphiques (Figure 40) montrent les signaux obtenus pour l'ion H_3O^+ 123 dans les blancs flacons pour la journée du 9 Juin 2020. Ces résultats montrent que les trois premiers flacons sont contaminés et non l'instrument, sinon l'ensemble des signaux seraient élevés et potentiellement supérieurs au seuil limite. En utilisant la base de données labsyft, l'ion H_3O^+ 123 pourrait correspondre au diethyl disulfure, à l'acide benzoïque ou encore au 4-éthyl-phénol. Ces résultats montrent à quel point le contrôle des blancs est un processus important pour obtenir des données de qualité et fiables.

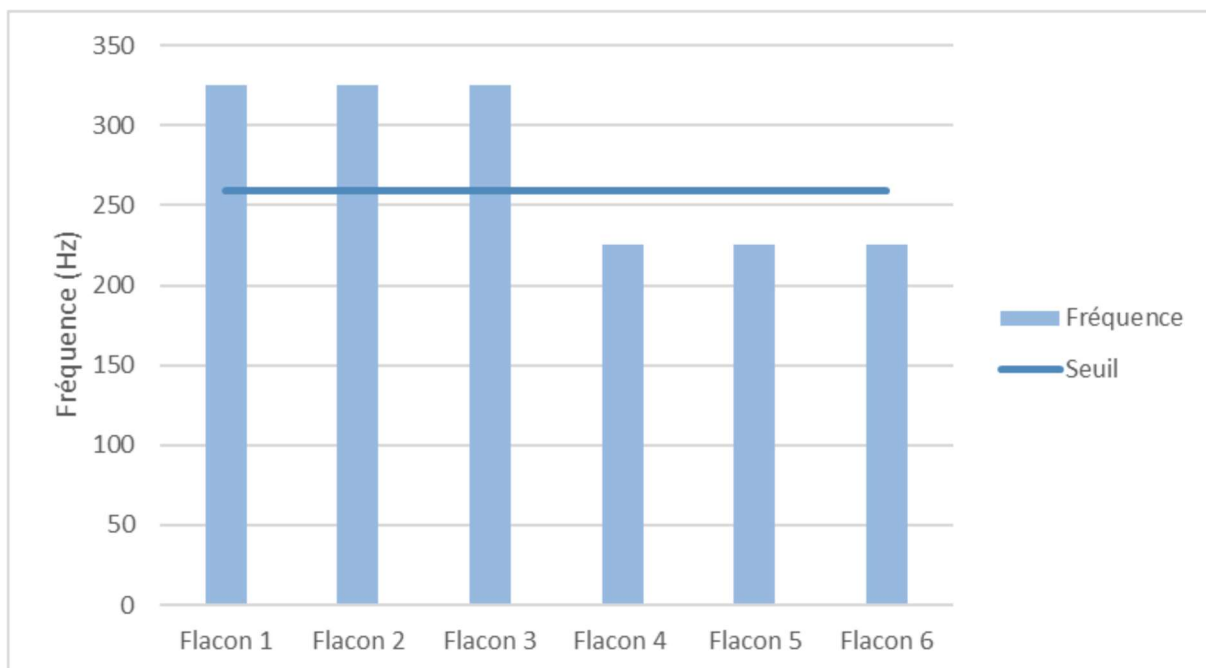


Figure 40 : Histogramme des valeurs de fréquence obtenus pour l'ion H_3O^+ 123

c. Correction du signal

Divers facteurs de correction ont été étudiés pour corriger les fréquences acquises. L'objectif de la correction du signal est de pouvoir réaliser des campagnes de mesures étendues (plusieurs mois entre différentes campagnes), comparer des résultats entre deux instruments ou pallier les effets des maintenances instrumentales sur les résultats. Les facteurs étudiés sont :

- L'impact de la quantité d'ions précurseurs. En effet, la génération des ions précurseurs est dépendante de l'état de l'instrument et plus particulièrement du plasma.
- L'ICF, facteur de correction instrumentale qui est déterminé quotidiennement avant chaque utilisation de l'instrument.
- Le blanc : Les résultats obtenus lors de l'analyse des blancs est important, cette mesure permet de soustraire le bruit de fond lié au flacon, à la mesure SIFT-MS ou à de potentielles contaminations.

Les différentes hypothèses pour corriger le calcul des fréquences sont présentées dans le Tableau 20 ci-après :

Hypothèses	Calculs
1. Sample	F_{sample}
2. Sample ICF	$F_{sample} \times ICF_{sample}$
3. Default	$F_{sample} - F_{blank}$
4. Default ICF	$(F_{sample} \times ICF_{sample}) - (F_{blank} \times ICF_{blank})$
5. Corrected	$\frac{(F_{sample} \times ICF_{sample}) - (F_{blank} \times ICF_{blank})_{ions\ produits}}{(F_{sample} \times ICF_{sample}) - (F_{blank} \times ICF_{blank})_{ions\ précurseurs}}$
6. Normed Blank	$\frac{(F_{sample} \times ICF_{sample}) - (F_{blank} \times ICF_{blank})_{ions\ produits}}{(F_{blank} \times ICF_{blank})_{ions\ précurseurs}}$
7. Normed Sample	$\frac{(F_{sample} \times ICF_{sample}) - (F_{blank} \times ICF_{blank})_{ions\ produits}}{(F_{sample} \times ICF_{sample})_{ions\ précurseurs}}$
8. Ratio Blank	$\frac{(F_{sample} \times ICF_{sample})_{ions\ produits}}{(F_{blank} \times ICF_{blank})_{ions\ précurseurs}}$
9. Ratio Sample	$\frac{(F_{sample} \times ICF_{sample})_{ions\ produits}}{(F_{sample} \times ICF_{sample})_{ions\ précurseurs}}$

Tableau 20 : Hypothèse de calculs pour corriger la fréquence

L'hypothèse 1 consiste à n'appliquer aucun facteur de correction et à simplement étudier les signaux de fréquences bruts obtenus. L'hypothèse 2 prend en considération le facteur de correction de l'instrument l'ICF en le multipliant à la valeur de fréquence. L'hypothèse 3 est la soustraction des fréquences obtenues lors du blanc à celle de l'échantillon correspondant. L'hypothèse 4 prend en considération la soustraction du blanc qui est multiplié par le facteur de correction instrument ICF. Les hypothèses 5, 6, 7, 8 et 9 prennent en compte la correction par le signal des ions précurseurs. L'hypothèse 5 consiste à soustraire les blancs des échantillons en appliquant le facteur ICF en divisant par le signal de l'ion précurseur considéré. Les hypothèses 6 et 7 normalisent le signal obtenu en soustrayant le signal du blanc à celui de l'échantillon en multipliant par l'ICF par le signal de l'ion précurseur soit du blanc soit de l'échantillon. Les hypothèses 8 et 9 sont identiques à celle 6 et 7 mais ne tiennent pas compte de la soustraction par le blanc.

Après application des hypothèses, une analyse non-supervisée en composante principale (ACP) (§ V.A.) a été réalisée sur les jeux de données obtenus par l'analyse de différents thés après avoir testé les différentes hypothèses. Les échantillons sont regroupés et colorés en fonction de leur date d'analyse (AnnéeMoisJour). L'ACP permet d'étudier la répartition spontanée du jeu de données et la discrimination sur les composantes.

La première hypothèse prenant en compte la fréquence brute obtenue donne de bons résultats concernant l'étude du profil volatil et la discrimination des échantillons, cependant la présence du

bruit de fond impacte la fiabilité de la discrimination et met en évidence des variables dont la fréquence est très faible. L'hypothèse 2 tenant compte de l'ICF ne modifie pas le résultat obtenu avec l'hypothèse 1. L'hypothèse 3 permet de s'affranchir du bruit de fond et donc des ions d'intensité relativement faible. L'hypothèse 4 permet d'accentuer la discrimination des échantillons en tenant compte de la soustraction du bruit de fond et du facteur de correction instrumental ICF (Figure 41-A). Les calculs effectués pour les hypothèses 5, 6 et 7 normalisent et centralisent fortement les données et par conséquent rassemblent les échantillons en masquant les grandes différences pouvant être observées sur certains signaux (Figure 41-B). De ce fait, la discrimination des échantillons est fortement impactée. Les hypothèses 8 et 9 regroupent les échantillons en fonction de leur date d'analyse (Figure 41-C), ce qui semble indiquer que la soustraction du blanc permet de pallier la variation temporelle de la réponse. En effet, la quantité d'ions précurseurs est impactée par l'état de l'instrument le jour de l'analyse mais également par les éventuelles maintenances qui peuvent (ont pu) avoir lieu sur l'instrument.

Notre choix pour corriger les valeurs de fréquences se porte sur l'hypothèse 4 qui prend en compte deux facteurs de correction : la soustraction du blanc pour réduire le bruit de fond et éliminer les potentielles contaminations et la multiplication par l'ICF permettant de tenir compte de l'état de l'instrument le jour de la mesure. Cette correction permet de lisser la variabilité instrumentale pouvant impacter les résultats obtenus sur des campagnes de mesures éloignées (plusieurs mois entre deux ou plusieurs campagnes) mais également de pouvoir comparer des résultats obtenus sur divers instruments.

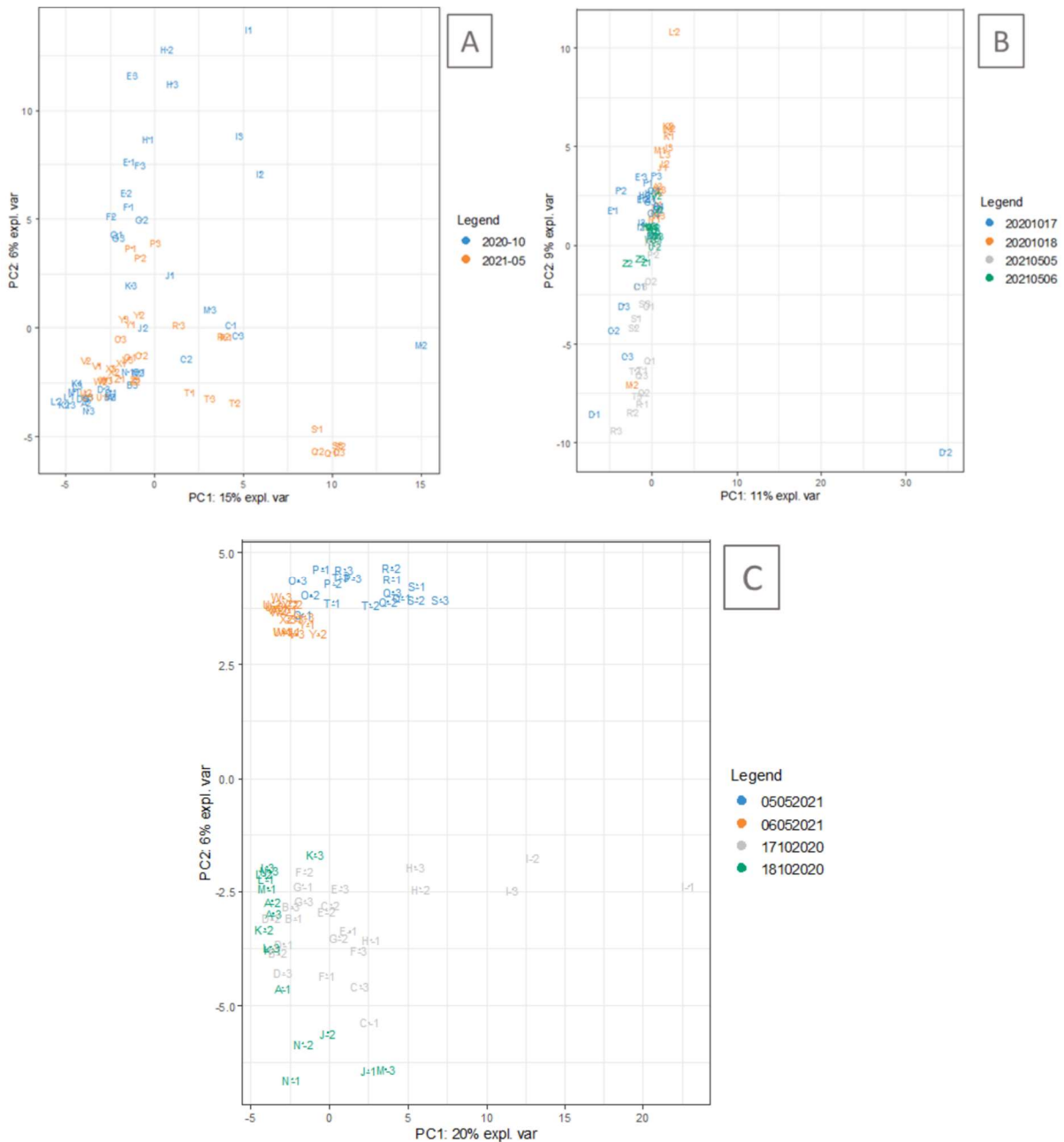


Figure 41 : ACP sur les données des thés dont les fréquences ont été corrigées par l'hypothèse 4 (A), 5 (B) et 9 (C). Les codes couleurs correspondent aux dates des analyses (AnnéeMoisJour).

Le facteur instrumental ICF est calculé quotidiennement lors des étapes de calibration de l'instrument. D'après le constructeur du SYFT-MS, l'ICF est le facteur qui permet de transposer les résultats d'un instrument à un autre. Cet ICF est un paramètre clé pour la quantification de molécules puisqu'il intervient dans le calcul. Une comparaison des ICF de deux instruments le même jour a été réalisée (Figure 42). On observe que le comportement de l'ICF en fonction du rapport masse sur charge des ions produits est le même pour les deux instruments. Pour les deux instruments, les profils d'ICF obtenus ne sont pas les mêmes pour les ions positifs que pour les ions négatifs. Une diminution de la valeur d'ICF est observée au fur et à mesure que le rapport m/z augmente jusqu'à atteindre une valeur stable qui augmente légèrement à partir d'un m/z de 150. A noter que la diminution de la valeur ICF des ions négatifs (de 15 à 30 m/z) est plus lente que pour les ions positifs (de 15 à 20 m/z). Cependant,

si les profils s'avèrent être les mêmes pour les deux instruments, les valeurs d'ICF quant à elles sont très différentes d'un instrument à l'autre, principalement pour les ions de faibles rapports m/z (< 30). Cette différence est particulièrement grande pour les ions négatifs de rapport m/z 15 et 16. Cette observation soulève de nombreuses questions comme la génération de ces ions par le plasma, le système de détection, l'impact sur le rapport signal sur bruit, la justesse et la limite de sensibilité de la quantification. Pour les ions précurseurs H_3O^+ , NO^+ , O^+ , OH^- avec des m/z inférieurs à 30, lors de la multiplication du signal par l'ICF, une surexpression de la quantité d'ions précurseurs pourrait être observée et impacter le calcul de concentrations. L'ICF est déterminé par les calibrations quotidiennes effectuées à l'aide d'une bouteille de gaz étalon (7 molécules) fournie par le constructeur. A noter que les 7 molécules ne sont quantifiées que par les ions H_3O^+ , $O_2^{+\bullet}$, NO^+ , O^+ et OH^- qui sont calibrés. En revanche, les ions $O_2^{+\bullet}$, NO_2^- et NO_3^- ne sont donc pas calibrés.

L'ICF étant le facteur permettant de transposer les résultats d'un instrument à un autre, une étude des variations d'ICF sur plusieurs instruments serait judicieuse pour comprendre l'impact de ces variations sur la quantification en surveillant les paramètres instrumentaux, de génération et de détection.

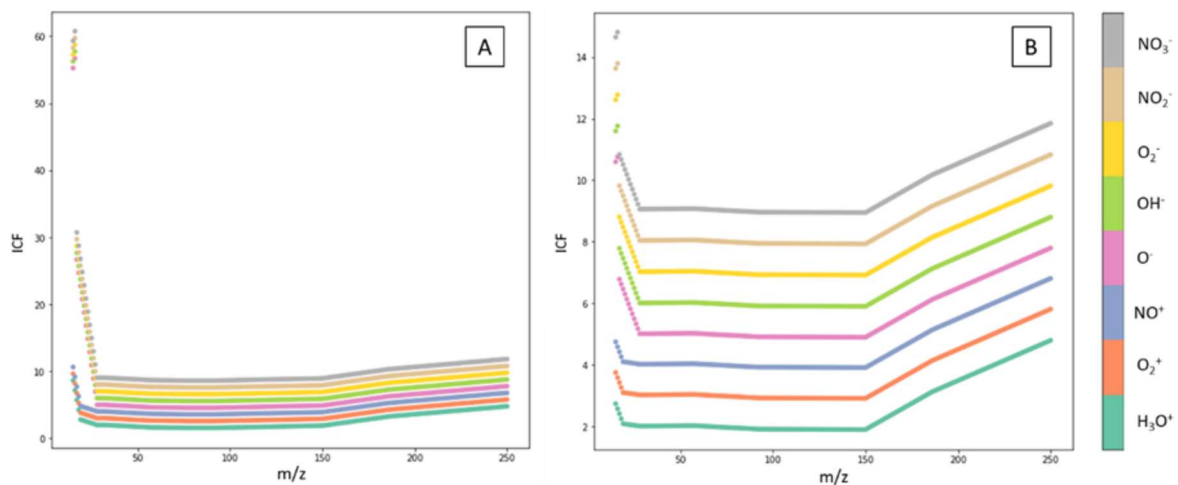


Figure 42 : courbes de la valeur de l'ICF pour les 8 ions précurseurs en fonction du rapport masse sur charge des ions détectés

Pour effectuer l'analyse statistique et mettre en évidence les différences entre les échantillons, les données sont ensuite nettoyées. Le nettoyage des données consiste à supprimer les variables quantitatives, qui sont constantes parmi tous les échantillons dont les variances sont égales à zéro. La soustraction du blanc crée des valeurs négatives dans le jeu de données, correspondant à des signaux plus élevés pour le blanc que pour l'échantillon. Ces valeurs négatives ne sont pas supprimées et sont donc considérées. Elles correspondent pour la plupart aux ions précurseurs puisque dans l'échantillon les ions précurseurs sont consommés pour réagir avec les analytes alors que dans les blancs, ils ne le sont pas. Elles peuvent également correspondre à des groupes d'ions précurseurs de l'eau. Ces valeurs négatives donnent également des informations sur l'échantillon.

V. Traitement statistique

A partir de ces données extraites des analyses SIFT-MS et HS-SPME-GC-MS/FID et prétraitées, nous recherchons des profils communs corrélés à des descripteurs (par exemple : le pays et la région de production, le procédé de transformation...). Les jeux de données obtenus présentent des variations d'intensité (forte versus faible) et des variances inhomogènes. La nature-même des données nécessite donc une transformation des données afin d'attribuer le même poids à toutes les variables. Dans cette partie, les traitements statistiques utilisés pour comparer les différentes empreintes volatiles obtenues sont présentés.

D'un point de vue statistique, il est communément admis que le nombre d'échantillons doit être significativement plus grand que le nombre de variables, mais ce n'est pas le cas avec les nouvelles techniques analytiques générant des données volumineuses comme pour les méthodes omiques (génomique, protéomique, métabolomique ou, dans notre cas, volatolomique). Selon Cunningham et al¹⁴, avec un problème appelé "big p small n", l'analyse nécessite une réduction de dimension pour améliorer l'efficacité et la précision de l'analyse des données. Pour cela, nous avons décidé d'utiliser le langage R¹⁵ avec le package " MixOmics "¹⁶ pour représenter la dispersion et la discrimination des échantillons.

MixOmics est un package R convivial dédié à l'exploration, l'exploitation, l'intégration et la visualisation de grands ensembles de données. Il offre de nombreuses fonctionnalités telles que (i) des visualisations perspicaces avec réduction de la dimension, (ii) l'identification de signatures moléculaires et (iii) une utilisation améliorée avec des profils communs à toutes les méthodes de visualisation et d'évaluation des performances. Nous avons utilisé une méthode non supervisée (Analyse en Composantes Principales clairsemée, Sparse ACP) pour étudier la variabilité des échantillons et une méthode supervisée pour mettre en évidence des variables discriminantes (Analyse Discriminante Partielle des Moindres Carrés clairsemée, Sparse PLS-DA). Ces deux méthodes permettent de transformer les données (centrer et réduire) afin de lisser les variances des valeurs présentes dans le jeu de données.

A. Analyse non supervisée exploratoire : la sparse ACP

L'analyse en composante principale (ACP) est une technique populaire de traitement des données et de réduction des dimensions, avec de nombreuses applications. La sparse ACP est une variante de l'ACP spécialisée dans l'apprentissage automatique qui est utilisée dans l'analyse statistique, en particulier lors de l'analyse de données multivariées. C'est une méthode non supervisée utilisée pour mettre en évidence la dispersion de l'échantillon. Cette méthode est un outil efficace dans le cas d'un ensemble de données constitué de dimensions multiples et d'un faible nombre d'échantillons. Elle consiste à réduire la dimensionnalité des données en introduisant des structures éparses dans les variables d'entrée. En effet, Johnstone & Lu ont montré que si p/n ne converge pas vers zéro, l'ACP classique n'est pas cohérente, mais la sparse ACP reste cohérente même si $p \gg n$ ¹⁷.

La sparse ACP est consacrée à la recherche de combinaisons linéaires, appelées composantes, qui ne contiennent que quelques variables d'entrée significatives pour expliquer la variabilité entre les échantillons. Cette méthode effectue une réduction de dimension en projetant les données dans un sous-espace plus petit tout en sélectionnant et en mettant en évidence les sources de variation des données les plus importantes du système étudié¹⁶.

Cette méthode s'applique pour réduire la dimension d'une matrice X dont une information Y est ajoutée en tant qu'information illustrative et descriptive pour discriminer des échantillons.

Réduire la dimension de l'ensemble de données implique de choisir un petit nombre de composantes qui représentent autant que possible la variabilité des données (appelée inertie). Le choix du nombre de composantes se fait donc logiquement en étudiant le pourcentage d'inertie expliqué par chacune. Dans un graphique où deux composantes sont utilisées, l'inertie de chacune est agrégée afin d'identifier le pouvoir informatif du graphique ; plus ce pourcentage cumulé est proche de 100%, meilleure est l'interprétation du phénomène. Le choix du nombre de composantes, arbitrairement limité, est dépendant du jeu de données.

B. Analyse supervisée exploratoire et confirmatoire : la sparse PLS-DA

Pour rappel, l'objectif des méthodes supervisées est principalement de définir des règles permettant de classer des individus à partir d'un ensemble de données étiquetées, l'étiquette provenant d'une variable descriptive cible. L'apprentissage supervisé est constitué de variables d'entrée (X) et d'une variable de sortie (Y). Dans cette approche, l'algorithme fait des prédictions itératives sur les données d'apprentissage pour discriminer les groupes étiquetés.

L'analyse discriminante par moindres carrés partiels éparses (sparse PLS-DA) est une adaptation des méthodes de régression PLS au problème du clustering supervisé et permet d'identifier et de quantifier la pertinence de la discrimination¹⁸. La variante clairsemée (sparse PLS-DA) permet de sélectionner les caractéristiques les plus prédictives ou discriminantes dans les données pour classer des échantillons. La première étape de la méthode PLS-DA sparse est l'application d'un modèle de régression PLS sparse sur des variables qui sont des indicateurs des groupes. La méthode PLS est utilisée pour trouver les relations fondamentales entre deux matrices (X et Y), c'est-à-dire une approche de variable latente pour modéliser les structures de covariance dans ces deux espaces. La deuxième étape du PLS-DA sparse consiste à classer les observations à partir des résultats de la régression PLS sparse sur des variables indicatrices¹⁹. L'analyse PLS-DA sparse vise à identifier un petit sous-ensemble de variables qui discriminent le mieux les classes.

C. Représentation graphique

La représentation des résultats obtenus par les deux méthodes (sparse ACP et sparse PLS-DA) sont principalement des graphiques comme dans toutes les méthodes descriptives multivariées. On distingue deux groupes de graphiques, ceux mettant en évidence la variabilité individuelle et ceux représentant les liens entre les variables.

Le "graphique individuel" fait partie de ceux expliquant les variabilités individuelles : il représente les échantillons sous forme de points placés en fonction de leur projection dans le plus petit sous-espace couvert par les composantes. Cette représentation permet de visualiser les similitudes et les différences entre les échantillons en se basant sur la distance entre les points.

Le "diagramme circulaire de corrélation" montre les relations entre les variables. Dans ce graphique, les coordonnées des variables sont obtenues en calculant la corrélation entre chaque variable originale et les composantes. Cependant, dans notre cas, le grand volume de données rend cette représentation

illisible si l'on considère toutes les variables. Il est possible d'indiquer un nombre limité de variables à prendre en considération pour constituer le graphique.

La contribution de chaque variable pour chaque composante est représentée dans un diagramme en barres appelé "LoadingsPlot" où chaque longueur de barre, qui peut être positive ou négative, correspond à l'importance de la variable dans la construction de la composante donnée. Dans le cas de l'analyse supervisée Sparse PLS-DA, les barres du diagramme en bâtons sont de la même couleur que celle du groupe d'échantillons dans lequel l'élément est le plus "abondant"²⁰.

D. Tests statistiques confirmatoires

D'autres outils statistiques, ont été utilisés pour interpréter les résultats obtenus. Après avoir mis en évidence des variables discriminantes, des tests statistiques confirmatoires étudiant l'homogénéité des variances (test de Levene²¹) et la signification de la comparaison de moyennes (test de Wilcoxon²²) ont été appliqués.

Le test de Levene étudie l'homogénéité des variances en évaluant l'égalité des variances pour une variable déterminée pour deux ou plusieurs groupes d'échantillons. Le test examine l'hypothèse nulle selon laquelle les variances de la population sont dites homogènes et comparera la variance de k échantillons, où k peut être des échantillons multiples.

Le test de Wilcoxon est un test non-paramétrique de comparaison de moyennes de deux échantillons indépendants ou appariés. Il teste l'hypothèse selon laquelle les distributions de chacun de deux groupes de données sont proches.

VI. Conclusion

Ce développement analytique nous a permis de développer l'acquisition et le traitement des empreintes volatiles. En effet, les conditions d'échantillonnage spécifiques aux deux matrices étudiées, le thé et le fromage, ont été déterminées afin d'obtenir les profils volatils les plus exhaustifs. L'analyse des profils a été réalisée en mode d'acquisition scan pour les deux méthodes (SIFT-MS et GC-MS/FID). Ces analyses génèrent de grand volume de donnée ainsi des systèmes de traitement des données ont été développés. Pour les données SIFT-MS, un algorithme a été développé, via le langage de programmation python, pour faciliter l'extraction et le traitement des données afin d'obtenir un tableau facile d'utilisation. Pour les données GC-MS/FID, de nombreux logiciels sont disponibles ainsi nous avons choisi d'utiliser le logiciel ChromCompare +. Enfin, des outils statistiques spécialisés pour l'analyse de grands volumes de données ont été choisis. Notre choix s'est porté sur l'utilisation du package Mioxmics du langage R, des méthodes multivariées seront utilisées.

Les chapitres 3 et 4 suivants présentent les différentes applications de cette nouvelle approche, les empreintes volatiles, pour l'analyse de produits agroalimentaires ainsi que les avantages et les inconvénients des profils volatils par rapport aux analyses ciblées mais également entre chaque méthode analytique.

REFERENCES BIBLIOGRAPHIQUES

1. Ghislain, M. *et al.* High frequency air monitoring by selected ion flow tube-mass spectrometry (SIFT-MS): Influence of the matrix for simultaneous analysis of VOCs, CO₂, ozone and water. *Microchem. J.* **153**, 104435 (2020).
2. Spanel, P. & Smith, D. Influence of water vapour on selected ion flow tube mass spectrometric analyses of trace gases in humid air and breath. *Rapid Commun. Mass Spectrom.* **14**, 1898–1906 (2000).
3. Fontanals, N., Marcé, R. M. & Borrull, F. New materials in sorptive extraction techniques for polar compounds. *J. Chromatogr. A* **1152**, 14–31 (2007).
4. Merkle, S., Kleeberg, K. K. & Fritsche, J. Recent Developments and Applications of Solid Phase Microextraction (SPME) in Food and Environmental Analysis—A Review. *Chromatography* **2**, 293–381 (2015).
5. Balasubramanian, S. & Panigrahi, S. Solid-Phase Microextraction (SPME) Techniques for Quality Characterization of Food Products: A Review. *Food Bioprocess Technol.* **4**, 1–26 (2011).
6. Feng, Z. *et al.* Tea aroma formation from six model manufacturing processes. *Food Chem.* **285**, 347–354 (2019).
7. Lin, S.-Y., Lo, L.-C., Chen, I.-Z. & Chen, P.-A. Effect of shaking process on correlations between catechins and volatiles in oolong tea. *J. Food Drug Anal.* **24**, 500–507 (2016).
8. Qi, D., Ma, C., Wang, W., Zhang, L. & Li, J. Gas Chromatography-Mass Spectrometry Analysis as a Tool to Reveal Differences Between the Volatile Compound Profile of Royal Jelly Produced from Tea and Pagoda Trees. *Food Anal. Methods* **14**, 616–630 (2021).
9. Souza Silva, É. A. *et al.* Development of a HS-SPME-GC/MS protocol assisted by chemometric tools to study herbivore-induced volatiles in *Myrcia splendens*. *Talanta* **175**, 9–20 (2017).
10. Larroque, V., Desauziers, V. & Mocho, P. Comparison of two solid-phase microextraction methods for the quantitative analysis of VOCs in indoor air. *Anal. Bioanal. Chem.* **386**, 1457–1464 (2006).
11. Nicolle, J., Desauziers, V. & Mocho, P. Solid phase microextraction sampling for a rapid and simple on-site evaluation of volatile organic compounds emitted from building materials. *J. Chromatogr. A* **1208**, 10–15 (2008).
12. Bourdin, D., Mocho, P., Desauziers, V. & Plaisance, H. Formaldehyde emission behavior of building materials: On-site measurements and modeling approach to predict indoor air pollution. *J. Hazard. Mater.* **280**, 164–173 (2014).
13. Spadafora, N. D., Mascrez, S., McGregor, L. & Purcaro, G. Exploring multiple-cumulative trapping solid-phase microextraction coupled to gas chromatography–mass spectrometry for quality and authenticity assessment of olive oil. *Food Chem.* **383**, 132438 (2022).
14. Cunningham, P. Dimension Reduction. in *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval* (eds. Cord, M. & Cunningham, P.) 91–112 (Springer, 2008). doi:10.1007/978-3-540-75171-7_4.
15. The R Core Team. *R - A Language and Environment for Statistical Computing.* (2020).
16. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLOS Comput. Biol.* **13**, e1005752 (2017).
17. Johnstone, I. M. & Lu, A. Y. On Consistency and Sparsity for Principal Components Analysis in High Dimensions. *J. Am. Stat. Assoc.* **104**, 682–693 (2009).
18. Cao, L. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *PLoS Comput. Biol.* **7**, e1001302 (2011).

19. Chevallier, S., Bertrand, D., Kohler, A. & Courcoux, P. Application of PLS-DA in multivariate image analysis. *J. Chemom.* **20**, 221–229 (2006).
20. Trendafilov, N. T. & Adachi, K. Sparse Versus Simple Structure Loadings. *Psychometrika* **80**, 776–790 (2015).
21. Yitnosumarto, S. & O’Neill, M. E. On Levene’s Tests of Variance Homogeneity. *Aust. J. Stat.* **28**, 230–241 (1986).
22. Noether, G. E. Efficiency of the Wilcoxon Two-Sample Statistic for Randomized Blocks. *J. Am. Stat. Assoc.* **58**, 894–898 (1963).

CHAPITRE 3 : Empreinte volatile des produits alimentaires utilisant des données SIFT-MS non ciblées couplées à des méthodes mixOmics pour la discrimination des profils : application au fromage

Les résultats présentés dans ce chapitre font l'objet d'une publication publiée en 2022 dans le journal Food Chemistry.

Pour améliorer le bien-être sanitaire des chèvres et brebis lactantes, des plantes à métabolites secondaires bioactifs peuvent être ajoutées dans la ration alimentaire. Cependant il est nécessaire de s'assurer que ce changement bénéfique pour l'animal, ne provoque pas de modifications organoleptiques du produit final, le fromage. Dans le cadre du projet Bionachol, les partenaires (la chambre d'agriculture 64, l'INRA de Lusignan et les plateformes technologiques Errecart et Actalia) ont étudié l'effet du changement de régime alimentaire sur la santé des chèvres et des brebis en lactation tout en contrôlant les qualités organoleptiques des fromages produits. Ainsi, dans leur plan d'expérience sur le territoire pyrénéen, deux éleveurs de brebis (E1 et E2) ont été sélectionnés, l'un nourrissant principalement ses animaux en bergerie et l'autre exclusivement en pâture. Le protocole alimentaire appliqué consistait à nourrir les animaux avec une ration alimentaire habituelle dite « contrôle » pendant 2 semaines, suivi d'une ration alimentaire enrichie en plantes à métabolites secondaires dite « modifiée » pendant les 2 semaines suivantes et un retour à une ration alimentaire « contrôle » pendant les 2 dernières semaines. Durant chaque période, le dernier lait collecté est transformé en fromage puis affiné pendant 6 à 9 mois. Ce protocole a été répété pendant 2 ans. Pour évaluer l'impact du changement alimentaire des brebis sur les fromages, des analyses sensorielles ont été réalisées par les partenaires. Pour renforcer l'étude des qualités organoleptiques des produits, l'approche de caractérisation par les empreintes volatiles des produits a été envisagée. L'objectif était de mettre en évidence la variabilité des échantillons de fromage et de les discriminer en fonction des producteurs et du régime alimentaire des brebis.

Le potentiel de la technologie SIFT-MS en mode de détection full scan positif et négatif associé à l'utilisation d'outils chimiométriques adaptés aux grands ensembles de données a donc été évalué pour acquérir les empreintes volatiles des fromages et exploiter les résultats.

Douze échantillons de fromages ont été analysés, en triplicats, par HS-SIFT-MS selon les conditions d'échantillonnage et d'analyse développées précédemment (chapitre 2). Les données obtenues ont été prétraitées par l'algorithme développé. L'exploitation statistique et la visualisation des données ont été réalisées en utilisant les méthodes multivariées non supervisées et supervisées du package MixOmics du langage R.

L'analyse non supervisée (Sparse ACP) a d'abord été appliquée sur l'ensemble des données pour évaluer la variabilité entre les différents échantillons. Une différence entre les deux producteurs a été observée avec une bonne homogénéité des fromages du producteur E1 et une plus grande hétérogénéité des fromages du producteur E2, pouvant être liée aux conditions d'élevage (bergerie pour E1 et pâture pour E2). Pour aller plus loin dans la mise en évidence des différences entre les producteurs, une analyse supervisée (Sparse PLS-DA) a été réalisée, montrant une séparation claire des deux groupes d'échantillons. Une identification des molécules responsables de la discrimination permettrait de remonter aux voies enzymatiques impliquées. C'est pourquoi deux stratégies de post-traitement des données ont été étudiées : (i) en recherchant des composés spécifiques, déjà identifiés dans la littérature, parmi les ions SIFT-MS les plus pertinents et (ii) en recherchant les composés potentiels à partir des ions les plus pertinents pour la discrimination. La première stratégie n'a pas permis d'identifier les molécules responsables et suggère que ce sont d'autres composés qui sont responsables de la discrimination. La deuxième stratégie a permis d'identifier l'hexanoate de propyle, un composé odorant, plus exprimé dans les fromages du producteur E2 que dans ceux du producteur E1. Des mesures SPME-GC-MS/FID réalisées en parallèle sur les mêmes échantillons (données non présentées) ont confirmé que l'hexanoate de propyle était bien présent et surexprimé dans les fromages du producteur E2.

La grande hétérogénéité des 12 fromages, caractéristique des diversités aromatiques des fromages Ossau-Iraty, ne permet cependant pas de voir l'effet de l'utilisation des plantes à métabolites bioactifs

secondaires. Ainsi, une sparce ACP a été appliquée sur le jeu de données contenant uniquement les échantillons du producteur E1. Des différences de profils sont observées selon le régime alimentaire mais également selon l'année de production. La même stratégie d'exploitation post-traitement des données a été utilisée pour rechercher les différences entre les régimes modifiés et les régimes témoins dans les fromages E1. Les ions produits discriminants sont principalement des ions négatifs et possèdent une valeur de m/z élevée, majoritairement supérieure à 100 et principalement 137. Cela suggère que la modification de la ration alimentaire implique des changements pour des molécules pouvant correspondre à des terpènes de masse moléculaire de 136 g.mol^{-1} . Cependant, le niveau élevé d'interférences dans les ions produits des terpènes et l'absence de base de données d'ionisation négative ne permet pas d'obtenir une identification.

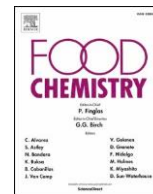
L'objectif principal de ce travail était de montrer le potentiel de l'analyse SIFT-MS non ciblée en mode full scan couplée à des outils statistiques adaptés dédiés aux jeux de données "omiques", pour étudier les composés volatils libérés par le fromage. Cette approche a permis de discriminer la typicité des fromages des deux producteurs et d'étudier l'impact du changement alimentaire des brebis sur le produit fini. Ces résultats ont confirmé notre hypothèse concernant les empreintes volatiles : l'analyse non ciblée améliore la capacité à détecter des composés inattendus. La complémentarité de cette approche avec des techniques d'analyse classiques peut permettre l'identification de molécules clés. Cette étude montre le potentiel des sciences des big data qui permettront aux chercheurs et aux producteurs de trouver des corrélations entre les empreintes volatiles et la typicité des produits ou l'impact d'un changement de procédé.

Pour aller plus loin, la construction d'une base de données d'empreintes SIFT-MS de l'AOP Ossau-Iraty est en cours, ce qui pourrait aider les producteurs à résoudre les questions de typicité, de diversité et de pratiques agricoles. La faible quantité d'échantillons de cette étude n'a pas permis de corrélérer les résultats des empreintes avec ceux de l'analyse sensorielle. En revanche, les échantillons permettant de construire la cartographie de l'AOP seront également dégustés par un jury expert. Les corrélations entre les qualités sensorielles des fromages et leurs empreintes volatiles seront alors étudiées.



Contents lists available at ScienceDirect

Food Chemistry

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

Volatile fingerprint of food products with untargeted SIFT-MS data coupled with mixOmics methods for profile discrimination: Application case on cheese

Marine Reyrolle ^a, Mylène Ghislain ^a, Noëlle Bru ^c, Germain Vallverdu ^a, Thierry Pigot ^a, Valérie Desauziers ^a, Mickael Le Bechec ^{a,*}

^a Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, IMT Mines Ales, IPREM, Pau, France, Institut des sciences analytiques et de Physicochimie pour l'environnement et les Matériaux, UMR5254, Hélioparc, 2 avenue Président Angot, 64053 PAU cedex 9, France

^c Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, LMAP, Anglet, France, Laboratory of Mathematics and its Applications of Pau – MIRA, UMR5142, All'ée du Parc Montaury, Anglet, France

ARTICLE INFO

Keyword:
SIFT-MS
MixOmics
Volatile fingerprints
Typicality of agro-food products ewe cheeses

ABSTRACT

Volatile organic compounds (VOCs) emitted by food products are decisive for the perception of aroma and taste. The analysis of gaseous matrices is traditionally done by detection and quantification of few dozens of characteristic markers. Emerging direct injection mass spectrometry technologies offer rapid analysis based on a soft ionisation of VOCs without previous separation. The recent increase of selectivity offered by the use of several precursor ions coupled with untargeted analysis increases the potential power of these instruments. However, the analysis of complex gaseous matrix results in a large number of ion conflicts, making the quantification of markers difficult, and in a large volume of data. In this work, we present the exploitation of untargeted SIFT-MS volatile fingerprints of ewe PDO cheeses in a real farm model, using mixOmics methods allowing us to illustrate the typicality, the manufacturing processes reproducibility and the impact of the animals' diet on the final product.

1. Introduction

Volatile organic compounds (VOCs) are low molecular weight organic molecules able of evaporating or subliming at room temperature thanks to their high vapour pressure and low boiling point. Depending on the concentration and the nature of the exposure, VOCs may have positive effect (pleasant smell) or negative impact (pollutants). In food, the VOCs composition contributes to the development of odour and flavour, which mainly affects the food acceptability by the consumer (Lytou et al., 2019). Thus, a wide variety of volatile chemical compounds with different concentration ranges are emitted by biological processes in food products (enzymatic or metabolic pathways). Recent trends in VOCs analysis in the food industry, supported by improved analytical techniques, have focused on three main areas:

- (i) Aroma profiling: allowing a better understanding of factors that give rise to the aroma profiles (Sousa et al., 2020),

- (ii) Food safety: tracking the origin of food deterioration or contaminations (Castro-Puyana & Herrero, 2013),
- (iii) Food consistency and quality: ensuring production-line consistency of products from batch to batch (Cecchi et al., 2018).

In addition, depending on the product, the aroma profile can be correlated to several parameters: geographic origin, seasonal variations or storage conditions (Santos & Oliveira, 2017). For example, with cheeses, various factors can impact aromatic profile of cheeses (Boltar et al., 2019), such as the animal's breed (Ferreira et al., 2009) and their diet (Ianni et al., 2020), the ferments or enzymes used during processing (Feutry et al., 2012), the pasteurisation of milk (Jiang et al., 2019), the ripening conditions and the seasonal variation (Boltar et al., 2015). Contamination during manufacturing with bacteria or fungi may also alter the flavour of cheeses.

To assess the authenticity of food products, a wide range of analytical approaches is currently used such as mass spectrometry techniques

* Corresponding author at: Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, IMT Mines Ales, IPREM, Pau, France, Institut des sciences analytiques et de Physicochimie pour l'environnement et les Matériaux, UMR5254, Hélioparc, 2 avenue Président Angot, 64053, PAU cedex 9, France. E-mail address: mickael.lebechec@univ-pau.fr (M. Le Bechec). <https://doi.org/10.1016/j.foodchem.2021.130801> Received 5 May 2021; Received in revised form 26 July 2021; Accepted 4 August 2021 Available online 8 August 2021 0308-8146/© 2021 Elsevier Ltd. All rights reserved.

coupled with liquid and gas chromatography (Ch et al., 2021), spectroscopic techniques (Infra-red, NMR) or sensory analysis (Luykx & van Ruth, 2008). The differentiation of products is mostly achieved with identified and quantified variables (markers/compounds) in a three-step methodology: 1) sample preparation with solvent extraction or headspace techniques (the latter often associated with solid-phase microextraction) for VOC pre-concentration; 2) separation of markers; 3) identification and quantification of markers.

However, there is a growing interest in the notion of volatile fingerprints, where the goal is not to identify every molecule of the matrix, but to compare profiles of products (Balkir et al., 2021). This leads to the definition of the term “volatilome” to describe all the volatile compounds emitted by an organism, an ecosystem or a product (such as food) and of the term volatilomics (Cumeras, 2017) as a subdomain of metabolomics.

Recent years have also seen the development of some direct injection mass spectrometer instruments (DIMS) (Biasioli et al., 2011; Deucher et al., 2019) such as proton-transfer reaction mass spectrometry (PTR-MS) or selected ion flow tube mass spectrometry (SIFT-MS) (Smith & Spanel, 2005). Such technologies use a soft chemical ionisation of the VOCs, leading to product ions detected and quantified without any previous separation step. The selectivity of these instruments is based on the chemical reactivity between a precursor ion and an analyte and no longer requires separation of the latter. These two technologies (PTR-MS and SIFT-MS) can be used for targeted detection and quantification of markers (Castada et al., 2014) with single ion monitoring mode (SIM) or non-targeted analysis with a SCAN detection mode, which gives the signal of all the product ions according to their mass to charge (m/z) ratio. Depending on the number of precursor ions used and the number of VOCs in the sample, these technologies generate a large amount of data that can be sometimes difficult to interpret.

A recent SIFT-MS instrument Voice 200Ultra (SYFT®) with a dual polarity plasma source offers up to 8 different precursor ions (H_3O^+ , O_2^{+} , NO^+ , O^{+} , OH^- , O_2^{+} , NO_2^- et NO_3^-), thus increasing the potential selectivity of the technique and increasing the data volume in the SCAN mode (up to 3 080 values of product ions with m/z ratios from 15 to 400). The dataset obtained by SIFT-MS analysis in full scan mode constitutes a real volatile fingerprint of the product. Samples can then be classified according to their volatile profile reflecting differences between groups. However, the substantial volume of data requires dedicated chemometric methods already applied in many domains of omics (Bajoub et al., 2018) and food science (Farag et al., 2020). Chemometric analysis is the application of mathematical and statistical tools to analyse and extract the maximum information from chemical data (Bergamaschi et al., 2020).

This study therefore proposes an original approach by coupling a DIMS analysis (SIFT-MS) and an exploitation of the results by adapted statistical methods. The volatile profiles emitted by ewe cheeses have been measured with SIFT-MS in positive and negative full scan detection modes and the resulted profiles have been analysed with chemometric tools to highlight the sample variability and to discriminate samples based on producers. Precisely, the multivariate mixOmics methods (Rohart et al., 2017) were used on the scan dataset as tools for representing the dispersion and the discrimination of samples. MixOmics is a user-friendly R package dedicated to the exploration, mining, integration and visualisation of large data sets. It provides attractive functionalities such as (i) insightful visualisations with dimension reduction, (ii) identification of molecular signatures and (iii) improved usage with common calls to all visualisation and performance assessment methods.

2. Material and method

2.1. Samples

Ossau-Iraty PDO cheese, a famous uncooked pressed cheese from the French Pyrenees, was provided by the partners of agricultural research

Project BioNACHOL (financially supported by the Conseil Régional Nouvelle Aquitaine, France). The specifications of the Ossau-Iraty PDO include six main rules: 1) the milk has to be produced in Bearn and French Basque Country (southwest of France on the Pyrenees mountain range), 2) by the only 3 local ewe breeds (Manech ginger head, Manech black head and Basco-Bearnaise), 3) fed in stables or in fields with local fodder and small amount of cereals but without GMO. 4) The ewes are milked only part of the year (from December to August), 5) the manufacturing has to respect the traditional methodology and, finally, 6) the ripening step has to be longer than 80 days. These cheeses take the form of a generously shaped tomme with a natural rind, the colour of which varies from orange-yellow to grey depending on the maturing conditions. The aromas are present and varied and the texture is supple to firm without being either sticky or dry. The taste of Ossau-Iraty is delicately ewe-like, with a slight hazelnut flavour.

The objective of this project was to study the impact of plants with secondary bioactive metabolites on lactating ewe's health (limitation of parasite development) and on the organoleptic qualities of cheeses. Thus, two farm models with two ewe breeders (E1 and E2) have been selected: E1 mostly fed his animals in the farm's stables and E2 mostly fed his animal in fields. A feeding protocol was applied: two weeks with control diet, two weeks with modified diet (supplied with bioactive secondary metabolites plants; sainfoin or chicory) and finally two weeks again with control diet. The composition of the diets is detailed in Supplementary Tables 1 & 2. Ossau-Iraty cheeses were then produced with the last collected milk of each period and ripened for 6 to 9 months. This protocol was repeated over two years (Table 1). The samples were identified with a code E__ with the first number corresponding to the producer and the second number to the sample (6 samples per producer).

2.2. Sample preparation

Cheese (15 g) was cut into 1-cm cubes and then introduced into a 1-L bottle. The bottle was fitted with a polypropylene screw cap with 2 tight connection ports fitted with 0.6 cm PFA tube. The first one was connected to SIFT-MS and the second one to a 1-L Tedlar® (Supelco, Bellefonte, PA) bag, filled with zero dry air (ZeroAir Alliance ZA1500, F-DGSI, Evry France) to compensate for the volume taken by the SIFT-MS analysis. The closed bottle was incubated for 2 h at $22 \pm 2^\circ C$ before performing the positive and negative SIFT-MS full scan analysis. This optimised preparation step allows the emission of volatile compounds from the solid sample to the gas phase. Thus SIFT-MS measurements were performed on the gas phase.

2.3. SIFT-MS analysis

A SIFT-MS Voice 200 Ultra (SYFT Technologies, Christchurch, New Zealand) equipped with a dual source producing positive and negative soft ionizing precursor ions (H_3O^+ , O_2^{+} , NO^+ , O^{+} , OH^- , O_2^{+} , NO_2^- and NO_3^-) in a single scan was used (Hera et al., 2017; Ghislain et al., 2019; Smith et al., 2020).

Table 1

Sample	Batch/Producers	Diet	Year
E11	E1	Control	1
E12	E1	Modified (sainfoin)	1
E13	E1	Control	1
E14	E1	Control	2
E15	E1	Modified (sainfoin)	2
E16	E1	Control	2
E21	E2	Control	1
E22	E2	Modified (chicory)	1
E23	E2	Control	1
E24	E2	Control	2
E25	E2	Modified (chicory)	2
E26	E2	Control	2

Samples description.

This instrument uses nitrogen as carrier gas, the sample was introduced through a temperature (110 °C) and flow controlled (20 mL min⁻¹) sample line (High Performance Inlet HPI®). The instrument was daily calibrated with a standard gas (Scott™ gas mixtures, Air Liquid USA; composition in [Supplementary Table 3](#)) containing standards at 2.0 ppmV in nitrogen (Air Liquide, Alphagaz 2). A blank experiment with an empty bottle was performed before each triplicate. The full scan raw data files containing product ion intensities with 15–250 *m/z* range were collected for a compilation and pre-processing step.

2.4. Compilation and pre-processing data

The compilation of the full scan raw data was carried out using Python programming language in which the product ions were coded into numerical variables of the XAB type with X a constant letter, A ranging from 1 to 8 indicating the precursor ion used and B the value of *m/z* ranging from 15 to 250. The plasma source of the used instrument produces a very small amount of precursor ion NO₃⁻ making the ion-analyte reaction inefficient with this ion. As a consequence, we limited the exploitation of the full scan data to seven other precursor ions, that is 1,652 variables.

The data pre-processing consisted first in subtracting the background noise (signal of empty bottle) for every replicate, then in averaging the triplicates for each sample to obtain a single value of signal intensity for each ion. The final dataset was organised in matrix denoted X with *n* rows (*n* = 12) corresponding to the analysed cheese samples and *p* columns (*p* = 1,652) corresponding to the intensities (mean values of the triplicate signals for each sample) of the product ions. Qualitative variables associated with the experimental design (batch, diet, years...) were also added in the matrix denoted Y.

To perform statistical analysis and to highlight differences between samples, the dataset was cleaned by suppressing the quantitative variables, which are constant among all samples (variance equal to zero). Negative values of the dataset, corresponding to higher signals with the blank than with a sample, were not suppressed. These negatives values mostly correspond to precursor ions and water clusters of precursor ions and also give information on the sample. In addition, information about the origin (producer, diet and year) are associated to each sample as qualitative variables.

Finally, after the pre-processing step and the cleaning step, the dataset was analysed with statistical tools. The matrix contains two groups of variables: quantitative variables (denoted X) corresponding to signal intensities and qualitative variables (denoted Y) corresponding to product descriptions. One can notice the dimensionality (number of variables) is greater than the number of samples. This dataset is then a wide matrix of omics data, and requires specially developed tools.

2.5. Statistical analysis

The main objective of this work was to draw the volatile fingerprint of cheese samples according to intensity levels (ion count per second) of product ions with SIFT-MS full scan analysis. From these volatile fingerprints, we looked for common profiles correlated to descriptors such as producer or animal diet. From a statistical point of view, it is widely accepted that the number of samples has to be significantly greater than the number of variables, but this is no more the case with high throughput analytical technics than for omics (genomic, proteomic or, in our case, volatilomic). According to Cunningham, with an issue called “big *p* small *n*”, the analysis requires dimension reduction to improve efficiency and accuracy of data analysis (Cunningham, 2008). Thus, the R language (R. Core Team, 2018) was used with “MixOmics” package (Rohart et al., 2017) to represent the dispersion and discrimination of the samples. So, we used a non-supervised method (sparse principal component analysis, sparse PCA) and a supervised one (sparse partial least squares-discriminant analysis, SPLS-DA).

a. Representation of a volatile fingerprint of Ossau-Iraty Cheese

The volatile fingerprint corresponds to the global response of a cheese analysis with SIFT-MS Voice 200-ultra in full scan mode after blank subtraction. Fig. 1 includes the signal (ion counts per second) of all ions, to visualise precursor ions and product ions. All the information is useful since it represents both the molecules present in the matrix and the overall concentration of analytes reacting with the precursor ions. However, the comparison of volatile fingerprints requires adapted statistical tools to identify correlations.

b. Reducing the dimensionality using an unsupervised method: sparse PCA

Sparse principal component analysis (sparse PCA) is a non-supervised method used to show the dispersion of the sample according to a large number of variables. Derived from classical PCA, this method is an efficient tool in the case of high-dimensional dataset and low sample size for the reduction of data dimensionality by introducing sparse structures to the input variables. Indeed, Johnstone & Lu showed that if *p/n* does not converge to zero, the classical PCA is not consistent, but the sparse PCA remains consistent even if *p* >> *n* (Johnstone & Lu, 2009).

Sparse PCA is devoted to finding linear combinations, called components, that contain just a few meaningful input variables to explain the variability between samples. This method performs dimension reduction by projecting the data into a smaller subspace while capturing and highlighting the largest sources of data variation, resulting in a powerful visualisation of the system under study (Rohart et al., 2017).

This method was applied to reduce the dimension of the X matrix and the information of Y was added as an illustrative information to detect potential heterogeneity of the cheese samples.

As in all the multivariate descriptive methods, the results consist mainly in graphical outputs: one highlighting the individual variability and the other the links between variables:

- The “individual plot” represents the samples as points placed according to their projection in the smallest subspace covered by the components. This representation makes it possible to visualise the

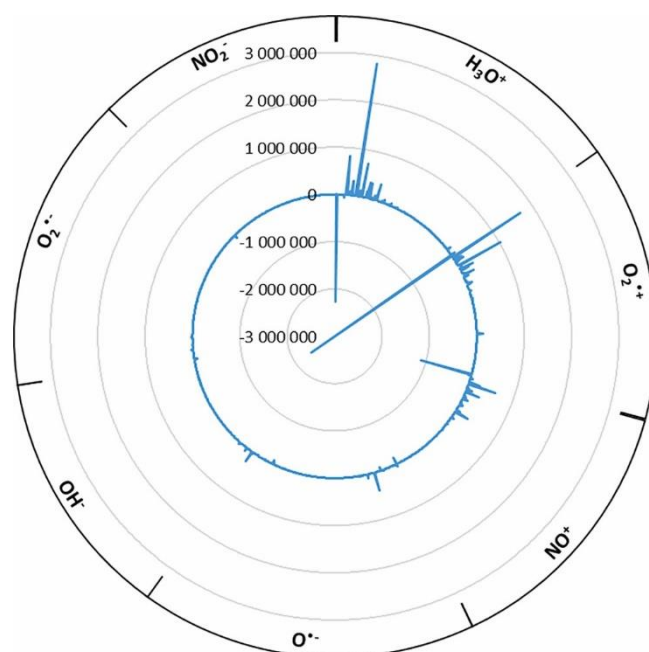


Fig. 1. Graphical representation of volatile fingerprint of the sample E11.

similarities and the dissimilarities between samples through the distance between points.

- The “correlation circle plot” shows the relationships between variables. In this graph, the coordinates of the variables are obtained by calculating the correlation between each original variable and the components. However, in our case, the large volume of data makes this representation unreadable if we consider all the variables.
- The contribution of each variable for each component is represented in a bar plot where each bar length corresponds to the importance of the variable in the construction of the given component, which can be positive or negative.

Reducing the dimension of the dataset involves choosing a small number of components that capture as much of the variability in the data (called inertia) as possible. The choice of the number of components is therefore logically made by looking at the percentage of inertia explained by each. In a graph where two components are used, the inertia of each is aggregated in order to identify the informative power of the graph; the closer this cumulated percentage is to 100% the better the interpretation of the phenomenon. The number of components presented in this paper is limited arbitrarily to 2.

c. Discrimination of sub-populations using a supervised method: sPLS-DA

The objective of supervised methods is mainly to define rules making it possible to classify individuals from a labelled data set, the label coming from a target qualitative variable. Supervised learning consists of input variables (X) and an output variable (Y). In this approach, the algorithm makes iterative predictions on the learning data to discriminate labelled groups (Guerra et al., 2011).

Sparse partial least squares discriminant analysis (sPLS-DA) is an adaptation of PLS regression methods to the problem of supervised clustering and allows identification and quantification of the discrimination relevance (Lê Cao et al., 2011). The first step of sparse PLS-DA is the application of a sparse PLS regression model on variables which are indicators of the groups. PLS is used to find the fundamental relations between two matrices (X and Y), i.e., a latent variable approach to model the covariance structures in these two spaces. The second step of sparse PLS-DA is to classify observations from the results of sparse PLS regression on indicator variables (Chevallier et al., 2006). The sparse PLS-DA analysis was applied to the datasets where X is the matrix containing n lines (n = 12) corresponding to the sample and p columns (p = 1,652) corresponding to the ion signals and Y a single qualitative variable such as producers, feeding, etc.

The sparse PLS-DA analysis aims to identify a small subset of variables that best discriminates the classes. A cross-validation procedure of 3-fold CV repeated 10 times was performed to determine the number of components to retain and the optimal number of explanatory variables for each.

This supervised method gave the same graphical outputs as the sparse PLS-DA with the “correlation circle plot” and the plot of individuals. For each component of sparse PLS-DA analysis, the importance of each variable was represented in a bar plot named the LoadingsPlot where each bar value corresponds to the coefficient affected to the corresponding variable to construct the corresponding component (as a linear combination). It can be positive or negative (Trendafilov & Adachi, 2015). For a discriminant analysis, the colour of the bars in the bar plot corresponds to the sample group in which the element is most “abundant”.

The results can also be represented in “clustered image maps” (CIM) or “heatmap” which graphically corresponds to a two-dimensional coloured image. CIM is based on hierarchical clustering operating simultaneously on the rows and columns of a matrix. Dendrograms indicate proximity between variables or samples and represent hierarchical grouping for samples (left) and for variables (top). The colour of the heat map indicates the nature of the correlation between the subsets of variables (positive, negative, strong, or weak). The advantage of this representation is that it carries a classification to

discuss the similarity between samples and groups of variables. CIM is a visualisation tool to observe (highlight) correlations between groups of subsets of different types of variables: between for example quantitative variables (ion) and qualitative variables (sample).

3. Results

After the pre-processing treatment of the full scan data, an example of the volatile fingerprint of a cheese is presented in Fig. 1. The negative values in the radar graphic illustrate the consumption of precursor ions by sample composition (high precursor ion consumption resulting in a high negative signal value) whereas the positive ones correspond to product ions. This graphical representation also gives a global information on the sample and the instrument. In this case, some general comments can be highlighted: i) it appears clearly that O_2^{+} is the main precursor ion consumed, ii) the most intense signal is obtained from the reaction of H_3O^+ precursor ion with water and iii) the ion production rate of the instrument source is lower in negative mode than in positive mode, as a consequence the intensities are much lower for detected anions (note that NO_3^- data were not presented in this work).

3.1. Sample distribution according to volatile profile

The principal component sparse analysis was first applied on the whole dataset to assess the variability between the different samples. The sparse PCA individual plot (Fig. 2) shows that principal component 1 (PC1) and principal component 2 (PC2) account for 37% of the total variability (1,652 variables). With both these dimensions, a slight differentiation between the two producers (E1 and E2) appears: E1 samples are located close together in the lower left corner whereas E2 samples are more to the right and also more distributed on the vertical axis. This means that PC1 reflects the maximum variability between producers in the data set, whereas PC2 highlights the dispersion of the cheeses of producer E2. The volatile fingerprints of control experiments (E11 and E13, E14 and E16) are very close for producer 1 and no clear trend was observed with the modified experiments (E12 and E15). For producer E2's cheeses, the control experiments (E21 and E23, E24 and E26) are too dispersed to allow a measurement of diet modification effect (E23 and E25). This first non-supervised analysis indicates that the global impact of animal diet seems to be less predominant than the variability between two producers or the variability between two production years. This observation can be linked to several causes: the lactation period of ewes (the second cheese control is made 4 weeks after the first), the grass evolution during time (and year), the reproducibility of cheese making and the ripening conditions. The homogeneity of cheeses from producer E1 and the heterogeneity of cheeses from producer E2 can furthermore be explained by the rearing conditions and by the ripening parameters. Indeed, producer E1 mostly breeds his ewes in a stable with dry hay, and ripens the cheeses in a controlled chamber, whereas producer 2 usually grazes his ewes in pastures and uses a home-made ripening chamber in his farm. These results are consistent with the diversity of cheeses from Protected Designation of Origin (PDO) Ossau-Iraty: raw milk or heat-treated milk, from ewes in stables or in fields from December to August and with different ripening times. Cheese consumers appreciate this diversity of typicality within this PDO: there is a cheese for every taste.

Sparse PCA shows the dispersion and the proximity of the cheese samples with each other. However, the use of different plants with secondary bioactive metabolites (sainfoin or chicory) by both farmers prevents a clear identification of markers. We therefore applied the principal component sparse analysis on a dataset containing only the samples of producer E1 (Fig. 2B), for which a better process control allows to study more precisely the variations in the composition of the volatile fingerprints. First, a year effect on volatile fingerprints is observed: the first year's samples are located to the left and the second year's sample to the right. Second, the samples produced with modified diet are located on the upper side of Fig. 2B. Note that this diet effect appears more significant for the first year than the second. To go further in

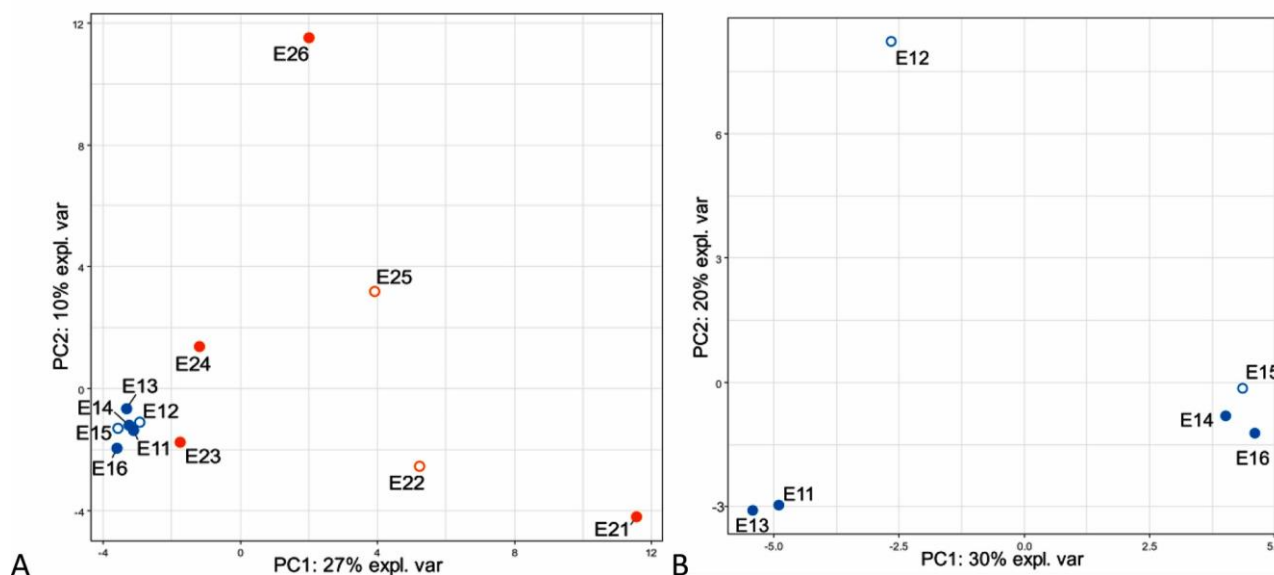


Fig. 2. (A) Individual plot of sparse PCA of volatile fingerprints of both producer's samples. (B) Principal component sparse analysis on producer E1 samples only. (Blue: producer 1, orange: producer 2; filled circles: control diet; open circles: modified diet). For a better readability, only mean values of samples ($n = 3$) are represented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

highlighting differences between samples according to labels (e.g., producer/diet) other statistical tools were applied to the dataset.

3.2. Producers discrimination according to the volatile profile

The supervised analysis with sparse PLS-DA was applied on the dataset to discriminate cheeses from the two different producers. According to cross-validation, the number of components selected for the sparse PLS-DA discriminating producers was 2 and the optimal number of variables to select was 70 for component 1 and 40 for component 2. The individual plot (Fig. 3A) of PC1 and PC2 shows a clear separation of the two sample groups. PC1 (26% of total variability) allows a clear discrimination of one producer from the other, whereas PC2 (13% of total variability) also illustrates the dispersion of producer E2 cheeses and the clustering of producer E1 cheeses. The

difference with the previous non-supervised analysis rests on the distinction of groups with the label "producer" leading to identify different variables permitting this discrimination.

The 40 most discriminant variables (ions) between the two producers were represented in the plot loading (Fig. 4A). The orange bars correspond to variables discriminating producer E2 and the blue bars correspond to variables which mostly discriminate E1. According to the product ion coding, 7 product ions in the top 40 were obtained with H_3O^+ precursor ion, 13 with NO^+ , 13 with O_2^+ , 1 with O^- , 5 with OH^- and 1 with NO_2^- . This distribution illustrates the importance of chemical selectivity provided by the SIFT-MS instrument. H_3O^+ is the most used precursor ion as a quantification marker with DIMS instruments (PTR-MS and SIFT-MS), due to its prior art, but is not the principal precursor ion allowing the discrimination between producers.

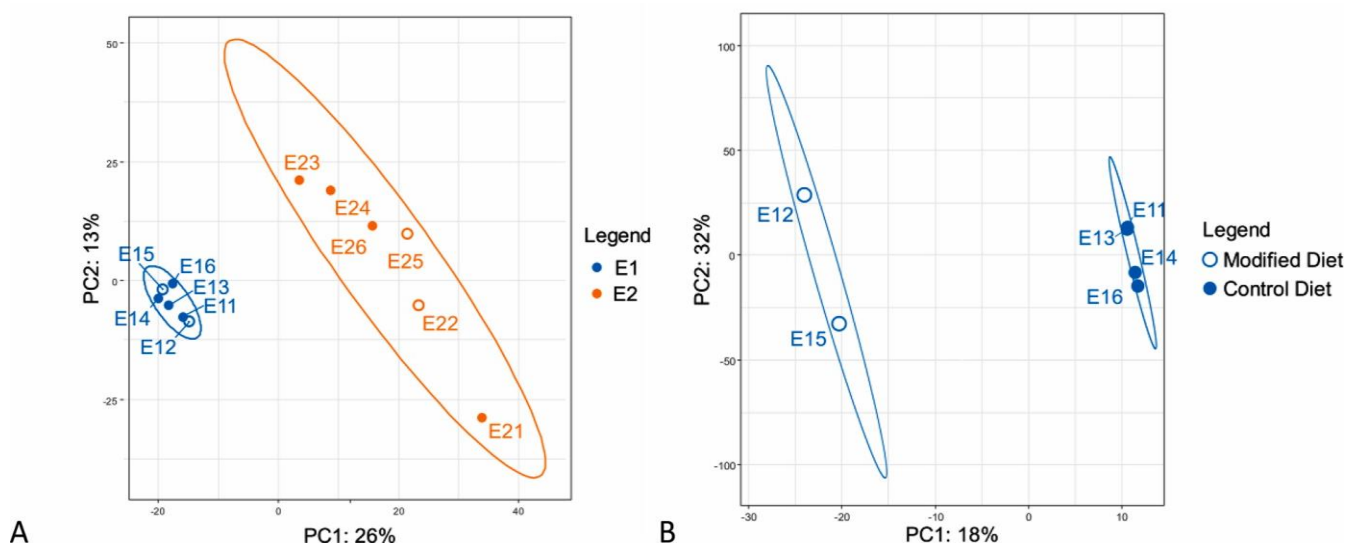


Fig. 3. Individual plot of sparse PLS-DA for the full scan mode data with the first two principal components. The different colours indicate the different batch/ producers (Blue: E1 and orange: E2) A: samples distribution of both producers according sparse PLS-DA. B: focus on sample distribution of the Producer 1 according sparse PLS-DA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

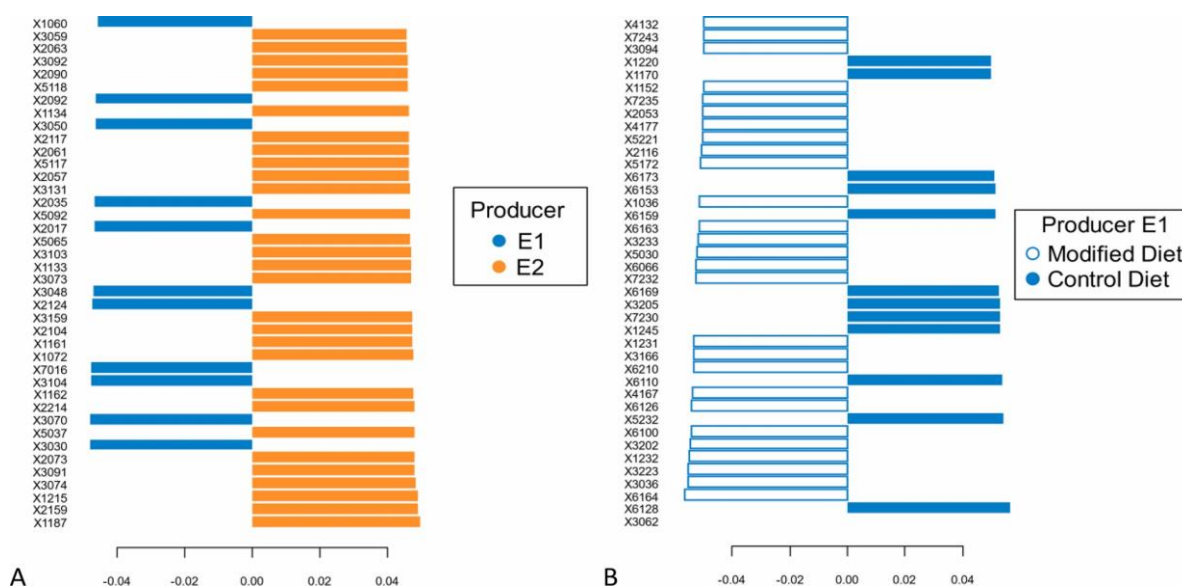


Fig. 4. A: Plot Loadings of PC1 from sparse PLS-DA method to discriminate the two producers. The colours indicate the different producers (Blue: E1 and orange: E2). **B:** Plot Loadings of PC1 from sparse PLS-DA method to discriminate the animal diet on cheeses of producer 1 samples (coloured bars: control diet, empty bars: modified diet). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As presented in the previous section, the impact of animal diet was not predominant on the whole volatile fingerprint distribution. Accordingly, a second supervised analysis was conducted only on the volatile fingerprints of the cheeses of the producer E1, in order to determine the variables that discriminate the impact of sainfoin on cheese (Fig. 3B). The impact of year and diet appears clearly with this statistical analysis, where PC1 discriminates the diet and PC2 discriminates the year of cheeses from producer E1. With the same strategy, a plot loading containing the 40 most discriminant ions along PC1 is presented in Fig. 4B. The significant product ions come also from both positive and negative precursor ions, as for the discrimination between producers (7 with H_3O^+ precursor ion, 8 with NO^+ , 2 with O_2^+ , 3 with O^+ , 4 with OH^- , 12 with O_2^{*-} and 1 with NO_2^-). However, we can observe that in this test, the ions produced with O_2^- are more represented and that the m/z ratios are also higher (7 product ions with $m/z < 100$, 18 between $100 < m/z < 200$ and 15 between $200 < m/z < 250$). This suggests that the most discriminant ions for the animal diet involve high molar mass molecules.

The CIM representation of the volatile fingerprint (Fig. 5) aims to illustrate the sample classification and, at the same time, the most relevant ions classification with the principal component 1 of the sparse PLS-DA. A clear distinction between cheeses from producers E1 and E2 with colour differences appears within the heatmap indicating that some variables are more expressed in one group than in the other. The variables in the upper right corner are stronger (red) for the E2 cheeses than the E1 ones and conversely for the variables on the lower left which are stronger in E1 than in E2. The dendrogram of samples (on the left), shows that the sample E1 cheeses are relatively close (short lines), which confirms the homogeneity of this producer samples compared to E2 ones. In addition, for producer E1, we observe a clustering by year of production: year 1 (E11-E12-E13) and year 2 (E14-E15-E16).

4. Discussions

The non-targeted analysis with SIFT-MS Voice200ultra enabled drawing the volatile fingerprint of cheese products with 1,652 variables. A direct analysis of volatile fingerprint (Fig. 1) with such

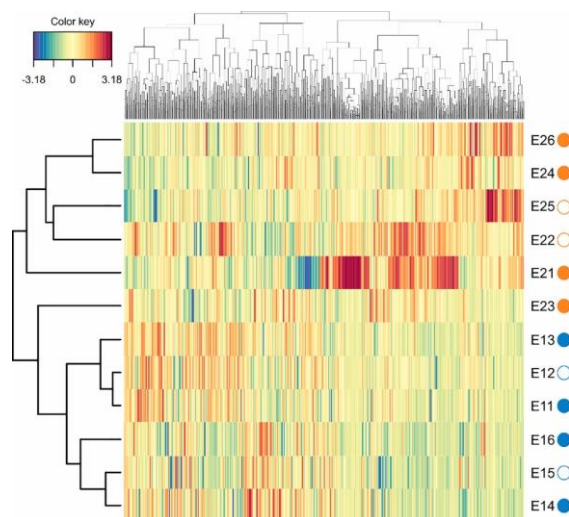


Fig. 5. Clustered image maps (CIM) on the dataset with sPLS-DA method to discriminate the producers (legend on right) according to the most discriminated variables (ions) with principal component 1. The dendrograms represent hierarchical grouping for samples (left) and for variables (top). The colour key indicates the negative (in blue scale) or positive (in red scale) intensity of the coefficients of each sample on the first component. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

volume of data is quite difficult to achieve. Thus, to compare these complex volatile fingerprints and to address similarity and distribution issues, multivariate methods of the mixOmics package in R software was applied on a dataset of twelve ewe cheeses (analysed in triplicate) from two producers. Unsupervised method sparse PCA, adapted to those datasets where variables are more important than the number of samples, made it possible to visualise the distribution and the variability of the samples in relation to each other. A different distribution of volatile fingerprints of cheeses appeared, according to the producers. The cheeses from producer 1 form a cluster, meaning a high similarity between samples, whereas cheeses from producer 2 spread out over the map, and

that is an indicator for higher heterogeneity. On one hand, according to PC1 and PC2 of the sparse PCA analysis, samples E23 and E24 are close to the E1 cluster. On the other hand, it appears from this data set that differences between producers or between years of production have more impact on the overall volatile fingerprints than on the effect of the change in ewe diet. This is consistent with the diversity of ewe uncooked pressed cheeses from the French Pyrenees (Feutry et al., 2012; Millet & Casabianca, 2019).

For an accurate discrimination of samples according to the qualitative variable “producer” and thus for looking for the origin of their typicality, a supervised sparse PLS-DA method was applied on the dataset (Fig. 3A). 26% of variables in PC1 led to an excellent separation between producers and allowed the extraction of the most important variables (product ions) in a Plot Loading figure (Fig. 4A). Hence, it is possible to identify the cheese’s typicality with these 430 variables. PC2 led to a discrimination between samples and illustrated the diversity in the same batch.

The clustered image map (CIM, Fig. 5) with sparse PLS-DA shows the classification of samples and the classification of the most relevant variables combined with an ion strength colour code. This tool was very useful to represent the volatile fingerprint of all cheese and to understand the trends. In fact, the distinction between the E1 group and the E2 group was clearly observed with a high homogeneity of the E1 cheese.

The comparison of the twelve cheeses produced by 2 breeders and with diet supplementations with two different plants do not permit identification of a global effect on cheese. However, when focusing only on samples of producer E1, a clear trend was observed according to the diet and the year of production. This is a first positive preliminary result encouraging breeders to study the secondary metabolites of plants on ewes’ welfare. Nevertheless, such a study on real farm model should be consolidated by increasing the number of samples and a better control of experiment design, based on two homogenous groups of ewes: one control and one with modified diet in a laboratory farm model.

The most relevant variables for the classification of samples correspond to product ions from the reaction between a precursor ion and an analyte. The aim of this study was to demonstrate that untargeted SIFT-MS analysis coupled with mixOmics supervised method allowed the discrimination of cheese samples. However, to go further, it could be very interesting to identify molecules responsible for a sample’s specificity and to go back to the involved enzymatic pathways, especially in the case of food products. Actually, the quadrupole mass detector of the SIFT-MS, is not a high-resolution mass detector and does not allow identification of the exact chemical formula of every product ion. A chemical compound emitted by a cheese sample may react with a precursor ion and give several product ions. In addition, two different chemical compounds may give rise to the same m/z product ion that makes the interpretation of SIFT-MS analysis challenging in complex matrices. It is all the truer because in a food product, the metabolic pathways give rise to families of chemical compounds such as alcohols, acids and esters that have a similar reactivity with precursor ions and then lead automatically to conflict ions, complicated to elucidate (Ghislain et al., 2021).

In most of the publications in the scientific field of DIMS analysis of food products, the authors (Castada et al., 2019; Taylor et al., 2013) based their strategies on the quantification of a few dozen known compounds. This hard work can be reinforced by our approach which demonstrated that there are more discriminating ions for classification. Moreover, the conflict ions with known and unknown compounds of the matrix make such quantifications difficult.

Knowing these experimental limitations, we have nevertheless tried to identify key compounds for the classification of sheep cheeses based on two post-treatment strategies: (i) first, by looking for product ions of known compounds, already identified in the literature: are these ions relevant in the list of 40 most relevant ions? (ii) second, by attempting attribution of product ions: which molecule could correspond to the most relevant ions?

- i) We have established a SIM method with 58 main compounds derived from the literature on volatile compounds released by cheeses, (Supplementary Information Table 4) and present in the software database (Labsyft®) for the positive ionisation (with H_3O^+ , O_2^+ and NO^+). These 58 compounds are known to produce 410 different product ions useful for the quantification, among which 329 are conflict ions. For example, O_2^+ precursor ion reacts with 8 different compounds from the list giving rise to a product ion with the same m/z 43 (SI Table 1). Thus, the precise determination of a compound concentration from a conflict ion is quite challenging. The remaining 81 ions with no conflict allow the measurements of the concentration of 36 compounds of the list without being confident on the accuracy of these calculations due to the possible presence of unexpected compounds. From the 40 most discriminating product ions obtained during the sPLS-DA analysis differentiating the cheese producers (Fig. 4A), 13 product ions from the 58 compounds of the SIM method were found. This clearly shows that other compounds participate in the discrimination and that the volatile fingerprint provides more information than the SIM method. This SIM method is only based on targeted compounds, usually identified by GC-MS and excluding inorganic compounds detected by SIFT-MS.
- ii) We approached the problem from another angle to identify all the key compounds. The 40 most relevant product ions providing discrimination of the producers VOC, were searched in the instrument database to identify the already known compounds. Approximately 300 compounds were thus identified, corresponding to at least 1 of these 40 ions. Some of these compounds can correspond to several product ions. For example, 3 product ions (m/z 92 with NO^+ , m/z 61 with O_2^+ , m/z 117 with O_2^+) discriminating producer E2 may correspond to propyl hexanoate, an already identified compound in cheeses (Bertolino et al., 2011; Bosset & Gauch, 1993; Di Cagno et al., 2003). This indicates that propyl hexanoate, an odorous compound, can be a more expressed marker in producer E2 cheeses than in producer E1 ones. Additional SPME-GC-MS experiments were carried out on the same samples (data not shown) and confirmed that propyl hexanoate was overexpressed in producer E2 cheeses.

The same data mining strategy was used to search differences between modified and control diets within E1 cheeses. The discrimination between sample categories was very clear in Figure 3B and a large number of the 40 most relevant product ions in Fig. 4B have a high m/z value. One of these ions generated from the H_3O^+ precursor ion at m/z 137, could be a very interesting candidate because of its involvement in a large number of terpene molecules. However, other characteristic ions of terpenes were not found in the top 40 ions list. This could be explained by a high level of interferences in product ions of terpenes. Several negative product ions were also highlighted by this analysis but the lack of a SIFT negative ionisation database will not enable this research to further develop. The obvious separation of samples in SIFT-MS full scan mode confirmed our hypothesis about the volatile fingerprints: untargeted analysis enhances the ability to detect unexpected compound. Nevertheless, the complementarity of this approach

with more traditional techniques can allow the identification of key molecules.

5. Conclusion

The main objective of this work was to show the great potential of unsupervised SIFT-MS analysis in full scan mode, to study the volatile compounds released by cheese. The use of adapted statistical tools, dedicated to “omics” datasets, allowed some characterisation of cheese producers and the effect of animal diet modifications. As for sensory analysis, where several molecules may fix on receptors and give different synergic or antagonist nervous signals, the raw measurement of all product ions (without any determination of molecule concentration) gave a fingerprint of the matrix. The construction of a large database of volatile fingerprints will enable the instrument to correlate a profile to a product specificity. The high frequency mass spectrometers like those in SIFT-MS generate a large database with a large number of variables and opens a new way of chemical analysis. In this work, the data mining and the statistical tools were adapted to generate a data set for a small number of ewe’s cheeses. It was demonstrated that the volatile fingerprints of samples make it possible to identify the cheese of each producer and highlighted the diversity of samples in the same batch. Cheese is a well-known matrix in food analytical science but a large number of parameters and manufacturing steps can impact the final product: the breed of animals, their lactation period, the composition of the nutritional part, the conditions of milk collection and at last, the transformation and maturation processes. To address issues like the impact of nutrition on the final product, it requires a perfect control of all the steps/parameters. Changing the mind by looking through big-data science will allow researchers and producers to find correlations between fingerprints and products’ typicality. To improve the present work, a construction of an Ossau-Iraty PDO SIFT-MS database is underway, which might help producers to solve question about typicality, diversity and agricultural practices.

CRedit authorship contribution statement

Marine Reyrolle: Methodology, Investigation, Software, Writing - original draft. **Mylène Ghislain:** Conceptualization, Methodology. **Noelle Bru:** Formal analysis. **Germain Vallverdu:** Software, Data curation. **Thierry Pigot:** Funding acquisition, Project administration. **Valerie Desauziers:** Validation. **Mickaël Le Bechec:** Writing - review & editing, Supervision.

Declaration of Competing Interest

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

Acknowledgements

Authors thank Conseil Régional Nouvelle Aquitaine (CRNA) and Communauté d’Agglomération Côte Basque (CAPB) for their financial support. Authors also thank the partners of BioNAchol project and in particular Xavier Recondo from Institute *Jean Errecart* (Saint Palais, France), the chamber of agriculture of Pyrénées Atlantique (France), the Protected Designation of Origin (PDO) Ossau-Iraty and the producers: *la ferme Oros Mirassou* (Arette, France) and *la ferme Bacquet* (Orègue, France).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2021.130801>.

References

- Bajoub, A., Medina-Rodríguez, S., Ajal, E. A., Cuadros-Rodríguez, L., Monasterio, R. P., Vercammen, J., Fernandez-Gutiérrez, A., & Carrasco-Pancorbo, A. (2018). A metabolic fingerprinting approach based on selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics: A reliable tool for Mediterranean origin-labeled olive oils authentication. *Food Research International*, 106, 233–242. <https://doi.org/10.1016/j.foodres.2017.12.027>.
- Balkir, P., Kemahlioglu, K., & Yuçel, U. (2021). Foodomics: A new approach in food quality and safety. *Trends in Food Science & Technology*, 108, 49–57. <https://doi.org/10.1016/j.tifs.2020.11.028>.
- Bergamaschi, M., Cipolat-Gotet, C., Cecchinato, A., Schiavon, S., & Bittante, G. (2020). Chemometric authentication of farming systems of origin of food (milk and ripened cheese) using infrared spectra, fatty acid profiles, flavor fingerprints, and sensory descriptions. *Food Chemistry*, 305, 125480. <https://doi.org/10.1016/j.foodchem.2019.125480>.
- Bertolino, M., Dolci, P., Giordano, M., Rolle, L., & Zeppa, G. (2011). Evolution of chemico-physical characteristics during manufacture and ripening of Castelmagno PDO cheese in wintertime. *Food Chemistry*, 129(3), 1001–1011. <https://doi.org/10.1016/j.foodchem.2011.05.060>.
- Biasioli, F., Yeretian, C., Mark, T. D., Dewulf, J., & Van Langenhove, H. (2011). Direct-² injection mass spectrometry adds the time dimension to (B)VOC analysis. *TrAC Trends in Analytical Chemistry*, 30(7), 1003–1017. <https://doi.org/10.1016/j.trac.2011.04.005>.
- Boltar, I., Canzek Majhenić, A., Jarni, K., Jug, T., & Bavcon Kralj, M. (2015). Volatile compounds in Nanos cheese: Their formation during ripening and seasonal variation. *Journal of Food Science and Technology*, 52(1), 608–623. <https://doi.org/10.1007/s13197-014-1565-6>.
- Boltar, I., Majhenić, A. C., Jarni, K., Jug, T., & Kralj, M. B. (2019). Research of volatile compounds in cheese affected by different technological parameters. *Journal of Food and Nutrition Research*, 58, 10.
- Bosset, J. O., & Gauch, R. (1993). Comparison of the volatile flavour compounds of six European ‘AOC’ cheeses by using a new dynamic headspace GC-MS method. *International Dairy Journal*, 3(4–6), 359–377. [https://doi.org/10.1016/0958-6946\(93\)90023-S](https://doi.org/10.1016/0958-6946(93)90023-S).
- Castada, H., Hanas, K., & Barringer, S. (2019). Swiss cheese flavor variability based on correlations of volatile flavor compounds, descriptive sensory attributes, and consumer preference. *Foods*, 8(2), 78. <https://doi.org/10.3390/foods8020078>.
- Castada, H. Z., Wick, C., Taylor, K., & Harper, W. J. (2014). Analysis of selected volatile organic compounds in split and nonsplit swiss cheese samples using selected-ion flow tube mass spectrometry (SIFT-MS): Analysis of selected volatile organic... *Journal of Food Science*, 79(4), C489–C498. <https://doi.org/10.1111/1750-3841.12418>.
- 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>.
- Cecchi, T., Sacchini, L., & Felici, A. (2018). First investigation on the shelf life of Mediterranean Mussels (*Mytilus galloprovincialis*) on the basis of their volatiles profiles. *Food Analytical Methods*, 11(5), 1451–1456. <https://doi.org/10.1007/s12161-017-1129-2>.
- Ch, R., Chevallier, O., McCarron, P., McGrath, T. F., Wu, D., Nguyen Doan Duy, L., Kapil, A. P., McBride, M., & Elliott, C. T. (2021). Metabolomic fingerprinting of volatile organic compounds for the geographical discrimination of rice samples from China, Vietnam and India. *Food Chemistry*, 334, 127553. <https://doi.org/10.1016/j.foodchem.2020.127553>.
- Chevallier, S., Bertrand, D., Kohler, A., & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, 20(5), 221–229. <https://doi.org/10.1002/cem.v20:510.1002/cem.994>.
- Cumeras, R. (2017). Volatilome metabolomics and databases, recent advances and needs. *Current Metabolomics*, 5(2), 79–89. <https://doi.org/10.2174/2213235X05666170502103408>.
- Cunningham, P. (2008). Dimension Reduction. In M. Cord, & P. Cunningham (Eds.), *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval* (pp. 91–112). Springer. https://doi.org/10.1007/978-3-540-75171-7_4.
- Deuscher, Z., Andriot, I., Sémon, E., Repoux, M., Preys, S., Roger, J.-M., Boulanger, R., Labouré, H., & Le Quééré, J.-L. (2019). Volatile compounds profiling by using proton transfer reaction-time of flight-mass spectrometry (PTR-ToF-MS). The case study of dark chocolates organoleptic differences. *Journal of Mass Spectrometry*, 54(1), 92–119. <https://doi.org/10.1002/jms.4317>.
- Di Cagno, R., Banks, J., Sheehan, L., Fox, P. F., Brechany, E. Y., Corsetti, A., & Gobetti, M. (2003). Comparison of the microbiological, compositional, biochemical, volatile profile and sensory characteristics of three Italian PDO ewes’ milk cheeses. *International Dairy Journal*, 13(12), 961–972. [https://doi.org/10.1016/S0958-6946\(03\)00145-6](https://doi.org/10.1016/S0958-6946(03)00145-6).

- Farag, M. A., Hegazi, N., Dokhalahy, E., & Khattab, A. R. (2020). Chemometrics based GC-MS aroma profiling for revealing freshness, origin and roasting indices in saffron spice and its adulteration. *Food Chemistry*, 331, 127358. <https://doi.org/10.1016/j.foodchem.2020.127358>.
- Ferreira, I. M. P. L. V. O., Pinho, O., & Sampaio, P. (2009). Volatile fraction of DOP "Castelo Branco" cheese: Influence of breed. *Food Chemistry*, 112(4), 1053–1059. <https://doi.org/10.1016/j.foodchem.2008.06.048>.
- Feutry, F., Torre, P., Arana, I., García, S., Desmasures, N., & Casalta, E. (2012). Lactococcus lactis strains from raw ewe's milk samples from the PDO Ossau-Iraty cheese area: Levels, genotypic and technological diversity. *Dairy Science & Technology*, 92(6), 655–670. <https://doi.org/10.1007/s13594-012-0084-3>.
- Ghislain, M., Costarramone, N., Sotiropoulos, J.-M., Pigot, T., Van Den Berg, R., Lacombe, S., & Le Behec, M. (2019). Direct analysis of aldehydes and carboxylic acids in the gas phase by negative ionization selected ion flow tube mass spectrometry: Quantification and modelling of ion–molecule reactions. *Rapid Communications in Mass Spectrometry*, 33(21), 1623–1634. <https://doi.org/10.1002/rcm.v33.2110.1002/rcm.8504>.
- Ghislain, M., Reyrolle, M., Sotiropoulos, J.-M., Pigot, T., & Le Behec, M. (2021). Chemical ionization of carboxylic acids and esters in negative mode selected ion flow tube – Mass spectrometry (SIFT-MS). *Microchemical Journal*, 169, 106609. <https://doi.org/10.1016/j.microc.2021.106609>.
- Guerra, L., McGarry, L. M., Robles, V., Bielza, C., Larranaga, P., & Yuste, R. (2011). Comparison between supervised and unsupervised classifications of neuronal cell types: A case study. *Developmental Neurobiology*, 71(1), 71–82. <https://doi.org/10.1002/dneu.20809>.
- Hera, D., Langford, V., McEwan, M., McKellar, T., & Milligan, D. (2017). Negative reagent ions for real time detection using SIFT-MS. *Environments*, 4(1), 16. <https://doi.org/10.3390/environments4010016>.
- Ianni, A., Bennato, F., Martino, C., Grotta, L., & Martino, G. (2020). Volatile flavor compounds in cheese as affected by ruminant diet. *Molecules*, 25(3), 461. <https://doi.org/10.3390/molecules25030461>.
- Jiang, Y., Li, N., Wang, Q., Liu, Z., Lee, Y.-K., Liu, X., Zhao, J., Zhang, H., & Chen, W. (2019). Microbial diversity and volatile profile of traditional fermented yak milk. *Journal of Dairy Science*, 103(1), 87–97. <https://doi.org/10.3168/jds.2019-16753>.
- Johnstone, I. M., & Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486), 682–693. <https://doi.org/10.1198/jasa.2009.0121>.
- Lê Cao, K.-A., Boitard, S., & Besse, P. (2011). Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*, 12(1), 253. <https://doi.org/10.1186/1471-2105-12-253>.
- Luyck, D. M. A. M., & van Ruth, S. M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry*, 107(2), 897–911. <https://doi.org/10.1016/j.foodchem.2007.09.038>.
- Lytou, A. E., Panagou, E. Z., & Nychas, G.-J. (2019). Volatilomics for food quality and authentication. *Current Opinion in Food Science*, 28, 88–95. <https://doi.org/10.1016/j.cofs.2019.10.003>.
- Millet, M., & Casabianca, F. (2019). Sharing values for changing practices, a lever for sustainable transformation? The case of farmers and processors in interaction within localized cheese sectors. *Sustainability*, 11(17), 4520. <https://doi.org/10.3390/su11174520>.
- R. Core Team. (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. www.R-project.org.
- Rohart, F., Gautier, B., Singh, A., Lê Cao, K.-A., & Schneidman, D. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, 13(11), e1005752. <https://doi.org/10.1371/journal.pcbi.1005752>.
- Santos, J., & Oliveira, M. B. P. P. (2017). Chromatography: Introduction to Chromatography - Techniques. In C. A. Georgiou, & G. P. Danezis (Eds.), *Food Authentication* (pp. 199–232). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118810224.ch7a>.
- Smith, D., McEwan, M. J., & Špaňel, P. (2020). Understanding gas phase ion chemistry is the key to reliable selected ion flow tube-mass spectrometry analyses. *Analytical Chemistry*, 13.
- Smith, D., & Špaňel, P. (2005). Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. *Mass Spectrometry Reviews*, 24(5), 661–700. <https://doi.org/10.1002/mas.v24:510.1002/mas.20033>.
- Sousa, A., Vareda, J., Pereira, R., Silva, C., Camara, J. S., & Perestrelo, R. (2020). Geographical differentiation of apple ciders based on volatile fingerprint. *Food Research International*, 137, 109550. <https://doi.org/10.1016/j.foodres.2020.109550>.
- Taylor, K., Wick, C., Castada, H., Kent, K., & Harper, W. J. (2013). Discrimination of Swiss cheese from 5 different factories by high impact volatile organic compound profiles determined by odor activity value using selected ion flow tube mass spectrometry and odor threshold: Swiss cheese factory VOC variability.... *Journal of Food Science*, 78(10), C1509–C1515. <https://doi.org/10.1111/1750-3841.12249>.
- Trendafilov, N. T., & Adachi, K. (2015). Sparse versus simple structure loadings. *Psychometrika*, 80(3), 776–790. <https://doi.org/10.1007/s11336-014-9416-y>.

Supplementary Information

	Control experiment	Modified diet
Dry hay	8,6%	8,6%
Dry second growth	34,3%	34,3%
Dry Alfalfa hay	29,7%	18,2%
Dry Sainfoin	0,0%	11,6%
Grain maize	10,9%	10,9%
Concentrated Nitrogen	16,5%	16,5%

Table SI 1 : Producer E1 feeding protocol

	Control experiment	Modified diet
Pasture	29,3%	29,3%
Dry second growth	34,1%	0,0%
Chicoree	0,0%	34,1%
Hay	19,5%	19,5%
Concentrated Nitrogen	17,1%	17,1%

Table SI 2 : Producer E2 Feeding protocol

Component	Concentration (Mole %)	Analytical uncertainty
1,2,3,4,tetrafluoro Benzene	2.100	± 5%
Benzene	2.110	± 5%
Ethylbenzene	2.040	± 5%
Isobutane	2.020	± 5%
Octafluorotoluene	2.110	± 5%
P Xylene	2.100	± 5%
Toluene	2.130	± 5%

Table SI 3: Calibration tank composition

Compound	CAS	Reagent Ion	Product Ion	m/z	Conflicts	References
<i>1-butanol</i>	71-36-3	O ₂ ⁺	CH ₅ O ⁺	33		(Attaie, 2009; Matteo Bergamaschi et al., 2016; Bertolino et al., 2011; Boltar et al., 2019; Bosset & Gauch, 1993)
	71-36-3	O ₂ ⁺	C ₃ H ₆ ⁺	42	X	
	71-36-3	O ₂ ⁺	C ₃ H ₇ ⁺	43	X	
	71-36-3	NO ⁺	C ₄ H ₉ ⁺	57	X	
	71-36-3	H ₃ O ⁺	C ₄ H ₉ ⁺	57	X	
	71-36-3	O ₂ ⁺	C ₄ H ₈ ⁺	56	X	
	71-36-3	NO ⁺	C ₄ H ₉ O ⁺	73	X	
	71-36-3	H ₃ O ⁺	C ₄ H ₁₁ O ⁺	75	X	
<i>1-hexanol</i>	111-27-3	O ₂ ⁺	C ₃ H ₆ ⁺	42	X	(M. Bergamaschi & Bittante, 2018; Bertolino et al., 2011; Bezerra et al., 2017; Boltar et al., 2019; Curioni & Bosset, 2002)
	111-27-3	O ₂ ⁺	C ₃ H ₇ ⁺	43	X	
	111-27-3	O ₂ ⁺	C ₄ H ₈ ⁺	56	X	
	111-27-3	O ₂ ⁺	C ₅ H ₁₀ ⁺	70	X	
	111-27-3	H ₃ O ⁺	C ₆ H ₁₃ ⁺	85	X	
	111-27-3	NO ⁺	C ₆ H ₁₃ O ⁺	101		
	111-27-3	O ₂ ⁺	C ₆ H ₁₂ ⁺	84	X	
<i>1-octanol</i>	111-87-5	H ₃ O ⁺	C ₄ H ₉ ⁺	57	X	(M. Bergamaschi & Bittante, 2018; Bertolino et al., 2011; Bezerra et al., 2017; Bontinis et al., 2012; Dahl et al., 2000)
	111-87-5	NO ⁺	C ₅ H ₈ ⁺	68	X	
	111-87-5	O ₂ ⁺	C ₄ H ₈ ⁺	56	X	
	111-87-5	H ₃ O ⁺	C ₅ H ₁₁ ⁺	71	X	
	111-87-5	O ₂ ⁺	C ₅ H ₈ ⁺	68	X	
	111-87-5	O ₂ ⁺	C ₅ H ₉ ⁺	69		
	111-87-5	O ₂ ⁺	C ₅ H ₁₀ ⁺	70	X	
	111-87-5	O ₂ ⁺	C ₆ H ₁₀ ⁺	82	X	
	111-87-5	O ₂ ⁺	C ₆ H ₁₁ ⁺	83	X	
	111-87-5	O ₂ ⁺	C ₆ H ₁₂ ⁺	84	X	
	111-87-5	H ₃ O ⁺	C ₈ H ₁₇ ⁺	113	X	
	111-87-5	NO ⁺	C ₈ H ₁₇ O ⁺	129		
<i>1-pentanol</i>	71-41-0	O ₂ ⁺	C ₃ H ₆ ⁺	42	X	(Attaie, 2009; M. Bergamaschi et al., 2015, 2016; M. Bergamaschi & Bittante, 2018; Bezerra et al., 2016)
	71-41-0	NO ⁺	C ₅ H ₉ ⁺	69	X	
	71-41-0	NO ⁺	C ₅ H ₁₁ ⁺	71	X	
	71-41-0	O ₂ ⁺	C ₄ H ₇ ⁺	55	X	
	71-41-0	H ₃ O ⁺	C ₅ H ₁₁ ⁺	71	X	
	71-41-0	O ₂ ⁺	C ₅ H ₁₀ ⁺	70	X	
	71-41-0	NO ⁺	C ₅ H ₁₁ O ⁺	87	X	
	71-41-0	O ₂ ⁺	C ₅ H ₁₁ ⁺	71	X	
<i>1-propanol</i>	71-23-8	O ₂ ⁺	CH ₃ O ⁺	31		(Attaie, 2009; Bosset & Gauch, 1993; Curioni & Bosset, 2002; Delgado et al., 2011; Imhof et al., 1995)
	71-23-8	H ₃ O ⁺	C ₃ H ₇ ⁺	43	X	
	71-23-8	O ₂ ⁺	C ₃ H ₆ ⁺	42	X	
	71-23-8	NO ⁺	C ₃ H ₇ O ⁺	59		
	71-23-8	H ₃ O ⁺	C ₃ H ₉ O ⁺	61		
<i>2,3-butanedione</i>	431-03-8	NO ⁺	C ₂ H ₃ O ⁺	43	X	(Attaie, 2009; Avsar et al., 2004; Boltar et al., 2019; Bontinis et al., 2012; Castada et al., 2014)
	431-03-8	O ₂ ⁺	C ₂ H ₃ O ⁺	43	X	
	431-03-8	NO ⁺	C ₄ H ₆ O ₂ ⁺	86	X	
	431-03-8	H ₃ O ⁺	C ₄ H ₇ O ₂ ⁺	87	X	

	431-03-8	O ₂ ⁺	C ₄ H ₆ O ₂ ⁺	86	X	
<i>2-butanol</i>	78-92-2	O ₂ ⁺	C ₃ H ₆ ⁺	42	X	(Bertolino et al., 2011; Bezerra et al., 2016; Bosset & Gauch, 1993; Delgado et al., 2010, 2011)
	78-92-2	H ₃ O ⁺	C ₄ H ₉ ⁺	57	X	
	78-92-2	O ₂ ⁺	C ₂ H ₅ O ⁺	45	X	
	78-92-2	NO ⁺	C ₄ H ₉ O ⁺	73	X	
	78-92-2	O ₂ ⁺	C ₃ H ₇ O ⁺	59	X	
	78-92-2	H ₃ O ⁺	C ₄ H ₁₁ O ⁺	75	X	
<i>2-heptanol</i>	543-49-7	H ₃ O ⁺	C ₄ H ₉ ⁺	57	X	(Battelli et al., 2019; Bertolino et al., 2011; Boltar et al., 2019; Buchin et al., 1998; Bugaud et al., 2001)
	543-49-7	O ₂ ⁺	C ₂ H ₅ O ⁺	45	X	
	543-49-7	O ₂ ⁺	C ₆ H ₁₁ ⁺	83	X	
	543-49-7	NO ⁺	C ₇ H ₁₅ O ⁺	115		
	543-49-7	H ₃ O ⁺	C ₇ H ₁₅ ⁺	99	X	
	543-49-7	O ₂ ⁺	C ₇ H ₁₄ ⁺	98	X	
	543-49-7	O ₂ ⁺	C ₆ H ₁₃ O ⁺	101		
<i>2-heptanone</i>	110-43-0	O ₂ ⁺	C ₂ H ₃ O ⁺	43	X	(Attaie, 2009; Battelli et al., 2019; M. Bergamaschi et al., 2015; M. Bergamaschi & Bittante, 2018; Bezerra et al., 2016)
	110-43-0	O ₂ ⁺	C ₄ H ₉ ⁺	57	X	
	110-43-0	O ₂ ⁺	C ₃ H ₆ O ⁺	58	X	
	110-43-0	O ₂ ⁺	C ₅ H ₁₁ ⁺	71	X	
	110-43-0	O ₂ ⁺	C ₆ H ₁₃ ⁺	85	X	
	110-43-0	H ₃ O ⁺	C ₇ H ₁₄ OH ⁺	115	X	
	110-43-0	NO ⁺	C ₇ H ₁₄ O.NO ⁺	144		
	110-43-0	O ₂ ⁺	C ₇ H ₁₄ O ⁺	114	X	
<i>2-methylbutanal</i>	96-17-3	H ₃ O ⁺	C ₂ H ₅ O ⁺	45		(Avsar et al., 2004; M. Bergamaschi et al., 2015; Bontinis et al., 2012; Castada et al., 2014; Curioni & Bosset, 2002)
	96-17-3	NO ⁺	C ₄ H ₉ ⁺	57	X	
	96-17-3	H ₃ O ⁺	C ₅ H ₉ ⁺	69	X	
	96-17-3	O ₂ ⁺	C ₄ H ₉ ⁺	57	X	
	96-17-3	O ₂ ⁺	C ₃ H ₆ O ⁺	58	X	
	96-17-3	NO ⁺	C ₅ H ₉ O ⁺	85	X	
	96-17-3	H ₃ O ⁺	C ₅ H ₁₀ .H ⁺	87	X	
	96-17-3	O ₂ ⁺	C ₅ H ₁₀ O ⁺	86	X	
<i>2-methylbutanoic acid</i>	116-53-0	NO ⁺	C ₄ H ₉ ⁺	57	X	(Attaie, 2009; Bontinis et al., 2012; Delgado et al., 2010, 2011; Di Cagno et al., 2003)
	116-53-0	H ₃ O ⁺	C ₄ H ₉ ⁺	57	X	
	116-53-0	O ₂ ⁺	C ₄ H ₉ ⁺	57	X	
	116-53-0	NO ⁺	C ₅ H ₉ O ⁺	85	X	
	116-53-0	H ₃ O ⁺	C ₅ H ₉ O ⁺	85	X	
	116-53-0	O ₂ ⁺	C ₄ H ₁₀ O ⁺	74	X	
	116-53-0	O ₂ ⁺	C ₅ H ₁₁ O ⁺	87		
	116-53-0	H ₃ O ⁺	C ₅ H ₁₀ O ₂ .H ⁺	103	X	
	116-53-0	NO ⁺	C ₅ H ₁₀ O ₂ .NO ⁺	132	X	
<i>2-methylpropanal</i>	78-84-2	O ₂ ⁺	C ₃ H ₇ ⁺	43	X	(Avsar et al., 2004; M. Bergamaschi & Bittante, 2018; Bontinis et al., 2012; Castada et al., 2014, 2015)
	78-84-2	NO ⁺	C ₄ H ₇ O ⁺	71	X	
	78-84-2	H ₃ O ⁺	C ₄ H ₉ O ⁺	73	X	
	78-84-2	O ₂ ⁺	C ₄ H ₇ O ⁺	71	X	
	78-84-2	O ₂ ⁺	C ₄ H ₈ O ⁺	72	X	
<i>2-nonanone</i>	821-55-6	O ₂ ⁺	C ₃ H ₆ O ⁺	58	X	(Attaie, 2009; Battelli et al., 2019; M. Bergamaschi et al., 2015; M. Bergamaschi & Bittante, 2018; Bezerra et al., 2017)
	821-55-6	O ₂ ⁺	C ₃ H ₇ O ⁺	59	X	
	821-55-6	O ₂ ⁺	C ₅ H ₁₁ ⁺	71	X	

	821-55-6	O ₂ ⁺	C6H13+	85	X	
	821-55-6	NO ⁺	C9H18O+	142		
	821-55-6	H ₃ O ⁺	C9H18OH+	143	X	
	821-55-6	NO ⁺	C9H18O.NO+	172	X	
	821-55-6	O ₂ ⁺	C9H18O+	142		
<i>2-nonenal</i>	2463-53-8	O ₂ ⁺	C4H9+	57	X	
	2463-53-8	O ₂ ⁺	C4H10+ or C3H6O+	58	X	
	2463-53-8	O ₂ ⁺	C4H6O+	70	X	
	2463-53-8	O ₂ ⁺	C5H7O+	83	X	
	2463-53-8	O ₂ ⁺	C5H8O+	84	X	
	2463-53-8	NO ⁺	C9H15O+	139		
	2463-53-8	O ₂ ⁺	C7H12+	96	X	
	2463-53-8	O ₂ ⁺	C6H10O+	98	X	
	2463-53-8	O ₂ ⁺	C9H14+	122	X	
	2463-53-8	H ₃ O ⁺	C9H17O+	141		
	2463-53-8	NO ⁺	C9H15O.NO+	169		
	2463-53-8	O ₂ ⁺	C9H15O+	139		(Avsar et al., 2004; M. Bergamaschi & Bittante, 2018; Bontinis et al., 2012; Castada et al., 2015; Curioni & Bosset, 2002)
<i>2-octanone</i>	111-13-7	O ₂ ⁺	C2H3O+	43	X	
	111-13-7	O ₂ ⁺	C3H6O+	58	X	
	111-13-7	O ₂ ⁺	C3H7O+	59	X	
	111-13-7	O ₂ ⁺	C5H11+	71	X	
	111-13-7	O ₂ ⁺	C6H13+	85	X	
	111-13-7	H ₃ O ⁺	C8H16OH+	129	X	
	111-13-7	NO ⁺	C8H16O.NO+	158		
	111-13-7	O ₂ ⁺	C8H16O+	128		
<i>2-pentanol</i>	6032-29-7	O ₂ ⁺	C2H4O+	44	X	
	6032-29-7	O ₂ ⁺	C2H5O+	45	X	
	6032-29-7	O ₂ ⁺	C4H7+	55	X	
	6032-29-7	NO ⁺	C5H11+	71	X	
	6032-29-7	H ₃ O ⁺	C5H11+	71	X	
	6032-29-7	NO ⁺	C5H11O+	87	X	
	6032-29-7	O ₂ ⁺	C5H10+	70	X	
	6032-29-7	O ₂ ⁺	C5H11+	71	X	(Bertolino et al., 2011; Boltar et al., 2019; Bosset & Gauch, 1993; Curioni & Bosset, 2002; Delgado et al., 2011)
<i>2-pentanone</i>	107-87-9	O ₂ ⁺	C2H3O+	43	X	
	107-87-9	O ₂ ⁺	C3H6O+	58	X	
	107-87-9	O ₂ ⁺	C4H7O+	71	X	
	107-87-9	H ₃ O ⁺	C5H11O+	87	X	
	107-87-9	NO ⁺	NO+.C5H10O	116		
	107-87-9	O ₂ ⁺	C5H10O+	86	X	(Battelli et al., 2019; M. Bergamaschi et al., 2015; Bezerra et al., 2016, 2017; Boltar et al., 2019)
<i>2-phenylethanol</i>	60-12-8	NO ⁺	C8H9O+	121	X	
	60-12-8	NO ⁺	C8H10O+	122		
	60-12-8	H ₃ O ⁺	C8H9+	105	X	
	60-12-8	O ₂ ⁺	C7H7+	91	X	
	60-12-8	O ₂ ⁺	C7H8+	92	X	
	60-12-8	O ₂ ⁺	C8H10O+	122	X	(Avsar et al., 2004; Bertolino et al., 2011; Bontinis et al., 2012; Curioni & Bosset, 2002; Delgado et al., 2010)
<i>2-undecanone</i>	112-12-9	O ₂ ⁺	C3H6O+	58	X	
	112-12-9	O ₂ ⁺	C5H11+	71	X	

	112-12-9	O ₂ ⁺	C6H13+	85	X	(Boltar et al., 2019; Curioni & Bosset, 2002; Delgado et al., 2011; Di Cagno et al., 2003; Dirinck & De Winne, 1999)
	112-12-9	O ₂ ⁺	C8H14+	110	X	
	112-12-9	O ₂ ⁺	C9H19+	127		
	112-12-9	NO ⁺	C11H22O.NO+	200		
	112-12-9	O ₂ ⁺	C11H20+	152		
	112-12-9	O ₂ ⁺	C11H22O+	170		
	112-12-9	H ₃ O ⁺	C11H22OH+	171		
<i>3-methyl-1-butanol</i>	123-51-3	NO ⁺	C5H9+	69	X	(Battelli et al., 2019; M. Bergamaschi et al., 2015; M. Bergamaschi & Bittante, 2018; Matteo Bergamaschi et al., 2016; Bezerra et al., 2016)
	123-51-3	NO ⁺	C5H11+	71	X	
	123-51-3	H ₃ O ⁺	C5H11+	71	X	
	123-51-3	O ₂ ⁺	C3H6O+	58	X	
	123-51-3	O ₂ ⁺	C3H7O+	59	X	
	123-51-3	NO ⁺	C5H11O+	87	X	
	123-51-3	O ₂ ⁺	C4H9O+	73	X	
<i>3-methyl-3-buten-1-ol</i>	763-32-6	NO ⁺	C3H8O+	60		(M. Bergamaschi et al., 2015; M. Bergamaschi & Bittante, 2018; Bosset & Gauch, 1993; Delgado et al., 2011; Di Cagno et al., 2003)
	763-32-6	NO ⁺	C5H8+	68	X	
	763-32-6	H ₃ O ⁺	C5H9+	69	X	
	763-32-6	O ₂ ⁺	C4H8+	56	X	
	763-32-6	NO ⁺	C5H9O+	85	X	
	763-32-6	O ₂ ⁺	C5H7+	67		
	763-32-6	O ₂ ⁺	C5H8+	68	X	
	763-32-6	NO ⁺	C5H10O+	86	X	
	763-32-6	O ₂ ⁺	C4H7O+	71	X	
	763-32-6	NO ⁺	C5H8.NO+	98	X	
	763-32-6	H ₃ O ⁺	C5H10O.H+	87	X	
<i>3-methylbutanal</i>	590-86-3	O ₂ ⁺	C2H4O+	44	X	(Avsar et al., 2004; M. Bergamaschi et al., 2015; Bezerra et al., 2016; Boltar et al., 2019; Castada et al., 2014)
	590-86-3	H ₃ O ⁺	C5H9+	69	X	
	590-86-3	O ₂ ⁺	C3H6O+	58	X	
	590-86-3	NO ⁺	C5H9O+	85	X	
	590-86-3	O ₂ ⁺	C4H7O+	71	X	
	590-86-3	H ₃ O ⁺	C5H11O+	87	X	
	590-86-3	O ₂ ⁺	C5H10O+	86	X	
<i>3-methylbutanoic acid</i>	503-74-2	NO ⁺	C5H9O+	85	X	(Battelli et al., 2019; M. Bergamaschi et al., 2015; Bertolino et al., 2011; Boltar et al., 2019; Bugaud et al., 2001)
	503-74-2	H ₃ O ⁺	C5H9O+	85	X	
	503-74-2	H ₃ O ⁺	C5H11O2+	103	X	
	503-74-2	NO ⁺	C5H10O2.NO+	132	X	
<i>acetoin</i>	513-86-0	O ₂ ⁺	C2H5O+	45	X	(Attaie, 2009; Battelli et al., 2019; Bezerra et al., 2016; Boltar et al., 2019)
	513-86-0	H ₃ O ⁺	C4H8O2.H+	89	X	
	513-86-0	NO ⁺	C4H8O2.NO+	118	X	
	513-86-0	O ₂ ⁺	C4H8O2+	88	X	
<i>acetone</i>	67-64-1	O ₂ ⁺	C2H3O+	43	X	(Battelli et al., 2019; M. Bergamaschi et al., 2015; Bosset & Gauch, 1993; Chin et al., n.d.; Curioni & Bosset, 2002)
	67-64-1	H ₃ O ⁺	C3H7O+	59		
	67-64-1	O ₂ ⁺	C3H6O+	58	X	
	67-64-1	NO ⁺	NO+.C3H6O	88		
<i>alpha-pinene</i>	80-56-8; 2437-95-8	NO ⁺	C7H8+	92	X	(Abilleira et al., 2011; M. Bergamaschi & Bittante, 2018; Bezerra et al., 2016, 2017; Boltar et al., 2019)
	80-56-8; 2437-95-8	H ₃ O ⁺	C6H9+	81	X	

	80-56-8; 2437-95-8	NO ⁺	C7H9 ⁺	93	X	
	80-56-8; 2437-95-8	O ₂ ⁺	C6H8 ⁺	80	X	
	80-56-8; 2437-95-8	O ₂ ⁺	C7H8 ⁺	92	X	
	80-56-8; 2437-95-8	O ₂ ⁺	C7H9 ⁺	93	X	
	80-56-8; 2437-95-8	NO ⁺	C10H16 ⁺	136	X	
	80-56-8; 2437-95-8	O ₂ ⁺	C8H11 ⁺	107	X	
	80-56-8; 2437-95-8	O ₂ ⁺	C9H13 ⁺	121	X	
	80-56-8; 2437-95-8	H ₃ O ⁺	C10H17 ⁺	137	X	
	80-56-8; 2437-95-8	O ₂ ⁺	C10H16 ⁺	136	X	
<i>benzaldehyde</i>	100-52-7	NO ⁺	C7H5O ⁺	105		(Battelli et al., 2019; M. Bergamaschi & Bittante, 2018; Bezerra et al., 2016; Boltar et al., 2019; Bontinis et al., 2012)
	100-52-7	H ₃ O ⁺	C7H7O ⁺	107		
	100-52-7	O ₂ ⁺	C7H5O ⁺	105		
	100-52-7	O ₂ ⁺	C7H6O ⁺	106		
<i>butanoic acid</i>	107-92-6	O ₂ ⁺	C2H3O ⁺	43	X	(Attaie, 2009; Avsar et al., 2004; Battelli et al., 2019; M. Bergamaschi & Bittante, 2018; Bertolino et al., 2011)
	107-92-6	NO ⁺	C3H7CO ⁺	71	X	
	107-92-6	H ₃ O ⁺	C3H7CO ⁺	71	X	
	107-92-6	O ₂ ⁺	CH3COOH ⁺	60	X	
	107-92-6	O ₂ ⁺	C2H4COOH ⁺	73	X	
	107-92-6	H ₃ O ⁺	C3H7COOH2 ⁺	89	X	
	107-92-6	NO ⁺	NO+.C3H7COOH	118	X	
<i>butanone</i>	78-93-3	O ₂ ⁺	C2H3O ⁺	43	X	
	78-93-3	O ₂ ⁺	C3H5O ⁺	57	X	
	78-93-3	H ₃ O ⁺	C4H9O ⁺	73	X	
	78-93-3	O ₂ ⁺	C4H8O ⁺	72	X	
	78-93-3	NO ⁺	NO+.C4H8O	102		
<i>butyl butanoate</i>	109-21-7	O ₂ ⁺	C3H7 ⁺	43	X	(Bertolino et al., 2011; Bontinis et al., 2012; Bosset & Gauch, 1993; Delgado et al., 2011; Di Cagno et al., 2003)
	109-21-7	O ₂ ⁺	C4H8 ⁺	56	X	
	109-21-7	NO ⁺	C4H7O ⁺	71	X	
	109-21-7	O ₂ ⁺	C4H7O ⁺	71	X	
	109-21-7	H ₃ O ⁺	C4H7O2H2 ⁺	89	X	
	109-21-7	NO ⁺	C8H16O2.NO ⁺	174	X	
	109-21-7	H ₃ O ⁺	C8H16O2H ⁺	145	X	
<i>decanal</i>	112-31-2	O ₂ ⁺	C5H8 ⁺	68	X	(M. Bergamaschi & Bittante, 2018; Boltar et al., 2019; Bontinis et al., 2012; Curioni & Bosset, 2002; Genovese et al., 2019)
	112-31-2	O ₂ ⁺	C5H10 ⁺	70	X	
	112-31-2	O ₂ ⁺	C5H11 ⁺	71	X	
	112-31-2	O ₂ ⁺	C6H9 ⁺	81	X	
	112-31-2	O ₂ ⁺	C6H10 ⁺	82	X	
	112-31-2	O ₂ ⁺	C7H12 ⁺	96	X	
	112-31-2	O ₂ ⁺	C8H14 ⁺	110	X	
	112-31-2	O ₂ ⁺	C8H16 ⁺	112	X	
	112-31-2	H ₃ O ⁺	C10H19 ⁺	139		
	112-31-2	NO ⁺	C10H19O ⁺	155		

	112-31-2	O ₂ ⁺	C10H18+	138		
	112-31-2	H ₃ O ⁺	C10H21O+	157		
<i>dimethyl disulfide</i>	624-92-0	O ₂ ⁺	CH2S+	46	X	(Battelli et al., 2019; M. Bergamaschi et al., 2015; Bertolino et al., 2011; Bontinis et al., 2012; Castada et al., 2014)
	624-92-0	O ₂ ⁺	CH3CH2S+	61	X	
	624-92-0	NO ⁺	(CH3)2S2+	94	X	
	624-92-0	O ₂ ⁺	CH3S2+	79		
	624-92-0	H ₃ O ⁺	(CH3)2S2.H+	95	X	
	624-92-0	O ₂ ⁺	(CH3)2S2+	94	X	
<i>dimethyl sulfide</i>	75-18-3	NO ⁺	(CH3)2S+	62		(Avsar et al., 2004; M. Bergamaschi & Bittante, 2018; Bontinis et al., 2012; Castada et al., 2014, 2015)
	75-18-3	O ₂ ⁺	CH2S+	46	X	
	75-18-3	O ₂ ⁺	CH3S+	47	X	
	75-18-3	H ₃ O ⁺	(CH3)2S.H+	63		
	75-18-3	O ₂ ⁺	(CH3)2S+	62		
<i>dimethyl trisulfide</i>	3658-80-8	O ₂ ⁺	C2H5S+	61	X	(Avsar et al., 2004; Bontinis et al., 2012; Bosset & Gauch, 1993; Castada et al., 2015; Curioni & Bosset, 2002)
	3658-80-8	O ₂ ⁺	CH2S2+	78		
	3658-80-8	O ₂ ⁺	CH3S2+	79		
	3658-80-8	O ₂ ⁺	CH4S2+	80	X	
	3658-80-8	NO ⁺	C2H6S3+	126		
	3658-80-8	H ₃ O ⁺	C2H6S3H+	127	X	
	3658-80-8	O ₂ ⁺	CH3S3+	111		
	3658-80-8	O ₂ ⁺	C2H6S3+	126		
<i>ethanol</i>	64-17-5	NO ⁺	C2H5O+	45		(Attaie, 2009; Battelli et al., 2019; Bertolino et al., 2011; Bezerra et al., 2016, 2017)
	64-17-5	H ₃ O ⁺	C2H7O+	47		
	64-17-5	O ₂ ⁺	C2H5O+	45	X	
	64-17-5	O ₂ ⁺	C2H6O+	46	X	
<i>ethyl acetate</i>	141-78-6	O ₂ ⁺	CH3CO+	43	X	(Battelli et al., 2019; M. Bergamaschi & Bittante, 2018; Bertolino et al., 2011; Bezerra et al., 2016; Boltar et al., 2019)
	141-78-6	O ₂ ⁺	C2H5O+	45	X	
	141-78-6	O ₂ ⁺	C2H5O2+	61	X	
	141-78-6	H ₃ O ⁺	CH3COOC2H5.H+	89	X	
	141-78-6	NO ⁺	NO+.CH3COOC2H5	118	X	
	141-78-6	O ₂ ⁺	C4H8O2+	88	X	
<i>ethyl butanoate</i>	105-54-4	O ₂ ⁺	C3H7+	43	X	(Avsar et al., 2004; Battelli et al., 2019; M. Bergamaschi et al., 2015; M. Bergamaschi & Bittante, 2018; Matteo Bergamaschi et al., 2016)
	105-54-4	NO ⁺	C4H7O+	71	X	
	105-54-4	H ₃ O ⁺	C4H7O+	71	X	
	105-54-4	O ₂ ⁺	C4H7O+	71	X	
	105-54-4	H ₃ O ⁺	C6H12O2.H+	117	X	
	105-54-4	NO ⁺	C6H12O2.NO+	146	X	
	105-54-4	O ₂ ⁺	C6H12O2+	116		
<i>ethyl decanoate</i>	110-38-3	H ₃ O ⁺	C2H5O.H+	46		(M. Bergamaschi & Bittante, 2018; Bertolino et al., 2011; Bezerra et al., 2016, 2017; Bontinis et al., 2012)
	110-38-3	NO ⁺	C4H9+	57	X	
	110-38-3	H ₃ O ⁺	C4H8.H+	57	X	
	110-38-3	O ₂ ⁺	C4H8+	56	X	
	110-38-3	NO ⁺	C5H11+	71	X	
	110-38-3	O ₂ ⁺	C4H9+	57	X	
	110-38-3	H ₃ O ⁺	C5H10.H+	71	X	
	110-38-3	NO ⁺	C6H13+	85	X	
	110-38-3	O ₂ ⁺	C5H10+	70	X	

	110-38-3	O ₂ ⁺	C5H11+	71	X	
	110-38-3	H ₃ O ⁺	C6H12.H+	85	X	
	110-38-3	NO ⁺	C8H16+	112		
	110-38-3	O ₂ ⁺	C6H12+	84	X	
	110-38-3	NO ⁺	C8H17+	113	X	
	110-38-3	O ₂ ⁺	C6H13+	85	X	
	110-38-3	H ₃ O ⁺	C7H14.H+	99	X	
	110-38-3	NO ⁺	C9H19+	127	X	
	110-38-3	H ₃ O ⁺	C8H16.H+	113	X	
	110-38-3	H ₃ O ⁺	C9H18.H+	127	X	
	110-38-3	NO ⁺	C12H24O2.NO+	230		
	110-38-3	O ₂ ⁺	C12H24O2+	200		
	110-38-3	H ₃ O ⁺	C12H24O2.H+	201		
<i>ethyl hexanoate</i>	123-66-0	NO ⁺	C6H11O+	99	X	(Avsar et al., 2004; Battelli et al., 2019; M. Bergamaschi et al., 2015; M. Bergamaschi & Bittante, 2018; Matteo Bergamaschi et al., 2016)
	123-66-0	NO ⁺	C8H16O2.NO+	174	X	
	123-66-0	H ₃ O ⁺	C8H16O2.H+	145	X	
<i>ethyl octanoate</i>	106-32-1	NO ⁺	C8H15O+	127	X	(M. Bergamaschi & Bittante, 2018; Bertolino et al., 2011; Bezerra et al., 2016; Boltar et al., 2019; Buchin et al., 1998)
	106-32-1	H ₃ O ⁺	C8H15O+	127	X	
	106-32-1	NO ⁺	C10H20O2.NO+	202		
	106-32-1	H ₃ O ⁺	C10H21O2+	173		
<i>gamma-octalactone</i>	104-50-7	H ₃ O ⁺	C8H14O2.H+	143	X	(Avsar et al., 2004; Bertolino et al., 2011; Boltar et al., 2019; Bontinis et al., 2012; Curioni & Bosset, 2002)
	104-50-7	NO ⁺	C8H14O2.NO+	172	X	
<i>heptanal</i>	111-71-7	O ₂ ⁺	C2H4O+	44	X	(M. Bergamaschi & Bittante, 2018; Bezerra et al., 2017; Boltar et al., 2019; Bontinis et al., 2012; Curioni & Bosset, 2002)
	111-71-7	O ₂ ⁺	C5H10+	70	X	
	111-71-7	O ₂ ⁺	C4H7O+	71	X	
	111-71-7	NO ⁺	C7H13O+	113	X	
	111-71-7	O ₂ ⁺	C5H10O+	86	X	
	111-71-7	H ₃ O ⁺	C7H13+	97		
	111-71-7	H ₃ O ⁺	C7H15O+	115	X	
	111-71-7	O ₂ ⁺	C7H12+	96	X	
<i>hexanal</i>	66-25-1	O ₂ ⁺	C2H4O+	44	X	(Avsar et al., 2004; M. Bergamaschi et al., 2015; M. Bergamaschi & Bittante, 2018; Matteo Bergamaschi et al., 2016; Curioni & Bosset, 2002)
	66-25-1	O ₂ ⁺	C4H8+	56	X	
	66-25-1	O ₂ ⁺	C3H5O+	57	X	
	66-25-1	H ₃ O ⁺	C6H11+	83	X	
	66-25-1	NO ⁺	C6H11O+	99	X	
	66-25-1	O ₂ ⁺	C5H12+	72	X	
	66-25-1	H ₃ O ⁺	C6H13O+	101		
	66-25-1	O ₂ ⁺	C6H9O+	97		
<i>hexanoic acid</i>	142-62-1	NO ⁺	C6H11O+	99	X	(Attaie, 2009; Avsar et al., 2004; Battelli et al., 2019; M. Bergamaschi et al., 2015; Bertolino et al., 2011)
	142-62-1	H ₃ O ⁺	C6H11O+	99	X	
	142-62-1	H ₃ O ⁺	C6H12O2.H+	117	X	
	142-62-1	NO ⁺	C6H12O2.NO+	146	X	
<i>isobutanoic acid</i>	79-31-2	O ₂ ⁺	(CH ₃) ₂ CH+	43	X	(Battelli et al., 2019; Bertolino et al., 2011; Boltar et al., 2019; Bugaud et al., 2001; Curioni & Bosset, 2002)
	79-31-2	NO ⁺	(CH ₃) ₂ CHCO+	71	X	
	79-31-2	H ₃ O ⁺	(CH ₃) ₂ CHCO+	71	X	
	79-31-2	O ₂ ⁺	CH ₃ CHCOOH+	73	X	
	79-31-2	H ₃ O ⁺	(CH ₃) ₂ CHCOOH ₂ ⁺	89	X	

	79-31-2	NO ⁺	NO+. (CH ₃) ₂ CHCOO H	118	X	
	79-31-2	O ₂ ⁺	(CH ₃) ₂ CHCOOH ⁺	88	X	
<i>Limonene</i>	5989-27-5	O ₂ ⁺	C ₅ H ₈ ⁺	68	X	(Abilleira et al., 2011; M. Bergamaschi & Bittante, 2018; Bertolino et al., 2011; Bezerra et al., 2017; Boltar et al., 2019)
	5989-27-5	H ₃ O ⁺	C ₆ H ₉ ⁺	81	X	
	5989-27-5	NO ⁺	C ₇ H ₈ ⁺	92	X	
	5989-27-5	NO ⁺	C ₇ H ₉ ⁺	93	X	
	5989-27-5	NO ⁺	C ₇ H ₁₀ ⁺	94	X	
	5989-27-5	O ₂ ⁺	C ₆ H ₈ ⁺	80	X	
	5989-27-5	H ₃ O ⁺	C ₇ H ₁₁ ⁺	95	X	
	5989-27-5	NO ⁺	C ₉ H ₁₃ ⁺	121	X	
	5989-27-5	O ₂ ⁺	C ₇ H ₈ ⁺	92	X	
	5989-27-5	NO ⁺	C ₁₀ H ₁₅ ⁺	135		
	5989-27-5	NO ⁺	C ₁₀ H ₁₆ ⁺	136	X	
	5989-27-5	O ₂ ⁺	C ₇ H ₉ ⁺	93	X	
	5989-27-5	O ₂ ⁺	C ₇ H ₁₀ ⁺	94	X	
	5989-27-5	O ₂ ⁺	C ₈ H ₁₁ ⁺	107	X	
	5989-27-5	O ₂ ⁺	C ₉ H ₁₃ ⁺	121	X	
		3268-49-3	O ₂ ⁺	C ₃ H ₈ S ⁺	76	
<i>methional</i>	3268-49-3	NO ⁺	C ₄ H ₈ OS ⁺	104	X	
	3268-49-3	H ₃ O ⁺	C ₄ H ₈ OS.H ⁺	105	X	
	3268-49-3	NO ⁺	C ₄ H ₈ OS.NO ⁺	134		
	3268-49-3	O ₂ ⁺	C ₄ H ₈ OS ⁺	104		
<i>methyl mercaptan</i>	74-93-1	H ₃ O ⁺	CH ₄ S.H ⁺	49		(Matteo Bergamaschi et al., 2016; Castada et al., 2014, 2015; Curioni & Bosset, 2002; Delgado et al., 2011)
	74-93-1	O ₂ ⁺	CH ₃ S ⁺	47	X	
	74-93-1	O ₂ ⁺	CH ₄ S ⁺	48		
<i>nonanal</i>	124-19-6	O ₂ ⁺	C ₃ H ₇ ⁺	43	X	(M. Bergamaschi et al., 2016; Boltar et al., 2019; Bontinis et al., 2012; Bosset & Gauch, 1993; Curioni & Bosset, 2002)
	124-19-6	O ₂ ⁺	C ₄ H ₉ ⁺	57	X	
	124-19-6	O ₂ ⁺	C ₅ H ₈ ⁺	68	X	
	124-19-6	O ₂ ⁺	C ₅ H ₁₀ ⁺	70	X	
	124-19-6	O ₂ ⁺	C ₅ H ₁₁ ⁺	71	X	
	124-19-6	O ₂ ⁺	C ₄ H ₈ O ⁺	72	X	
	124-19-6	O ₂ ⁺	C ₆ H ₉ ⁺	81	X	
	124-19-6	O ₂ ⁺	C ₆ H ₁₀ ⁺	82	X	
	124-19-6	O ₂ ⁺	C ₅ H ₉ O ⁺	85	X	
	124-19-6	NO ⁺	C ₉ H ₁₇ O ⁺	141		
	124-19-6	O ₂ ⁺	C ₇ H ₁₂ ⁺	96	X	
	124-19-6	H ₃ O ⁺	C ₉ H ₁₇ ⁺	125		
	124-19-6	O ₂ ⁺	C ₆ H ₁₀ O ⁺	98	X	
	124-19-6	O ₂ ⁺	C ₇ H ₁₄ O ⁺	114	X	
	124-19-6	O ₂ ⁺	C ₉ H ₁₆ ⁺	124		
	124-13-0	O ₂ ⁺	C ₅ H ₈ ⁺	68	X	(M. Bergamaschi et al., 2015; M. Bergamaschi & Bittante, 2018; Boltar et al., 2019; Bontinis et al., 2012; Curioni & Bosset, 2002)
<i>octanal</i>	124-13-0	O ₂ ⁺	C ₆ H ₁₀ ⁺	82	X	
	124-13-0	O ₂ ⁺	C ₆ H ₁₂ ⁺	84	X	
	124-13-0	O ₂ ⁺	C ₆ H ₁₃ ⁺	85	X	

	124-13-0	NO ⁺	C8H15O ⁺	127	X	
	124-13-0	H ₃ O ⁺	C8H15 ⁺	111		
	124-13-0	O ₂ ⁺	C8H14 ⁺	110	X	
	124-13-0	H ₃ O ⁺	C8H17O ⁺	129	X	
<i>octanoic acid</i>	124-07-2	H ₃ O ⁺	C8H16O ₂ .H ⁺	145	X	(Attaie, 2009; Avsar et al., 2004; Battelli et al., 2019; M. Bergamaschi et al., 2015; Matteo Bergamaschi et al., 2016)
<i>pentanoic acid</i>	109-52-4	O ₂ ⁺	CH ₃ COOH ⁺	60	X	(Attaie, 2009; Avsar et al., 2004; M. Bergamaschi et al., 2015; Bertolino et al., 2011; Bezerra et al., 2016)
	109-52-4	NO ⁺	C ₄ H ₉ CO ⁺	85	X	
	109-52-4	H ₃ O ⁺	C ₄ H ₉ CO ⁺	85	X	
	109-52-4	O ₂ ⁺	CH ₂ CH ₂ COOH ⁺	73	X	
	109-52-4	H ₃ O ⁺	C ₄ H ₉ COOH ₂ ⁺	103	X	
	109-52-4	NO ⁺	NO+.C ₄ H ₉ COOH	132	X	
<i>phenylacetaldehyde</i>	122-78-1	NO ⁺	C ₇ H ₇ ⁺	91		(Avsar et al., 2004; Boltar et al., 2019; Castada et al., 2014, 2015; Curioni & Bosset, 2002)
	122-78-1	NO ⁺	C ₈ H ₈ O ⁺	120		
	122-78-1	O ₂ ⁺	C ₇ H ₇ ⁺	91	X	
	122-78-1	O ₂ ⁺	C ₇ H ₈ ⁺	92	X	
	122-78-1	H ₃ O ⁺	C ₈ H ₈ O.H ⁺	121		
	122-78-1	NO ⁺	C ₈ H ₈ O.NO ⁺	150		
	122-78-1	O ₂ ⁺	C ₈ H ₈ O ⁺	120		
<i>propanoic acid</i>	79-09-4	NO ⁺	C ₂ H ₅ CO ⁺	57	X	(Avsar et al., 2004; Battelli et al., 2019; M. Bergamaschi et al., 2015; Bertolino et al., 2011; Bontinis et al., 2012)
	79-09-4	H ₃ O ⁺	C ₂ H ₅ CO ⁺	57	X	
	79-09-4	O ₂ ⁺	C ₂ H ₄ CO ⁺	56	X	
	79-09-4	O ₂ ⁺	C ₂ H ₅ CO ⁺	57	X	
	79-09-4	H ₃ O ⁺	C ₂ H ₅ COOH ₂ ⁺	75	X	
	79-09-4	O ₂ ⁺	C ₂ H ₅ COO ⁺	73	X	
	79-09-4	O ₂ ⁺	C ₂ H ₅ COOH ⁺	74	X	
	79-09-4	NO ⁺	NO+.C ₂ H ₅ COOH	104	X	
<i>propyl butanoate</i>	105-66-8	NO ⁺	C ₃ H ₇ ⁺	43	X	(Bosset & Gauch, 1993; Curioni & Bosset, 2002; Delgado et al., 2010, 2011; Di Cagno et al., 2003)
	105-66-8	H ₃ O ⁺	C ₃ H ₇ ⁺	43	X	
	105-66-8	O ₂ ⁺	C ₃ H ₇ ⁺	43	X	
	105-66-8	H ₃ O ⁺	C ₄ H ₉ ⁺	57	X	
	105-66-8	NO ⁺	C ₄ H ₇ O ⁺	71	X	
	105-66-8	H ₃ O ⁺	C ₄ H ₇ O ⁺	71	X	
	105-66-8	O ₂ ⁺	C ₄ H ₇ O ⁺	71	X	
	105-66-8	H ₃ O ⁺	C ₄ H ₇ O ₂ H ₂ ⁺	89	X	
	105-66-8	O ₂ ⁺	C ₄ H ₉ O ₂ ⁺	89		
	105-66-8	H ₃ O ⁺	C ₇ H ₁₄ O ₂ H ⁺	131		
	105-66-8	NO ⁺	C ₇ H ₁₄ O ₂ .NO ⁺	160		
<i>toluene</i>	108-88-3	NO ⁺	C ₇ H ₈ ⁺	92	X	(M. Bergamaschi & Bittante, 2018; Bezerra et al., 2016, 2017; Bontinis et al., 2012; Curioni & Bosset, 2002)
	108-88-3	H ₃ O ⁺	C ₇ H ₈ .H ⁺	93		
	108-88-3	O ₂ ⁺	C ₇ H ₈ ⁺	92	X	

Table SI 4 : Volatile Organic Compounds of cheese matrixes cited in the literature; with reactive ions, product ions and m/z ratio available in the labyft® database.

- Abilleira, E., Virto, M., Nájera, A. I., Albisu, M., Pérez-Elortondo, F. J., de Gordo, J. C. R., de Renobales, M., & Barron, L. J. R. (2011). Effects of seasonal changes in feeding management under part-time grazing on terpene concentrations of ewes' milk. *Journal of Dairy Research*, *78*(2), 129–135. <https://doi.org/10.1017/S0022029910000750>
- Attaie, R. (2009). Quantification of volatile compounds in goat milk Jack cheese using static headspace gas chromatography. *Journal of Dairy Science*, *92*(6), 2435–2443. <https://doi.org/10.3168/jds.2008-1732>
- Avsar, Y. K., Karagul-Yuceer, Y., Drake, M. A., Singh, T. K., Yoon, Y., & Cadwallader, K. R. (2004). Characterization of Nutty Flavor in Cheddar Cheese. *Journal of Dairy Science*, *87*(7), 1999–2010. [https://doi.org/10.3168/jds.S0022-0302\(04\)70017-X](https://doi.org/10.3168/jds.S0022-0302(04)70017-X)
- Battelli, G., Scano, P., Albano, C., Cagliani, L. R., Brasca, M., & Consonni, R. (2019). Modifications of the volatile and nonvolatile metabolome of goat cheese due to adjunct of non-starter lactic acid bacteria. *LWT*, *116*, 108576. <https://doi.org/10.1016/j.lwt.2019.108576>
- Bergamaschi, M., Biasioli, F., Cappellin, L., Cecchinato, A., Cipolat-Gotet, C., Cornu, A., Gasperi, F., Martin, B., & Bittante, G. (2015). Proton transfer reaction time-of-flight mass spectrometry: A high-throughput and innovative method to study the influence of dairy system and cow characteristics on the volatile compound fingerprint of cheeses. *Journal of Dairy Science*, *98*(12), 8414–8427. <https://doi.org/10.3168/jds.2015-9803>
- Bergamaschi, M., & Bittante, G. (2018). From milk to cheese: Evolution of flavor fingerprint of milk, cream, curd, whey, ricotta, scotta, and ripened cheese obtained during summer Alpine pasture. *Journal of Dairy Science*, *101*(5), 3918–3934. <https://doi.org/10.3168/jds.2017-13573>
- Bergamaschi, M., Cipolat-Gotet, C., Stocco, G., Valorz, C., Bazzoli, I., Sturaro, E., Ramanzin, M., & Bittante, G. (2016). Cheesemaking in highland pastures: Milk technological properties, cream, cheese and ricotta yields, milk nutrients recovery, and products composition. *Journal of Dairy Science*, *99*(12), 9631–9646. <https://doi.org/10.3168/jds.2016-11199>
- Bergamaschi, Matteo, Cecchinato, A., Biasioli, F., Gasperi, F., Martin, B., & Bittante, G. (2016). From cow to cheese: Genetic parameters of the flavour fingerprint of cheese investigated by direct-injection mass spectrometry (PTR-ToF-MS). *Genetics Selection Evolution*, *48*(1), 89. <https://doi.org/10.1186/s12711-016-0263-4>
- Bertolino, M., Dolci, P., Giordano, M., Rolle, L., & Zeppa, G. (2011). Evolution of chemico-physical characteristics during manufacture and ripening of Castelmagno PDO cheese in wintertime. *Food Chemistry*, *129*(3), 1001–1011. <https://doi.org/10.1016/j.foodchem.2011.05.060>
- Bezerra, T. K. A., Araújo, A. R. R., Arcanjo, N. M. de O., Silva, F. L. H. da, Queiroga, R. de C. R. do E., & Madruga, M. S. (2016). Optimization of the HS-SPME-GC/MS technique for the analysis of volatile compounds in caprine Coalho cheese using response surface methodology. *Food Science and Technology*, *36*(1), 103–110. <https://doi.org/10.1590/1678-457X.0035>
- Bezerra, T. K. A., Arcanjo, N. M. de O., Araújo, A. R. R. de, Queiroz, A. L. M. de, Oliveira, M. E. G. de, Gomes, A. M. P., & Madruga, M. S. (2017). Volatile profile in goat coalho cheese supplemented with probiotic lactic acid bacteria. *LWT - Food Science and Technology*, *76*, 209–215. <https://doi.org/10.1016/j.lwt.2016.03.041>
- Boltar, I., Majhenič, A. Č., Jarni, K., Jug, T., & Kralj, M. B. (2019). Research of volatile compounds in cheese affected by different technological parameters. *J. Food Nutr. Res.*, *58*, 10.
- Bontinis, Th. G., Mallatou, H., Pappa, E. C., Massouras, Th., & Alichanidis, E. (2012). Study of proteolysis, lipolysis and volatile profile of a traditional Greek goat cheese (Xinotyri) during ripening. *Small Ruminant Research*, *105*(1–3), 193–201. <https://doi.org/10.1016/j.smallrumres.2012.01.003>
- Bosset, J. O., & Gauch, R. (1993). Comparison of the volatile flavour compounds of six European 'AOC' cheeses by using a new dynamic headspace GC-MS method. *International Dairy Journal*, *3*(4–6), 359–377. [https://doi.org/10.1016/0958-6946\(93\)90023-S](https://doi.org/10.1016/0958-6946(93)90023-S)
- Buchin, S., Duboz, G., Le Quéré, J.-L., & Grappin, R. (1998). Développement des caractéristiques biochimiques et sensorielles des fromages de chèvre. Étude interspèce par échange de la matière grasse et du lait écrémé de laits de vache et de chèvre. *Le Lait*, *78*(6), 673–687. <https://doi.org/10.1051/lait:1998659>
- Bugaud, C., Buchin, S., Hauwuy, A., & Coulon, J.-B. (2001). Relationships between flavour and chemical composition of Abondance cheese derived from different types of pastures. *Le Lait*, *81*(6), 757–773. <https://doi.org/10.1051/lait:2001162>
- Castada, H. Z., Wick, C., Harper, W. J., & Barringer, S. (2015). Headspace quantification of pure and aqueous solutions of binary mixtures of key volatile organic compounds in Swiss cheeses using selected ion flow tube mass spectrometry: Headspace quantification of VOCs using SIFT-MS. *Rapid Communications in Mass Spectrometry*, *29*(1), 81–90. <https://doi.org/10.1002/rcm.7089>
- Castada, H. Z., Wick, C., Taylor, K., & Harper, W. J. (2014). Analysis of Selected Volatile Organic Compounds in Split and Nonsplit Swiss Cheese Samples Using Selected-Ion Flow Tube Mass Spectrometry (SIFT-MS): Analysis of selected volatile organic.... *Journal of Food Science*, *79*(4), C489–C498. <https://doi.org/10.1111/1750-3841.12418>
- Chin, H. W., Bernhard, R. A., & Rosenberg, M. (n.d.). *Solid Phase Microextraction for Cheese Volatile Compound Analysis*. 6.
- Curioni, P. M. G., & Bosset, J. O. (2002). Key odorants in various cheese types as determined by gas chromatography-olfactometry. *International Dairy Journal*, *12*(12), 959–984. [https://doi.org/10.1016/S0958-6946\(02\)00124-3](https://doi.org/10.1016/S0958-6946(02)00124-3)
- Dahl, S., Tavaría, F. K., & Malcata, F. X. (2000). Relationships between flavour and microbiological profiles in Serra da Estrela cheese throughout ripening. *International Dairy Journal*, *8*.
- Delgado, F. J., González-Crespo, J., Cava, R., García-Parra, J., & Ramírez, R. (2010). Characterisation by SPME–GC–MS of the volatile profile of a Spanish soft cheese P.D.O. Torta del Casar during ripening. *Food Chemistry*, *118*(1), 182–189. <https://doi.org/10.1016/j.foodchem.2009.04.081>
- Delgado, F. J., González-Crespo, J., Cava, R., & Ramírez, R. (2011). Formation of the aroma of a raw goat milk cheese during maturation analysed by SPME–GC–MS. *Food Chemistry*, *129*(3), 1156–1163. <https://doi.org/10.1016/j.foodchem.2011.05.096>
- Di Cagno, R., Banks, J., Sheehan, L., Fox, P. F., Brechany, E. Y., Corsetti, A., & Gobbetti, M. (2003). Comparison of the microbiological, compositional, biochemical, volatile profile and sensory characteristics of three Italian PDO ewes' milk cheeses. *International Dairy Journal*, *13*(12), 961–972. [https://doi.org/10.1016/S0958-6946\(03\)00145-6](https://doi.org/10.1016/S0958-6946(03)00145-6)
- Dirinck, P., & De Winne, A. (1999). Flavour characterisation and classification of cheeses by gas chromatographic–mass spectrometric profiling. *Journal of Chromatography A*, *847*(1–2), 203–208. [https://doi.org/10.1016/S0021-9673\(99\)00193-4](https://doi.org/10.1016/S0021-9673(99)00193-4)
- Genovese, A., Marrazzo, A., De Luca, L., Romano, R., Manzo, N., Masucci, F., Di Francia, A., & Sacchi, R. (2019). Volatile Organic Compound and Fatty Acid Profile of Milk from Cows and Buffaloes Fed Mycorrhizal or Nonmycorrhizal Ensiled Forage. *Molecules*, *24*(8), 1616. <https://doi.org/10.3390/molecules24081616>
- Imhof, R., Glättli, H., & Bosset, J. O. (1995). Volatile organic compounds produced by thermophilic and mesophilic single strain dairy starter cultures. *LWT - Food Science and Technology*, *28*(1), 78–86. [https://doi.org/10.1016/S0023-6438\(95\)80016-6](https://doi.org/10.1016/S0023-6438(95)80016-6)

CHAPITRE 4 : Discrimination de la zone géographique et de la couleur de 26 thés à partir des empreintes volatiles obtenues par deux méthodes : HS-SPME-GC-MS/FID et HS-SIFT-MS.

Les résultats présentés dans ce chapitre font l'objet d'une publication soumise à « *Journal of Agricultural and Food Chemistry* ».

Après avoir démontré le potentiel des empreintes volatiles obtenues par SIFT-MS pour discriminer des échantillons de fromage (Chapitre 3), cette approche nécessite d'être comparée avec les méthodes couramment citées dans la littérature pour l'analyse des COV dans des matrices alimentaires. La méthode très largement utilisée pour l'analyse des COV est la chromatographie gazeuse couplée à la spectrométrie de masse (GC-MS). D'après la littérature, différentes approches sont utilisées, l'approche ciblée consistant à rechercher des marqueurs spécifiques et l'approche non ciblée, plus récente et basée sur le traitement de l'intégralité du signal.

L'objectif de cette étude est de comparer les performances des deux méthodes analytiques, HS-SIFT-MS et HS-SPME-GC-MS/FID, développées au chapitre 2 et appliquées à des analyses ciblées et non ciblées pour discriminer des échantillons de thé. Le thé a été choisi car il s'agit d'une matrice sèche sans activité microbienne, donc plus stable. Pour cette étude, une experte en thé, Lydia Gauthier, a collecté 26 échantillons aux origines certifiées. Ces 26 thés proviennent de 7 pays (11 régions) et ont 5 couleurs différentes correspondant à différents procédés de transformation.

Les analyses full scan HS-SPME-GC-MS/FID et HS-SIFT-MS des 26 échantillons de thés ont été réalisées selon les conditions déterminées dans le chapitre 2. Pour le suivi des marqueurs, les molécules identifiées par GC-MS avec un pourcentage de correspondance supérieur à 70 % ont été retenues et les aires des pics chromatographiques ont été mesurées à partir du signal FID. Les molécules identifiées ont été utilisées pour construire la liste des molécules à quantifier par SIFT-MS. Pour ces molécules, les ions produits à partir des 8 ions précurseurs ont été suivis à l'aide du logiciel de base de données (Labsyft®). Pour l'analyse des profils volatils, les données full scan SIFT-MS ont été extraites avec l'algorithme développé et les données HS-SPME-GC-MS ont été traitées avec le logiciel Chromcompare+. Les jeux de données ainsi obtenus ont été traités avec le langage R en utilisant une méthode supervisée (sparse PLS-DA) du package « MixOmics » pour représenter la dispersion et la discrimination des échantillons de thé selon leur origine géographique et leur couleur.

Concernant les marqueurs, les résultats ont montré que le suivi par HS-SPME-GC-MS/FID (signal FID) permettait de distinguer la zone géographique de production avec un impact du cultivar sur certains échantillons mais ne permettait pas de distinguer la couleur. En revanche, le suivi par HS-SIFT-MS n'a montré aucune discrimination à partir des mêmes molécules. Ces résultats soulèvent les inconvénients de l'instrument SIFT-MS, la problématique des ions conflits et la difficulté de déterminer si la quantification d'un ion est spécifique à un composé ou si elle est due à un mélange de composés isomères. L'exclusion des ions conflictuels pour le suivi peut être envisagé mais dans le cas de matrices complexes comme les aliments, d'autres molécules non ciblées mais présentes dans la matrice peuvent générer des ions conflictuels sans que l'on puisse en avoir connaissance. Ceci rend la méthode de suivi SIFT-MS complexe pour la quantification.

En ce qui concerne les profils volatils, les résultats montrent que les deux méthodes analytiques permettent de discriminer partiellement la zone géographique de production avec une mise en évidence de l'importance des cultivars utilisés sur les COV émis. Les profils volatils obtenus par les deux techniques sont également capables de retracer le procédé de transformation et de séparer les échantillons selon la couleur du thé.

Les résultats de cette étude montrent que la comparaison des profils volatils sans ciblage de composés spécifiques s'avère être une approche prometteuse pour discriminer des échantillons de thés. L'analyse non ciblée des COV, exploitant l'ensemble du signal, couplée à des outils chimiométriques adaptés, permet d'obtenir une réponse rapide quant à la discrimination des échantillons.

Ce chapitre confirme le potentiel de l'analyse SIFT-MS de matrices complexes mais démontre aussi la capacité des profils volatils obtenus par GC-MS à discriminer des échantillons alimentaires à partir des

COV émis. Les deux méthodes s'avèrent donc pertinentes et cela ouvre des perspectives pour le post-traitement de mesures GC-MS déjà réalisées. Cependant, elles ont chacune leurs avantages et leurs limites concernant le temps d'analyse, l'identification des composés, la quantification, leur sélectivité physico-chimique ou encore leur sensibilité. Ces paramètres vont avoir un rôle clé dans le choix de la méthode analytique en fonction du besoin.

Performances of monitoring and profiling approaches with SPME-GC-MS and SIFT-MS for tea traceability.

Marine Reyrolle¹, Thierry Pigot¹, Valerie Desauziers¹ and Mickael Le Behec¹

¹ Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, IMT Mines Ales, IPREM, Pau, France, Institut des sciences analytiques et de Physicochimie pour l'environnement et les Matériaux, UMR5254, Hélioparc, 2 avenue Président Angot, 64053, PAU cedex 9, France

The performances of a separative technique and a direct injection mass spectrometry (DIMS) technique for volatile organic compounds (VOCs) measurement were compared to discriminate 26 tea samples. Two strategies were evaluated to trace the origins or the transformation processes of tea: first, 50 different markers were identified with head space, solid phase micro extraction gas chromatography, mass spectrometry (HS-SPME-GC-MS) and quantified with a flame ionization detector (FID) on the same instrument. This set of markers permitted to separate the different origins of samples but do not allowed to group samples with the color. The same list of markers was measured with a selected ion flow tube mass spectrometry (SIFT-MS) instrument without success for either origin separation nor color differentiation. Second, a post processing with an untargeted approach was applied on the GC-MS and SIFT-MS data set. 400 features were selected for GC-MS and resulted in the separation of a part of countries and the separation of the 5 different colors, corresponding to different transformation processes. 1 880 product ions were extract from the full scan SIFT-MS and also permitted to separate some of countries and all colors of the tea set. Interestingly, the impact of cultivar was highlighted for the origin discrimination with both techniques. This preliminary work, showed that the volatile profiles of tea may be analyzed with untargeted approach with separative and DIMS techniques and highlight the importance of the sample set composition on the separation rate.

INTRODUCTION

The increase of food fraud and the multiplication of food scandals with intoxications, falsifications and adulterations has resulted in the emergences of the need for consumer information and the need to secure the supply chains. Sanitary controls and quality labels are up to now the main tools to reinforce the control of traceability and authenticity and to restore the consumer confidence. Food traceability includes all the declarative information on geographical course of the product, from its production to its consumption. Whereas, food authentication is the global process of verifying the conformity of a product with its label description, including origin information (species, geographical area...), production techniques (conventional, organic agriculture, ...) and processing methods.

Tea, produced from buds and leaves of *Camelia sinensis* plants, is one the most consumed drink in the world, thanks to the diversity of its flavors and its health benefits. Historically originating from China, tea is now produced in a wide number of countries, as Turkey, Kenya, Pakistan, Russia and USA but

still mostly produced in Asia (China, India, Japan...)¹. The soil, the climate, the associated plants, and the finesse of the plucking have a crucial role in the development of aromatic potential of tea. Numerous transformation processes of the tea leaves have led to the creation of several tea grades^{2,3} with different oxidation, fermentation, or roasting rates which result in different colors of tea (green, Oolong or blue green, black, dark, red, white and yellow). Tea is a very old product consumed all over the world and is highly impacted by fraud⁴. For the assessment of food authenticity and traceability, control agency rely on analytical techniques such as separation techniques coupled with mass spectrometry (LC/MS, GC/MS), direct injection mass spectrometry (SIFT/MS, PTR/MS), near infrared spectrometry (NIRS) and isotope ratio mass spectrometry (IRMS) have been developed in recent years⁵.

In food industry, volatile organic compounds (VOCs), a diverse and ubiquitous set of chemicals, have a great importance due to their contribution in the development of odor, the first attribute for food acceptability by consumers. The volatile profile of a food product may be affected by

geographical origin, seasonal variation, manufacturing processes or storage conditions. This profile includes a large number of different compounds - sometimes hundreds - of varying chemical classes that may be present at very different concentrations. Identification and quantification of VOCs has always been a challenge due to their diversity and the complexity of food matrices.

Gas chromatography coupled with mass spectrometry allows the analysis of complex mixtures of compounds, with a wide diversity in chemical nature and volatility. This technique allows the separation of compounds based on their physicochemical properties and their affinity for the capillary column⁶. To increase the sensitivity of GC-MS analysis, sample preparation involving the use of extraction and pre-concentration techniques is a key step. Headspace (HS) solid phase micro-extraction (SPME) coupled with gas chromatography-mass spectrometry (GC-MS) is one of the most used method for the analysis of VOCs in food matrices^{7,8}. This method has been used for VOC analysis in tea to study different steeping temperatures⁹, different varieties of cultivars^{10,11}, different fermentation processes^{12,13} and several geographical origins¹⁴.

In recent years, direct injection mass spectrometry (DIMS) instruments such as proton transfer reaction mass spectrometer (PTR-MS) or selected ion flow tube mass spectrometry (SIFT-MS) have been developed for rapid, non-invasive and direct on-line measurement of volatile organic compounds (VOCs)¹⁵⁻¹⁷. These technologies do not base their selectivity on the separation of analytes but on a soft chemical ionization of the VOCs emitted by products. These instruments are becoming increasingly popular due to their high sensitivity and speed. SIFT-MS technique has been used as a tool to authenticate olive oils of Mediterranean origin¹⁸ or their adulteration¹⁹, for geographic traceability of Moroccan argan oils^{20,21}, to discriminate volatolome from *Vitis vinifera* berries²² or to discriminate cheeses²³⁻²⁵. Up to now, only few studies compared the performances of DIMS and GC-MS for VOC monitoring²⁶⁻²⁹. In this work, targeted (monitoring) and untargeted (fingerprint) analysis with HS-SPME-GC-MS and HS-SIFT-MS were performed on 26 tea samples with certified origin. The objective of this study was to evaluate the potential of the both analytical techniques with the two modes of data treatment (monitoring and fingerprint) to discriminate tea according to geographical area of production (country and region) and color.

MATERIEL AND METHODS

Samples

A tea expert collected 26 tea samples with certified origins directly from factory. The 26 teas come from seven different countries: Nepal (n=3), China (n=9), Vietnam (n=5), India (n=3), Taiwan (n=2), Sri Lanka (n=2) and Japan (n=2), twelve different regions and have five different colors: green (n=6), black (n=10), dark (n=3), white (n=4) and oolong (n=3) (table 1).

Head Space – Solid Phase Micro Extraction – Mass Spectrometry – Flame Ionization Detector (HS-SPME-GC-MS-FID) analysis

1 g of dry tea was placed into a 10mL headspace vial immediately sealed. After a headspace equilibration phase at 60 °C during 10 minutes, VOC were extracted and pre-concentrated with a SPME fiber (30/50 µm PDMS/DVB/CAR Supelco, St. Louis, MO, USA) to extract a maximum of compounds in a wide range of physicochemical properties. SPME fiber was exposed to tea sample headspace during 12 min at 60°C and desorbed during 5 min at 250°C into the GC-MS/FID injector.

A gas chromatograph (Agilent 7890B) coupled with a MS detector (Agilent 5977B, Agilent, Santa Clara, CA, USA) and a Flame ionization detector and connected to a multifunction autosampler (Combi-Pal, CTC Analytics, Zwingen, Switzerland) was used to analyze the volatile compounds in tea. Separation of volatile compounds was achieved on ELITE-5MS capillary column ((5%-Phenyl)-methylpolysiloxane, 30 m × 250 µm, 0.25 µm film thickness, Perkin-Elmer, Waltham, MA, USA). Helium 6.0 (Air Liquide, France) was used as carrier gas with a constant flow rate of 1 mL min⁻¹. The injection port was equipped with a 0.8 mm internal diameter liner and maintained at 250°C in splitless mode. The oven temperature was initially held at 40°C for 4 min, followed by a first temperature rise to 90°C at a rate of 15°C/min, held at 250°C for 4 min and a second temperature rise to 250°C at a rate of 10°C/min, and held at 250°C for 5 min. The total time of analysis was 32 min 33 s.

Tea	Country	Region	Color
1	Nepal	Ilam Valley	White
2	Nepal	Ilam Valley	White
3	Nepal	Ilam Valley	White
4	China	Zhejiang	Green
5	China	Zhejiang	Green
6	China	Zhejiang	Green
7	Japan	Kyushu	Green
8	Japan	Kyushu	Green
9	China	Jiangsu	Black
10	China	Jiangsu	Black
11	China	Jiangsu	Black
12	India	Darjeeling	Black
13	India	Darjeeling	Black
14	India	Darjeeling	Black
15	Taiwan	Ali Mounts	Black
16	Taiwan	Ali Mounts	Oolong
17	Vietnam	Lai Chau	Oolong
18	Vietnam	Lai Chau	Oolong
19	Vietnam	Lai Chau	White
20	Vietnam	Lai Chau	Black
21	Vietnam	Lao Cai	Green
22	China	Yunnan	Dark
23	China	Yunnan	Dark
24	China	Yunnan	Dark
25	Sri Lanka	Ratnapura	Black
26	Sri Lanka	Matara	Black

Table 1: Description of tea samples

The mass spectrometer was operated using Electronic Impact (EI) mode with electron energy of 70 eV. The ion source, transfer line and quadrupole temperature were set at 280°C, 250°C and 150°C, respectively. Acquisition was carried out in scan mode with mass ranging from 15 to 250 atomic mass unit (amu). The volatile compounds were tentatively identified using the standard NIST 14 library for mass spectra. Each tea sample was analyzed in triplicate.

Table 2: Description of tea samples

Head Space-SIFT-MS

10 g of dry tea were introduced into a 1-litre bottle. The bottle was fitted with a polypropylene screw cap with 2 tight connection ports fitted with 0.6 cm PFA Tube. The first one was connected to SIFT-MS and the second one to a 1L Tedlar © bag (Supelco, Bellefonte, PA, USA), filled with zero dry air (ZeroAir Alliance ZA1500, F-DGSI, Evry France) to compensate the volume taken by the SIFT-MS analysis. The closed bottle was previously incubated for 1 hour at 60°C before performing the positive and negative SIFT-MS full scan analysis.

A SIFT-MS Voice 200 Ultra (SYFT Technologies, Christchurch, New Zealand) equipped with a dual source producing positive and negative soft ionizing precursor ions (H_3O^+ , O_2^{*+} , NO^+ , O^+ , OH^+ , O_2^{*+} , NO_2^- and NO_3^-) in a single scan was used^{30,31}. Nitrogen (Air Liquide, Paris, France) was used as carrier gas, the sample was introduced through a temperature (110°C) and flow controlled (20 mL min⁻¹) sample line (High Performance Inlet HPI®). The instrument was daily calibrated with a standard gas containing standards at 2.0 ppmV in nitrogen (Scott™ gas mixtures, Air Liquid USA). A blank measure of each empty bottle was performed before sample introduction. The scan raw data files containing product ion intensities within 15-250 m/z range were collected.

Data processing and statistical analysis

The molecules identified by GC-MS, with a NIST database percentage match higher than 70%, were retained as markers and the corresponding peak areas were measured using FID signal. This list of markers was used to construct a quantification method for SIFT-MS measurement. Thus, the product ions corresponding to these markers were identified in the data software Labsyft®. After removing the conflict ions, corresponding to product ions with same m/z ratio from different molecules, a Multi Ion Monitoring method was developed and applied on the full scan measure to quantify these markers.

Untargeted analysis of HS-SPME-GC-MS was performed with Chromcompare+ software (Markes International, Bridgend, United Kingdom). According to Spadafora *et al*³², a data alignment algorithm was applied to one the 78 chromatograms. Secondly, an untargeted tile-based approach was done on the aligned chromatograms, dividing each chromatogram into 4 000 tiles. An Initial filtering process was applied with a minimum absolute intensity of 1 000 and minimum label frequency of 100 %. Then a feature discovery algorithm was applied by selecting a descriptive label (country,

region or color). The 400 most discriminative features were then extracted before statistical analysis and graphical representations.

Compounds	CAS	MM (g/mol)	Family	SIFT-MS
1-Penten-3-ol	616-25-1	86.13	alcohol	✓*
1-Pentanol	71-41-0	88.15	alcohol	✓
2-Penten-1-ol	1576-95-0	86.13	alcohol	✓*
3-Hexen-1-ol	544-12-7	100.16	alcohol	✓*
2-Hexen-1-ol	928-95-0	100.16	alcohol	✓
1-Hexanol	111-27-3	102.17	alcohol	✓
1-Octen-3-ol	3391-86-4	128.21	alcohol	✓*
Benzyl alcohol	100-51-6	108.14	alcohol	✓
Linalool oxide	5989-33-3	170.25	alcohol	x
trans Linalool oxide	34995-77-2	170.25	alcohol	x
Phenylethyl alcohol	60-12-8	122.16	alcohol	✓
trans-Linalool 3,7 oxide	39028-58-5	170.25	alcohol	x
beta-Myrcene	123-35-3	136.23	terpene	✓
D-Limonene	5989-27-5	136.23	terpene	✓*
beta-cis-Ocimene	3338-55-4	136.23	terpene	✓*
Linalool	78-70-6	154.25	terpene	✓
beat-Cyclocitral	432-25-7	152.23	terpene	✓
Linalyl acetate	115-95-7	196.29	terpene	✓
D-Carvone	2244-16-8	150.22	terpene	✓
Geraniol	106-24-1	154.25	terpene	✓
alpha-Longipinene	5989-08-2	204.35	terpene	x
beta-Ionone	14901-07-6	192.3	terpene	✓
beta-Ionone epoxide	23267-57-4	208.3	terpene	x
Pentanal	110-62-3	86.13	aldehyde	✓*
Hexanal	66-25-1	100.16	aldehyde	✓*
2-Hexenal	6728-26-3	98.14	aldehyde	✓
Heptanal	111-71-7	114.19	aldehyde	✓
Benzaldehyde	100-52-7	106.12	aldehyde	✓
Octanal	124-13-0	128.21	aldehyde	✓
2,4-Heptadienal	4313-03-5	110.15	aldehyde	✓
Nonanal	124-19-6	142.24	aldehyde	✓*
3-Penten-2-one, 4-methyl	141-79-7	98.14	ketone	x
2-Heptanone	110-43-0	114.19	ketone	✓
Butyrolactone	96-48-0	86.09	ketone	✓
5-Hepten-2-one, 6-methyl	110-93-0	126.2	ketone	✓
3,5-Octadien-2-one	30086-02-3	124.18	ketone	x
Methyl salicylate	119-36-8	152.15	ester	✓
Hexanoic acid, methyl ester	106-70-7	130.18	ester	✓
Butanoic acid, 2-methyl-, hexyl ester	10032-15-2	186.29	ester	x
Decane	124-18-5	142.28	alkane	✓
Undecane, 3-methyl	1002-43-3	170.33	alkane	x
Dodecane	112-40-3	170.33	alkane	✓
Tetradecane	629-59-4	198.39	alkane	✓
Acetic acid	64-19-7	60.05	carboxylic acid	✓
Hexanoic acid	142-62-1	116.16	carboxylic acid	✓
Butanoic acid, 4-hydroxy	591-81-1	104.1	carboxylic acid	x
Furan, 2-pentyl	3777-69-3	138.21	furan	✓
Caffeine	58-08-2	194.19	heterocyclic compound	x

Table 3: Identified molecules by HS-SPME-GC-MS. The cross indicates the unknown compounds in the Labsyft® database, the check mark indicates the known molecules and the asterisk indicates the molecules that was excluded because of conflict ions.

For volatile profiles by SIFT-MS, the algorithm developed in our previous work was used to extract the 1 888 ions produced by the full scan analysis²⁵.

The R language³³ was used with “corrplot” package³⁴ for correlation visualization and “MixOmics” package³⁵ to represent the dispersion and discrimination of the samples. Supervised methods (Partial Least Squares-Discriminant Analysis, PLS-DA) were applied on the datasets.

RESULTS/DISCUSSION

Targeted markers monitoring strategy

The objective was first to compare the performance of the both VOC analytical techniques (HS-SPME-GC-MS and SIFT-MS) with the monitoring of markers, to discriminate the tea samples according to origin or transformation processes. Thus 48 compounds with a NIST score over 70% were identified by GC-MS and quantified thanks to the FID signal (Table 3). These markers included 12 alcohols, 11 terpenes, 8 aldehydes, 5

ketones, 3 esters, 4 alkanes, 3 acids and 2 other compounds (Furan, 2-pentyl and caffeine). 37 of the 48 molecules were found in the Labsyft database with 138 corresponding product ions. However, among these product ions, 90 were conflict ions, meaning that two or more molecules can generate an ion with the same m/z ratio. These signals have to be removed for the quantification of markers. Finally, only 48 product ions were not suspected to be interfered by another molecule of the list. Hence 9 compounds cannot be quantified (marked with an asterisk in table 2). Sparse PLS-DA analysis were applied on each dataset (GC-FID pic area and SIFT-MS concentrations) to try to discriminate the two descriptive parameters “Country_region” and “Color”.

Targeted origin discrimination

The mean values of triplicates of FID peak areas of the 50 different markers were calculated for every sample. A sPLS-DA was then performed on the dataset according to the descriptive parameter “Country_region” to look for correlations. The results are presented in the Figure 1 A according to the first and the second components of the sPLS-DA.

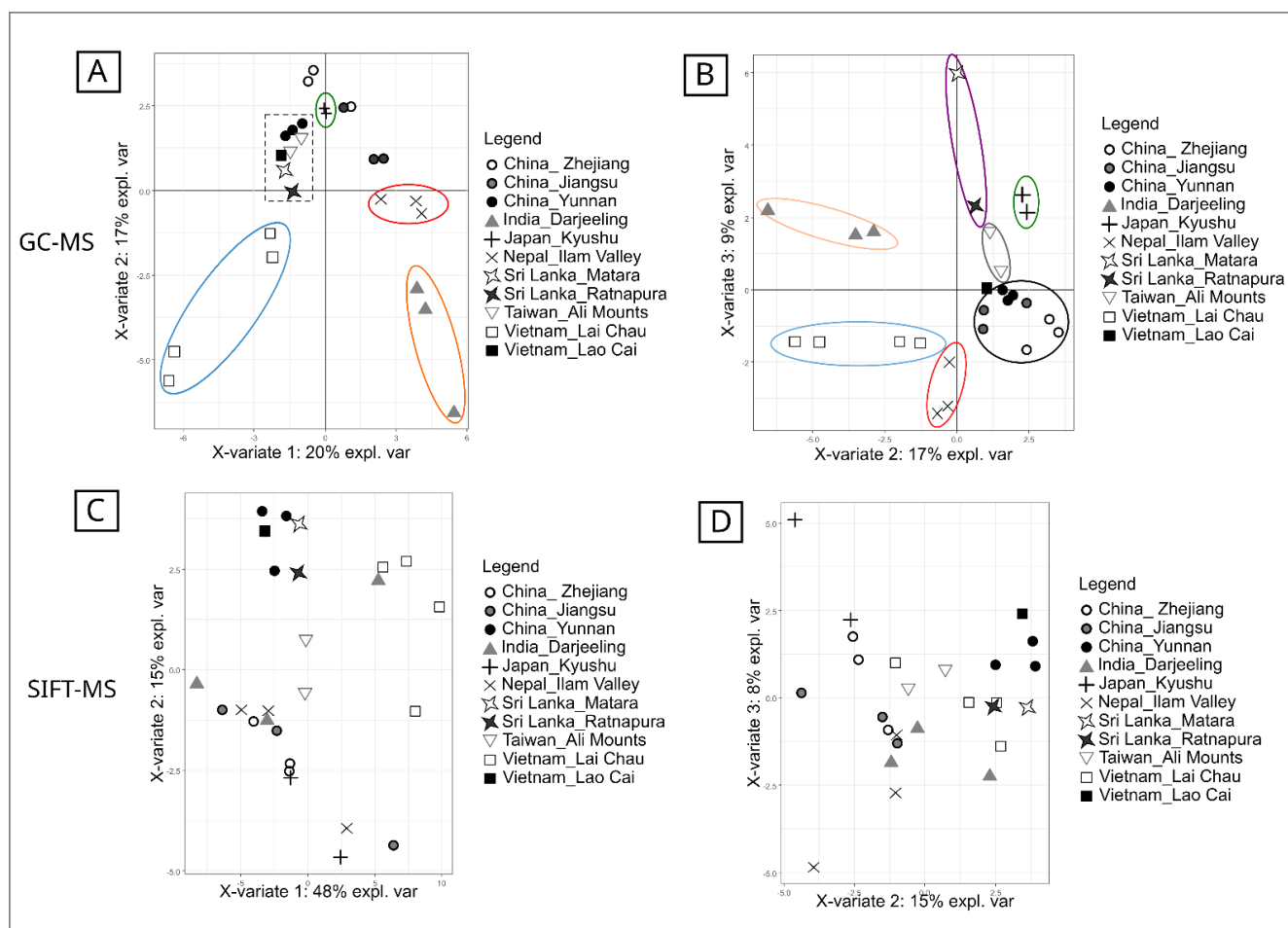


Figure 1 : Individual plot of sparse PLS-DA for country discrimination. A) PC1 and PC2 plot for GC-FID data. B) PC2 and PC3 plot for GC-FID data. C) PC1 and PC2 plot for SIFT-MS data. D) PC1 and PC2 plot for SIFT-MS data.

4 different classes of samples are clearly separated: Japan_Kyushu (circled in green), Nepal_IllamValley (circled in red), Vietnam_Lai Chau (circled in blue) and India_Darjeeling (circled in orange). A group of 8 samples from different origins (China_Yunnan, Taiwan_Ali Mounts, Sri Lanka_Matara and Sri Lanka_Ratnapura, (framed with dot lines) were not separated and presented strong similarities according to this test. Interestingly, these tea samples shared a same type of plant cultivar: large leaf *Camelia sinensis*. With the second representation according to the second and the third components of the sPLS-DA (Figure 1 B), three other origins may be separated (Taiwan_Ali Mounts circled in grey, the two regions of Sri Lanka together, circled in purple and the 3 regions of China, circled in black). Only one sample from Vietnam_Lao Cai was not successfully separated from the other samples with this test. These results confirmed that tea samples may be sorted according to their origin with GC-MS peak areas as it was previously demonstrated on green teas by Ye *et al*¹² or on black teas by Yun *et al* 2021³⁵. However, this study showed that this classification may be also achieved with a collection of tea from different regions and different grade (colors).

With the same approach, the mean values of triplicates of HS-SIFT-MS measurements in SIM mode were also calculated for every samples. A sPLS-DA was performed on the dataset but no clear separation according the region origin was highlight neither with components 1 and 2, nor with components 2 and 3 (Figure 1 C and D) of the test. On the contrary, the different samples were spread over the representation indicating a high level of heterogeneity. This result may be explained by the weak number of ions that have been used for this statistical test. Indeed, the conflict ions inherent to DIMS technology, greatly limit the exploitation of these instruments for the quantification of markers on complex matrices. However, Yener *et al* succeeded to find statistical correlations of VOC profiles of green and black teas according to their origin with a PTR-ToF-MS³⁷. The authors selected 257 mass peaks before applying statistical tools to

trace the origin of green teas and black teas. With the high-resolution instrument, the authors also succeeded to tentatively identify some chemical compounds (61) corresponding to the most discriminant ions.

Targeted transformation processes discrimination

The same statistical analysis on mean values of markers signals with HS-SPME-GC-MS/FID and HS-SIFT-MS were also performed to look for correlations with the descriptive value "color", reflecting the different transformation processes of tea. The results presented in the Figure 2 show a high dispersity of the samples with same color. With separate technique (figure 2A), the 3 oolong teas, the 5 greens teas and 10 black teas tend to separate from each other. Unfortunately, dark and white tea samples were placed in a median position and make the segregation of the 5 grades difficult. No more explicative information was provided with the third component of the sPLS-DA test. Several studies already tried to characterize the transformation process of tea with markers. Thus, Alasalvar *et al*³⁸ investigated the characterization of different quality grades of black teas from Turkey with several techniques of which HS-GC-MS. They concluded on the importance of the concentration levels of some of the 57 volatiles compounds, without a clear distinction between the 7 grades. Lin *et al*¹¹ analyzed 26 VOCs emitted by different varieties of oolong teas with HS-SPME-GC-MS and obtained good separations with a limit in the miss classification of cheaper products. Zhang *et al*¹² analyzed several grades of tea (green, oolong and black) with two dimensional GC-ToF-MS after simultaneous distillation extraction and well succeed to separate sample colors thanks to 3 000 detected peaks of which 450 were tentatively identified. Finally, 74 differential compounds allow the discrimination between the three grades.

This work with markers was done with the 50 identified molecules. These molecules were chosen, on the basis of the most concentrated and better separated by GC, allowing a good match score with the NIST database.

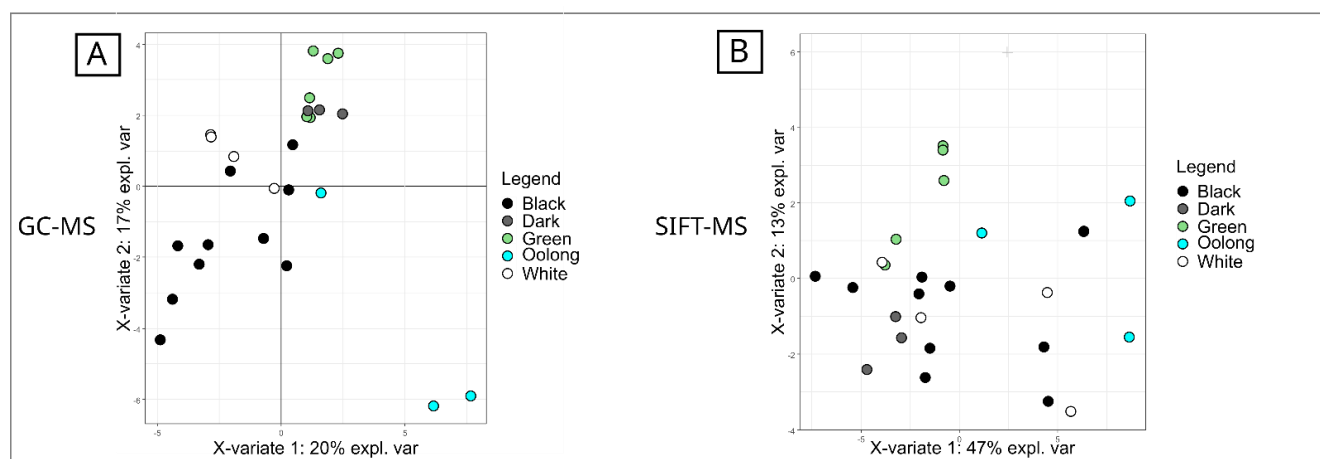


Figure 2 : Individual plot of sparse PLS-DA for color discrimination. A) PC1 and PC2 plot for GC-FID data. B) PC1 and PC2 plot for SIFT-MS data.

Contrary to Whang et al, we didn't succeed to sort the different samples' colors. Nevertheless, the choice of markers was oriented by the nature of the SPME fiber and the phase of the GC column selected. These choices were done according to the literature and the aim to detect the widest range of volatile chemical compounds. Was it the best choice to discriminate the origins or colors of our samples? Clearly not. With GC-MS and even more with SIFT-MS SIM mode, the number and the choice of key markers are essential. The emerging approach of untargeted analysis starts on the hypothesis that the key molecules are unknown and not necessary the most concentrated.

Untargeted analysis of volatile profiles of tea

The global signals of HS-SPME-GC-MS and HS-SIFT-MS were analyzed to investigate the contribution of the volatile profiles to discriminate tea samples according to geographical area of production or color. The post treatment of the entire raw Total Ion Chromatogram (TIC) trace of the chromatograms with Chromcompare+ included an alignment protocol, a filtering step and a feature discovering according to the region or the color parameter. The 400 features were then further explored to highlight correlations. For the SIFT-MS full scan measurements, the entire signals for the product ions from 15 to 250 m/z ratio for the 8 different precursors (1 880 ions) were analyzed to discriminate the two descriptive parameters (country_region and color).

Untargeted origin discrimination

The sPLS-DA analysis of volatile profiles obtained by GC-MS allow a good discrimination of the geographical areas (Figure 3A), especially for teas from Taiwan_Ali Mounts (circled in grey), Vietnam_Lai Chau (circled in red) and Japan_Kyushu (circled in green). The other samples were not separated by regions but 3 separated clusters can be drawn. The first (square 1) contained the samples from China_Yunnan, Vietnam_Lao Cai and the two provinces of Sri Lanka. The proximity of these samples cannot be explained with geographical proximity nor their proximity with the nearby group Vietnam_Lai Chau. However, the knowledge of the products allowed to propose another reason that could explain these proximities. All these teas were obtained from large leaf cultivars of *Camelia sinensis*, meaning that the cultivar may have a great impact on the volatile profile of tea. This cluster was nevertheless successfully separated according to the third component (figure 3 B). The two Chinese provinces Jiangsu and Zhejiang located in the east of China also presented a high statistical proximity of their volatile profiles. The geographical proximity may explain this observation and moreover, the cultivars used in both provinces are small leaf teas (square 2). Finally, the teas from India_Darjeeling have very similar volatile fingerprints to those from Nepal_Ilam Valley and formed a third cluster (square 3). Indeed, these two regions are geographically very close but the cultivars used for these teas are also very similar. In addition, the cultivars used

for these two regions are mixtures of small leaf teas like those used for China_Jiangsu and China_Zhejiang, so it is consistent to find a proximity between these tea samples. The untargeted analysis of HS-SPME-GC-MS allowed to trace some of tea origins as for targeted analysis with 50 VOCs but this approach also highlighted the statistical proximity of GC-MS fingerprints with the type of cultivars.

The results of the sparse PLS-DA analysis from the volatile profiles obtained by SIFT-MS (Figure 3C and D) also showed a good discrimination of tea samples from Japan_Kyushu and Vietnam_Lai Chau as previously shown with untargeted GC-MS. The other discriminated group were China_Zhejiang, China_Jiangsu and China_Yunnan with the first and the second component of the test. Interestingly, with the representation of components 2 and 3 of the test, the large leaf *Camelia sinensis* samples grouped together in a cluster and the samples from Taiwan_Ali Mounts were separated. However, no clear separation was observed for India_Darjeeling samples. Thus, the proximity of small leaf *Camelia sinensis* samples, highlighted with GC-MS was not clear according to the 3 first components of the sPLS-DA test on SIFT-MS fingerprints.

Untargeted analysis according to transformation processes.

Contrary to targeted analysis with the 50 VOCs, the untargeted data post treatment of GC-MS and SIFT-MS full scan analysis allowed a good discrimination of tea colors with the component 1, 2 and 3 of the sPLS-DA (Figure 4). Thus, Oolong and dark teas were clearly separated with the component 1 and 2 of the GC-MS analysis whereas the 3 other grades were grouped with the components 2 and 3 (figure 4B). With SIFT-MS analysis, the most different grade of tea was the green (figure 4C). The 4 other grades were separated with the 2 and 3 components of the sPLS-DA (figure 4D). The color of tea is related to the processing of the tea leaves after harvesting³⁹. The different stages of processing will lead to different levels of oxidation, fermentation and roasting. Oolong teas are the most oxidized, dark teas are fermented, green teas are those that undergo the least amount of processing, they are neither oxidized nor fermented. Black teas are widely produced with very heterogeneous processes depending on the cultures leading to heterogeneous levels of roasting. These transformation processes of tea, that modified their chemical composition, were clearly visible with untargeted analysis either with HS-SPME-GC-MS nor SIFT-MS.

The aim of this study was to compare the performances of two VOCs analytical techniques (HS-SPME-GC-MS/FID and HS-SIFT-MS) using a marker monitoring method and a volatile profile method to discriminate tea samples with descriptive variables, e.g. geographical area of production, or transformation processes.

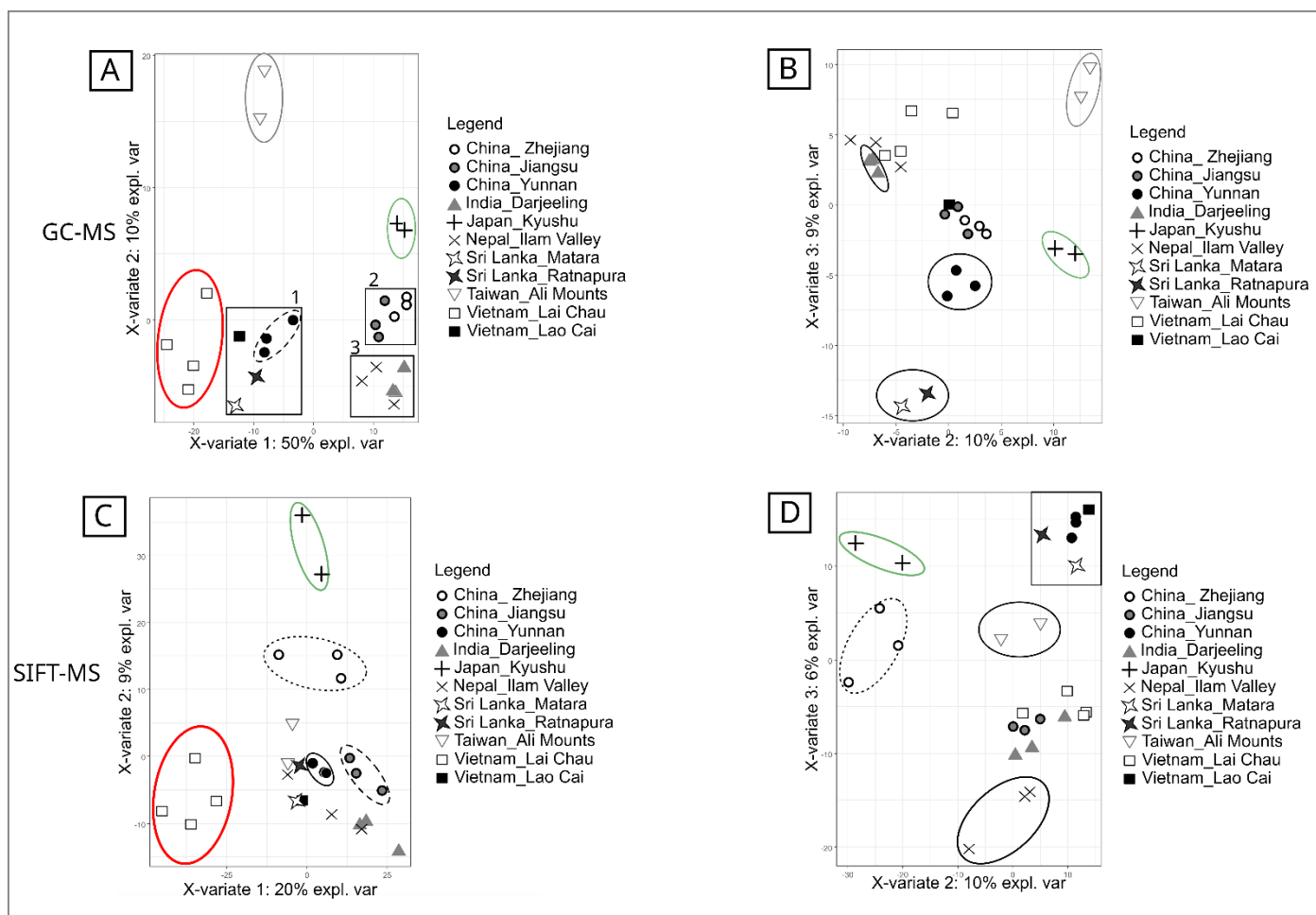


Figure 3 : Individual plot of sparse PLS-DA for country discrimination. A) PC1 and PC2 plot for volatile profile obtained by GC-MS/FID. B) PC2 and PC3 plot for volatile profile obtained by GC-MS/FID. C) PC1 and PC2 plot for volatile profile obtained by SIFT-MS. D) PC2 and PC3 plot for volatile profile obtained by SIFT-MS.

First, our results show that markers monitoring with HS-SPME-GC-MS/FID can discriminate the geographical area of production of our sample set with highlighting the impact of the cultivars but does not allowed the discrimination according the color of tea. But, for this method, the selection of the key compounds was done once the chromatograms were already obtained, with adapted protocols of HS-SPME extraction and GC separation of the sample database, thus influencing the choice of compound selection. It seems that the 50 selected compounds were more discriminating of geographical area and cultivar than of the process (color). Second, with SIFT-MS, no discrimination was obtained, with a SIM method established with the same list of molecules, either with origin nor color. Nevertheless, a lot of conflict ions were found inside the quantification Labsyft® method and then, reduced the number of analyzed signals (Table1 SI). Finally, only 28 of the 50 products were analyzed with only 48 non-conflict product ions of the 138 possible ones. SIFT-MS instrument has some well-known drawbacks due to the ion-molecule reaction in soft chemical ionization⁴⁰. Isomeric compounds will form same product ions with the same m/z signal, which will prevent their quantification.^{41,42} Knowing the molecules searched, a withdrawal of the conflict ions for the follow-up can be envisaged. On the other hand, in the case of complex matrices such as foodstuffs, unknown molecules but present in the matrix can generate conflict ions without

having knowledge of it. This limits the SIFT-MS monitoring for quantification in complex matrices.

Another way to use the data of VOCs measurements was the untargeted analysis. We applied post-treatments on HS-SPME-GC-MS and HS-SIFT-MS full scan data to look for correlations. We demonstrated that untargeted analysis of GC-MS measurements allowed to distinguish 5 geographical origins, and grouped several samples that share same type of cultivars. This suggested that the 400 most discriminant features selected by the software Chromcompare+, do not correspond to the total information given by the 50 selected compounds with the targeted approach. Indeed, among these 400 features, only 24 molecules of the list were found. The identification of the other molecules was difficult because of a weak correspondence with NIST Database. This observation may guide the scientists on the great importance of the markers used for the discrimination. With SIFT-MS, the untargeted analysis permitted to find correlations that were not observed with the SIM method. It seems obvious with these results, that the quantification of markers without separation, results in a loss of performance and that a profile comparison is relevant. 5 different tea origins were separated with a sPLS-DA analysis of full scan data. The impact of large leaf tea cultivar was also highlighted by this methodology. As with untargeted GC-MS, it is possible to evaluate the transformation process of tea with untargeted SIFT-MS.

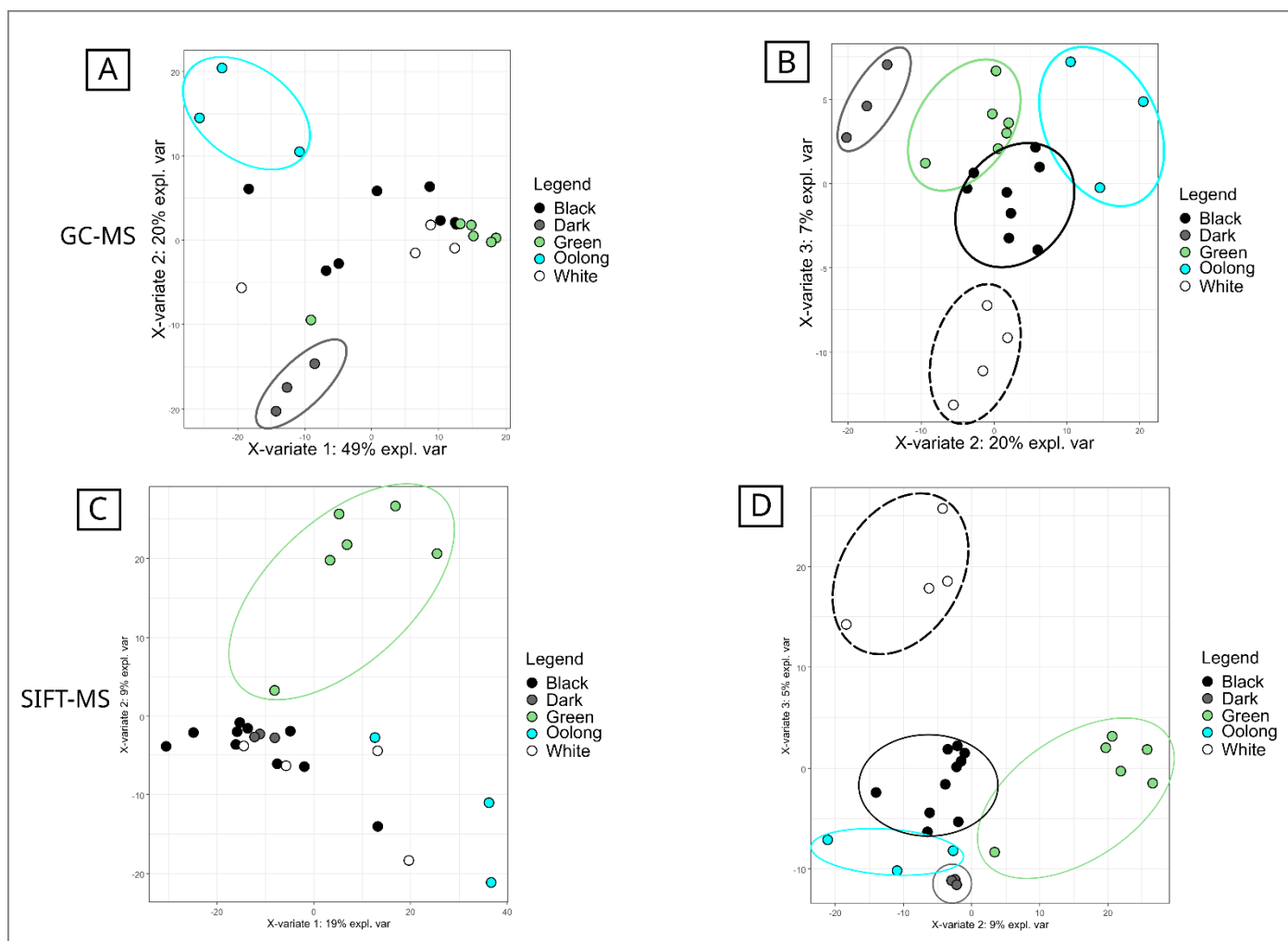


Figure 4: Individual plot of sparse PLS-DA for color discrimination. A) PC1 and PC2 plot for volatile profile obtained by GC-MS/FID. B) PC1 and PC2 plot for volatile profile obtained by SIFT-MS. C) PC2 and PC3 plot for volatile profile obtained by GC-MS/FID. D) PC1 and PC2 plot for volatile profile obtained by SIFT-MS data.

Both analytical methods are relevant for the analysis of VOCs in food matrices. However, they each have their advantages and drawbacks. The comparison of these two methods on the four following parameters was summarized in Table 3.

Speed: Time is an important factor in food industry where a lot of samples have to be classified. HS-SPME-GC-MS/FID analysis doesn't require a difficult preparation step and can be easily automatized. However, it takes several tens of minutes for one sample. HS-SIFT-MS analysis can also be easily automatized and doesn't require long preparation, but a full scan record takes 3 to 6 minutes depending on the number of precursor ions. This high-speed instrument allows kinetic studies with high throughput.

Selectivity: HS-SPME-GC-MS/FID bases its selectivity on the composition of the SPME fiber and of the GC column that drive the detected chemical compounds. Separation protocols (time, temperature...) are generally adapted on a selection of molecules that will also drive the resolution. The mass spectrometry detection allows the identification with databases (NIST or other) whereas the FID detection allows a simple and robust quantification. SIFT-MS presents a chemical selectivity with precursor ions. Depending on the ions, some chemical families will be addressed by this instrument. However, the absence of separation leads to the risk of conflict ions, coming from several analytes giving the same m/z ratio ions. With complex matrices, these conflict ions are inevitable and require the elimination of these signals impossible to explain. This data filtering step limits the number of molecules that can be

measured in the same time. The unitary mass resolution of the instrument, the soft chemical ionizations and these ions conflicts limit the identification of unknown molecules in the matrix.

Sensitivity: SPME is a sampling technique that increase a lot the limit of detection of one molecule by increasing the time and the temperature of the fiber exposure. However, with complex matrixes where a lot of different molecules with different concentration ranges are mixed, SPME may be saturated by the most concentrated molecule and then, not detect the less concentrated ones. Nevertheless, the global sensitivity of this technique is in the range of ppbV for most of VOC⁴³. The sensitivity of SIFT-MS is driven by the chemical reactivity inside the flow tube and the kinetic constant k given in the software database. Only the 400 molecules already recorded in the database could be directly quantified with SIFT-MS. If the molecule of interest is unknown, the database may be updated after calibrations. Most of the VOC can be detected at a range of ten ppbV with SIFT-MS^{44,45}. **Fingerprints:** both technologies may be exploited with untargeted strategies. For GC-MS, several software solutions are available on the market. To our knowledge, only lab-made software solutions are available for SIFT-MS full scan treatment.

CONCLUSION

This preliminary work shows the potential of the profile/fingerprint VOC analysis, to discriminate origin and

transformation processing of tea samples. First, it clearly appears that the choice of markers

	SPME-GC-FID-MS	SIFT-MS
Analysis time	- (35 min)	++ (3 min)
Physico-chemical selectivity	Depends on SPME and GC column phases	Depends on reactivity of precursor ions
Compound identification	++	+/- (unit mass resolution)
Quantification in complex matrices	++ (possible with internal or external calibration)	- (conflict ions)
Volatile profile discrimination	++	++

Table 3: Comparison of the two techniques (GC-MS-FID and SIFT-MS)

drives the performances of the targeted GC-MS analysis. In this study, the 50 chosen molecules allowed the origin discrimination but not the transformation process.

Second, the quantification of a long list of markers with SIFT-MS in SIM mode, leads to a high number of conflict ions and then limits the resolution of the discrimination. Third, the post-treatment of raw data GC-MS highlights new correlations with the transformation process that was not explained with markers. The choice of markers or features (in the case of Chromcompare+ software) was oriented by the composition of the sample set. The untargeted analysis of SIFT-MS full scan data, make it possible to analyze complex matrices with statistical profile comparisons. These analyses based on volatile profiles could be techniques of choice for discriminating products with a label such as a Protected Indication of Origin (PGI), a Traditional Speciality Guaranteed (TSG) or a red label defining a superior quality. In the current study, a limited number of samples were used but these results are promising for further study on a larger number of samples. For example, by focusing on the construction of a sample bank of black teas, the discrimination of geographical origins would be probably more easily achieved. However, the knowledge of the sample is also a key parameter that allowed us to understand the clustering according to cultivar proximity. Finally, learning models could be envisaged in order to highlight fraud by comparing volatile profiles.

1) Majumder, A. B.; Bera, B.; Rajan, A. Tea Statistics: Global Scenario. *2012*, 4.

(2) Feng, Z.; Li, Y.; Li, M.; Wang, Y.; Zhang, L.; Wan, X.; Yang, X. Tea Aroma Formation from Six Model Manufacturing Processes. *Food Chemistry* **2019**, 285, 347–354. <https://doi.org/10.1016/j.foodchem.2019.01.174>.

(3) Shitandi, Anakalo. A.; Muigai Nguere, F.; Mahungu, S. M. Tea Processing and Its Impact on Catechins, Theaflavin and Thearubigin Formation. In *Tea in Health and Disease Prevention*; Elsevier, 2013; pp 193–205. <https://doi.org/10.1016/B978-0-12-384937-3.00016-1>.

(4) Kennedy, S. P. Chapter 8. Coffee and Tea Fraud. *Food Fraud* 12.

(5) Luykx, D. M. A. M.; van Ruth, S. M. An Overview of Analytical Methods for Determining the Geographical Origin of Food Products. *Food Chemistry* **2008**, 107 (2), 897–911. <https://doi.org/10.1016/j.foodchem.2007.09.038>.

(6) Pennerman, K. K.; AL-Maliki, H. S.; Lee, S.; Bennett, J. W. Fungal Volatile Organic Compounds (VOCs) and the Genus *Aspergillus*. In *New and Future Developments in Microbial Biotechnology and Bioengineering*; Elsevier, 2016; pp 95–115. <https://doi.org/10.1016/B978-0-444-63505-1.00007-5>.

(7) Wang, Z.; Han, B.; Jing, W.; Yi, Z.; Zhang, Y.; Ren, D.; Yi, L. Effects of Different Steeping Temperatures on the Leaching of Aroma Components in Black Tea by SPME–GC–MS Coupled with Chemometric Method. *J AOAC Int* **2019**, 102 (6), 1834–1844. <https://doi.org/10.5740/jaoacint.18-0405>.

(8) Guo, X.; Schwab, W.; Ho, C.-T.; Song, C.; Wan, X. Characterization of the Aroma Profiles of Oolong Tea Made from Three Tea Cultivars by Both GC–MS and GC-IMS. *Food Chemistry* **2022**, 376, 131933. <https://doi.org/10.1016/j.foodchem.2021.131933>.

(9) Lin, J.; Zhang, P.; Pan, Z.; Xu, H.; Luo, Y.; Wang, X. Discrimination of Oolong Tea (*Camellia Sinensis*) Varieties Based on Feature Extraction and Selection from Aromatic Profiles Analysed by HS-SPME/ GC–MS. *Food Chemistry* **2013**, 7.

(10) Zhang, W. HS-SPME and GC/MS Volatile Component Analysis of Yinghong No. 9 Dark Tea during the Pile Fermentation Process. *Food Chemistry* **2021**, 10.

(11) Wu, H. GC–MS-Based Metabolomic Study Reveals Dynamic Changes of Chemical Compositions during Black Tea Processing. *Food Research International* **2019**, 9.

(12) Ye, N.; Zhang, L.; Gu, X. Discrimination of Green Teas from Different Geographical Origins by Using HS-SPME/GC–MS and Pattern Recognition Methods. *Food Anal. Methods* **2012**, 5 (4), 856–860. <https://doi.org/10.1007/s12161-011-9319-9>.

(13) Biasioli, F.; Yeretzian, C.; Märk, T. D.; Dewulf, J.; Van Langenhove, H. Direct-Injection Mass Spectrometry Adds the Time Dimension to (B)VOC Analysis. *TrAC Trends in Analytical Chemistry* **2011**, 30 (7), 1003–1017. <https://doi.org/10.1016/j.trac.2011.04.005>.

(14) Deuscher, Z.; Andriot, I.; Sémon, E.; Repoux, M.; Preys, S.; Roger, J.-M.; Boulanger, R.; Labouré, H.; Le Quéré, J.-L. Volatile Compounds Profiling by Using Proton Transfer Reaction-Time of Flight-Mass Spectrometry (PTR-ToF-MS). The Case Study of Dark Chocolates Organoleptic Differences. *J Mass Spectrom* **2019**, 54 (1), 92–119. <https://doi.org/10.1002/jms.4317>.

(15) Smith, D.; Španěl, P. Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) for on-Line Trace Gas Analysis. *Mass Spectrometry Reviews* **2005**, 24 (5), 661–700. <https://doi.org/10.1002/mas.20033>.

(16) Castada, H. Z.; Wick, C.; Taylor, K.; Harper, W. J. Analysis of Selected Volatile Organic Compounds in Split and Nonsplit Swiss Cheese Samples Using Selected-Ion Flow Tube Mass Spectrometry (SIFT-MS): Analysis of Selected Volatile Organic.... *Journal of Food Science* **2014**, 79 (4), C489–C498. <https://doi.org/10.1111/1750-3841.12418>.

(17) Bajbou, A.; Medina-Rodríguez, S.; Ajal, E. A.; Cuadros-Rodríguez, L.; Monasterio, R. P.; Vercammen, J.; Fernández-Gutiérrez, A.; Carrasco-Pancorbo, A. A Metabolic Fingerprinting Approach Based on Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) and Chemometrics: A Reliable Tool for Mediterranean Origin-Labeled Olive Oils Authentication. *Food Research International* **2018**, 106, 233–242. <https://doi.org/10.1016/j.foodres.2017.12.027>.

(18) Ozcan-Sinir, G. Detection of Adulteration in Extra Virgin Olive Oil by Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) and Chemometrics. *Food Control* **2020**, 118, 107433. <https://doi.org/10.1016/j.foodcont.2020.107433>.

(19) Kharbach, M.; Kamal, R.; Mansouri, M. A.; Marmouzi, I.; Viaene, J.; Cherrah, Y.; Alaoui, K.; Vercammen, J.; Bouklouze, A.; Vander Heyden, Y. Selected-Ion Flow-Tube Mass-Spectrometry (SIFT-MS) Fingerprinting versus Chemical Profiling for Geographic Traceability of Moroccan Argan Oils. *Food Chemistry* **2018**, 263, 8–17. <https://doi.org/10.1016/j.foodchem.2018.04.059>.

(20) Kharbach, M.; Yu, H.; Kamal, R.; Marmouzi, I.; Alaoui, K.; Vercammen, J.; Bouklouze, A.; Vander Heyden, Y. Authentication of Extra Virgin Argan Oil by Selected-Ion Flow-Tube Mass-Spectrometry Fingerprinting and Chemometrics. *Food Chemistry* **2022**, 383, 132565. <https://doi.org/10.1016/j.foodchem.2022.132565>.

(21) Baerenzung dit Baron, T.; Yobrégat, O.; Jacques, A.; Simon, V.; Geffroy, O. A Novel Approach to Discriminate the Volatilome of *Vitis Vinifera*

- Berries by Selected Ion Flow Tube Mass Spectrometry Analysis and Chemometrics. *Food Research International* **2022**, 111434. <https://doi.org/10.1016/j.foodres.2022.111434>.
- (22) Taylor, K.; Wick, C.; Castada, H.; Kent, K.; Harper, W. J. Discrimination of Swiss Cheese from 5 Different Factories by High Impact Volatile Organic Compound Profiles Determined by Odor Activity Value Using Selected Ion Flow Tube Mass Spectrometry and Odor Threshold: Swiss Cheese Factory VOC Variability.... *Journal of Food Science* **2013**, *78* (10), C1509–C1515. <https://doi.org/10.1111/1750-3841.12249>.
- (23) Castada, H.; Hanas, K.; Barringer, S. Swiss Cheese Flavor Variability Based on Correlations of Volatile Flavor Compounds, Descriptive Sensory Attributes, and Consumer Preference. *Foods* **2019**, *8* (2), 78. <https://doi.org/10.3390/foods8020078>.
- (24) Reyrolle, M.; Ghislain, M.; Bru, N.; Vallverdu, G.; Pigot, T.; Desauziers, V.; Le Behec, M. Volatile Fingerprint of Food Products with Untargeted SIFT-MS Data Coupled with MixOmics Methods for Profile Discrimination: Application Case on Cheese. *Food Chemistry* **2022**, *369*, 130801. <https://doi.org/10.1016/j.foodchem.2021.130801>.
- (25) Olivares, A.; Dryahina, K.; Navarro, J. L.; Smith, D.; Španěl, P.; Flores, M. SPME-GC-MS versus Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) Analyses for the Study of Volatile Compound Generation and Oxidation Status during Dry Fermented Sausage Processing. *J. Agric. Food Chem.* **2011**, *59* (5), 1931–1938. <https://doi.org/10.1021/jf104281a>.
- (26) Olivares, A.; Dryahina, K.; Španěl, P.; Flores, M. Rapid Detection of Lipid Oxidation in Beef Muscle Packed under Modified Atmosphere by Measuring Volatile Organic Compounds Using SIFT-MS. *Food Chemistry* **2012**, *135* (3), 1801–1808. <https://doi.org/10.1016/j.foodchem.2012.06.075>.
- (27) Cauchi, M.; Fowler, D. P.; Walton, C.; Turner, C.; Waring, R. H.; Ramsden, D. B.; Hunter, J. O.; Teale, P.; Cole, J. A.; Bessant, C. Comparison of GC-MS, HPLC-MS and SIFT-MS in Conjunction with Multivariate Classification for the Diagnosis of Crohn's Disease in Urine. *Anal. Methods* **2015**, *7* (19), 8379–8385. <https://doi.org/10.1039/C5AY01322D>.
- (28) Hastie, C.; Thompson, A.; Perkins, M.; Langford, V. S.; Eddleston, M.; Homer, N. ZM. Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS) as an Alternative to Gas Chromatography/Mass Spectrometry (GC/MS) for the Analysis of Cyclohexanone and Cyclohexanol in Plasma. *ACS Omega* **2021**, *6* (48), 32818–32822. <https://doi.org/10.1021/acsomega.1c03827>.
- (29) Hera, D.; Langford, V.; McEwan, M.; McKellar, T.; Milligan, D. Negative Reagent Ions for Real Time Detection Using SIFT-MS. *Environments* **2017**, *4* (1), 16. <https://doi.org/10.3390/environments4010016>.
- (30) Ghislain, M.; Costarramone, N.; Pigot, T.; Reyrolle, M.; Lacombe, S.; Le Behec, M. High Frequency Air Monitoring by Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS): Influence of the Matrix for Simultaneous Analysis of VOCs, CO₂, Ozone and Water. *Microchemical Journal* **2020**, *153*, 104435. <https://doi.org/10.1016/j.microc.2019.104435>.
- (31) Spadafora, N. D.; Mascrez, S.; McGregor, L.; Purcaro, G. Exploring Multiple-Cumulative Trapping Solid-Phase Microextraction Coupled to Gas Chromatography–Mass Spectrometry for Quality and Authenticity Assessment of Olive Oil. *Food Chemistry* **2022**, *383*, 132438. <https://doi.org/10.1016/j.foodchem.2022.132438>.
- (32) R. Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. www.r-project.org **2018**.
- (33) Wei, T.; Simko, V.; Levy, M.; Xie, Y.; Jin, Y.; Zemla, J. *Corrplot: Visualization of a Correlation Matrix*; 2017.
- (34) Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. MixOmics: An R Package for 'omics Feature Selection and Multiple Data Integration. *PLoS Comput Biol* **2017**, *13* (11), e1005752. <https://doi.org/10.1371/journal.pcbi.1005752>.
- (35) Yun, J.; Cui, C.; Zhang, S.; Zhu, J.; Peng, C.; Cai, H.; Yang, X.; Hou, R. Use of Headspace GC/MS Combined with Chemometric Analysis to Identify the Geographic Origins of Black Tea. *Food Chemistry* **2021**, *360*, 130033. <https://doi.org/10.1016/j.foodchem.2021.130033>.
- (36) Yener, S.; Sánchez-López, J. A.; Granitto, P. M.; Cappellin, L.; Märk, T. D.; Zimmermann, R.; Bonn, G. K.; Yeretizian, C.; Biasioli, F. Rapid and Direct Volatile Compound Profiling of Black and Green Teas (*Camellia Sinensis*) from Different Countries with PTR-ToF-MS. *Talanta* **2016**, *152*, 45–53. <https://doi.org/10.1016/j.talanta.2016.01.050>.
- (37) Alasalvar, C.; Topal, B.; Serpen, A.; Bahar, B.; Pelvan, E.; Gökmen, V. Flavor Characteristics of Seven Grades of Black Tea Produced in Turkey. *J. Agric. Food Chem.* **2012**, *60* (25), 6323–6332. <https://doi.org/10.1021/jf301498p>.
- (38) Yun, J. Use of Headspace GC/MS Combined with Chemometric Analysis to Identify the Geographic Origins of Black Tea. *Food Chemistry* **2021**, *9*.
- (39) Feng, Z.; Li, Y.; Li, M.; Wang, Y.; Zhang, L.; Wan, X.; Yang, X. Tea Aroma Formation from Six Model Manufacturing Processes. *Food Chemistry* **2019**, *285*, 347–354. <https://doi.org/10.1016/j.foodchem.2019.01.174>.
- (40) Smith, D.; McEwan, M. J.; Španěl, P. Understanding Gas Phase Ion Chemistry Is the Key to Reliable Selected Ion Flow Tube-Mass Spectrometry Analyses. *Anal. Chem.* **2020**, *92* (19), 12750–12762. <https://doi.org/10.1021/acs.analchem.0c03050>.
- (41) Pysanenko, A.; Spánel, P.; Smith, D. A Study of Sulfur-Containing Compounds in Mouth- and Nose-Exhaled Breath and in the Oral Cavity Using Selected Ion Flow Tube Mass Spectrometry. *Journal of Breath Research* **2008**, *2*. <https://doi.org/10.1088/1752-7155/2/4/046004>.
- (42) Smith, D.; Sovová, K.; Španěl, P. A Selected Ion Flow Tube Study of the Reactions of H₃O⁺, NO⁺ and O₂⁺ with Seven Isomers of Hexanol in Support of SIFT-MS. *International Journal of Mass Spectrometry* **2012**, *319–320*, 25–30. <https://doi.org/10.1016/j.ijms.2012.03.009>.

CHAPITRE 5 : Authenticité du thé par signatures multi-élémentaires, rapport isotopique et empreintes volatiles

Les résultats présentés dans ce chapitre font l'objet d'une publication soumise au journal « *Food Chemistry* ».

Dans ce travail, la complémentarité de l'analyse inorganique et de l'analyse organique a été étudiée pour tracer l'authenticité des 26 échantillons de thé étudiés au chapitre 4. Pour rappel, ces thés sont caractérisés par 5 couleurs, des origines certifiées et proviennent de 7 pays différents dont 11 régions. Chaque thé a été analysé en triplicat par SIFT-MS pour la détermination de son empreinte volatile et par ICP-MS (Spectrométrie de masse par plasma à couplage inductif) pour la détermination de sa signature multi-élémentaire et la mesure du rapport isotopique Strontium (Sr) par MC-ICP-MS (Spectrométrie de masse à plasma à couplage inductif et à collecteur multiple). Des méthodes statistiques multivariées non supervisées et supervisées ont été utilisées pour explorer les données obtenues.

La détermination du rapport $^{87}\text{Sr}/^{86}\text{Sr}$ dans les échantillons de thé a permis de distinguer 6 des 11 sites de collecte, confirmant le pouvoir de cette technique pour déduire l'origine géographique. Les rapports mesurés sont cohérents avec ceux fournis par les cartes géologiques récemment publiées. Ces résultats renforcent le lien entre les plantes et le rapport $^{87}\text{Sr}/^{86}\text{Sr}$ biodisponible dans le sol où elles sont cultivées. La signature multi-élémentaire, quant à elle, a permis de discriminer 10 des 11 sites en mettant en évidence des éléments permettant de les différencier. La signature élémentaire vient renforcer la discrimination géographique déterminée par le rapport isotopique. Par exemple, les deux provinces situées à l'est de la Chine ne sont pas discriminées par leur rapport isotopique. En revanche, un site présente une concentration élevée en certains éléments associés à la présence de 12 mines dans cette région. Ainsi, les signatures élémentaires sont de bons indicateurs reflétant des informations sur le climat, l'environnement et les pratiques agricoles. Les analyses inorganiques apportent des résultats précis et justes grâce à l'utilisation de matériaux références permettant la comparaison de résultats. En revanche, ces méthodes sont relativement longues, utilisent des solvants organiques et dégradent l'échantillon.

L'analyse non supervisée réalisée sur les empreintes volatiles des différents thés a montré que la composition en COV était plus influencée par le cultivar et le terroir que par le pays. Le terroir n'est pas seulement défini par la zone géographique mais également par les personnes, les pratiques agricoles et de transformation. Une analyse statistique supervisée (sPLS-DA) a été réalisée pour évaluer la contribution du procédé de transformation sur les empreintes. Les résultats ont montré une nette discrimination des thés verts et Oolong qui est expliquée par leurs niveaux d'oxydation. Le thé vert correspond au thé le moins transformé et donc le moins oxydé alors que le thé cyan est un thé oxydé issu de nombreuses étapes de transformation. Le thé sombre (« Pu Erh ») est également différencié des autres couleurs de thés de par la fermentation qu'il subit avant maturation. En revanche le thé noir et le thé blanc n'ont pas été discriminés, potentiellement à cause de la faible quantité d'échantillon et la mixité des cultivars et terroirs. L'analyse chimiométrique des empreintes volatiles a permis de mettre en évidence deux molécules, l'hexanol et l'hexanal, surexprimées dans les thés Oolong et pouvant être responsable de la discrimination de ces thés. Cependant, la problématique des conflits d'ions intrinsèques à la technique SIFT-MS et la résolution unitaire en masse de l'instrument empêchent d'aller plus loin dans l'identification des molécules discriminantes. Cette méthode présente tout de même des avantages comme la simplicité de préparation des échantillons sans dégradation physique et la rapidité d'analyse.

Ces trois méthodes analytiques différentes, la signature multi-élémentaire, le rapport isotopique du strontium et les empreintes volatiles sont des outils très intéressants pour assurer la traçabilité et l'authenticité du thé tout au long de sa chaîne de production, de la culture au produit final. La signature isotopique est influencée par la scène géologique, l'impact anthropique, agricole et environnemental. Le rapport $^{87}\text{Sr}/^{86}\text{Sr}$ peut certifier le lieu de production, grâce à sa stabilité et à son niveau connu dans le sol. Les empreintes volatiles permettent de répondre à des questions d'authenticité et de traçabilité

en différenciant le « terroir », le cultivar et les procédés de transformation. Pour vérifier l'hypothèse concernant la différence des profils aromatiques du thé en lien avec le cultivar utilisé, des analyses phylogénétiques ultérieures seront réalisées sur cette collection de thé. Cette étude montre le fort potentiel de la combinaison de plusieurs analyses instrumentales couplées à des outils chimiométriques pour améliorer les systèmes de traçabilité pour faire face aux pratiques frauduleuses dans le secteur alimentaire et réguler la qualité et la sécurité des produits tout au long de la chaîne d'approvisionnement.

Authenticity of tea by multi-elemental signatures, isotopic ratio and volatile compounds profiling

Marine Reyrolle¹, Gilles Bareille¹, Ekaterina N. Epova², Julien Barre², Sylvain Bérail², Thierry Pigot¹, Valerie Desauziers¹, Lydia Gautier³ and Mickael Le Behec¹

¹ Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, IMT Mines Ales, IPREM, Pau, France, Institut des sciences analytiques et de Physicochimie pour l'environnement et les Matériaux, UMR5254, Hélioparc, 2 avenue du Président Angot, 64053, Pau cedex 9, France

² Advanced Isotopic Analysis (A.I.A.), Hélioparc, 2 avenue du Président Angot, 64000, Pau, France

³ Tea expert, T Edition, 63 rue Vercingétorix, Paris, France

Abstract

This study aimed to combine multi-elemental analysis, Strontium isotopic ratio and volatile compound profiling with untargeted Selected Ion Flow Tube Mass Spectrometer (SIFT-MS) analysis to discriminate the geoprovenance of 26 different teas and the authenticity of samples. Ten of the twelve regions of the tea collection were significantly identified by their inorganic pattern whereas the type of cultivar (large leaf or small leaf) was highlighted as the main contributing factor of the volatile fingerprints. Supervised analysis of the tea collection also allowed the classification of samples according to the transformation processes leading to different colors of teas. Green teas, oolong teas and dark teas were clearly differentiated by this approach. Product ions corresponding to hexanal and hexanol were identified as contributing to the discrimination of semi-oxidized oolong teas to the other tea grades.

1. Introduction

Food provenance assessment is a global concern for sanitary and quality control. Consumers expect more and more guarantees of food quality i.e., information on composition, absence of health risks, information on product origin, quality and transformation processes with environmental concern. Thus, labels have been developed to meet this expectation, providing information on product origin (Protected Designation of Origin PDO, Protected Geographical Indication PGI), environmental and social impact (AB, European organic, USDA Organic, Fair Trade). Through these labels, delivered by authorities, the authenticity of agri-food products covers both the notion of product origin and an expected sensory quality (Belletti et al., n.d.). Unfortunately, these certified products are *de facto* more expensive and profitable, with the consequence of increasing of fraud and fake products. It is particularly the case for a very old product, consumed all over the world: the tea (Kennedy, n.d.).

For many centuries, people have consumed tea and enjoyed its typical taste and aroma. Originally produced in China, tea is now mainly produced in Asia (India, China, Sri Lanka, Vietnam, Japan), also in Africa (Kenia) and South America (FAO Intergovernmental Group on Tea, 2015). The tea plant "*Camellia Sinensis*" belongs to the family *Theaceae*. The first wild tea plants took root at the source of the Irrawaddy River in a triangle formed by Yunnan in China, northeast India, Burma and Tibet. Tea plants would then have spread naturally

along the rivers, giving three main varieties: *sinensis*, *assamica* and *cambodiensis*. National research institutes worked on new varieties adapted to local climatic conditions (resistant to certain pests, early or mature growth, etc.) and presenting a good balance between productivity and quality (Patel et al., 2019). As a result, a great number of varieties of cultivars were created through cuttings, grafting, layering or suckering. The terroir, i.e. the composition of the soil, the climate, the altitude, the latitude, the plants grown in association with the tea plant and the picking types and periods, play also an essential role in the development of the organoleptic qualities of tea.

Moreover, numerous modifications in the transformation process of *Camellia sinensis* leaves have led to the creation of several tea grades (Feng et al., 2019; Shitandi et al., 2013). Thus, according to the levels of drying, fixation, fermentation or enzymatic oxidation, teas can be classified into different colors : green, , black, dark, red, white and yellow (Zhang et al., 2019). Nevertheless, only few labels have been created as Darjeeling tea, a typical black tea from north of India, which is one of the most famous. Today, only 10,000 tonnes of Darjeeling tea are produced while 4 times as much are sold worldwide. This illustrates the challenge of traceability of high value food products. However, more than the product origin, consumers also want a typicality of the product.

Thus, to protect producers, manufacturers and consumers, labelled products have to be cultivated and transformed in a

defined geographic area. In addition to the declarative aspect and the control the supply chain to obtain a label, several fraud controls may also be conducted to assess the product origin. The development of new analytical techniques for determining the geographical origin and authenticity of products has been increased since the 1980s, such as stable isotope of carbon, nitrogen and oxygen, near infrared spectroscopy, nuclear magnetic resonance, electronic nose, multi-element analysis, chromatographic techniques, mass spectrometry techniques or sensory analysis (Danezis et al., 2016; Luykx & van Ruth, 2008).

Inductively coupled plasma mass spectrometry (ICP-MS) is widely used to obtain multi-element fingerprint for geographical assessment of a variety of plant food products such as olive oil (Beltrán et al., 2015; Nasr et al., 2021), coffee (H.-C. Liu et al., 2014), wine (Epova et al., 2020) and tea (W. Liu et al., 2021). Plants grow with organic and mineral matter from the soil and water. Mineral elements are transferred to the plant during cultivation and give rise to different multi-elemental signatures depending on their origin of production (underlying geology and soil composition) (W. Liu et al., 2021). Anthropogenic actions and environmental pollutions may also contribute to elemental signature of plants (Zhao et al., 2017). Inorganic elements such as Sr, Pb, B, Cd, Li, Na, Mg, Al, K, Ca, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Rb and Ba were found to be good markers of tea leaves before and after processing (Fernández-Cáceres et al., 2001; Gomes et al., 2019). Moreover, post-processing (drying or roasting) and storage methods are more likely to affect the active molecular compounds which may degrade or attenuate (Ni et al., 2018). Among metal elements, strontium (Sr) and more particularly the ratio of two of its four isotopes ($^{87}\text{Sr}/^{86}\text{Sr}$) has an advantage over elements contents. Indeed, the $^{87}\text{Sr}/^{86}\text{Sr}$ isotope ratio was demonstrated not to change during biological processes (no biological fractionation) and reflects in plants the environment of growth: mainly the bioavailable fraction of Sr derived from minerals weathering processes of bedrock and soil (Bin et al., 2022) although atmospheric and anthropogenic (mainly by fertilisers) inputs may sometime play a role (Bataille et al., 2020).

In addition to geographic origin, consumers also expect the tastes and the aroma of a labelled tea. This quality aspect of tea is closely related to the agricultural practices, including growing and harvesting conditions, but also to transformation processes and storage conditions. Tea experts are able to assess the awaited quality of labelled samples but it requires a long learning process. Analytical chemistry methods were thus developed to monitor characteristic markers of tea grades (Alasalvar et al., 2012; Lin et al., 2013; Zhang et al., 2013). Recently a new global approach based on the determination of volatile compounds fingerprint for product traceability have been developed with Proton Transfer Reaction - Mass Spectrometry (PTR-MS) (Yener et al., 2016). It corresponds to the non-targeted measurement of Volatile Organic Compounds (VOCs) emitted by a product with a rapid and non-destructive mass spectrometer with soft chemical ionization source. Selected Ion Flow Tube Mass Spectrometer (SIFT-MS), is also a Direct Injection Mass Spectrometer (DIMS), able to draw the volatile fingerprint of a product by monitoring the ions produced by volatile molecules and chemical precursors (Reyrolle et al., 2022).

In this work, we developed the measurement of volatile compounds signatures by SIFT-MS of 26 tea samples with certified origins. With statistical unsupervised method we first looked at what aspect guides the classification (origin of production, cultivar or transformation process). Then, with supervised method we extracted the information correlated to transformation process. Finally, we combined these results with multi-elemental signature and Sr isotope ratio to illustrate the complementarity of these analytical approaches for traceability of tea origin and the different contributing factors: origin of production, cultivar and transformation.

2. Materiel and methods

2.1. Samples

A tea expert has collected 26 tea samples (Table 1) with certified origins directly from producers or factories, from 2011 to 2020. The 26 teas came from seven different countries: Nepal (n=3), China (n=9), Vietnam (n=5), India (n=3), Taiwan (n=2), Sri Lanka (n=2) and Japan (n=2), corresponding to twelve different regions and have five different colors: green (n=6), black (n=10), dark (n=3), white (n=4) and oolong (n=3).

Sample	Country	Region	Color	Cultivars
Tea-1	Nepal	Ilam Valley	White	Mix Clonal Cultivars
Tea-2	Nepal	Ilam Valley	White	Mix Clonal Cultivars
Tea-3	Nepal	Ilam Valley	White	Mix Clonal Cultivars
Tea-4	China	Zhejiang	Green	Jiu Hang
Tea-5	China	Zhejiang	Green	Jiu Hang
Tea-6	China	Zhejiang	Green	Jiu Hang
Tea-7	Japan	Kyushu	Green	Yabukita
Tea-8	Japan	Kyushu	Green	Yabukita
Tea-9	China	Jiangsu	Black	Xiao Ye
Tea-10	China	Jiangsu	Black	Xiao Ye
Tea-11	China	Jiangsu	Black	Xiao Ye
Tea-12	India	Darjeeling	Black	China Jat (Mix)
Tea-13	India	Darjeeling	Black	China Jat (Mix)
Tea-14	India	Darjeeling	Black	China Jat (Mix)
Tea-15	Taiwan	Monts Ali	Black	Chin Shin
Tea-16	Taiwan	Monts Ali	Oolong	Chin Shin
Tea-17	Vietnam	Lai Chau	Oolong	Jin Xuan
Tea-18	Vietnam	Lai Chau	Oolong	Jin Xuan
Tea-19	Vietnam	Lai Chau	White	Jin Xuan
Tea-20	Vietnam	Lai Chau	Black	Jin Xuan
Tea-21	Vietnam	Lao Cai	Green	Da Ye age
Tea-22	China	Yunnan	Dark	Da Ye
Tea-23	China	Yunnan	Dark	Da Ye
Tea-24	China	Yunnan	Dark	Da Ye
Tea-25	Sri Lanka	Ratnapura	Black	Assamica
Tea-26	Sri Lanka	Matara	Black	Assamica

Table 1: Description of tea samples

For some samples, several years of production were analyzed. The samples were stored in sealed aluminum bags at room temperature in the dark.

2.2. Multi-elements concentrations and strontium isotopic analysis

A mass of three grams of tea was dry-grinded in a porcelain mortar. The obtained powder was transferred to a zipped polyethylene bag and stored in a dry, dark place until analysis. An amount of 0.2-0.4 g of tea powder was combined with 7 mL of sub-boiled nitric acid (HNO₃ 69.0-70.0%, Instra Analysed Reagent, J.T.Baker®, Fisher Scientific, France) and left to react overnight at room temperature. Then, 1 mL of hydrogen peroxide (H₂O₂ 30%, ULTREX® II Ultrapure Reagent, J.T.Baker®, Fisher Scientific, France) was added to the sample/acid solution. The resulted mixture was microwaved under the standard conditions. After cooling, the solution was transferred into 50 mL polypropylene tubes and diluted with ultrapure water up to 30 mL (Milli-Q system, resistivity 18.2 MΩ.cm, Millipore, Burlington, MA, USA).

2.2.1. Multi-elemental composition by quadrupole ICP-MS

Multi-elemental analysis of tea samples was performed on aliquot of the diluted solutions using an ICP-MS Plasma Quant Elite spectrometer (Analytik Jena, Jena, Germany). All samples were analyzed for determining the total concentration of 21 elements: Al, As, B, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Li, Mg, Mn, Na, Ni, Rb, Sr, Pb, V, and Zn. The instrumental operating conditions and measured isotopes are presented in Table SI2. The analysis was performed using two operating modes: a standard “no gaz” mode and an integrated collisional reaction cell (iCRC) mode with helium (He) and hydrogen (H₂) as collision and reaction gases respectively, to reduce spectral interferences. External calibration solutions at eight different concentrations ranging from 0.01 µg L⁻¹ to 500 µg L⁻¹ were prepared by appropriate dilution of the multi-elemental standard solutions CCS-4 and CCS-6 (Inorganic ventures, Christiansburg, VA, USA). Yttrium (Y), rhodium (Rh), and iridium (Ir) at a concentration of 2.5 µg L⁻¹ were used as internal standards for the low, medium, and heavy masses, respectively, in order to correct instrumental drifts. Samples concentrations were calculated after applying the corrections for the analytical and procedural blanks and internal standard. Accuracy and precision of the measurements were controlled by analyzing the following certified reference materials: NIST SRM 1643f (Trace elements in water, The National Institute of Standards and Technology, Gaithersburg, MD, USA) and CRM NIES 23 (The National Institute for Environmental Studies, Tsukuba-City, Ibaraki, Japan), the recoveries obtained were within the acceptance maximal limits of ± 10%.

2.2.2. Sr Isotopic analysis by MC-ICP-MS

An appropriate volume of each sample digested solution was evaporated to dryness to ultimately get 4 µg of Sr and redissolved in 3M HNO₃. Then Sr was separated from the rest of the matrix using columns containing Eichrom Sr-Spec resin and sequential elution with ultrapure water and 3M ultrapure HNO₃ according to the procedure described in Martin et al.

(2013) (Martin et al., 2013). The Sr isotopic analysis was performed on a Nu Plasma MC-ICP-MS (Nu Instruments, Wrexham, UK) under wet plasma conditions using a self-aspirating micro concentric nebulizer (200 µL min⁻¹) combined with a glass cyclonic spray chamber. Measurements were performed using a conventional “Sample-Standard-Bracketing” calibration sequence with the SRM 987 (Strontium carbonate isotopic standard, NIST, Gaithersburg, MD, United States) used as bracketing standard every two samples. After instrumental blank subtraction using the On Peak Zero approach (OPZ), the ⁸⁷Sr/⁸⁶Sr ratio was corrected from instrumental mass bias through an exponential law using the constant ratio ⁸⁶Sr/⁸⁸Sr of 0.1194, and from the potential remaining interferences from traces of ⁸⁷Rb using the monitored ⁸⁵Rb and a ratio ⁸⁵Rb/⁸⁷Rb of 2.5926. The value of ⁸⁷Sr/⁸⁶Sr ratio for SRM 987 applied for data processing was 0.710255, according to Waight, Baker, and Peate (2002)²⁷.

The ⁸⁷Sr/⁸⁶Sr ratio was normalized to the certified reference materials NIST CRM 987 according to the following equation $\delta^{87}\text{Sr} = [(\frac{{}^{87}\text{Sr}/{}^{86}\text{Sr}}{\text{sample}}/\frac{{}^{87}\text{Sr}/{}^{86}\text{Sr}}{\text{NIST987}}) - 1] \times 1000$ (‰) (Liu et al., 2012)(H.-C. Liu et al., 2012).

$$\delta^{87}\text{Sr} = \left[\frac{(\frac{{}^{87}\text{Sr}/{}^{86}\text{Sr}}{\text{sample}})}{(\frac{{}^{87}\text{Sr}/{}^{86}\text{Sr}}{\text{NIST987}})} - 1 \right] \times 1000 \text{ (‰)}$$

Equation 1

2.3. Analyse of volatile components by SIFT-MS

SIFT-MS measurements were performed on headspace of teas; 10 g of tea were introduced into a 1 litre glass bottle fitted with a polypropylene screw cap with 2 tight connection ports fitted with 0.6 cm PFA Tubes. The first one was directly connected to SIFT-MS while the second to a 1L Tedlar © bag (Supelco, Bellefonte, PA, USA), filled with zero dry air (ZeroAir Alliance ZA1500, F-DGSI, Evry France) to compensate the analysed volume during the SIFT-MS analysis. The closed bottle was incubated for 2 hours at 60°C to achieve an equilibrium in gas phase before performing the positive and negative SIFT-MS full scan analysis.

A SIFT-MS Voice 200 Ultra (SYFT Technologies, Christchurch, New Zealand) equipped with a dual source producing positive and negative soft ionizing precursor ions (H₃O⁺, O₂^{•+}, NO⁺, O^{•+}, OH⁺, O₂^{•+}, NO₂⁺ and NO₃⁺) in a single scan was used. This instrument used nitrogen (Air Liquide, Alphagaz 2) as carrier gas, the sample was introduced through a heated (110 °C) and flow controlled (20 mL min⁻¹) sample line (High Performance Inlet HPI®). The instrument was daily calibrated with a standard gas (Scott™ gas mixtures, Air Liquid USA) containing standards at 2.0 ppmV in nitrogen. A blank experiment for every empty bottle was performed before each triplicate. The full scan raw data files containing product ion intensities with 15-250 m/z range were collected before a compilation and pre-processing steps.

The compilation and pre-processing of the full scan raw data were done according to the approach of Reyrolle et al., 2022 using a Python language. Data was pre-processed by subtracting the background noise (signal of bottles without samples) for every replicate. Triplicates of each sample were averaged to get a single value of signal intensity for every ion. This value was multiplied by the ICF factor corresponding to

the instrument calibration correction. The dataset was cleaned by suppressing constant quantitative variables.

2.4. Statistical analysis

Statistical analysis, including non-supervised and supervised methods, were performed with the R language and “MixOmics” package. For multi-element analysis and Sr isotope, principal component analysis and partial least squares-Discriminant analysis were used whereas, for volatile fingerprint analysis, Sparse Principal Component Analysis (Sparse PCA) and Sparse Partial Least Squares-Discriminant Analysis (sPLS-DA) were used to show the dispersion or the discrimination of the sample according to a large number of variables. In order to combine inorganic elemental contents and the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio in the statistical analysis, the latter was normalized to the certified reference materials NIST CRM 987 ($\delta^{87}\text{Sr}_{\text{NIST987}}$) according to Equation 1.

3. Results

3.1. Elemental and Sr isotopic signatures

In order to infer similarities or differences in the chemical fingerprint of teas across sites and countries of origin, partial Least Square regression and Discriminant Analysis was performed and expressed as hierarchical clustering heatmap (Figure 1) with elemental concentrations and $\delta^{87}\text{Sr}$ as variables (in column) and regions and countries of origin as observations (in rows).

Among the eleven sites studied, ten origins can be clearly distinguished with a significance level of 0.05. Nine of them as site with unique geochemical signature China_Jiangsu, Vietnam_Lao Cai, India_Darjeeling, Nepal_Ilam Valley, Japan_Kyushu, Vietnam_Lai Chau, Taiwan_Mounts Ali, China_Zhejiang and China_Yunnan; and one group of two sites (the two Sri-Lanka provinces, Ratnapura and Matara districts). Four sites or groups of sites were primarily discriminated by their elemental content (China_Jiangsu; Sri Lanka_Ratnapura/Matara; Vietnam_Lai Chau; China_Zhejiang), three sites mainly by their $^{87}\text{Sr}/^{86}\text{Sr}$ ratio (Nepal_Ilam Valley; China_Yunnan; Taiwan_Mounts Ali) and

three sites according to both $^{87}\text{Sr}/^{86}\text{Sr}$ ratio ($\delta^{87}\text{Sr}$) and specific elemental concentrations (Vietnam_Lao Cai; India_Darjeeling; Japan_Kyushu) (Figure. 2 and Table S12).

A first cluster of region profiles gathered Sri Lanka_Ratnapura, Sri Lanka_Matara, China_Yunnan, and Vietnam_Lao Cai samples. The proximity of the volatile fingerprints of these four regions cannot be explained by geographical proximity. However, thanks to the knowledge of Tea expert, we know that the two cultivars subspecies used in these four provinces are *Da Ye* and *Assamica* (Table 1), large leafed teas, which are very similar since *Assamica* is a clone of *Da Ye*.

3.2. Volatile fingerprints

3.2.1. Unsupervised analysis of tea samples grouped by region of origin

Volatile fingerprint corresponds to all the VOCs emitted by tea products and detected with SIFT-MS. This chemical pattern can be related to the odour and flavour perception by human, with the exception that the selectivity and sensitivity were not similar. The power of the volatile fingerprint approach lies in the search for correlation with descriptive values. An unsupervised analysis (sparse PCA) was performed on the mean volatile fingerprints obtained for 26 tea samples by grouping them with regions’ origins and then represented in the cluster image map (Figure 3).

No spontaneous classification of tea samples according to the country appeared (China, Vietnam, or Sri Lanka samples were not grouped). We observed that Vietnam_Lai Chau tea presented a pattern completely different to the rest of the dataset with a high number of product ions overexpressed compared to the whole dataset. Surprisingly, this profile corresponds to a mix of different grades of tea (2 oolong, 1 white, and 1 black). Japan_Kyushu tea, corresponding to 2 different samples of green tea from the same producer, presented also another cluster of product ions that are over-expressed compared to the whole dataset and that ward these samples from the rest of the samples.

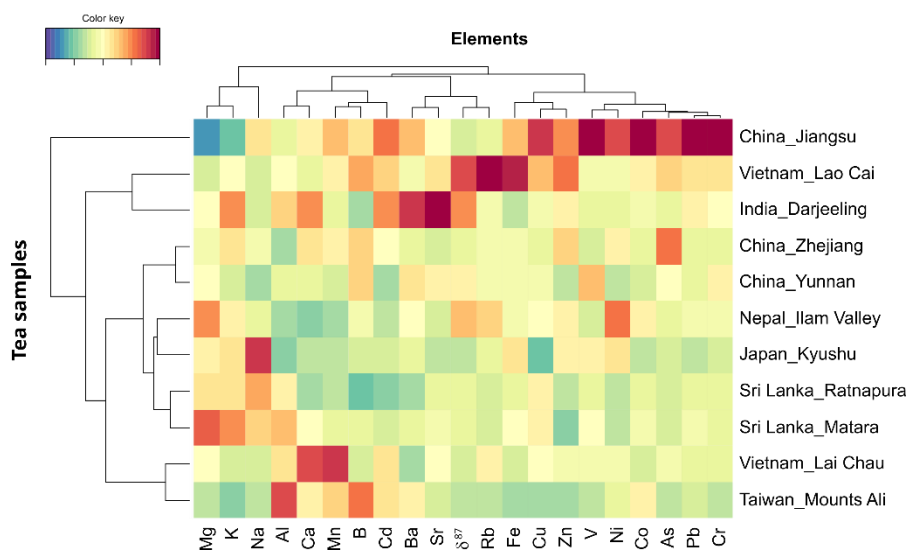


Figure 1: A clustered image map (CIM) based on PLS-DA according to the parameter “country_region of origin”. The dendrogram represents hierarchical grouping for regions (horizontal) and for elements (vertical). The color key indicates the under expression (in blue scale) or overexpression (in red scale) of the element in each sample group relative to the others.

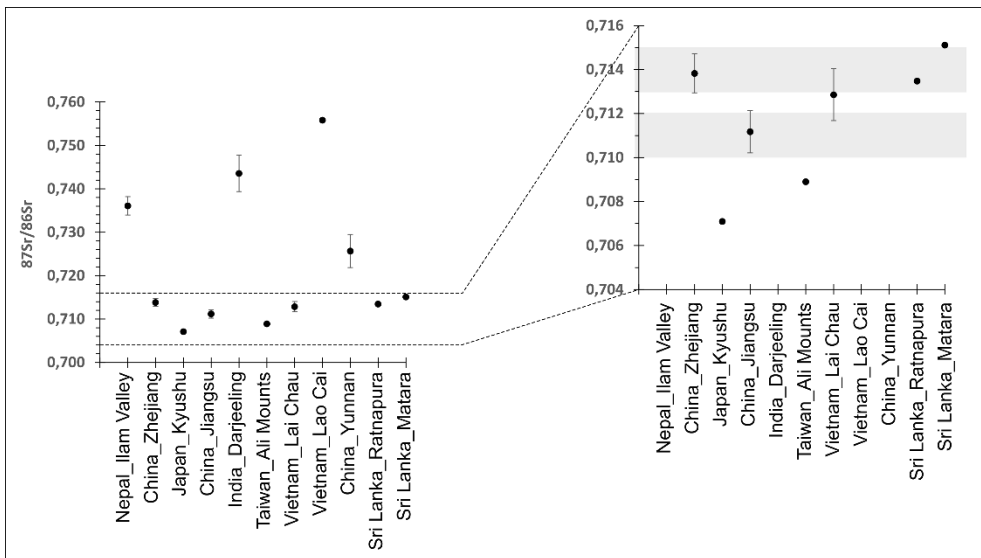


Figure 2: Strontium isotope ratios ($^{87}\text{Sr}/^{86}\text{Sr}$) according to the regions. Mean values are represented with circles and the standard deviation with vertical lines. Figure 2B represent a zoom of figure 2A between 0.704 and 0.716 of $^{87}\text{Sr}/^{86}\text{Sr}$.

The volatile fingerprints of tea samples from China_Zhejiang and China-Jiangsu were very close, which can be explained by the geographical proximity of the two production areas and also by the use of similar cultivars (*Jiu Hang* and *Xiao Ye* respectively), small-leafed *Camellia sinensis var sinensis* tea bushes. With the same trend, the volatile fingerprints of the India_Darjeeling and Nepal_Ilam Valley tea samples were very close also due to the geographical proximity of the two production areas and the use of similar mix of cultivars: a mixture of small leaf teas from eastern China and *Assamica* (large leaf tea). The proximity of the volatile fingerprints obtained between the Darjeeling and Ilam Valley tea samples and those from Jiangsu and Zhejiang was not explained by geographical proximity but by the use of the same cultivar base

(small leaf tea). The small leaf tea plants in the mix of cultivar seems to contribute more to the typicality of the Darjeeling and Ilam Valley tea in a clone mix compared to the large leaf tea plant.

1.1.1. Volatile fingerprints from processing methods

The tea expert selected different colors of tea corresponding to several types of transformation processes of tea leaves. To highlight the impact of these transformation processes, we performed a supervised analysis (sparse PLS-DA) on the volatile fingerprints of the 26 tea and then represented them in a cluster image map (Figure 4). This statistical tool allowed us to look for what is common inside the volatile fingerprints and

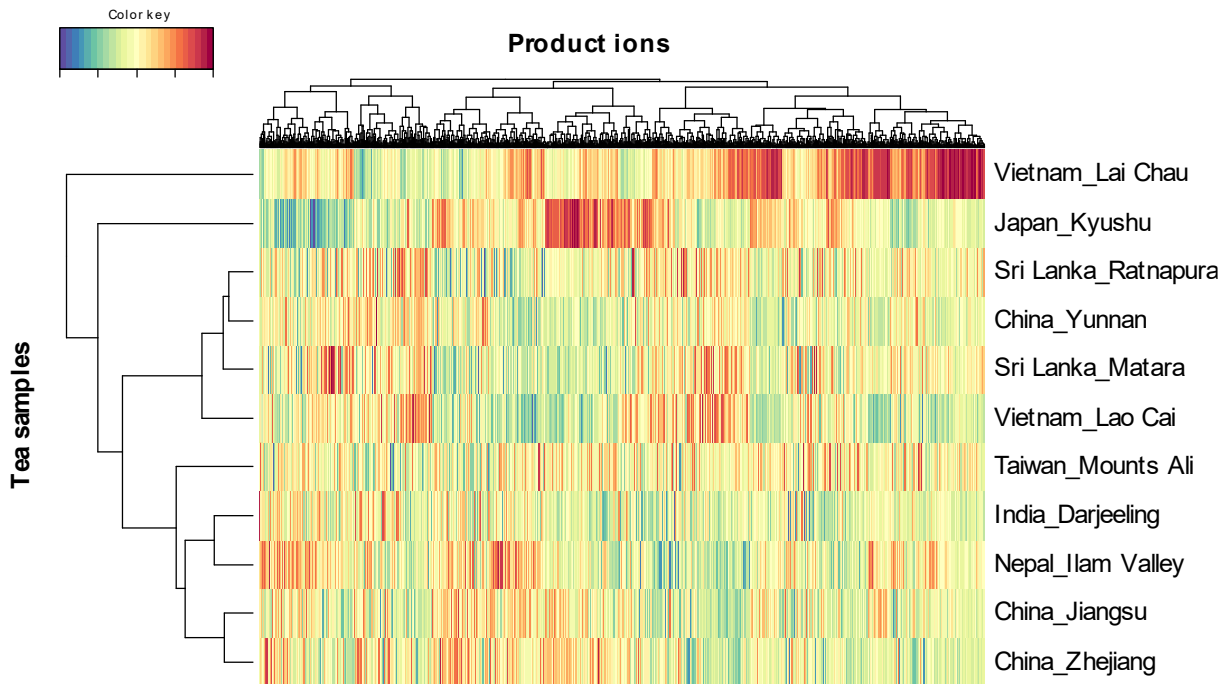


Figure 3: A clustered image map (CIM) based on Sparse PCA analysis of volatile fingerprint to show the dispersion of tea region (or regions?). The horizontal dendrogram represents the hierarchical grouping for tea region and the vertical dendrogram for product ions. The colour key indicates the negative (in blue scale) or positive (in red scale) intensity of the product ion in each group sample compared to the other samples.

what makes the difference between classes. The dendrogram on the left of the CIM corresponds to the classification of tea based on volatile profiles and the dendrogram at the top corresponds to the classification of product ions.

The supervised analysis of volatile fingerprints of the dataset according to the colors of teas shows a good discrimination of green and oolong teas. All green teas were grouped together and were the most distant to oolong tea cluster. The 3 different dark tea samples are located together in the center of the CIM representation (Figure 4), indicating a strong similarity in their volatile fingerprints.

Black and white teas were mixed into two groups, some of which are closer to the oolong group and others to the green group. In our dataset, volatile fingerprints did not clearly discriminate black and white teas in two different clusters. Other factors such as terroir, cultivar (as seen above), storage conditions, etc. may influence the volatile fingerprints.

Volatile fingerprints allowed us to discriminate between dark (fermented), cyan (semi-oxidised), and green (non-oxidised) teas, which are very different. The plotloadings figure (Fig. SI 3) shows the 40 most discriminating product ions on the first component of the sPLS-DA. Due to the unitary mass resolution of the instrument and the possible ion conflicts, it might be challenging to identify molecules among this list. Nevertheless, we found that 5 product ions corresponding to hexanal were founded for the positive discrimination of oolong (H_3O^+ m/z 101, NO^+ m/z 99, O^+ m/z 99, O_2^+ m/z 99, and OH^+ m/z 99). These particular teas are semi-oxidized with a higher level of oxygenated molecules such as aldehydes or alcohols among which hexanal and hexanol were identified by Zhou et al. (Zhou et al., 2020). The boxplot (Figure 5) shows the intensity of one ion produced by hexanal (NO^+ m/z 99) and by Hexanol (NO^+ m/z 101) in the different tea categories. Lower signals were observed for green and dark teas, which are consistent since they are not oxidized whereas the highest signals were observed for oolong teas. Black and white teas presented an intermediate level for these two product ions.

2. Discussion

In this work, we analysed 26 different teas, selected by a tea expert from 11 regions corresponding to 7 countries and with 5 different colors. These tea samples are perfectly traced from the field to the cup. We performed inorganic, strontium isotopic and volatile organic compounds analysis on that sample database to look for correlations with origin and/or quality, that are awaited by consumers, producers and sectors. The different results were exploited with a global approach and the inorganic profiles and the volatile fingerprints of the tea were drawn.

Studies on geoprovenance of food products by using multi-tracers assignment approaches (trace element fingerprints, stable isotopic signatures) are increasing steadily in recent years (Bandoniene et al., 2018). Trace element fingerprints successfully allowed the determination of the origin of products, especially in plants because they inherit their element composition primarily from the bioavailable element fraction underlying the soils (Bataille et al., 2020; Song et al., 2015). Strontium isotopes, the radiogenic ^{87}Sr vs the stable ^{86}Sr ($^{87}Sr/^{86}Sr$), have been the focus of interest to infer geographical provenance and thus largely applied in fishery sciences (Brennan et al., 2015), archaeological sciences (Alexander Bentley, 2006), and food sciences (Epova et al., 2019). Applying the $^{87}Sr/^{86}Sr$ ratio presents several advantages. First, it is transferred to biological materials in a conservative manner, i.e. without any biological fractionation across the food chain (Alexander Bentley, 2006; Walther & Limburg, 2012). Second, it mainly depends on the bioavailable Sr in soil (Bataille et al., 2018, 2020), fraction which is largely controlled by the differential weathering intensity of minerals from the parent bedrock and soil having variable Sr and Rb concentrations (Rb/Sr ratio) (Bataille et al., 2020; Walther & Limburg, 2012). Ultimately, the combination of lithology, features of weathering processes and considering the Sr additional sources (atmospheric, anthropogenic) conducts to considerable variability of $^{87}Sr/^{86}Sr$ available fraction in the geosphere and thus the biosphere (Bataille et al., 2020).

In this study, using the $^{87}Sr/^{86}Sr$ ratio in tea samples we clearly discriminated 6 of 11 collection sites confirming the power of this techniques to infer geographical origin as previously mentioned in the scientific literature (Coelho et al., 2017; Epova et al., 2019). Globally, the $^{87}Sr/^{86}Sr$ ratio found in tea samples studied here agrees with recently published $^{87}Sr/^{86}Sr$ isoscapes (Aoyama et al., 2017; Bataille et al., 2020; X. Wang & Tang, 2020) for provenance studies adding strength to the link between plants and the bioavailable $^{87}Sr/^{86}Sr$ ratio in the soil their grown (Bataille et al., 2020; Song et al., 2015). Different $^{87}Sr/^{86}Sr$ ratios were found in teas sample from Yunnan (China), Darjeeling (India), Ilam Valley (Nepal), Lao Cai (Vietnam), Monts Ali (Taiwan) and Kuysyu (Japan).

Obviously, the high $^{87}Sr/^{86}Sr$ ratio we found in Darjeeling tea (average 0.744 ± 0.004 , $n=3$) is in agreement with values obtained by Lagad et al., 2013 (average 0.745 ± 0.029 , range: 0.726-0.829, $n=9$) for the same region, as well as with data obtained from local water (surface? Ground?) and soil reservoirs (Galy et al., 1999). Valley (Nepal: 0.736 ± 0.002 , $n=3$) and Lao Cai (Vietnam: 0.7558, $n=1$). Indeed, combining the $^{87}Sr/^{86}Sr$ ratio and trace elements Ca, Sr,

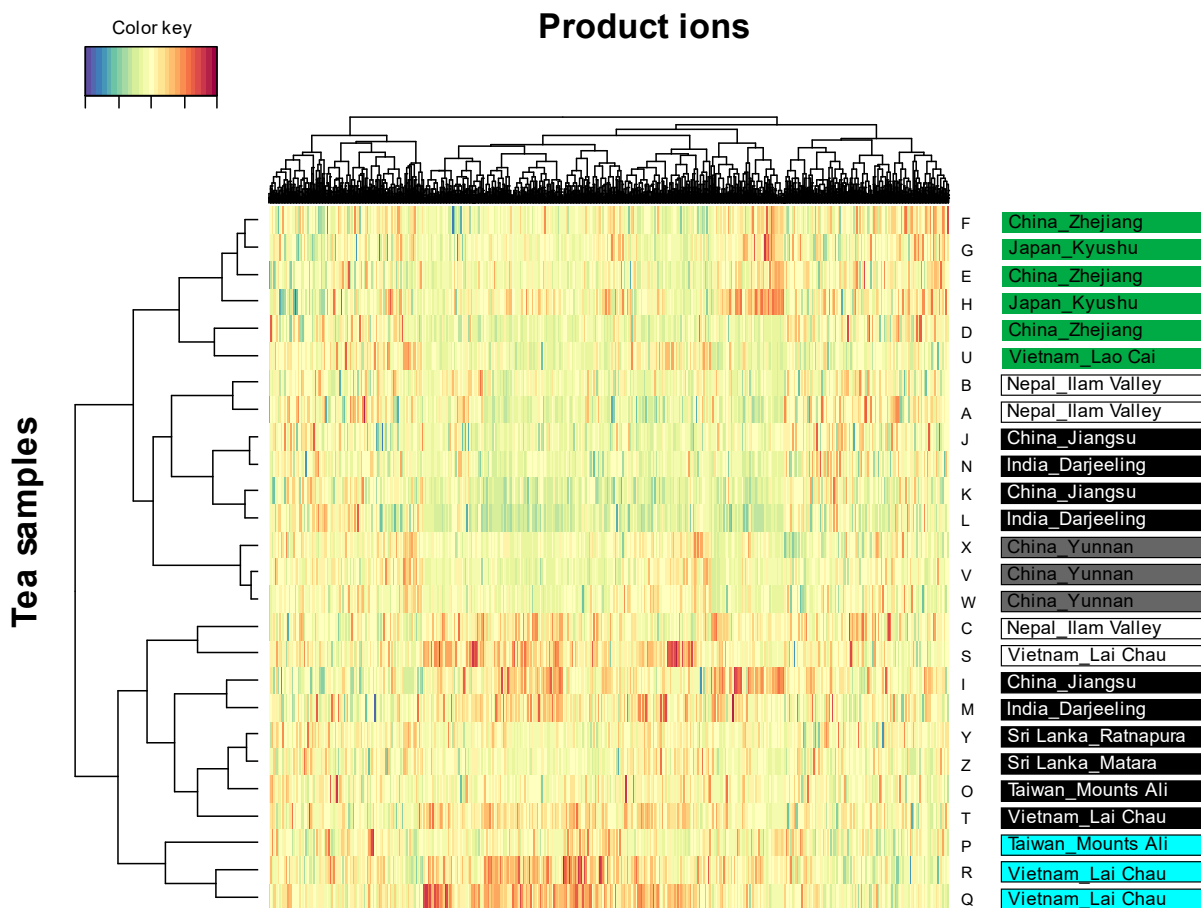


Figure 4: Clustered image maps (CIM) based on sparse PLS-DA analysis of volatile fingerprint to discriminate color/processing of tea. Green teas were highlighted with green, Oolong teas with blue, black teas highlighted with black, Dark teas highlighted with grey and white teas highlighted with white. The dendrograms represent hierarchical grouping for tea samples (horizontal) and for variables (vertical). The colour key indicates the negative (in blue scale) or positive (in red scale) intensity of the product ion in each group sample compared to the other samples.

Ba, Cd and K allowed to clearly discriminate Darjeeling from Ilam Valley and Lao Cai teas. Moreover, Lao Cai teas were found to be enriched in Rb and Fe compared to those of Darjeeling allowing a more confident discrimination of these two sources. To protect Darjeeling tea, a Geographical Indication was given by the Tea Board of India to preserve of its treasured icon of India's cultural heritage (*The Official Website of Tea Board India*, n.d.). Thus, a combination of strontium isotopic ratio and trace element would have the potential to prove the authenticity of the product and fight against fake product using the labelling "Darjeeling Tea". Further, because of the volcanic substratum of the Kyushu province from Japan the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio easily discriminates these teas from other sources in the world (Aoyama et al., 2017). Additionally, their relatively large level in Na add further insight in their differentiation against other teas.

Although $^{87}\text{Sr}/^{86}\text{Sr}$ clearly allowed some sites discrimination, it would not be meaningful for 5 sites. Indeed, $^{87}\text{Sr}/^{86}\text{Sr}$ ratio may overlap between the following sites, Ratnapura and Matara from Sri Lanka, Lai Chau (Vietnam) and Zhejiang (China) on one hand, and Jiangsu (China) and Lai Chau (Vietnam) on the other hand. Nonetheless, we clearly identified inorganic patterns between these sites allowing differentiating them, except between the two Sri Lanka sites. For instance, Jiangsu tea samples presented high levels of vanadium, cobalt,

chromium, lead and copper compared to the four others sites, as well as those already discriminated with the Sr isotopic ratio. This specificity would be associated to surrounding mines and industrial activities. This high elemental signature in heavy metals can be related to the presence of 12 mines listed by the United States Geological Survey (USGS), whose main mining products are iron, copper, zinc, lead, chromium, vanadium and silver. Moreover, GEM, one of the largest battery recyclers in China, process battery scrap and have an output of 10,000 tonnes of cobalt tetroxide in Jiangsu. These elements are therefore good markers of anthropogenic impact on multielement signature and allow identification of teas from this specific region. Further, Lai Chau, Zhejiang and Ratnapura-Matara would differ from each other by their different concentrations in Ca/Mn, As and Na/Mg, respectively. Our results confirm thus that multi-element signatures are useful indicators of food products, and especially for tea authentication, as they can reflect information on climate, environment, agricultural practices (Camin et al., 2017) (Anderson et al., 1999). Finally, combining Sr isotopic ratio and elemental fingerprints resulted in almost complete discrimination of the studied sites (10 of 11) as only the two sites from Sri Lanka are inseparable.

Studies on volatile organic compounds emitted by food products have been carried out for a long time, since

VOCs are good tracers of the metabolisms of plants, animals and microbial involved in the food production and transformation. Nevertheless, it remains difficult to rely taste or aroma to specific compounds because of the high complexity of the human perception. New approaches, based on unspecific analysis of food, combined to statistical tools allow to look for correlations between qualitative attributes and chemical composition. Direct injection mass spectrometers, like PTR-MS and SIFT-MS, are powerful instruments to investigate the volatile fraction, thanks to their high chemical selectivity. Thus, it is possible to investigate the specific biogenesis of volatile molecules in plant products such as olive oil (Bajoub et al., 2018; Ozcan-Sinir, 2020) or saffron (Nenadis et al., 2016), to control processes like drying (Olivares et al., 2011), roasting (Dryahina et al., 2018), oxidation (Davis & McEwan, 2007) and fermentation (Van Kerrebroeck et al., 2015) or to illustrate the typicality of products (Deuscher et al., 2019; Reyrolle et al., 2022). Traceability of food products can be in some case correlated with taste and flavor as for wine (Cravero, 2019), but it requires experts with high experience.

In this work, we studied the potential of the volatile profiles of tea samples to correlated to country production origin. The untargeted analysis of tea headspace with SIFT-MS Voice 200-ultra allowed us to draw the volatile fingerprints with 1 880 variables. Non-supervised method (Sparse PCA) have shown that, more than the country, the volatile fingerprints of tea seem to be primarily impacted

by the cultivar and the terroir. In the particular old product as for tea, the terroir doesn't mean the country but the people who live there and the habits exchange with neighbours: the plants, the agriculture practices and transformation processes (H. Wang et al., 2021). Thus, tea samples with geographical proximity and similar cultivar types, develop very close volatile fingerprints, as for Chinese teas from Jiangsu and Zhejiang, or for India and Nepal teas. The statistical proximity of volatile fingerprints of Sri Lanka_Ratnapura, Sri Lanka_Matara, China_Yunnan and Vietnam_Lao Cai samples cannot be explained by geographical proximity but by cultivar subspecies proximities. Moreover, regarding the impact of the cultivar and the terroir on the volatile fingerprints, particular profiles of tea from Japan_Kyushu and Vietnam_Lai Chau might be tentatively explained. In the case of the Japanese teas, it seems consistent due to a specific terroir that is very different from other countries, but also because Japanese research institutes have developed a new variety of *Camelia sinensis* adapted to their climatic conditions: Yabukita. Teas from Lai Chau province are produced from a tea plant called Jin Xuan, which was developed in Taiwan by the Tea Research and Extension Station and subsequently imported to this region. The secret of genetic selection of tea cultivars combined to the number of cultivars, makes the understanding of phylogenetic relationships complex.

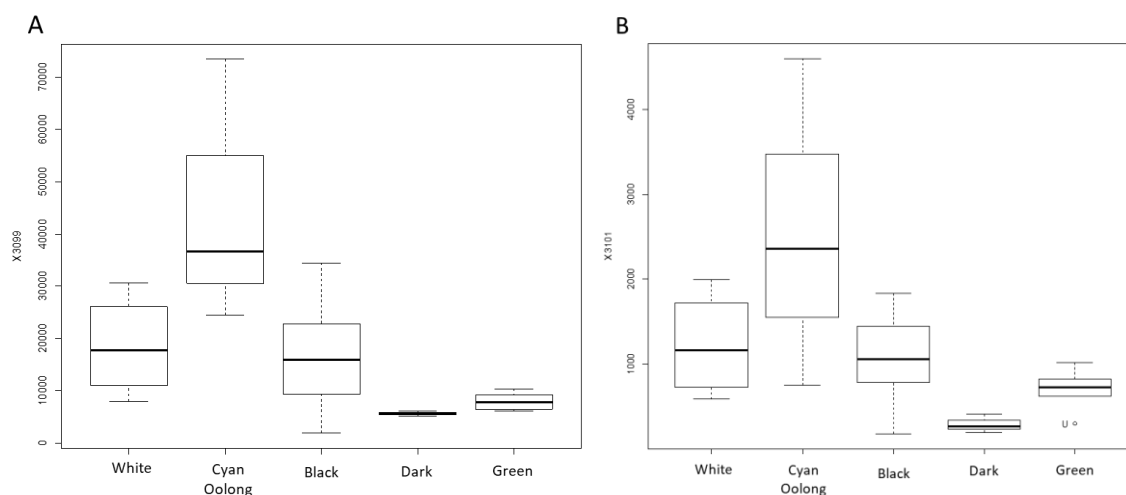


Figure 5: Boxplot of the intensity of product ions X3099 (NO^+ m/z 99) corresponding to Hexanal (A) and X3101 (NO^+ m/z 101) (corresponding to Hexanol (B)) for each color of tea.

As can be seen from the figure 3, the difference in cultivar between the two Vietnamese provinces has a clear impact on the volatile profile. Guo et al., 2022 have shown that aroma profiles of oolong tea from several cultivars were significantly different. Furthermore, one would expect the province of Lai Chau (Vietnam) to have a volatile fingerprint close to that of Mont Ali (Taiwan), but the production area, i.e. the growing area, also seems to have an impact and thus differentiates these two regions.

In addition to biogenic volatile profiles of *Camelia* plants, tea samples undergo complex transformations processes leading to several grades and colours of tea. To evaluate the contribution of the process on the volatile fingerprints, we applied supervised statistical analysis (sPLS-DA). Thus, we first observed a clear separation of two groups: green teas and oolong teas. Green teas are not oxidised while cyan teas are semi-oxidised (Feng et al., 2019; Zhang et al., 2019). A third cluster with dark teas from Yunnan province in China was observed in the sPLS-DA.

These teas are also called “Pu Erh”, which is post-fermented tea. Thus, these three processes giving oolong, dark and green teas were clearly separated in our dataset and correspond to the most particular processes, from the simplest (green tea) to the more complex (Pu Erh) giving the 3 typical tastes (Zhang et al., 2013). Black tea is the most produced tea with various process stages allowing different degrees of leaves’ oxidation and roasting (Alasalvar et al., 2012). Finally, white tea is the least known, produced from buds or young leaves and undergoing a process with prolonged withering (allowing slight enzymatic oxidation) and drying process (Chen et al., 2019; Dai et al., 2017).

Black and White tea of our collection were not discriminated with SIFT-MS. This may be explained by the worldwide diversity of cultivars, agricultural practices and processing techniques for that tea grades. Are there really similarities between black teas from China, India, Vietnam or Sri Lanka? In our tea collection, oolong teas presented the most discriminative volatile fingerprints in which we tried to identify some characteristic molecules. Hexanal and hexanol are good molecular candidates over expressed in oolong teas. But, the ion conflicts, intrinsic to direct injection mass spectrometry technique, and the unitary mass resolution of the instrument, prevent us from absolute identification. However, the main advantages of this technique rely on the sample preparation’s simplicity and the high speed of analysis. With an appropriate sample collection in size and in composition, this global approach with SIFT-MS volatile fingerprints may highlight the most discriminative pattern for a qualitative description, such as typicality or sensory properties.

3. Conclusion

These three different analytical methods, multi-elemental signature, strontium isotope ratio and volatile fingerprints are very interesting tools to ensure the traceability and genuineness of tea throughout its production chain from cultivation to the final product. The multi-elemental analysis of tea will answer to questions related to the growing environment (soil composition, climatic conditions, and anthropogenic activity). Combined inorganic analysis, i.e. elemental and strontium isotopic ratio ($^{87}\text{Sr}/^{86}\text{Sr}$), can certify the production place, thanks to its stability and its known level in soil. This geographical certification can answer to specific questions on the production place, for the protection of PGI’s. In addition to the inorganic analyses, volatile fingerprint analysis will address questions on genuineness, conformity and organoleptic aspect of the final product and will be able to discriminate the “terroir”, the cultivar and the transformation process, probably due to producers’ habits and environmental conditions. In this work, we highlighted the potential impact of cultivar on the volatile fingerprint. This hypothesis would be addressed in a further work with phylogenetic analysis of the tea collection. In conclusion, this study shows the high potential of combining instrumental and chemometric analyses to improve traceability systems to regulate product quality and safety along the supply chain. The application of this analytical

strategy on a collection of tea samples adapted in number and quality will make it possible to answer questions relating to the traceability and authenticity of products.

REFERENCES

- Alasalvar, C., Topal, B., Serpen, A., Bahar, B., Pelvan, E., & Gökmen, V. (2012). Flavor Characteristics of Seven Grades of Black Tea Produced in Turkey. *Journal of Agricultural and Food Chemistry*, 60(25), 6323–6332. <https://doi.org/10.1021/jf301498p>
- Alexander Bentley, R. (2006). Strontium Isotopes from the Earth to the Archaeological Skeleton: A Review. *Journal of Archaeological Method and Theory*, 13(3), 135–187. <https://doi.org/10.1007/s10816-006-9009-x>
- Aoyama, K., Nakano, T., Shin, K.-C., Izawa, A., & Morita, S. (2017). Variation of strontium stable isotope ratios and origins of strontium in Japanese vegetables and comparison with Chinese vegetables. *Food Chemistry*, 237, 1186–1195. <https://doi.org/10.1016/j.foodchem.2017.06.027>
- Bajoub, A., Medina-Rodríguez, S., Ajal, E. A., Cuadros-Rodríguez, L., Monasterio, R. P., Vercammen, J., Fernández-Gutiérrez, A., & Carrasco-Pancorbo, A. (2018). A metabolic fingerprinting approach based on selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics: A reliable tool for Mediterranean origin-labeled olive oils authentication. *Food Research International*, 106, 233–242. <https://doi.org/10.1016/j.foodres.2017.12.027>
- Bandoniene, D., Meisel, T., Rachetti, A., & Walkner, C. (2018). A tool to assure the geographical origin of local food products (glasshouse tomatoes) using labeling with rare earth elements: A tool to assure the geographical origin of local food products. *Journal of the Science of Food and Agriculture*, 98(12), 4769–4777. <https://doi.org/10.1002/jsfa.9124>
- Bataille, C. P., Crowley, B. E., Wooller, M. J., & Bowen, G. J. (2020). Advances in global bioavailable strontium isoscapes. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 555, 109849. <https://doi.org/10.1016/j.palaeo.2020.109849>
- Bataille, C. P., Holstein, I. C. C. von, Laffoon, J. E., Willmes, M., Liu, X.-M., & Davies, G. R. (2018). A bioavailable strontium isotope for Western Europe: A machine learning approach. *PLOS ONE*, 13(5), e0197386. <https://doi.org/10.1371/journal.pone.0197386>
- Belletti, G., Burgassi, T., Manco, E., Marescotti, A., Pacciani, A., & Scaramuzzi, S. (n.d.). The impact of geographical indications (PDO and PGI) on the internationalisation process of agro-food products. 26.
- Beltrán, M., Sánchez-Astudillo, M., Aparicio, R., & García-González, D. L. (2015). Geographical traceability of virgin olive oils from south-western Spain by their multi-elemental composition. *Food Chemistry*, 169, 350–357. <https://doi.org/10.1016/j.foodchem.2014.07.104>
- Bin, L., Wang, C., Liu, Z., He, W., Zhao, D., Fang, Y., Li, Y., Zhang, Z., Chen, P., Liu, W., & Rogers, K. M. (2022). Geographical origin traceability of muskmelon from Xinjiang province using stable isotopes and multi-elements with chemometrics. *Journal of Food Composition and Analysis*, 106, 104320. <https://doi.org/10.1016/j.jfca.2021.104320>
- Brennan, S. R., Zimmerman, C. E., Fernandez, D. P., Cerling, T. E., McPhee, M. V., & Wooller, M. J. (2015). Strontium isotopes delineate fine-scale natal origins and migration histories of Pacific salmon. *Science Advances*, 1(4), e1400124. <https://doi.org/10.1126/sciadv.1400124>
- Camin, F., Boner, M., Bontempo, L., Fauth-Hasek, C., Kelly, S. D., Riedl, J., & Rossmann, A. (2017). Stable isotope techniques for verifying the declared geographical origin of food in legal cases. *Trends in Food Science & Technology*, 61, 176–187. <https://doi.org/10.1016/j.tifs.2016.12.007>
- Chen, Q., Zhu, Y., Dai, W., Lv, H., Mu, B., Li, P., Tan, J., Ni, D., & Lin, Z. (2019). Aroma formation and dynamic changes during white tea processing. *Food Chemistry*, 274, 915–924. <https://doi.org/10.1016/j.foodchem.2018.09.072>
- Coelho, I., Castanheira, I., Bordado, J. M., Donard, O., & Silva, J. A. L. (2017). Recent developments and trends in the application of strontium and its isotopes in biological related fields. *TRAC Trends in Analytical Chemistry*, 90, 45–61. <https://doi.org/10.1016/j.trac.2017.02.005>

- Cravero, M. C. (2019). Wine Traceability. MDPI - Multidisciplinary Digital Publishing Institute.
- Dai, W., Xie, D., Lu, M., Li, P., Lv, H., Yang, C., Peng, Q., Zhu, Y., Guo, L., Zhang, Y., Tan, J., & Lin, Z. (2017). Characterization of white tea metabolome: Comparison against green and black tea by a nontargeted metabolomics approach. *Food Research International*, 96, 40–45. <https://doi.org/10.1016/j.foodres.2017.03.028>
- Danezis, G. P., Tsagkaris, A. S., Camin, F., Brusci, V., & Georgiou, C. A. (2016). Food authentication: Techniques, trends & emerging approaches. *TrAC Trends in Analytical Chemistry*, 85, 123–132. <https://doi.org/10.1016/j.trac.2016.02.026>
- Davis, B. M., & McEwan, M. J. (2007). Determination of Olive Oil Oxidative Status by Selected Ion Flow Tube Mass Spectrometry. *Journal of Agricultural and Food Chemistry*, 55(9), 3334–3338. <https://doi.org/10.1021/jf063610c>
- Deuscher, Z., Andriot, I., Sémon, E., Repoux, M., Preys, S., Roger, J.-M., Boulanger, R., Labouré, H., & Le Quééré, J.-L. (2019). Volatile compounds profiling by using proton transfer reaction-time of flight-mass spectrometry (PTR-ToF-MS). The case study of dark chocolates organoleptic differences. *Journal of Mass Spectrometry*, 54(1), 92–119. <https://doi.org/10.1002/jms.4317>
- Dryahina, K., Smith, D., & Španěl, P. (2018). Quantification of volatile compounds released by roasted coffee by selected ion flow tube mass spectrometry. *Rapid Communications in Mass Spectrometry*, 32(9), 739–750. <https://doi.org/10.1002/rcm.8095>
- Epova, E. N., Bérail, S., Séby, F., Barre, J. P. G., Vacchina, V., Médina, B., Sarthou, L., & Donard, O. F. X. (2020). Potential of lead elemental and isotopic signatures for authenticity and geographical origin of Bordeaux wines. *Food Chemistry*, 303, 125277. <https://doi.org/10.1016/j.foodchem.2019.125277>
- Epova, E. N., Bérail, S., Séby, F., Vacchina, V., Bareille, G., Médina, B., Sarthou, L., & Donard, O. F. X. (2019). Strontium elemental and isotopic signatures of Bordeaux wines for authenticity and geographical origin assessment. *Food Chemistry*, 294, 35–45. <https://doi.org/10.1016/j.foodchem.2019.04.068>
- FAO Intergovernmental Group on Tea. (2015). World tea production and trade. Current and future development (I4480E/1/03.15; p. 17). Food and Agriculture Organization of the United Nations.
- Feng, Z., Li, Y., Li, M., Wang, Y., Zhang, L., Wan, X., & Yang, X. (2019). Tea aroma formation from six model manufacturing processes. *Food Chemistry*, 285, 347–354. <https://doi.org/10.1016/j.foodchem.2019.01.174>
- Fernández-Cáceres, P. L., Martín, M. J., Pablos, F., & González, A. G. (2001). Differentiation of Tea (*Camellia sinensis*) Varieties and Their Geographical Origin According to their Metal Content. *Journal of Agricultural and Food Chemistry*, 49(10), 4775–4779. <https://doi.org/10.1021/jf0106143>
- Galy, A., France-Lanord, C., & Derry, L. A. (1999). The strontium isotopic budget of Himalayan rivers in Nepal and Bangladesh. *Geochimica et Cosmochimica Acta*, 63(13–14), 1905–1925. [https://doi.org/10.1016/S0016-7037\(99\)00081-2](https://doi.org/10.1016/S0016-7037(99)00081-2)
- Gomes, D. A. S., Alves, J. P. dos S., da Silva, E. G. P., Novaes, C. G., Silva, D. S., Aguiar, R. M., Araújo, S. A., dos Santos, A. C. L., & Bezerra, M. A. (2019). Evaluation of metal content in tea samples commercialized in sachets using multivariate data analysis techniques. *Microchemical Journal*, 151, 104248. <https://doi.org/10.1016/j.microc.2019.104248>
- Guo, X., Schwab, W., Ho, C.-T., Song, C., & Wan, X. (2022). Characterization of the aroma profiles of oolong tea made from three tea cultivars by both GC-MS and GC-IMS. *Food Chemistry*, 376, 131933. <https://doi.org/10.1016/j.foodchem.2021.131933>
- Kennedy, S. P. (n.d.). Chapter 8. Coffee and tea fraud. *Food Fraud*, 12.
- Lagad, R. A., Alamelu, D., Laskar, A. H., Rai, V. K., Singh, S. K., & Aggarwal, S. K. (2013). Isotope signature study of the tea samples produced at four different regions in India. *Analytical Methods*, 5(6), 1604. <https://doi.org/10.1039/c3ay26142e>
- Lin, J., Zhang, P., Pan, Z., Xu, H., Luo, Y., & Wang, X. (2013). Discrimination of oolong tea (*Camellia sinensis*) varieties based on feature extraction and selection from aromatic profiles analysed by HS-SPME/GC-MS. *Food Chemistry*, 141(1), 259–265. <https://doi.org/10.1016/j.foodchem.2013.02.128>
- Liu, H.-C., You, C.-F., Chen, C.-Y., Liu, Y.-C., & Chung, M.-T. (2014). Geographic determination of coffee beans using multi-element analysis and isotope ratios of boron and strontium. *Food Chemistry*, 142, 439–445. <https://doi.org/10.1016/j.foodchem.2013.07.082>
- Liu, H.-C., You, C.-F., Huang, K.-F., & Chung, C.-H. (2012). Precise determination of triple Sr isotopes ($\delta^{87}\text{Sr}$ and $\delta^{88}\text{Sr}$) using MC-ICP-MS. *Talanta*, 88, 338–344. <https://doi.org/10.1016/j.talanta.2011.10.050>
- Liu, W., Chen, Y., Liao, R., Zhao, J., Yang, H., & Wang, F. (2021). Authentication of the geographical origin of Guizhou green tea using stable isotope and mineral element signatures combined with chemometric analysis. *Food Control*, 125, 107954. <https://doi.org/10.1016/j.foodcont.2021.107954>
- Luyckx, D. M. A. M., & van Ruth, S. M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry*, 107(2), 897–911. <https://doi.org/10.1016/j.foodchem.2007.09.038>
- Martin, J., Bareille, G., Bérail, S., Pecheyran, C., Davaert, F., Bru, N., Tabouret, H., & Donard, O. (2013). Spatial and temporal variations in otolith chemistry and relationships with water chemistry: A useful tool to distinguish Atlantic salmon *Salmo salar* parr from different natal streams: Otolith chemistry of *Salmo salar*. *Journal of Fish Biology*, 82(5), 1556–1581. <https://doi.org/10.1111/jfb.12089>
- Nasr, E. G., Epova, E. N., de Diego, A., Souissi, R., Hammami, M., Abderrazak, H., & F. X. Donard, O. (2021). Trace Elements Analysis of Tunisian and European Extra Virgin Olive Oils by ICP-MS and Chemometrics for Geographical Discrimination. *Foods*, 11(1), 82. <https://doi.org/10.3390/foods11010082>
- Nenadis, N., Heenan, S., Tsimidou, M. Z., & Van Ruth, S. (2016). Applicability of PTR-MS in the quality control of saffron. *Food Chemistry*, 196, 961–967. <https://doi.org/10.1016/j.foodchem.2015.10.032>
- Ni, K., Wang, J., Zhang, Q., Yi, X., Ma, L., Shi, Y., & Ruan, J. (2018). Multi-element composition and isotopic signatures for the geographical origin discrimination of green tea in China: A case study of Xihu Longjing. *Journal of Food Composition and Analysis*, 67, 104–109. <https://doi.org/10.1016/j.jfca.2018.01.005>
- Olivares, A., Dryahina, K., Navarro, J. L., Smith, D., Španěl, P., & Flores, M. (2011). SPME-GC-MS versus Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) Analyses for the Study of Volatile Compound Generation and Oxidation Status during Dry Fermented Sausage Processing. *Journal of Agricultural and Food Chemistry*, 59(5), 1931–1938. <https://doi.org/10.1021/jf104281a>
- Ozcan-Sinir, G. (2020). Detection of adulteration in extra virgin olive oil by selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics. *Food Control*, 118, 107433. <https://doi.org/10.1016/j.foodcont.2020.107433>
- Patel, P. K., Zhang, D., Borthakur, D., Hazarika, M., Boruah, P., Barooah, R., Sabhapondit, S., Neog, N. J., & Gogoi, R. C. (2019). Quality Green Tea (*Camellia sinensis* L.) Clones Marked through Novel Traits. *Beverages*, 5(4), 63. <https://doi.org/10.3390/beverages5040063>
- Reyrolle, M., Ghislain, M., Bru, N., Vallverdu, G., Pigot, T., Desauziers, V., & Le Behec, M. (2022). Volatile fingerprint of food products with untargeted SIFT-MS data coupled with mixOmics methods for profile discrimination: Application case on cheese. *Food Chemistry*, 369, 130801. <https://doi.org/10.1016/j.foodchem.2021.130801>
- Shitandi, Anakalo, A., Muigai Ngure, F., & Mahungu, S. M. (2013). Tea Processing and its Impact on Catechins, Theaflavin and Thearubigin Formation. In *Tea in Health and Disease Prevention* (pp. 193–205). Elsevier. <https://doi.org/10.1016/B978-0-12-384937-3.00016-1>
- Song, B.-Y., Gautam, M. K., Ryu, J.-S., Lee, D., & Lee, K.-S. (2015). Effects of bedrock on the chemical and Sr isotopic compositions of plants. *Environmental Earth Sciences*, 74(1), 829–837. <https://doi.org/10.1007/s12665-015-4087-2>

The Official Website of Tea Board India. (n.d.). Retrieved April 30, 2022, from <https://www.teaboard.gov.in/>

Van Kerrebroeck, S., Vercammen, J., Wuyts, R., & De Vuyst, L. (2015). Selected Ion Flow Tube–Mass Spectrometry for Online Monitoring of Submerged Fermentations: A Case Study of Sourdough Fermentation. *Journal of Agricultural and Food Chemistry*, 63(3), 829–835. <https://doi.org/10.1021/jf505111m>

Waight, T., Baker, J., & Peate, D. (2002). Sr isotope ratio measurements by double-focusing MC-ICPMS: Techniques, observations and pitfalls. *International Journal of Mass Spectrometry*, 221(3), 229–244. [https://doi.org/10.1016/S1387-3806\(02\)01016-3](https://doi.org/10.1016/S1387-3806(02)01016-3)

Walther, B. D., & Limburg, K. E. (2012). The use of otolith chemistry to characterize diadromous migrations. *Journal of Fish Biology*, 81(2), 796–825. <https://doi.org/10.1111/j.1095-8649.2012.03371.x>

Wang, H., Cao, X., Yuan, Z., & Guo, G. (2021). Untargeted metabolomics coupled with chemometrics approach for Xinyang Maojian green tea with cultivar, elevation and processing variations. *Food Chemistry*, 352, 129359. <https://doi.org/10.1016/j.foodchem.2021.129359>

Wang, X., & Tang, Z. (2020). The first large-scale bioavailable Sr isotope map of China and its implication for provenance studies. *Earth-Science Reviews*, 210, 103353. <https://doi.org/10.1016/j.earscirev.2020.103353>

Yener, S., Sánchez-López, J. A., Granitto, P. M., Cappellin, L., Märk, T. D., Zimmermann, R., Bonn, G. K., Yeretzian, C., & Biasioli, F. (2016). Rapid and direct volatile compound profiling of black and green teas (*Camellia sinensis*) from different countries with PTR-ToF-MS. *Talanta*, 152, 45–53. <https://doi.org/10.1016/j.talanta.2016.01.050>

Zhang, L., Ho, C., Zhou, J., Santos, J. S., Armstrong, L., & Granato, D. (2019). Chemistry and Biological Activities of Processed *Camellia sinensis* Teas: A Comprehensive Review. *Comprehensive Reviews in Food Science and Food Safety*, 18(5), 1474–1495. <https://doi.org/10.1111/1541-4337.12479>

Zhang, L., Zeng, Z., Zhao, C., Kong, H., Lu, X., & Xu, G. (2013). A comparative study of volatile components in green, oolong and black teas by using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry and multivariate data analysis. *Journal of Chromatography A*, 1313, 245–252. <https://doi.org/10.1016/j.chroma.2013.06.022>

Zhao, H., Zhang, S., & Zhang, Z. (2017). Relationship between multi-element composition in tea leaves and in provenance soils for geographical traceability. *Food Control*, 76, 82–87. <https://doi.org/10.1016/j.foodcont.2017.01.006>

Zhou, Z., Wu, Q., Yao, Z., Deng, H., Liu, B., Yue, C., Deng, T., Lai, Z., & Sun, Y. (2020). Dynamics of ADH and related genes responsible for the transformation of C 6 -aldehydes to C 6 -alcohols during the postharvest process of oolong tea. *Food Science & Nutrition*, 8(1), 104–113. <https://doi.org/10.1002/fsn3.1272>

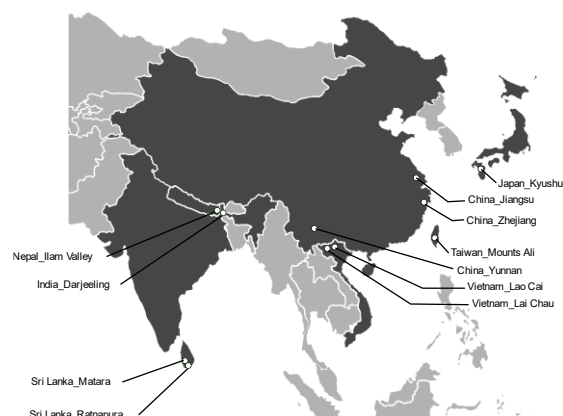


Figure S1: A schematic map of Asia map with indication of sample's origins.

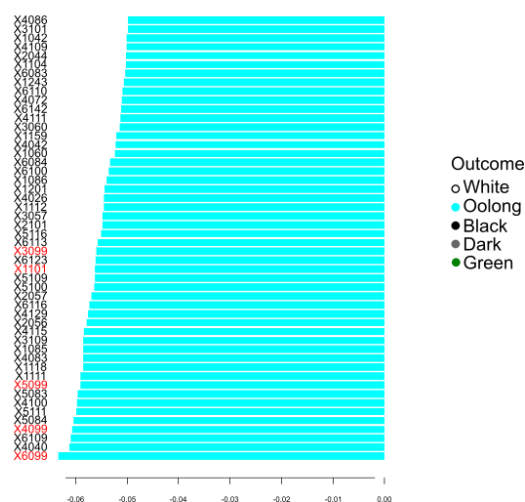


Figure S13: Plot Loading of the 40 most discriminative variables of the first component of the sPLS DA according the transformation process. Variables corresponding to hexanal product ions are in red (X6099: O²⁺- m/z 99, X4099: O⁺- m/z 99, X5099: OH- m/z 99, X1101: H3O⁺ m/z 101, X3099: NO⁺ m/z 99)

Sample	Country	Region	B (µg/kg)		Na (mg/kg)		Mg (mg/kg)		Al (mg/kg)		K (mg/kg)		Ca (mg/kg)		V (µg/kg)		Cr (µg/kg)		Mn (mg/kg)		Fe (mg/kg)		Co (µg/kg)		Ni (mg/kg)		Cu (mg/kg)	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Tea-1	Nepal	Ilam Valley	15521	466	20,4	0,5	2068	32	256	11	19285	482	1661	99	65,6	16,9	138	28	158	8	95,6	8,5	190	30	9,94	0,19	15,0	0,4
Tea-2	Nepal	Ilam Valley	15950	479	20,5	0,5	2141	33	300	13	18983	475	1670	100	49,1	12,7	171	34	186	9	107	10	213	33	10,8	0,2	16,3	0,5
Tea-3	Nepal	Ilam Valley	17337	520	19,3	0,5	2081	32	262	11	20061	502	1884	113	38,2	9,9	345	69	224	11	88,5	7,9	230	36	10,2	0,2	16,0	0,5
Tea-4	China	Zhejiang	17322	520	33,3	0,8	1840	28	301	13	18659	466	3100	185	87,9	22,7	152	30	289	15	98,7	8,8	111	17	7,46	0,15	11,8	0,4
Tea-5	China	Zhejiang	18920	568	20,1	0,5	1921	30	291	13	20145	504	2420	145	47,8	12,3	126	25	1007	111	118	11	137	22	6,45	0,13	17,7	0,5
Tea-6	China	Zhejiang	18182	545	20,4	0,5	1864	29	213	9	19654	491	2342	140	56,7	14,6	195	39	295	15	81,9	7,3	124	19	5,34	0,10	11,0	0,3
Tea-7	Japan	Kyushu	14208	425	114	3	1937	30	328	15	19478	489	1975	125	207	53	206	72	266	14	170	16	51,1	8,0	7,21	0,14	8,11	0,25
Tea-8	Japan	Kyushu	16438	224	20,6	0,5	1931	30	177	8	19811	495	1896	113	95,4	24,6	82,6	16,5	242	12	91,4	8,2	48,4	7,6	7,43	0,15	8,28	0,25
Tea-9	China	Jiangsu	19461	265	52,5	1,2	1521	24	332	14	14033	351	2889	173	523	135	857	171	809	41	174	16	805	127	12,2	0,2	24,3	0,7
Tea-10	China	Jiangsu	14968	204	20,7	0,5	1860	29	417	18	18685	467	2583	154	214	55	333	67	884	45	120	11	680	107	9,62	0,19	36,4	1,1
Tea-11	China	Jiangsu	18291	249	39,2	0,9	1422	22	230	10	15412	385	2097	125	489	126	1284	257	491	25	145	13	389	61	13,2	0,3	18,0	0,5
Tea-12	India	Darjeeling	13742	187	18,4	0,4	1843	29	433	19	20528	513	3024	181	141	36	142	28	347	18	90,5	8,1	134	21	4,37	0,09	13,9	0,4
Tea-13	India	Darjeeling	15462	210	21,4	0,5	2000	31	419	18	23721	593	3348	200	47,3	12,2	92,8	18,6	381	19	66,7	6,0	130	20	4,53	0,09	15,3	0,5
Tea-14	India	Darjeeling	15084	205	9,47	0,22	1869	29	378	16	19028	476	2619	156	< 26,0	/	574	115	425	22	72,1	6,4	133	21	5,83	0,11	14,8	0,4
Tea-15	Taiwan	Monts Ali	23972	326	12,6	0,3	1653	26	409	18	15725	393	2129	127	52,4	13,5	111	22	565	29	61,4	5,5	251	39	5,13	0,10	10,5	0,3
Tea-16	Taiwan	Monts Ali	16248	221	4,51	0,11	1910	30	671	29	17494	437	2806	168	< 26,0	/	97,4	19,5	786	40	54,4	4,9	225	35	3,99	0,08	8,7	0,3
Tea-17	Vietnam	Lai Chau	14738	201	13,2	0,3	2040	32	371	16	17761	444	3830	229	76,7	19,8	151	30	1601	148	80,6	7,2	120	19	5,68	0,11	15,0	0,5
Tea-18	Vietnam	Lai Chau	15858	216	10,5	0,2	1971	31	294	13	18433	461	2508	150	98,1	25,3	79,7	15,9	569	29	68,4	6,1	67,9	10,7	3,85	0,08	12,2	0,4
Tea-19	Vietnam	Lai Chau	15592	460	16,6	0,4	1869	29	478	21	17717	443	3704	221	137	35	261	52	573	29	104	9	80,2	12,6	5,95	0,12	15,0	0,4
Tea-20	Vietnam	Lai Chau	15820	475	15,0	0,4	1812	28	438	19	17185	430	3193	191	77,7	20,1	243	49	1320	140	86,6	7,7	100	16	5,94	0,12	21,3	0,6
Tea-21	Vietnam	Lao Cai	18997	570	16,8	0,4	1804	28	365	16	18964	474	2137	128	100	26	367	73	551	28	217	19	237	37	5,24	0,10	20,4	0,6
Tea-22	China	Yunnan	16597	498	7,40	0,18	1941	30	359	16	18647	466	1647	98	251	65	237	47	204	10	129	12	104	16	2,39	0,05	14,1	0,4
Tea-23	China	Yunnan	18086	543	10,5	0,2	1896	29	359	16	16389	410	2341	140	318	82	446	89	363	19	120	11	48,0	7,5	3,49	0,07	12,3	0,4
Tea-24	China	Yunnan	20293	609	6,87	0,16	1899	29	269	12	18323	458	2535	152	67,0	17,3	312	62	458	23	57,1	5,1	75,7	11,9	4,40	0,09	16,0	0,5
Tea-25	Sri Lanka	Ratnapura	13660	410	46,7	1,1	1983	31	381	16	19786	495	1797	107	84,7	21,9	156	31	279	14	93,4	8,3	115	18	3,58	0,07	16,6	0,5
Tea-26	Sri Lanka	Matara	16049	481	38,6	0,9	2164	33	442	19	21327	533	2327	139	117	30	189	38	356	18	115	10	144	23	3,52	0,07	16,4	0,5

Sample	Country	Region	Zn (mg/kg)		As (µg/kg)		Rb (mg/kg)		Sr (mg/kg)		Cd (µg/kg)		Ba (mg/kg)		Pb (µg/kg)		⁸⁷ Sr/ ⁸⁶ Sr	2SD**	Delta 87
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD			
Tea-1	Nepal	Ilam Valley	39,4	1,2	20,6	1,8	111	6	5,76	0,26	11,4	1,7	9,9	0,4	255	12	0,734680	0,000035	34,265
Tea-2	Nepal	Ilam Valley	40,6	1,2	21,2	1,9	112	6	6,57	0,29	9,62	1,39	10,9	0,4	225	11	0,735029	0,000035	34,756
Tea-3	Nepal	Ilam Valley	39,5	1,2	13,1	1,1	106,3	5,5	6,94	0,29	7,89	1,14	10,6	0,4	143	7	0,738525	0,000035	39,678
Tea-4	China	Zhejiang	47,0	1,4	57,6	5,0	52,0	2,7	6,55	0,28	15,9	2,3	7,97	0,31	164	8	0,714419	0,000035	5,742
Tea-5	China	Zhejiang	46,9	1,4	41,6	3,6	63,9	3,3	5,36	0,24	28,4	4,1	10,5	0,4	176	8	0,712806	0,000035	3,472
Tea-6	China	Zhejiang	40,9	1,2	193	17	62,0	3,2	5,01	0,20	14,6	2,1	5,81	0,23	161	8	0,714252	0,000035	5,507
Tea-7	Japan	Kyushu	41,6	1,3	19,6	1,7	41,6	1,9	5,78	0,22	11,4	1,6	7,11	0,28	63,4	3,0	0,707096	0,000035	-4,567
Tea-8	Japan	Kyushu	37,7	1,1	12,4	1,1	48,3	2,4	5,08	0,20	13,8	2,0	7,18	0,28	24,7	1,2	0,707092	0,000035	-4,573
Tea-9	China	Jiangsu	55,6	1,7	147	13	44,7	2,2	10,0	0,4	47,6	6,9	16,6	0,6	1182	56	0,711607	0,000035	1,783
Tea-10	China	Jiangsu	51,5	1,5	68,5	6,0	46,0	2,3	10,7	0,4	40,4	5,8	19,3	0,8	1049	50	0,710071	0,000035	-0,379
Tea-11	China	Jiangsu	53,6	1,6	134	12	45,9	2,4	7,24	0,31	22,6	3,3	13,7	0,5	787	38	0,711840	0,000035	2,111
Tea-12	India	Darjeeling	46,2	1,4	21,2	1,9	52,9	2,5	20,9	0,9	13,4	1,9	27,7	1,1	289	14	0,742317	0,000035	45,016
Tea-13	India	Darjeeling	31,3	0,9	15,3	1,3	60,5	3,0	20,2	0,9	11,4	1,6	25,9	0,9	376	18	0,740119	0,000035	41,923
Tea-14	India	Darjeeling	36,5	1,1	17,1	1,5	60,8	3,0	20,4	0,9	81,0	11,7	22,9	1,0	237	11	0,748200	0,000035	53,298
Tea-15	Taiwan	Monts Ali	24,2	0,7	5,63	0,49	14,1	0,7	5,13	0,22	26,2	3,8	9,65	0,37	92,2	4,4	0,708944	0,000035	-1,965
Tea-16	Taiwan	Monts Ali	21,8	0,7	6,94	0,61	16,3	0,8	7,19	0,32	24,8	3,6	15,3	0,6	89,6	4,3	0,708855	0,000035	-2,090
Tea-17	Vietnam	Lai Chau	32,2	1,0	21,1	1,8	118	6	11,1	0,5	25,5	3,7	4,15	0,16	103	5	0,713863	0,000035	4,960
Tea-18	Vietnam	Lai Chau	26,2	0,8	18,9	1,7	60,2	3,1	5,08	0,21	14,5	2,1	1,73	0,07	88,5	4,2	0,711842	0,000035	2,114
Tea-19	Vietnam	Lai Chau	39,2	1,2	49,6	4,3	82,6	4,4	8,67	0,39	38,8	5,6	2,45	0,10	210	10	0,711833	0,000035	2,102
Tea-20	Vietnam	Lai Chau	34,6	1,0	28,8	2,5	115	6	9,23	0,41	21,2	3,1	3,28	0,13	182	9	0,713880	0,000035	4,984
Tea-21	Vietnam	Lao Cai	56,2	1,7	60,0	5,2	249	12	10,9	0,5	27,2	3,9	9,4	0,4	383	18	0,755834	0,000035	64,046
Tea-22	China	Yunnan	24,9	0,7	43,2	3,8	51,8	2,6	6,32	0,27	7,03	1,02	12,6	0,5	317	15	0,729903	0,000035	27,541
Tea-23	China	Yunnan	26,4	0,8	49,9	4,4	51,0	2,5	12,5	0,5	7,65	1,11	10,9	0,4	84,8	4,0	0,722529	0,000035	17,160
Tea-24	China	Yunnan	27,5	0,8	12,2	1,1	61,0	3,2	11,5	0,5	7,32	1,06	15,0	0,6	95,0	4,5	0,724589	0,000035	20,060
Tea-25	Sri Lanka	Ratnapura	26,2	0,8	10,0	0,9	32,6	1,7	6,67	0,30	6,02	0,87	3,82	0,16	141	7	0,713474	0,000035	4,412
Tea-26	Sri Lanka	Matara District	21,5	0,6	11,3	1,0	35,0	1,8	8,32	0,35	12,9	1,9	7,02	0,27	185	9	0,715112	0,000035	6,717

Table SI2 : Element concentrations and isotope ratio of the 26 tea samples

CONCLUSIONS ET PERSPECTIVES

L'instrument SIFT-MS est un outil intéressant pour la quantification directe et rapide de COV dans l'air à un seuil de sensibilité bas ($\mu\text{g m}^{-3}$). Cependant, le risque d'ions conflits provenant de plusieurs analytes complique grandement la quantification des COV dans des matrices complexes. Ces travaux de thèse avaient donc pour but de développer une nouvelle utilisation de la technologie SIFT-MS par l'analyse non ciblée des COV couplée à des outils chimiométriques. Ce développement a permis l'acquisition et le traitement des « empreintes volatiles », constituées par l'ensemble des COV émis par des échantillons de fromage et de thé. La comparaison des empreintes volatiles a permis tout d'abord de valider cette approche en discriminant des échantillons de fromage Ossau-Iraty selon leur typicité et l'alimentation des brebis. L'exploitation des données a mis en évidence des composés potentiellement responsables de ces discriminations.

Dans un second temps, des mesures HS-SIFT-MS et HS-SPME-GC-MS/FID non ciblées ont été réalisées sur des échantillons de thé. Les empreintes volatiles acquises ont permis de les différencier en fonction de la zone de production associée à un cultivar mais également en fonction de leur couleur qui est liée au procédé de transformation. Des analyses ciblées « traditionnelles » par suivi de marqueurs ont également été réalisées sur les mêmes échantillons mais n'ont pas permis d'atteindre les mêmes performances pour discriminer les différents thés étudiés. Le choix des marqueurs est crucial et va orienter la résolution de la discrimination. Ainsi, l'analyse ciblée par HS-SPME-GC-MS/FID a bien permis de classer les échantillons par pays d'origine mais ces mêmes marqueurs n'ont pas permis de différencier les couleurs. L'analyse ciblée HS-SIFT-MS sur cette même liste de marqueurs n'a pas donné de résultats satisfaisants, confirmant les problématiques de quantification par cette technologie dans des matrices complexes riches en COV.

Cette étude comparative des analyses non ciblées nous a permis de mettre en évidence les avantages et les limites des deux instruments étudiés, dont les deux paramètres différenciants sont le temps de mesure et l'identification des molécules d'intérêt. En effet, le retraitement des profils volatils obtenus par HS-SPME-GC-MS/FID permet d'identifier les molécules discriminantes. Cependant, cette analyse est dix fois plus longue que celle réalisée par HS-SIFT-MS. Ainsi, l'identification des molécules clés est clairement en faveur de la technologie HS-SPME-GC-MS tandis que la rapidité de la mesure est en faveur de la technologie SIFT-MS. La force de l'un est la faiblesse de l'autre. Le choix de la méthode dépendra donc de la problématique à étudier.

Pour renforcer la traçabilité de la chaîne de production du thé, les empreintes volatiles des échantillons ont ensuite été combinées avec leurs signatures multi-élémentaires et leurs rapports isotopiques du strontium. L'analyse inorganique apporte des informations sur l'origine géographique, le lieu de production, les activités anthropiques et agricoles. Les empreintes volatiles SIFT-MS quant à elles, sont impactées par le « terroir », le cultivar et les procédés de transformation. Le couplage de ces trois analyses permet donc de répondre à des questions différentes sur la traçabilité et l'authenticité du thé.

Ces travaux de thèse ouvrent de nouvelles perspectives pour l'utilisation de la technologie SIFT-MS sur des matrices complexes. Le développement analytique réalisé permet d'obtenir des résultats intéressants, cependant des axes d'amélioration sont d'ores et déjà envisagés, notamment pour renforcer et automatiser les systèmes de contrôle de l'instrument, mais également pour comprendre

comment fusionner des données acquises sur une longue période ou sur plusieurs instruments. Une question se pose également quant à la gestion de ces « big data ». L'algorithme développé a fait l'objet d'un dépôt d'antériorité et pourrait être valorisé par la création d'un logiciel d'exploitation des données full scan SIFT-MS en implémentant les méthodes statistiques.

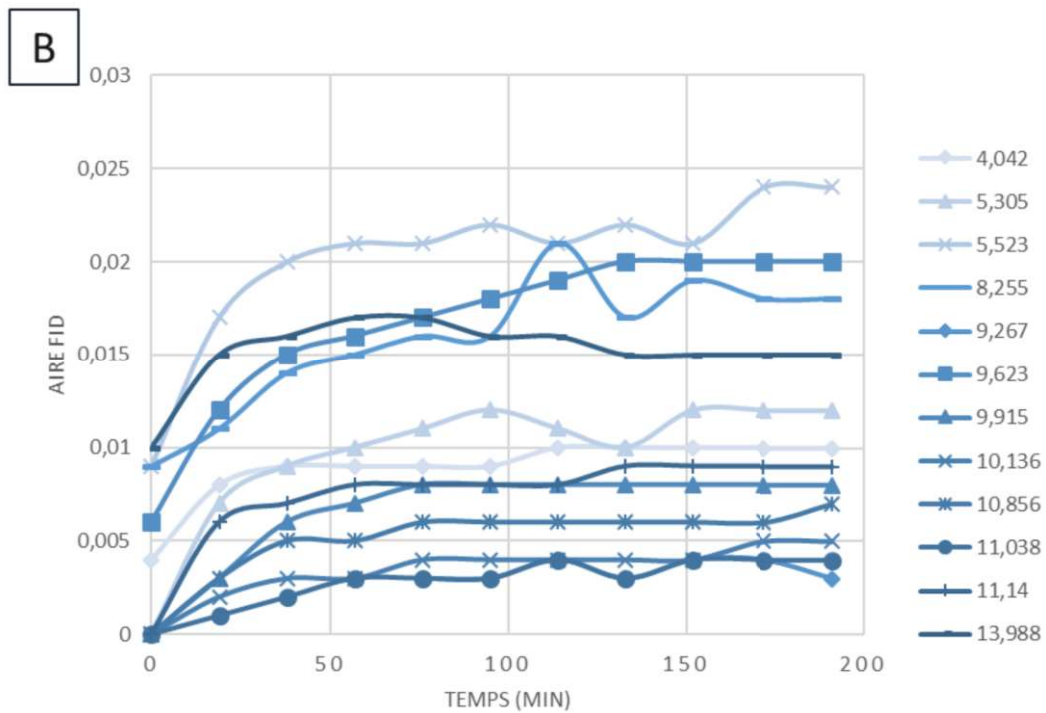
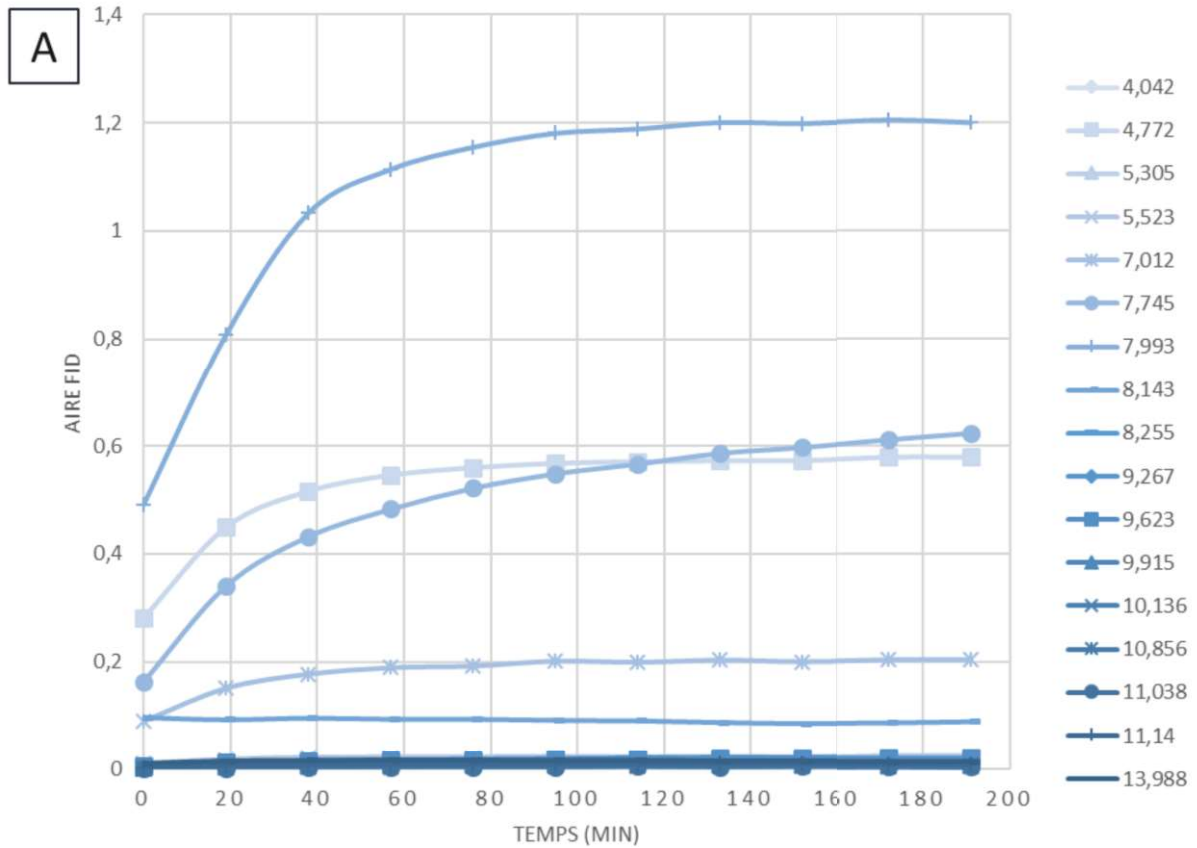
Les résultats obtenus sur les échantillons de thé ont mis en évidence l'impact prédominant des cultivars sur les empreintes volatiles qui pourrait être confirmé par des analyses phylogéniques. Le renforcement de la traçabilité du thé pourrait également être complété par d'autres techniques dont l'analyse des isotopes stables (C, H, O...), qui apporterait des informations sur les pratiques culturales.

Pour aller plus loin dans l'exploitation du potentiel des empreintes volatiles, des études avec de plus grandes banques d'échantillons seraient judicieuses. Ainsi, de nombreuses discussions sont en cours avec des acteurs du domaine agroalimentaire pour répondre à leurs problématiques. Par exemple, l'approche des empreintes volatiles développée dans ce travail est actuellement utilisée pour optimiser des paramètres de transformation de matrices végétales. Un autre axe d'application concerne la cartographie des typicités de fromages par leurs empreintes volatiles pour étudier leurs évolutions dans le temps et l'impact du réchauffement climatique sur la qualité du produit. La construction d'un modèle statistique pourrait être envisagée pour mettre en place un outil de prédiction et d'aide à la décision, venant renforcer les tests de conformité basés sur des analyses sensorielles.

Un des objectifs initiaux de cette thèse concernait la recherche de corrélation entre les empreintes volatiles et les profils sensoriels. Malheureusement, la réalisation d'analyses sensorielles solides par un jury expert est un défi qui n'a pas pu être réalisé pour être exploité dans le cadre de ces travaux. Une étude basée sur la construction d'un jury sensoriel expert permettrait de comparer ou de coupler les profils sensoriels aux empreintes volatiles pour une même banque d'échantillons.

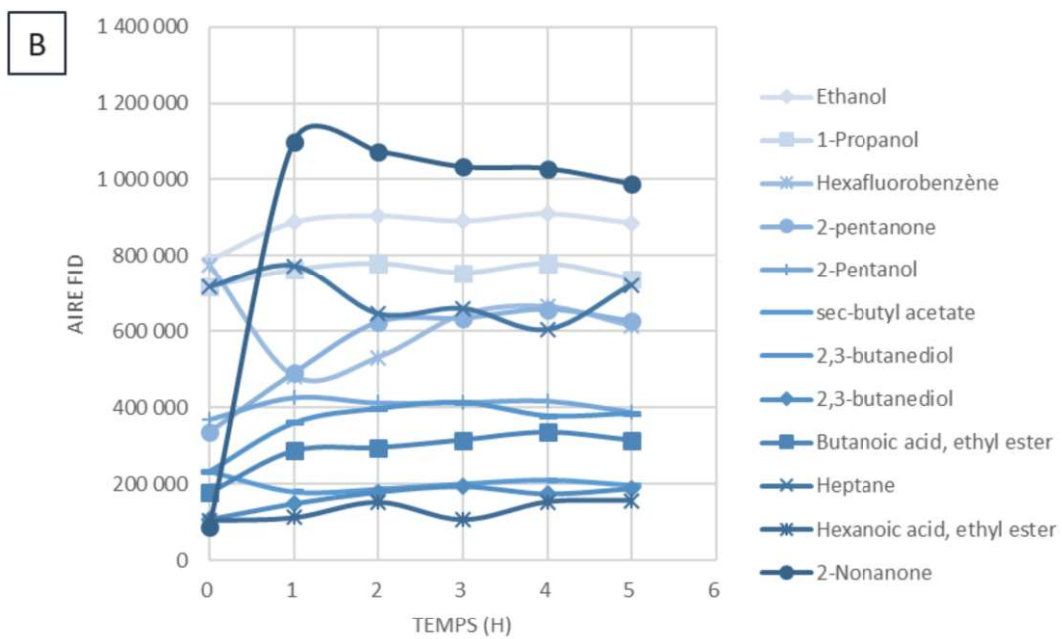
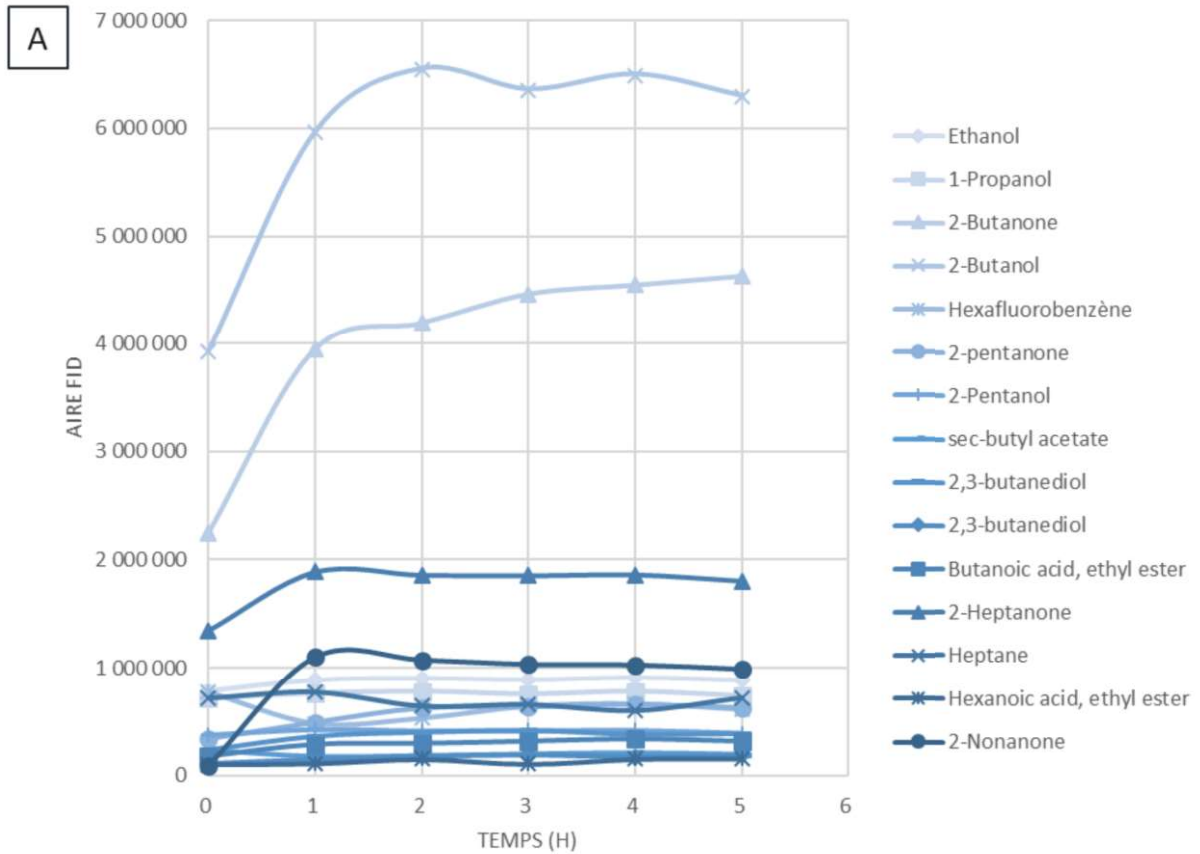
ANNEXE 1 : Résultats cinétique Ossau (VG-GC-FID)

A : Mesures cinétiques des COV émis par un échantillon de fromage Ossau-Iraty, par GC-FID équipé d'une Vanne à Gas. B : zoom sur les composés ayant une faible aire de pic FID.



ANNEXE 2 : Résultats cinétique Ossau (HS-SPME-GC-MS/FID)

A : Mesures cinétiques des COV émis par un échantillon de fromage Ossau-Iraty, par HS-SPME-GC-MS/FID. B : zoom sur les composés ayant une faible aire de pic FID.



ANNEXE 3 : Explication de l’algorithme de traitement des données SIFT-MS Full scan

Pour faciliter l’exploitation des données, un algorithme d’extraction nommé « **pysiftms** » a été développé. Il s’agit d’un algorithme programmé sous Python fournissant un ensemble d’outils pour lire et analyser les données brutes d’une série d’expériences SIFT-MS en mode full scan.

- **Cœur du module PYSIFTMS**

La classe « **pysiftms.core.Scan** » représente une expérience SIFT-MS en mode scan. Les masses des ions produits obtenues à partir des ions précurseurs positifs et négatifs ainsi que les blancs correspondants sont pris en compte. La classe a besoin d’un nom d’échantillon, d’accéder aux fichiers XML des données brutes correspondant aux scans du blanc et de l’échantillon, et à leurs identifiants. Les colonnes lues à partir des fichiers XML sont : "period", "datum", "icf", "reagent", "product", "num_id" (Tableau 1). Des colonnes supplémentaires peuvent être lues en les ajoutant dans l’algorithme, par exemple le temps de réaction : "reactionTime".

Paramètres	Signification
Period	Temps d’intégration de la mesure
Datum	Nombre moyen d’ion atteignant le détecteur en coups par seconde (cps)
ICF	Facteur de correction instrumentale quotidien
Reagent	Ion précurseur étudié
product	Ratio m/z de l’ion produit
num_id	Numéro identifiant les ions produits

Tableau 1: information sur les paramètres extrait des fichiers bruts

Le paramètre « num_id » est déterminé par une classe appelée « **pysiftms.core.Mass_ID** » utilisée pour définir un ID unique à chaque valeur de donnée pour un ion produit en fonction de l’ion réactif utilisé et du ratio masse/charge (m/z) mesuré. Chaque ion réactif est associé à un millier (Tableau 2). L’ID est obtenu en ajoutant à l’id du réactif la valeur du m/z sous forme de nombre entier. Par exemple, pour l’ion produit de m/z 107 avec l’ion réactif H₃O⁺, l’ID sera de 1 107 (1000 + 107). La fonction « **classmethod num_id()** » retourne l’identifiant numérique sous forme d’un nombre entier.

Ions réactifs	H ₃ O ⁺	O ₂ ⁺	NO ⁺	O ⁻	OH ⁻	O ₂ ⁻	NO ₂ ⁻	NO ₃ ⁻
ID	1000	2000	3000	4000	5000	6000	7000	8000

Tableau 2 : Valeur numérique en millier attribuée à chaque ion précurseur

Dans la classe « **pysiftms.core.Scan** », une suite de fonctions est appliquée afin d’obtenir un tableau de données.

La fonction « **_read_data()** » lit les données brutes des fichiers XML correspondant aux scans du blanc et de l’échantillon obtenus avec les ions positifs et négatifs .

La fonction « **property frequencies** » réalise le calcul de la fréquence pour chaque ion produit des blanc et des échantillons selon le calcul suivant : $F = \frac{datum}{Period} \times 10^6$.

La fonction « **property corrected_frequencies** » permet de calculer la fréquence corrigée. Divers calculs ont été étudiés et sont présentés dans la partie suivante : Correction du signal. Cette fonction retourne un cadre de données avec en colonne le nom des échantillons, les identifiants des ions produits en ligne et les fréquences corrigées comme valeurs.

La fonction « pivot_data(values_column) » a pour fonction de faire pivoter les colonnes et les lignes et retourne un jeu de données avec cette fois-ci en ligne le nom de l'échantillon, l'identifiant comme colonnes et les fréquences corrigées comme valeurs.

La fonction « static read_scan() » lit un fichier XML correspondant à un résultat d'expériences SIFT-MS en mode full scan et retourne un jeu de données contenant les valeurs brutes.

- **Déroulement du module PYSIFTMS**

Ce module fournit des classes permettant d'exporter et de recueillir des données sur une série d'expériences réalisées sur un ensemble d'échantillons. Les classes fournies sont :

- Classe Experiments, qui décrit une série d'expériences sur un ensemble d'échantillons
- Classe SampleScans, qui décrit un ensemble de scans SIFT-MS sur un échantillon donné
- Classe Replicat, qui contient les données associées à une réplique d'un échantillon.

La classe « pysiftms.workflow.Experiments() » décrit une série d'expériences sur plusieurs échantillons analysés par SIF-MS en fournissant l'emplacement des données brutes, le type d'expérience à considérer et d'autres métadonnées concernant des paramètres descriptifs sont fournis dans cette classe. Les différents paramètres à renseigner sont présentés dans le tableau 3 ci-dessous.

Paramètre	Signification
Path	chemin vers le stockage des données brutes
Name	Nom de l'expérience
Frequencies	Indiquer le calcul de fréquence à réaliser
Replicat	Vrai ou Faux, si Faux la moyenne des répliqués est calculée.

Tableau 3 : Information à compléter pour lire l'expérience

La classe « pysiftms.workflow.SampleScans » se focalise sur la description d'un échantillon pour lequel une expérience SIFT-MS en mode Scan a été effectuée avec des métadonnées qui peuvent être utilisées pour spécifier des informations sur l'expérience et sur l'échantillon. Les paramètres à renseigner pour chaque échantillon sont présentés dans le tableau 4

Paramètre	Correspondance
Name	Nom de l'échantillon
Path	chemin où sont stockées les données brutes
Répliqués	Listes des scans réalisés blanc et échantillons en ionisation positive et négative

Tableau 4 : paramètre à renseigner pour décrire l'échantillon

La classe « pysiftms.workflow.Replicat() » stocke les informations de chaque répliqué c'est-à-dire les numéros des fichiers de données brutes pour chaque analyse full scan (blanc et échantillon). Les fichiers bruts d'un ensemble de répliqués doivent être accessibles par un même chemin (Tableau 5).

Paramètres	Correspondance
Blank.pos	Numéro d'analyse du full scan du blanc en ionisation positive
Blank.neg	Numéro d'analyse du full scan du blanc en ionisation négative
Pos	Numéro d'analyse du full scan en ionisation positive
Neg	Numéro d'analyse du full scan de l'échantillon en ionisation négative

Tableau 5 : Attribution des numéros d'analyse pour chaque mesure

- **Construction du système**

Cette section décrit comment préparer un fichier d'entrée pour charger une série de données à partir de fichiers bruts, rassembler les données dans un tableau structuré puis les exporter en fichiers CSV.

La première étape concerne la construction de fichiers YAML (Yet Another Markup Language). Ces fichiers sont utilisés pour configurer les analyses et définir les options. Dans notre cas, il existe 2 fichiers :

- `samples.yml` : ce fichier est lié à une série d'expérience, il décrit les métadonnées des échantillons étudiés et les fichiers bruts XML associés à toutes les mesures full scan effectuées (Figure 1).

```
---
# The 7
name: G
path: The/
date: 17/10/2020
batch : Japon
diet : Kyushu
replicats:
  # 7-1
  - non-diluted:
    blank_pos: 36452
    blank_neg: 36453
    pos: 36500
    neg: 36501
  # 7-2
  - non-diluted:
    blank_pos: 36454
    blank_neg: 36455
    pos: 36502
    neg: 36503
  # 7-3
  - non-diluted:
    blank_pos: 36456
    blank_neg: 36457
    pos: 36504
    neg: 36505
---
```

Figure 1 : Exemple d'un script concernant les informations sur les échantillons et les analyses

- `build.yml` est le fichier d'entrée du script. Ce fichier rassemble l'ensemble des expériences réalisées en rassemblant toutes les options pour produire les fichiers CSV (Figure 2). Les paramètres disposant de plusieurs options sont « `replicats` » avec « `true` » pour avoir les valeurs des triplicats et « `false` » pour avoir les moyennes ; « `frequencies` » différents calculs ont été intégrés avec des niveaux de corrections différents. (voir paragraphe ci-dessous)

```
---
out_dir: data/
experiments:
  - path: Campagnes/Tea.yml
    name: Tea
    kind: non-diluted
    replicats : true
    frequencis : default
  - path: Campagnes/Tea.yml
    name: Tea
    kind: non-diluted
    replicats : False
    frequencis : default
```

Figure 2 : Exemple d'un script pour lire l'algorithme et extraire le jeu de données

Pour lire l'algorithme et obtenir le jeu de données final en CSV, il suffit d'exécuter la commande *build*, fournit lors de l'installation du paquet *pysiftms*, dans un terminal. Le jeu de données final est organisé en matrice notée X avec n lignes correspondant aux échantillons analysés et p colonnes correspondant aux intensités des ions du produit (valeurs de fréquence des triplicats ou moyenne des triplicats pour chaque échantillon). Des variables qualitatives associées au plan expérimental (par exemple, le pays, la région ou l'année de production, la couleur...) sont également ajoutées dans la matrice notée Y.

ABSTRACT

Gastronomy is an essential part of the French culture, with many authentic and high quality agri-food products. However, several sanitary scandals and an increase of fraud negatively impacted the consumer confidence, encouraging the sectors to improve the quality and traceability controls of their products. Volatile organic compounds (VOCs), naturally present in food products or generated during the transformation processes, are linked to the perception of odors and flavors by consumers. Thus, the VOCs composition of a food product is an interesting tool to characterize its authenticity, control its processing or evaluate the storage conditions. The aim of this thesis was to develop the volatile fingerprints approach, composed of all the VOCs emitted by a food product. This work was done with a selected ion flow tube mass spectrometer (SIFT-MS), which performances of speed, wide selectivity and high sensitivity were very interesting. A new approach based on the untargeted analysis of VOCs was developed on two different matrices: cheese and tea. The mining and processing of the large amounts of data required the development of an algorithm, combined with adapted chemometric tools. This work permitted to highlight the impact of the ewes' diets on the cheeses and the typicality provided by the two producers. The performances of the volatile fingerprints with SIFT-MS were then compared to those obtained with a separative techniques of gas chromatography coupled with mass spectrometry, on a selection of tea samples. The untargeted approach on both techniques allowed the discrimination of samples according to their origins, the cultivars and the transformation processes. The target measurements of markers were less successful, highlighting the crucial importance of the choice of these markers. Finally, this new approach of volatile fingerprints with SIFT-MS was combined with other traceability tools: inorganic multi element signature and strontium isotopic ratio. These analytical techniques provided supplementary information on the whole production and supply chain, strengthening the traceability and demonstrating the authenticity of the products.

The methodology of volatile fingerprints with SIFT-MS was successfully applied on complex matrices with a large number of VOCs. This approach is a new tool for decision-making aid for producers and sectors, to trace the authenticity and to control the transformation processes.

Keywords: Volatile Organic Compounds, VOCs, Mass Spectrometry, SIFT-MS, Volatile Fingerprints, Food science, Authenticity, Traceability.

RESUME

L'alimentation occupe une place essentielle dans la culture Française et s'appuie sur de nombreux produits agroalimentaires authentiques et de grande qualité. Cependant, plusieurs scandales sanitaires et l'augmentation des produits frauduleux ont ébranlé la confiance des consommateurs, incitant aujourd'hui les acteurs du secteur à renforcer les contrôles de la qualité et de la traçabilité de leurs produits. Les composés organiques volatils (COV), naturellement présents ou formés au cours de la transformation des produits agroalimentaires, sont responsables de leurs odeurs et de leurs saveurs, participant ainsi à la perception sensorielle par les consommateurs. Ainsi, déterminer la composition en COV d'un aliment, constitue donc un outil intéressant pour caractériser son authenticité, suivre un procédé et évaluer les conditions de stockage et de conservation. L'objectif de cette thèse est de développer l'approche des empreintes volatiles constituées de l'ensemble des COV émis par un produit alimentaire. Pour cela, un spectromètre de masse à injection directe (SIFT-MS Selected Ion Flow Tube Mass Spectrometer) a été choisi pour ses performances : sa rapidité d'analyse, sa capacité de détection et sa sensibilité. Une nouvelle exploitation d'analyse non ciblée des COV a été mise en place sur deux matrices : le fromage Ossau-Iraty et le thé. Pour traiter et exploiter les grands volumes de données obtenus par les empreintes volatiles, un algorithme a été mis en place ainsi que des outils chimiométriques adaptés. Ces travaux ont permis de mettre en évidence l'impact de la ration alimentaire des brebis sur le fromage ainsi que la typicité apportée par chacun des producteurs. Les performances de cette nouvelle approche ont ensuite été comparées à celles obtenues par une méthode séparative de chromatographie gazeuse couplée à la spectrométrie de masse, sur une collection d'échantillons de thé. L'approche non ciblée par ces deux techniques a permis de discriminer les différents thés en fonction de leurs origines, des cultivars et des procédés de transformation. Les mesures ciblées par suivi de marqueurs ont montré de moins bonnes performances, soulevant ainsi l'importance cruciale du choix des marqueurs étudiés. Enfin, cette nouvelle approche d'empreintes volatiles SIFT-MS a été combinée à d'autres outils de traçabilité : la signature multi élémentaire et le rapport isotopique du strontium. Ces techniques apportent des informations complémentaires sur l'ensemble de la chaîne de production du thé (origine de production, cultivar, procédé de transformation), permettant de renforcer la traçabilité et de démontrer l'authenticité des produits.

La méthodologie des empreintes volatiles par SIFT-MS a été appliquée avec succès sur des matrices complexes, riches en COV. Cette approche constitue un nouvel outil d'aide à la décision pour tracer l'authenticité et suivre des procédés de transformation afin d'accompagner les producteurs et les filières.

Mots clés : Composés Organiques Volatiles, COV, Spectrométrie de masse, SIFT-MS, Empreintes Volatiles, Sciences Agroalimentaires, authenticité, Traçabilité.