

École Doctorale Biologie Santé de Lille
THÈSE DE DOCTORAT

En vue de l'obtention du grade de
Docteur de l'Université

Soutenue publiquement par

Marie LENSKI

Le 27 juin 2023

**DÉVELOPPEMENT ET APPLICATION D'UNE APPROCHE DE
MÉTABOLOMIQUE POUR L'IDENTIFICATION DE MARQUEURS PRÉCOCES
DE TOXICITÉ DU TABAC, DE LA CIGARETTE ÉLECTRONIQUE
ET DU TABAC CHAUFFÉ**

Mention : Sciences de la Vie et de la Santé

Spécialité : Pharmacie en sciences du médicament et des autres produits de santé

Composition du jury :

Président

Pr David Balayssac

PU-PH – Université de Clermont-Ferrand

Rapporteurs

Pr Armelle Baeza

PU – Université de Paris Cité

Pr Patrick Emond

PU-PH – Université de Tours

Co-directeurs de thèse

Pr Delphine Allorge

PU-PH – Université de Lille

Dr Sébastien Anthérieu

MCF – Université de Lille

REMERCIEMENTS

Je remercie les membres du jury de thèse de m'avoir fait l'honneur d'accepter d'examiner ce travail. Je remercie Monsieur le Professeur David Balayssac, Madame le Professeur Armelle Baeza et Monsieur le Professeur Patrick Emond d'avoir accepté de juger mes travaux.

Je remercie mes directeurs de thèse Madame le Professeur Delphine Allorge et Monsieur le Docteur Sébastien Anthérieu de m'avoir aidée à mener à terme ce travail de recherche.

Je remercie Monsieur le Docteur Jean-Marc Lo-Guidice de m'avoir accueillie dans son équipe de recherche IMPECS (ULR4483) « Impact de l'environnement chimique sur la santé », ainsi que toute l'équipe d'IMPECS.

Je remercie Monsieur le Professeur Guillaume Garçon de m'avoir accueillie au sein du laboratoire de toxicologie et santé publique de la faculté de pharmacie de Lille. Merci également aux collègues enseignants.

Je remercie Monsieur le Professeur Aurélien Thomas et Monsieur le Professeur Jérôme Kluza avec qui j'ai eu la chance d'échanger lors de deux comités de suivi de thèse. Merci pour vos conseils avisés, vos critiques constructives et vos travaux inspirants.

Un merci particulier au Docteur Saïd Maallem pour son aide dans ce projet.

Je remercie l'ensemble des biologistes, ingénieurs, techniciens et secrétaires, de l'UF de toxicologie du Pôle de Biologie Pathologie Génétique CHU de Lille, pour leur soutien dans la réalisation de ces travaux.

Enfin, je remercie mes proches pour leur soutien.

TABLE DES MATIÈRES

TABLE DES FIGURES	9
TABLE DES TABLEAUX.....	11
ABRÉVIATIONS	13
VALORISATIONS SCIENTIFIQUES DES TRAVAUX	15
CONTEXTE	19
PARTIE I : INTRODUCTION.....	25
<i>CHAPITRE I : TABAGISME ET PRODUITS ALTERNATIFS.....</i>	<i>27</i>
1. Le tabagisme.....	27
2. La cigarette électronique	29
3. Le tabac chauffé.....	34
<i>CHAPITRE II : LA MÉTABOLOMIQUE.....</i>	<i>38</i>
1. Généralités.....	38
2. Approche ciblée et non ciblée	39
3. Techniques analytiques.....	41
3.1. Chromatographie liquide	42
3.2. Spectrométrie de masse	43
3.2.1. Source d'ionisation.....	43
3.2.2. Analyseur	44
3.2.3. Acquisition.....	44
3.3. La mobilité ionique	46
4. Sens biologique des données acquises	47
4.1. Traitement des données	48
4.1.1. Extraction des données.....	48
4.1.2. Pré-traitement des données.....	49
4.1.3. Analyses statistiques.....	50
4.2. Identification des composés détectés	51
4.3. Interprétation biologique.....	53
<i>CHAPITRE III : LA MÉTABOLOMIQUE EN TOXICOLOGIE</i>	<i>56</i>
1. Applications de la métabolomique.....	56
2. Métabolomique et exposition aux produits alternatifs du tabac	56
2.1. Métabolomique et exposition à la cigarette électronique.....	57
2.2. Métabolomique et exposition au tabac chauffé	64
<i>CHAPITRE IV : OBJECTIFS DE LA THÈSE.....</i>	<i>67</i>
PARTIE II : TRAVAUX PERSONNELS	69
<i>CHAPITRE I : DÉVELOPPEMENT ANALYTIQUE</i>	<i>71</i>
1. Mise au point d'une méthode d'UPLC-HRMS	71
1.1. Méthode chromatographique	74
1.1.1. Matériel	74
1.1.2. Type de pente	75
1.1.2.1. Méthode	75

1.1.2.2. Résultats	76
1.1.3. Durée du plateau en début de gradient.....	77
1.1.3.1. Méthode	77
1.1.3.2. Résultats	77
1.1.4. Composition de la phase mobile en début de gradient	78
1.1.4.1. Méthode	78
1.1.4.2. Résultats	78
1.1.5. Durée de la pente	79
1.1.5.1. Méthode	79
1.1.5.2. Résultats	79
1.2. Méthode de spectrométrie de masse	80
1.2.1. Matériel	80
1.2.2. Résultats	81
1.3. Conditions analytiques finales choisies	83
2. Création d'une librairie interne de référence	85
2.1. Matériel et méthode	85
2.1.1. Échantillons.....	85
2.1.2. Méthodologie d'indexation dans la librairie	86
2.2. Résultats	86
3. Création d'une librairie de métabolomique non ciblée	87
3.1. Matériel et méthode	89
3.1.1. Bases de données	89
3.1.2. Méthodologie de prédiction des CCS.....	89
3.1.3. Méthodologie de prédiction des temps de rétention.....	90
3.2. Résultats	91
3.2.1. Prédiction des CCS.....	91
3.2.2. Prédiction des temps de rétention.....	92
4. Discussion	94
CHAPITRE II : APPLICATION BIOLOGIQUE.....	98
1. Matériel et méthode.....	98
1.1. Cultures cellulaires et échantillonnage.....	98
1.1.1. Cultures cellulaires.....	98
1.1.2. Méthode d'exposition	99
1.1.2.1. Machine à fumer.....	99
1.1.2.2. Dispositifs d'exposition	100
1.1.2.3. Choix des durées d'exposition.....	100
1.1.3. Extraction des métabolites	101
1.2. Mesure de la cytotoxicité.....	102
1.3. Analyses de métabolomique	104
1.3.1. Analyse des échantillons.....	104
1.3.2. Pré-traitement des données en non ciblé.....	105
1.3.3. Statistiques	106
1.3.4. Sens biologique des données acquises	107
1.3.4.1. Identification	107
1.3.4.2. Interprétation biologique	108
2. Résultats	109
2.1. Mesure de la cytotoxicité.....	109
2.2. Métabolomique.....	110
2.2.1. Pré-traitement des données de métabolomique	110
2.2.1.1. Pré-traitement des données en mode d'ionisation positif	110

2.2.1.2. Pré-traitement des données en mode d'ionisation négatif.....	114
2.2.1.3. Association des données des deux modes d'ionisation.....	117
2.2.1.4. Statistiques.....	118
2.2.2. Sens biologique des données acquises.....	121
2.2.2.1. Identification.....	121
2.2.2.2. Interprétation biologique.....	126
3. Discussion.....	129
3.1. Composés exogènes.....	131
3.2. Composés endogènes.....	133
3.2.1. Stress oxydant.....	133
3.2.2. Métabolisme énergétique.....	134
3.2.3. Métabolisme lipidique.....	135
3.2.4. Autres métabolites.....	136
3.3. Conclusions.....	137
3.4. Limites.....	138
CONCLUSIONS GÉNÉRALES.....	141
RÉFÉRENCES.....	147
ANNEXE.....	167
PUBLICATION.....	177

TABLE DES FIGURES

Figure 1 : Evolution de la prévalence du tabagisme quotidien en France selon le sexe parmi les 18-75 ans, d'après (2)	27
Figure 2 : Schéma de l'atomiseur d'une cigarette électronique, d'après (11)	29
Figure 3 : Schéma des quatre générations d'e-cigarettes, d'après (13)	30
Figure 4 : Dispositif de tabac chauffé IQOS de PMI, d'après (31)	35
Figure 5 : La métabolomique dans la famille des « omiques » d'après (78).....	38
Figure 6 : Workflow d'une analyse de métabolomique ciblée et non ciblée.....	40
Figure 7 : Modules de l'UPLC-HRMS	42
Figure 8 : Schéma de l'électrospray d'après (86)	43
Figure 9 : Principe du mode d'acquisition MS ^E d'après (87).....	45
Figure 10 : Schéma de la séparation par mobilité ionique TWIMS d'après (92).....	46
Figure 11 : Configuration de l'analyseur à temps de vol Vion (Waters®) d'après (95)	47
Figure 12 : Etapes du traitement de données en métabolomique non ciblée	48
Figure 13 : Exemple de résultat obtenu lors de l'identification de la β-hydroxyéthylthéophylline A : chromatogramme, B : spectres à haute et basse énergie, C : les données acquises sont confrontées à celles de la base de données « maison » de référence	51
Figure 14 : Voies métaboliques de la base de données KEGG.....	54
Figure 15 : Performances chromatographiques évaluées	73
Figure 16 : Gradient de phase mobile initial et ses points d'optimisation	74
Figure 17 : Profils des pentes testés	76
Figure 18 : Influence de la pente sur le Tr des composés éluant après 4 min.....	76
Figure 19 : Influence de la durée du plateau initial sur la largeur à la base des pics (width).....	77
Figure 20 : Influence de la proportion de phase aqueuse/organique de la phase mobile sur le Tr	78
Figure 21 : Influence de la proportion de phase aqueuse/organique de la phase mobile sur la largeur à la base des pics (width)	79
Figure 22 : Influence de la durée de la pente sur le Tr	80
Figure 23 : Spectre de fragmentation en basse énergie (A) et haute énergie (B) de la kynurénine après analyse sur le G2-XS (A1 et B1) et le Vion (A2 et B2)	81
Figure 25 : Chromatogramme de l'arginine analysée avec le G2-XS (A) et le Vion (B)	82
Figure 25 : Taxonomie chimique des métabolites inclus dans la librairie interne de référence et proportion de chaque superclasse et classe	87
Figure 26 : Méthodologie de prédication de QSRR Automator (167)	91
Figure 27 : CCS prédit (Å ²) à l'aide de l'algorithme CCSBase en fonction du CCS mesuré (Å ²).....	92
Figure 28 : Tr (min) prédit à l'aide de l'algorithme QSRR Automator en fonction du Tr (min) mesuré	93
Figure 29 : Dispositif de génération de fumées et d'aérosols (Vitrocell®) et d'exposition des cultures cellulaires	99
Figure 30 : Méthode de mesure de l'activité de la LDH	102
Figure 31 : Abondance des composés pour chaque échantillon pour les données acquises en mode d'ionisation positif avant vs après pré-traitement des données	111
Figure 32 : Abondance des deux étalons internes pour les données acquises en mode d'ionisation positif avant vs après pré-traitement des données	112

Figure 33 : Score plots des analyses ACP pour les données acquises en mode d'ionisation positif	113
Figure 34 : Abondance des composés pour chaque échantillon pour les données acquises en mode d'ionisation négatif avant vs après pré-traitement	114
Figure 35 : Abondance de l'étalon interne pour les données acquises en mode d'ionisation négatif avant vs après pré-traitement	115
Figure 36 : Score plots des analyses ACP pour les données acquises en mode d'ionisation négatif	116
Figure 37 : Score plot de l'analyse ACP associant les données issues de l'acquisition en mode d'ionisation positif et négatif	117
Figure 38 : Score plot de l'analyse PLS-DA pour les groupes d'échantillons exposés (T1 et T2) des dispositifs MB-18W, MB-30W, IQOS et 3R4F	118
Figure 39 : Analyse heatmap des métabolites présentant un score VIP > 1,5	119
Figure 40 : Répartition des métabolites significatifs au test ANOVA parmi les différents types d'exposition	120
Figure 41 : Box plot de la nicotine pour les différents temps et dispositifs d'exposition	122
Figure 42 : Taxonomie chimique des métabolites identifiés pour l'exposition HTP et proportion de chaque superclasse et classe	126
Figure 43 : Diagramme de dispersion des voies métaboliques générées sur le logiciel MetaboAnalyst pour l'exposition HTP	127

TABLE DES TABLEAUX

Tableau 1 : Comparaison des approches analytiques utilisables en métabolomique	41
Tableau 2 : Quelques-unes des bases de données les plus utilisées pour l'annotation des composés issues d'analyses LC-MS non ciblées, adapté de (109)	52
Tableau 3 : Etudes de métabolomique traitant de l'exposition à la cigarette électronique et leurs conclusions principales	61
Tableau 4 : Etudes de métabolomique traitant de l'exposition au tabac chauffé et leurs conclusions principales	66
Tableau 5 : Gradient d'élution initialement choisi	71
Tableau 6 : Conditions analytiques de spectrométrie de masse	72
Tableau 7 : Composition de la solution de standards utilisée pour l'optimisation analytique	75
Tableau 8 : Comparaison des performances entre le Vion et le G2-XS	82
Tableau 9 : Gradient d'élution pour la méthode optimisée	84
Tableau 10 : Conditions analytiques de spectrométrie de masse pour la méthode optimisée	85
Tableau 11 : Performances du modèle SVR utilisant l'algorithme QSRR Automator	93
Tableau 12 : Protocole d'exposition	101
Tableau 13 : Ordre d'analyse des différents types d'échantillons d'après (109)	105
Tableau 14 : Critère d'identification pour chaque base de données	108
Tableau 15 : Cytotoxicité (%) en fonction du dispositif testé et de la durée d'exposition	109
Tableau 16 : Principe des analyses de PLS-DA et ANOVA et comparaison de leurs résultats	120
Tableau 17 : Métabolites identifiés et FC associés pour chaque exposition	123
Tableau 18 : Top 10 des voies métaboliques de la base KEGG obtenues par l'analyse fonctionnelle (MetaboAnalyst) des composés identifiés	127

ABRÉVIATIONS

3R4F	dispositif cigarette conventionnelle de référence
ACN	acétonitrile
ACP	analyse en composante principale
ANOVA	test statistique d'analyse de la variance
BPCO	bronchopneumopathie chronique obstructive
CCS	collision cross section ou section efficace de collision
e-cig	dispositif de cigarette électronique
ERO	espèce réactive de l'oxygène
ESI- ou ESI+	ionisation par électrospray négatif ou positif
ET	écart-type
FDR	false discovery rate ou taux de fausses découvertes
HCI	health canada intense
HILIC	chromatographie liquide à interaction hydrophile
HMDB	human metabolome database
HTP	heated tobacco product ou tabac chauffé
HAP	hydrocarbures aromatiques polycycliques
HO	hème oxygénase
HRMS	spectrométrie de masse haute résolution
ICC	corrélation intra-classe
IMS	spectrométrie de mobilité ionique
IQOS	dispositif de tabac chauffé commercialisé par Philip Morris International
KEGG	kyoto encyclopedia of genes and genomes
LDH	lactate déshydrogénase
LOESS	locally estimated scatterplot smoothing ou régression locale
LOQ	limite de quantification
LysoPC	lysophosphatidylcholine
LysoPE	lysophosphatidyléthanolamine
m/z	rapport masse/charge
MAD	écart absolu médian
MB-18W	dispositif modbox réglé à une puissance faible
MB-30W	dispositif modbox réglé à une puissance moyenne

MeOH	méthanol
MS	spectrométrie de masse
MS ^E	Mode d'acquisition des données de spectrométrie de masse
nAChR	récepteurs nicotiques de l'acétylcholine
OMS	organisation mondiale de la santé
PMI	Philip Morris International
PLA2	phospholipase A2
PLS-DA	partial least squares-discriminant analysis ou régression des moindres carrés partiels
QC	contrôle de qualité
QCdil	contrôle de qualité dilué
QSRR	quantitative structure retention relationships ou relation quantitative structure à rétention
RMN	résonance magnétique nucléaire
RSD	relative standard deviation ou coefficient de variation
SMILES	simplified molecular input line entry specification
SMPDB	the small molecule pathway database
S/N	ratio signal/bruit
TOF	time of flight ou détecteur à temps de vol
Tr	temps de rétention
TWIMS	spectrométrie de mobilité ionique à champ oscillant
UPLC	chromatographie liquide d'ultra-haute performance
VIP	variable importance in the prediction ou importance des variables dans les modèles prédictifs

VALORISATIONS SCIENTIFIQUES DES TRAVAUX

Articles et revues

1. **Lenski M**, Maallem S, Zarccone G, Garcon G, Lo Guidice JM, Anthérieu S, Allorge D. Prediction of a large-scale database of collision cross-section and retention time using machine learning to reduce false positive annotations in untargeted metabolomics, *Metabolites* 2023, 13, 282. <https://doi.org/10.3390/metabo13020282>.
2. **Lenski M**, Bruno C, Darrouzain F, Allorge D. Métabolomique : principes et applications en toxicologie biologique et médico-légale. *Toxicol Anal Clin.* 2023, 35 (2), 96–112. <https://doi.org/10.1016/j.toxac.2023.01.002>.
3. Zarrouk E, **Lenski M**, Bruno C, Thibert V, Contreras P, Privat K, Ameline A, Fabresse N. High-resolution mass spectrometry: Theoretical and technological aspects, *Toxicol Anal Clin.* 2022;34:3–18., <https://doi.org/10.1016/j.toxac.2021.11.002>.
4. **Lenski M**, Sidibé J, Gholam M, Hennart B, Dubath C, Augsburger M, Von Gunten A, Conus P, Allorge D, Thomas A, Eap CB. Metabolomic Alteration Induced by Psychotropic Drugs: Short-Term Metabolite Profile as a Predictor of Weight Gain Evolution. *Clin Transl Sci* 2021. <https://doi.org/10.1111/cts.13122>.

Communications orales

1. **Lenski M**, Zarccone G, Maallem S, Garcon G, Lo Guidice JM, Allorge D, Anthérieu S. Metabolomic strategy to identify markers of exposure and toxicity of electronic cigarettes, heated tobacco products and tobacco cigarettes, *International conference on the E-Cigarette: patterns of use and health impacts*, Paris, Décembre 2022.
2. **Lenski M**, Sidibé J, Gholam M, Hennart B, Dubath C, Augsburger M, Von Gunten A, Conus P, Allorge D, Thomas A, Eap CB. Metabolomic alteration induced by psychotropic drugs: Short-term metabolite profile as a predictor of weight gain evolution, *Congrès Français de Psychiatrie*, Lille, France, Décembre 2022.

3. **Lenski M.** Métabolomique ciblée ou non ciblée ? Exemple de la recherche de marqueurs précoces de toxicité des aérocontaminants émergents, *Journée des doctorants de l'ULR4483*, Avril 2022.
4. **Lenski M.** Développement et application d'une approche de métabolomique ciblée pour l'identification de biomarqueurs précoces de toxicité des aérocontaminants émergents, *Journée des doctorants de l'ULR4483*, Juin 2019.
5. **Lenski M**, Allorge D, Thomas A, Eap CB. Prise de poids induite par les psychotropes : contribution d'une méthode de métabolomique ciblée pour identifier de potentiels biomarqueurs prédictifs, *International congress of analytical, forensic and clinical toxicology*, Lille, Mai 2019.
6. **Lenski M**, La métabolomique, un outil prometteur en toxicologie, *Printemps des jeunes scientifiques de la SFTA*, Paris, Avril 2019.

Communications affichées

1. Platel A, Zarccone G, **Lenski M**, Martinez T, Talahari S, Garçon G, Nessler F, Lo Guidice JM, Anthérieu S. Comparison of cytotoxic, genotoxic and epigenetic effects of heated tobacco product, electronic cigarette and conventional cigarette emissions in human bronchial epithelial cells, *Congrès de la Société Française de Toxicologie*, Avignon, Novembre 2022.
2. **Lenski M**, Zarccone G, Maallem S, Garçon G, Lo Guidice JM, Allorge D, Anthérieu S. Metabolomic strategy to identify markers of exposure and toxicity of electronic cigarettes, heated tobacco products and tobacco cigarettes, *International Congress of Toxicology 2022*, Maastricht, Pays-Bas, Septembre 2022.
3. **Lenski M**, Maallem S, Anthérieu S, Allorge D. Collision cross-section and retention time prediction for untargeted metabolomic analyses, *Analytics 2022*, Nantes, Septembre 2022.
4. Zarccone G, **Lenski M**, Martinez T, Garçon G, Allorge D, Nessler F, Lo-Guidice J.-M, Platel A, Anthérieu S. Comparison of toxicity and transcriptomic effects of emissions from heated tobacco products, electronic cigarettes and tobacco cigarettes in human bronchial epithelial cells, *EuroTox 2021*, Visioconférence.

5. **Lenski M**, Allorge D, Anthérieu S. Development of a targeted metabolomic strategy to identify markers of exposure and toxicity of electronic cigarette and heated tobacco product, *Journée André Verbert*, Lille, Novembre 2020.
6. **Lenski M**, Allorge D, Thomas A, Eap CB. Weight gain induced by psychotropic drugs : contribution of a targeted metabolomic approach, *Congrès de la Société Française de Pharmacologie et Thérapeutique*, Lyon, Juin 2019.

CONTEXTE

CONTEXTE

Le tabagisme est un facteur de risque établi de diverses pathologies, dont le cancer du poumon, posant un problème majeur de santé publique. La mortalité attribuable au tabagisme est estimée à près de 7,7 millions de décès dans le monde chaque année. Le sevrage tabagique est aujourd'hui le seul moyen de supprimer ce facteur de risque. La cigarette électronique (e-cig) et le tabac chauffé (ou HTP pour Heated Tobacco Products) sont de nouveaux produits arrivés sur le marché depuis quelques années et parfois utilisés comme aides au sevrage tabagique.

Les e-cig sont des dispositifs sur batterie qui chauffent un liquide (ou e-liquide) pour générer un aérosol qui est inhalé par l'utilisateur. Les e-cig, dont le e-liquide peut contenir de la nicotine, sont utilisées en France métropolitaine de façon courante par environ 6,7 % des 18-75 ans (en 2021) et dans 99 % des cas par des fumeurs ou des ex-fumeurs. Cet usage est en augmentation chaque année depuis sa commercialisation en 2010, notamment car les e-cig sont généralement présentées comme moins nocives que les cigarettes conventionnelles du fait de l'absence de tabac et de combustion. La littérature est en faveur d'une toxicité de la e-cig moins importante par rapport à la cigarette conventionnelle, mais le risque absolu reste rarement évalué. Enfin, leur utilisation chez les jeunes reste une préoccupation majeure, avec notamment un risque de représenter un premier pas vers le tabagisme. Dans le cadre du sevrage tabagique, l'absence de tabac et de sensation de « hit » (picotement) au niveau de la gorge lors du passage des émissions encourage parfois à un retour à la cigarette.

L'industrie du tabac continue de développer de nouveaux produits, comme par exemple le tabac chauffé, pour procurer à l'utilisateur des sensations et goûts plus proches de la cigarette tout en fournissant un taux de nicotine satisfaisant pour les consommateurs. Philip Morris International (PMI) a notamment créé l'IQOS, commercialisée en France depuis 2017. Il s'agit d'un dispositif électronique permettant de chauffer un stick de tabac à une température bien inférieure à sa température de combustion, générant un aérosol qui est inhalé par l'utilisateur par aspiration. Ces produits sont vendus en tant que produits à risques réduits par rapport aux cigarettes conventionnelles du fait de l'absence de combustion du tabac. Bien que peu utilisé en France métropolitaine (environ 0,2 % de consommateurs parmi les 18-75 ans en

2018), il est plus largement utilisé dans d'autres pays d'Europe (comme l'Italie) ou d'Asie (comme la Corée du Sud ou le Japon). L'Organisation Mondiale de la Santé considère que toutes les formes de consommation de tabac sont nocives, y compris les produits du tabac chauffé. En outre, la littérature sur leur toxicité reste très insuffisante et la plupart des études sont issues de l'industrie du tabac ou du vapotage.

Ainsi, l'innocuité des e-cig ou du HTP n'est pas établie actuellement, du fait d'un manque d'études toxicologiques approfondies et indépendantes. Dans ce contexte, l'ULR4483 a développé depuis 2014 des méthodes d'analyse des principaux composants chimiques des fumées de cigarette et des aérosols d'e-cig ou de tabac chauffé générés par une machine à fumer, ainsi que des méthodes d'exposition de modèles cellulaires. Les études menées par l'équipe, qui ont reçu le soutien financier de l'Institut National du Cancer (INCa) et de l'Institut de Recherche en Santé Publique (IReSP), ont permis de mieux caractériser la composition chimique des aérosols et de déterminer leurs mécanismes de toxicité *in vitro* à court terme en comparaison avec la fumée de cigarette conventionnelle.

Les résultats issus de ces projets ont d'abord démontré que les émissions de tabac chauffé contenaient moins de composés toxiques (hydrocarbures aromatiques polycycliques (HAP), composés carbonylés et de métaux) que la fumée de cigarette conventionnelle, mais plus que les aérosols d'e-cig. Sur un modèle de cellules épithéliales bronchiques humaines (lignée BEAS-2B) cultivées à l'interface air-liquide, la cigarette est plus cytotoxique et induit plus de génotoxicité que le tabac chauffé, contrairement à l'e-cig qui n'a pas d'effet sur ces paramètres dans les conditions étudiées. L'e-cig peut cependant, comme les produits du tabac, induire un stress oxydant cellulaire et une réponse pro-inflammatoire. Concernant l'impact épigénétique de ces produits, aucune modification de la méthylation globale de l'ADN ni de modifications d'histones n'ont été mises en évidence. L'analyse pangénomique a révélé des profils transcriptomiques très semblables pour les cellules exposées aux deux produits du tabac. Sont notamment activées des voies de signalisation impliquées dans les processus de cancérogenèse (prolifération, invasion, survie cellulaire, glycolyse, stress oxydant...). L'e-cig, en revanche, induit moins de dérégulations transcriptomiques.

C'est dans ce contexte global multidisciplinaire, combinant des analyses toxicologiques par approches physico-chimiques, des expérimentations *in vitro* et des analyses « omiques », que se situe mon projet de thèse. L'utilisation d'approches globales telles que la métabolomique est complémentaire des approches déjà mises en œuvre. La métabolomique est définie comme l'étude qualitative et/ou quantitative de tous les métabolites mesurables produits par un organisme à la suite d'un stimulus physiologique, environnemental, pathologique, ou une variabilité génétique. La comparaison de l'expression des métabolites suite à l'exposition de cellules pulmonaires humaines exposées à des émissions de cigarette électronique, de tabac chauffé ou de cigarettes de tabac par rapport à des cellules non exposées permettra de définir un profil métabolomique spécifique de chaque exposition. L'exploration biologique de ces profils pourrait aider à mieux comprendre les mécanismes moléculaires et cellulaires de toxicité et à identifier des biomarqueurs pertinents d'effets et/ou d'exposition à ces aérocontaminants.

PARTIE I : INTRODUCTION

CHAPITRE I : TABAGISME ET PRODUITS ALTERNATIFS

1. Le tabagisme

Le *Global Burden of Disease Project* a estimé qu'environ 1,14 milliard de personnes fumaient en 2019 dans le monde (1). A l'échelle nationale, le tabagisme quotidien concernait 25,3 % des 18-75 ans en 2021 (27,8 % des hommes et 23 % des femmes) (2), plaçant la France comme l'un des pays avec les consommations les plus élevées parmi les pays occidentaux (3). La prévalence du tabagisme tend à diminuer dans le temps. Entre 2007 et 2019, elle était passée d'une moyenne mondiale de 22,7 % à 17,5 %, ce qui représente une réduction relative de 23 % sur 12 ans (4). A l'échelle nationale, les résultats du Baromètre Santé (Santé Publique France (2)) soulignent que la prévalence du tabagisme a également diminué en France ces dernières années, passant de 28,5 % de fumeurs quotidiens en 2014 à 24,0 % en 2019 (Figure 1). Depuis cette baisse, la prévalence s'est stabilisée en 2020 et 2021. Il est néanmoins à souligner que cette prévalence tend à augmenter chez les femmes, passant de 20,7 % en 2019 à 23 % en 2021.

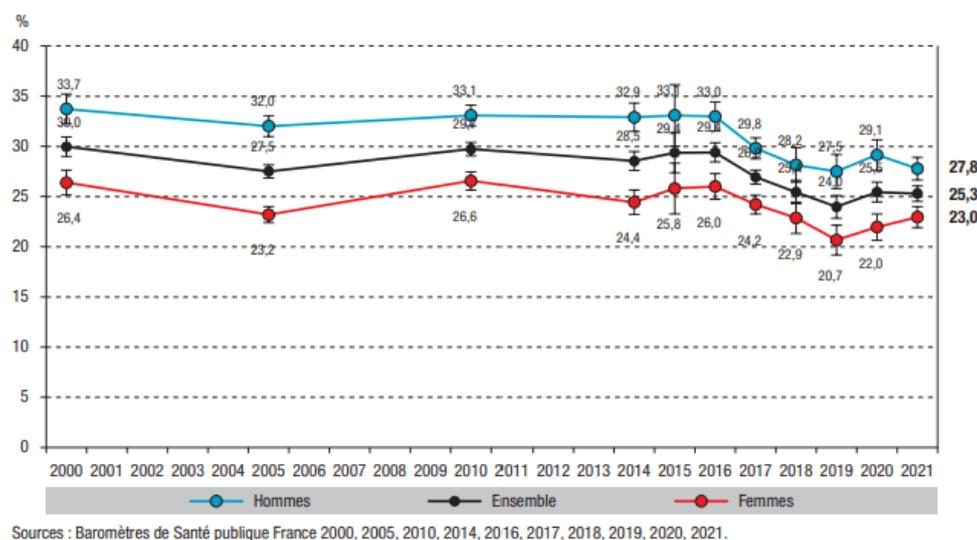


Figure 1 : Evolution de la prévalence du tabagisme quotidien en France selon le sexe parmi les 18-75 ans, d'après (2)

Le tabagisme est l'une des principales causes évitables de décès prématurés dans le monde, défini comme le décès avant 70 ans (5). Les fumeurs perdent en moyenne au moins une décennie de vie. La mortalité attribuable au tabagisme, proportionnelle à la durée et l'intensité des consommations passées, a été estimée à 7,7 millions de décès dans le monde en 2020, dont 1,2 million pour le tabagisme passif, et le tabagisme

provoque des maladies évitables chez 200 millions d'autres personnes (1). Au niveau national, la mortalité attribuable au tabagisme s'élevait à 75 000 décès en France en 2015, soit 13 % des décès survenus en France métropolitaine (6). Parmi les maladies attribuables au tabagisme figurent les maladies cardiaques et pulmonaires, les maladies respiratoires chroniques, le cancer et le diabète. Les principales causes de décès dus au tabac sont le cancer (poumon, voies aérodigestives supérieures, vessie, foie, rein), les pathologies cardiovasculaires (cardiopathies hypertensives et ischémiques, maladies cérébro-vasculaires) ou encore les maladies respiratoires (maladies chroniques des voies respiratoires inférieures) (6). En France, le tabagisme est le premier facteur de risque de cancer, responsable d'un cancer sur cinq et d'un décès sur trois par cancer, soit environ 45 000 décès par cancer par an.

Les cigarettes sont des dispositifs générant un aérosol après combustion, c'est-à-dire de la fumée, composée de plus de 6 000 produits chimiques (7,8). Une centaine de ces composés sont décrits comme toxiques, dont une cinquantaine sont classés comme cancérigènes, cancérigènes probables, ou cancérigènes possibles pour l'Homme (9) : composés carbonylés, composés organiques volatiles, nitrosamines, métaux lourds... Le sevrage tabagique est le seul moyen de supprimer ces facteurs de risque (5,10). Selon l'OMS, en 2019, plus de 60 % des fumeurs avaient l'intention d'arrêter et plus de 40 % avaient essayé d'arrêter au niveau mondial (11). En 2021, au niveau national, 30,3 % des fumeurs quotidiens avaient fait une tentative d'arrêt d'au moins une semaine au cours des 12 derniers mois (2). Le tabagisme est néanmoins difficile à inverser du fait du fort pouvoir addictogène de la nicotine délivrée. Trois classes de pharmacothérapie sont disponibles pour le sevrage tabagique : les substituts nicotiques qui sont des médicaments à base de nicotine (patchs, gommes, pastilles, mini-pastilles, vaporisateurs nasaux et inhalateurs à base de nicotine), la varénicline (un agoniste partiel des récepteurs nicotiques nAChR) et le bupropion (un inhibiteur de la recapture de la noradrénaline et de la dopamine, qui inhibe également la fonction des nAChR).

Depuis quelques années, de nouveaux produits de consommation imitant ou remplaçant la cigarette conventionnelle sont apparus sur le marché, comme la cigarette électronique ou le tabac chauffé. Ils sont souvent présentés par l'industrie du tabac ou du vapotage comme des solutions de substitution « plus sûres » ou « sans fumée » par rapport aux cigarettes classiques.

2. La cigarette électronique

Les e-cig (ou vapoteuses) sont des dispositifs sur batterie qui chauffent un liquide (ou e-liquide) pour générer un aérosol qui est inhalé par l'utilisateur. Au niveau mondial, la prévalence de son utilisation était estimée à 68 millions en 2020 (12). En 2021, en France métropolitaine, 38,7 % des adultes ont déclaré avoir déjà expérimenté l'e-cig. L'usage courant d'une vapoteuse a été déclaré par 6,7 % des 18-75 ans, et de façon quotidienne pour 5,0 % (5,7 % des hommes et 4,4 % des femmes), proportions en hausse continue depuis son lancement (notamment 2,7 % en 2017 *versus* 4,3 % en 2020) (2). 41,2 % des utilisateurs sont des fumeurs quotidiens (parmi lesquels 1/3 ne souhaite pas arrêter de fumer), 14,3 % des fumeurs occasionnels, 43,7 % des ex-fumeurs, et 0,9 % n'ont jamais fumé. Ainsi, 99 % des vapoteurs sont des fumeurs ou des ex-fumeurs.

L'e-cig est composée de 3 parties principales : une batterie, un réservoir et un atomiseur. La batterie permet de chauffer une résistance présente dans l'atomiseur (Figure 2). Le déclenchement de la chauffe peut être activé par un bouton pression ou par l'aspiration d'une bouffée. La résistance est en contact avec l'e-liquide présent dans le réservoir *via* un système de mèche. Ce dernier est chauffé et vaporisé, se mélange au flux d'air généré par l'utilisateur aspirant une bouffée au niveau de l'embout buccal, et se condense en aérosol.

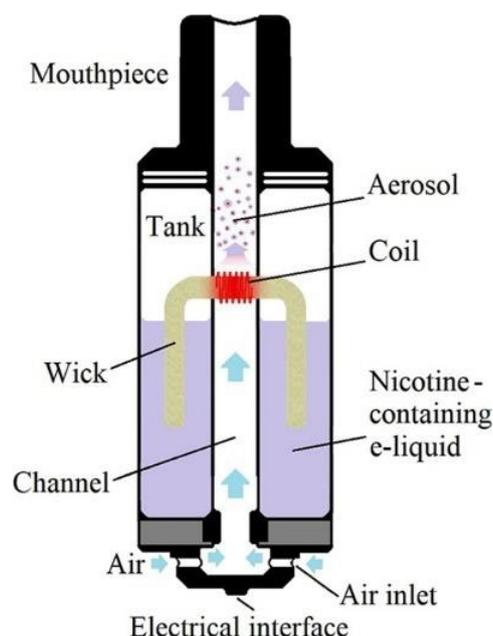


Figure 2 : Schéma de l'atomiseur d'une cigarette électronique, d'après (11)

Les e-cig varient considérablement en fonction de leur *design*, du voltage de la batterie, des e-liquides utilisés, ainsi que de leur capacité à moduler la délivrance de nicotine (Figure 3). À l'heure actuelle, plus de 30 000 marques d'inhalateurs électroniques de nicotine (dispositifs et e-liquides) sont commercialisées dans l'Union Européenne (4). Plusieurs « générations » d'e-cig existent et diffèrent selon la technologie utilisée et la manière dont elles sont utilisées.

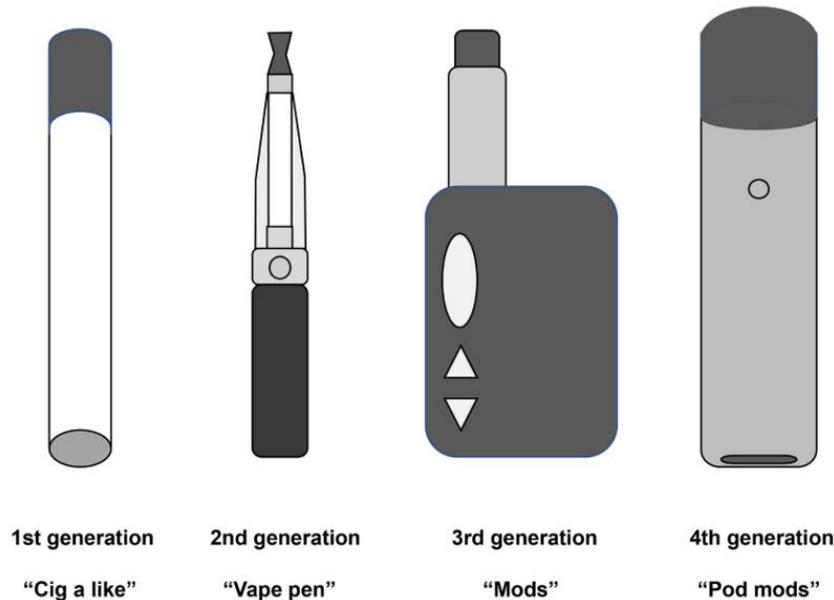


Figure 3 : Schéma des quatre générations d'e-cigarettes, d'après (13)

Les plus utilisés en France actuellement sont les modèles « Mod » qui sont modulables, aussi bien en termes de puissance que d'e-liquide à déposer dans un réservoir rechargeable. Les e-liquides utilisés contiennent ou non de la nicotine (mais pas de tabac), substance à fort potentiel addictif. Au niveau national, en 2021, 73,6 % des vapoteurs déclaraient utiliser ces dispositifs avec nicotine selon Santé Publique France (2). L'utilisateur peut également ajouter de la nicotine à un e-liquide formulé sans. En France, la nicotine des e-liquides est limitée à une concentration maximale de 20 mg/mL, et se présente sous forme de nicotine base (forme la plus fréquente actuellement), ou sous forme de sel de nicotine. Cette dernière permet de diminuer le pH du e-liquide par rapport à la nicotine base, plus proche du pH de la gorge, et ainsi offrir une meilleure tolérance à des concentrations plus importantes de nicotine. Plusieurs autres facteurs peuvent déterminer la quantité de nicotine à laquelle l'utilisateur est exposé. La puissance électrique générée dans le dispositif (variable par modification notamment du voltage et de la résistance) peut augmenter la quantité de nicotine contenue dans l'aérosol. Le profil de bouffée de l'utilisateur (durée, volume...)

a également le potentiel d'augmenter la prise de nicotine de l'utilisateur. Les autres composés principaux du e-liquide sont l'eau, le propylène glycol, le glycérol, et parfois des arômes. La caractérisation de la composition chimique des e-liquides et de leurs aérosols a été décrite précédemment (14,15), et a montré d'importantes variations pour les aérosols générés, en fonction des modalités d'utilisation (16). Au sein de l'ULR4483, 6 e-liquides ont été testés (17), confirmant la composition annoncée par le fabricant (en propylène glycol, glycérol et nicotine), mais aussi la présence d'éléments traces métalliques, pesticides, d'HAP et de composés carbonylés. Des analyses réalisées sur les aérosols générés par une e-cig de deuxième génération et avec ces 6 e-liquides ont également confirmé la présence d'éléments traces métalliques, d'HAP et de composés carbonylés à des concentrations 7 à 6100 fois inférieures à celles mesurées dans la fumée d'une cigarette de référence. Dans les conditions testées, l'e-cig exposerait donc les consommateurs à moins de composés potentiellement toxiques et en plus faible quantité que la cigarette conventionnelle. Il a également été démontré qu'avec un modèle Modbox de troisième génération, l'e-cig émettait plus de composés chimiques (HAP et métaux) à puissance moyenne (30 W), qu'à faible puissance (18 W) (18).

Une revue Cochrane suggère que les cigarettes électroniques avec nicotine augmentent plus les taux d'abandon du tabac (arrêt pendant au moins six mois) par rapport aux thérapies de substitution nicotinique et par rapport aux cigarettes électroniques sans nicotine (19). De plus, l'absence de combustion du tabac dans ces dispositifs expose le consommateur à moins de composés toxiques. Néanmoins, l'innocuité de ces produits, notamment à long terme, n'a pas été démontrée. Leur intérêt fait donc encore débat, aussi bien concernant leur impact sur la santé, la trajectoire de ses usagers, ou encore leur capacité à favoriser la renormalisation du tabagisme, c'est-à-dire 1) que fumer redevienne normal et désirable ; 2) que la visibilité des produits et comportements tabagiques augmente ; 3) que les habitudes tabagiques de la population soient maintenues et favorisées (20). Enfin leur utilisation chez les jeunes, et notamment celles des modèles jetables appelées « puffs », reste une préoccupation majeure, avec notamment un risque de représenter un premier pas vers le tabagisme (21).

Des évaluations expérimentales de la toxicité potentielle des aérosols ou des e-liquides sur des modèles *in vitro* et *in vivo* ont été réalisées (22). Dans ces deux cas,

une des limites principales est le manque de standardisation des types d'exposition, ne permettant pas toujours de comparer et généraliser les résultats obtenus (23).

Les études *in vitro* ont évalué la viabilité cellulaire, les dommages à l'ADN, la production d'espèces réactives de l'oxygène, ou encore de médiateurs pro-inflammatoires, après exposition aux aérosols en utilisant une variété d'approches et de lignées cellulaires, et diverses modalités de génération des aérosols (24). Des effets de réduction de la viabilité cellulaire ont pu être constatés (25). Certains auteurs ont pu décrire une augmentation de la sécrétion de médiateurs pro-inflammatoires (26–28) et du stress oxydant (29), ainsi que la réduction des défenses antimicrobiennes (30,31). Ceci est également observé avec les e-liquides à base d'arômes chimiques (32,33), ou plus spécifiquement avec les e-liquides contenant de la nicotine (34). Les dysfonctionnements endothéliaux entraînés par une augmentation de l'inflammation, du stress oxydant et de l'apoptose pourraient être des mécanismes participant à l'augmentation du risque de maladies cardiovasculaires (33) et être à l'origine de mutations et de l'induction de cancers (35). D'autres études ont également montré l'absence de différence entre cellules exposées et non exposées sur ces mêmes paramètres (36–39), mettant en évidence la difficulté de consensus sur ces effets.

Pour les études *in vivo*, des souris exposées à des aérosols d'e-cig durant leurs 10 premiers jours de vie présentaient une diminution de leur poids total et une diminution modérée de la prolifération cellulaire alvéolaire par rapport aux souris contrôles non exposées (40). Une autre étude sur des souris de 8 semaines exposées à des aérosols d'e-cig pendant 3 jours a montré une augmentation des cytokines pro-inflammatoires dans le liquide broncho-alvéolaire et une augmentation de marqueurs de stress oxydant au niveau du tissu pulmonaire (glutathion) (27). De façon complémentaire, l'exposition de souris aux aérosols d'e-cig pendant 4 semaines a montré une augmentation de la susceptibilité aux infections bactériennes, par diminution des défenses antimicrobiennes (31). Une équipe a étudié les effets cardiaques de l'exposition de souris aux aérosols d'e-cig pendant 2 semaines (41) et pendant 3 à 6 mois dans une autre étude (34). La première étude suggère une augmentation de l'angiogenèse, sans effet sur la contraction cardiaque ou la fibrose tissulaire. A contrario, l'exposition chronique a mis en évidence une augmentation des protéines pro-inflammatoires, une fibrose cardiaque, ainsi qu'une augmentation de la pression

sanguine et de la fréquence cardiaque. Ces études mettent ainsi en évidence les effets possibles de l'e-cig sur les systèmes respiratoire, cardiovasculaire et immunitaire. Le potentiel génotoxique et mutagène des aérosols d'e-cig testés à deux puissances (18 W et 30 W) et de la fumée de cigarette conventionnelle (3R4F) a été comparé sur un modèle murin au sein de l'ULR4483 (42). Plusieurs durées d'exposition ont permis d'évaluer les effets d'une exposition sub-aiguë (4 jours), sub-chronique (3 mois) et chronique (6 mois). Ainsi, l'e-cig à haute puissance et la cigarette 3R4F ont induit des dommages oxydatifs de l'ADN dans les poumons et le foie des souris exposées. Cette étude est en faveur d'un effet génotoxique des e-cig comparable à la cigarette conventionnelle.

Chez l'Homme, des effets sur la fonction pulmonaire (43,44), la fonction cardiaque (45) et une augmentation du stress oxydant et de l'inflammation (46) ont été mis en évidence. Dans certaines conditions particulières d'utilisation, des pneumopathies associées à l'utilisation de produits du vapotage (appelées EVALI) ont pu être observées en 2019 aux Etats-Unis (47). Il s'agit d'une pneumopathie aiguë pouvant engager le pronostic vital des patients. L'acétate de vitamine E, ajouté dans l'e-liquide d'huile de cannabis utilisée par la majorité des patients atteints, a été identifié comme le facteur causal. Néanmoins, au-delà de l'évaluation du risque absolu, l'évaluation du risque relatif par rapport aux fumeurs de cigarette a montré des résultats plus bénéfiques sur la fonction respiratoire (48), la bronchopneumopathie obstructive chronique (49), et l'asthme (50).

En conclusion, les effets *in vitro* ou *in vivo* restent partiellement étudiés et pas toujours concluants, justifiant la réalisation d'autres études pour examiner de manière plus approfondie les effets de ces aérosols sur les organismes vivants.

Du point de vue de l'utilisateur, les e-cig n'utilisent pas les véritables ingrédients du tabac et ne provoquent pas la sensation de « hit » au niveau de la gorge, encourageant parfois à un retour à la cigarette. Par conséquent, l'industrie du tabac continue de développer de nouveaux produits pour répondre aux besoins des consommateurs. La réponse la plus récente a été de lancer de nouveaux produits basés sur le chauffage et non la combustion du tabac, dits dispositifs de tabac chauffé.

3. Le tabac chauffé

Les HTP sont présentés comme étant moins nocifs que les cigarettes conventionnelles, tout en imitant les perceptions sensorielles classiques, offrant ainsi une alternative aux fumeurs. Il s'agit un dispositif électronique permettant de chauffer un stick de tabac à une température inférieure à sa température de combustion, générant un aérosol qui est inhalé par l'utilisateur par aspiration. Les HTP constituent une catégorie relativement nouvelle de produits du tabac, et seuls quelques produits sont disponibles sur le marché. Leur utilisation a émergé au Japon et en Corée du Sud, et s'est étendue progressivement en Europe (comme en Italie). Actuellement, les HTP sont commercialisés dans une quarantaine de pays et les systèmes IQOS, commercialisés par Philip Morris International (PMI), sont présents dans la plupart d'entre eux (51), et notamment en France. Dans ce dispositif, le tabac est chauffé jusqu'à 350°C au maximum (contre 900°C pour la cigarette conventionnelle). Il existe peu de données sur leur utilisation. En Europe, en 2020, 6,5 % des plus de 15 ans ont déclaré avoir déjà expérimenté l'HTP. L'usage actuel de ce dispositif y a été déclaré par 1,3 % des 18-75 ans, et de façon quotidienne pour 0,7 %, avec une prédominance pour les jeunes adultes (15-24 ans) (52). Une étude italienne parmi les plus de 15 ans a montré qu'en 2017, 45 % des utilisateurs italiens d'IQOS, et 51 % des personnes intéressées pour le tester, n'avaient jamais fumé (53). La trajectoire de ses usagers est donc une préoccupation majeure. Ainsi l'HTP se présenterait plutôt comme un produit d'entrée et de maintien du tabagisme plutôt qu'un produit de réduction des risques pour les fumeurs. En France métropolitaine, moins de 1 % des 18-75 ans déclarait en avoir déjà utilisé en 2018 (54). La prévalence de l'utilisation quotidienne en France en 2017 et 2018 était de l'ordre de 0,1 % à 0,2 % (54,55).

Le dispositif IQOS est composé de 3 éléments principaux (Figure 4A) : un stick de tabac, un « holder » et un chargeur qui sert de batterie portable. Le tabac se présente sous forme de bâtonnets/sticks appelés « heets », ressemblant fortement à des petites cigarettes traditionnelles. Il est composé d'un filtre en acétate de cellulose qui sert d'embout buccal, d'une zone de refroidissement de l'aérosol en acide polylactique (biopolymère), ainsi que de feuilles de tabac moulues et mixées à d'autres ingrédients (notamment du propylène glycol et du glycérol). Le tout est roulé pour obtenir un stick (Figure 4C). Le « holder » est composé d'une batterie permettant de chauffer une résistance en forme de lame (Figure 4B). Le déclenchement de la chauffe est activé

par un bouton pression. La résistance qui chauffe jusqu'à 350°C est en contact avec le bâtonnet de tabac. Ce dernier est chauffé et aérosolisé, et se mélange au flux d'air généré par l'utilisateur aspirant une bouffée au niveau de l'embout buccal. Après utilisation, le stick doit être enlevé et le « holder » rechargé avec le chargeur.

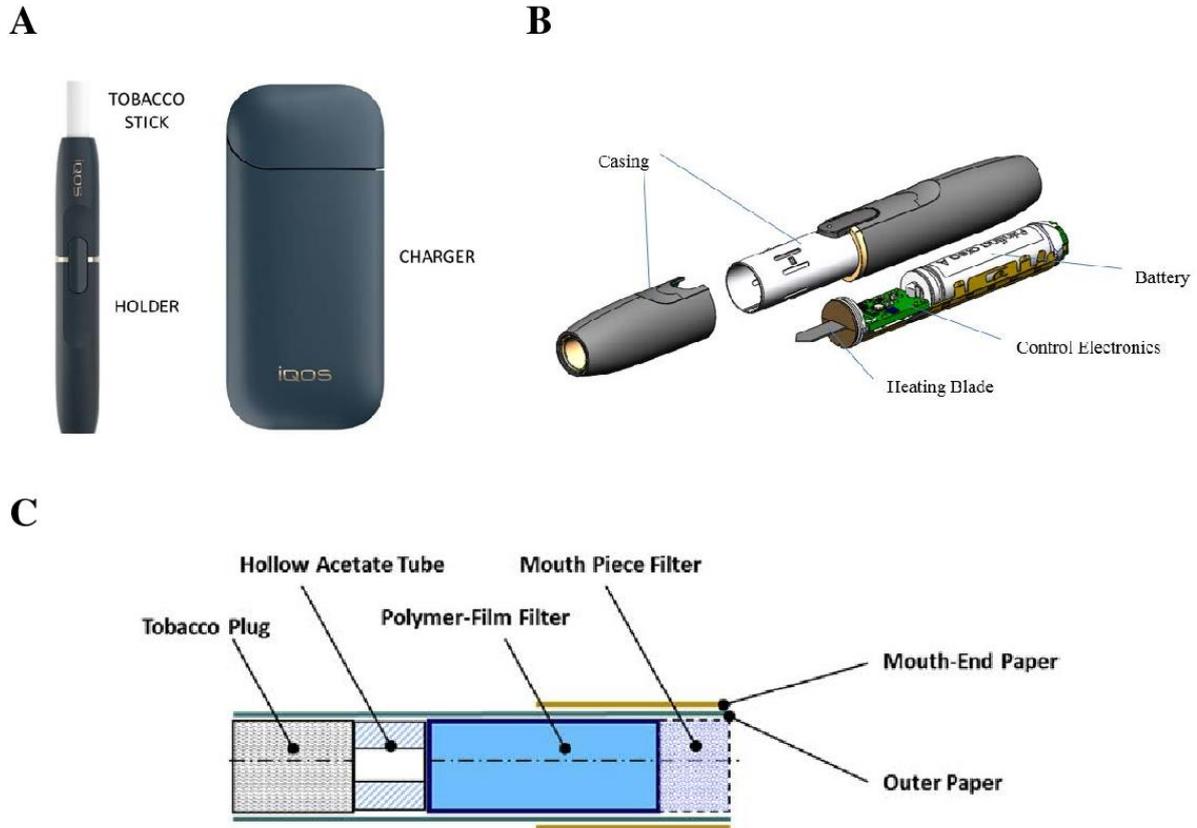


Figure 4 : Dispositif de tabac chauffé IQOS de PMI, d'après (31)

Le tabac contient de la nicotine et des additifs notamment pour aromatiser l'aérosol. Le profil de bouffées de l'utilisateur peut faire varier la quantité de nicotine apportée par l'HTP. Avec ce dispositif, les consommateurs sont incités à prendre des bouffées sur un temps très court avant l'arrêt automatique du dispositif (8 à 14 bouffées sur une durée d'environ 6 minutes). Ces bouffées répétées provoquent un pic de nicotine plus brutal et plus intense, conférant au HTP un fort potentiel addictif. La composition chimique des aérosols de tabac chauffé a été caractérisée précédemment (56,57). La plupart des composés toxiques ou potentiellement toxiques présents dans les aérosols d'HTP sont les mêmes que dans la fumée de cigarette conventionnelle, bien que la majorité soit en concentrations plus faibles et formés à des températures plus basses. L'HTP ne chauffe pas à plus de 350°C, mais les températures maximales atteintes au niveau du stick sont incertaines, et les points chauds locaux pourraient provoquer la formation de ces composés toxiques (58). Une étude de l'ULR4483 a montré que les

émissions d'HTP contiennent moins d'HAP, de composés carbonylés et de métaux que la fumée de cigarette conventionnelle, mais qu'elles en contiennent beaucoup plus que dans les aérosols d'e-cig (18). Une diminution des concentrations en composés toxiques ne peut pas être systématiquement traduite par une réduction proportionnelle du risque pour la santé humaine. En effet, des composés génotoxiques cancérigènes restent présents dans les émissions. De plus, une étude (indépendante de l'industrie du tabac) a montré que plus de 20 composés dangereux ou potentiellement dangereux sont retrouvés à des niveaux plus importants dans les aérosols d'HTP que dans la fumée de cigarette (59). L'Organisation Mondiale de la Santé considère que toutes les formes de consommation de tabac sont nocives, y compris les HTP (60). La société européenne de pneumologie conclut que, comme pour le tabagisme, les HTP créent une dépendance et sont cancérigènes pour l'Homme (61).

Une revue systématique des effets sur la santé a été publiée par Znyk *et al.* (62). Il a été démontré que l'exposition *in vitro* aux aérosols d'HTP induisait plus de cytotoxicité que pour les contrôles non exposés dans un modèle de cellules épithéliales bronchiques humaines cultivées en interface air-liquide (63). Cette cytotoxicité était plus importante que l'e-cig, mais moins importante que pour le tabac. D'autres études ont décrit une altération de la morphologie et de la fonction mitochondriale des cellules BEAS-2B exposées aux aérosols d'HTP, mais moins importante qu'après l'exposition aux fumées de cigarette conventionnelle (64,65). Ces altérations mitochondriales peuvent contribuer au stress oxydant et aux lésions cellulaires. L'étude de Sohal *et al.* montre que l'exposition de cellules BEAS-2B et de cellules musculaires lisses des voies respiratoires humaines aux aérosols d'HTP pendant 72 heures altère la fonction mitochondriale, aggrave l'inflammation des voies respiratoires (augmentation d'IL-8), et favorise le remodelage cellulaire (66). Les auteurs décrivent également une aggravation des infections respiratoires par augmentation de l'adhérence microbienne aux voies respiratoires, phénomène déjà décrit avec l'e-cig et le tabac.

Une étude *in vivo* (67) sur des souris exposées pendant 1,5 à 5 minutes à l'HTP ou à la cigarette conventionnelle a montré une altération comparable de la fonction vasculaire endothéliale des deux types d'exposition par rapport aux animaux non exposés (mesurée par le pourcentage de vasodilatation des artères en réponse à une augmentation du flux sanguin, indicateur clinique de la fonction endothéliale et de la santé cardiovasculaire). Ces résultats sont en faveur de potentiels effets

cardiovasculaires similaires entre l'exposition à la fumée de cigarette et aux aérosols d'HTP. L'exposition sub-aiguë (5 sticks sur 1 journée) aux aérosols d'HTP a provoqué des réactions de stress aigu dans le tissu pulmonaire de souris (68). Bhat *et al.* ont étudié l'effet de l'exposition de souris aux aérosols d'HTP ou de cigarette pendant 5 heures/jour sur 2 semaines (69). Ils ont observé une augmentation de nombreux marqueurs de lésions pulmonaires et d'inflammation, notamment l'albumine et des infiltrats de cellules immunitaires pulmonaires, les cytokines pro-inflammatoires et les chimiokines lors de l'exposition aux produits du tabac par rapport aux souris non exposées. Ces résultats montrent que l'inhalation sub-aiguë d'aérosols d'HTP provoque des lésions pulmonaires et un état pro-inflammatoire dans les poumons qui sont semblables à ceux provoqués par l'exposition au tabac. Une autre étude s'est intéressée à l'effet à plus long terme en exposant des souris 30 min/jour pendant 5 jours/semaine sur 6 mois à la fumée de cigarette ou aux aérosols d'HTP (70). Après exposition, les souris n'ont pas montré de gain de poids par rapport aux contrôles non exposés. L'exposition sub-chronique a induit une augmentation de la cellularité du liquide de lavage broncho-alvéolaire (neutrophiles et lymphocytes) et un emphysème pulmonaire principalement par des voies liées à l'apoptose chez les souris exposées à la fumée de cigarette et aux aérosols d'HTP par rapport au groupe contrôle.

Chez l'Homme, l'exposition à l'HTP a été associée à l'apparition de pneumopathies aiguës à éosinophiles (71,72), ou encore à l'asthme, à la rhinite allergique et à la dermatite atopique chez les adolescents (73).

Les études présentées, tant *in vitro* qu'*in vivo*, suggèrent une corrélation positive entre l'utilisation d'HTP et l'apparition de maladies respiratoires et cardiovasculaires.

Une revue Cochrane souligne qu'aucune étude n'a jusqu'alors porté sur le sevrage tabagique avec des HTP (74). Il est donc difficile d'évaluer l'effets de ces dispositifs pour le sevrage tabagique. Cette revue suggère également que les utilisateurs d'HTP sont moins exposés aux produits chimiques toxiques que les fumeurs de cigarettes, mais plus que les non-fumeurs. Cependant, près de 85 % des études incluses dans cette revue ont été financées par des fabricants de tabac, et sont donc susceptibles d'être biaisées en raison de conflits d'intérêts. Des recherches sur l'efficacité et la sécurité des produits des HTP financées de manière indépendante sont donc nécessaires.

CHAPITRE II : LA MÉTABOLOMIQUE

1. Généralités

Les métabolites sont de petites molécules de faible poids moléculaire (< 1500 Da) du système biologique, substrats et produits du métabolisme, constituants de nombreux autres composés biologiques (ADN, ARN, protéines...). Les métabolites se positionnent donc au centre du métabolisme et assurent des fonctions cellulaires essentielles, telles que la production et le stockage de l'énergie, la transduction des signaux ou encore l'apoptose (75). L'ensemble des métabolites forme le métabolome (76), niveau final du processus de régulation cellulaire, dont l'étendue est de plusieurs dizaines de milliers de métabolites. Des stimuli environnementaux, physiologiques, pathologiques, ou des modifications génétiques peuvent modifier l'expression du métabolome, régulé ou non par des mécanismes de contrôles homéostatiques. La métabolomique est définie par Nicholson *et al.* comme la mesure quantitative de la réponse métabolique multiparamétrique de systèmes vivants suite à ces stimuli (77). Cette approche fait partie intégrante des approches « omiques » (Figure 5), rassemblant la génomique, la transcriptomique, la protéomique et enfin la métabolomique.

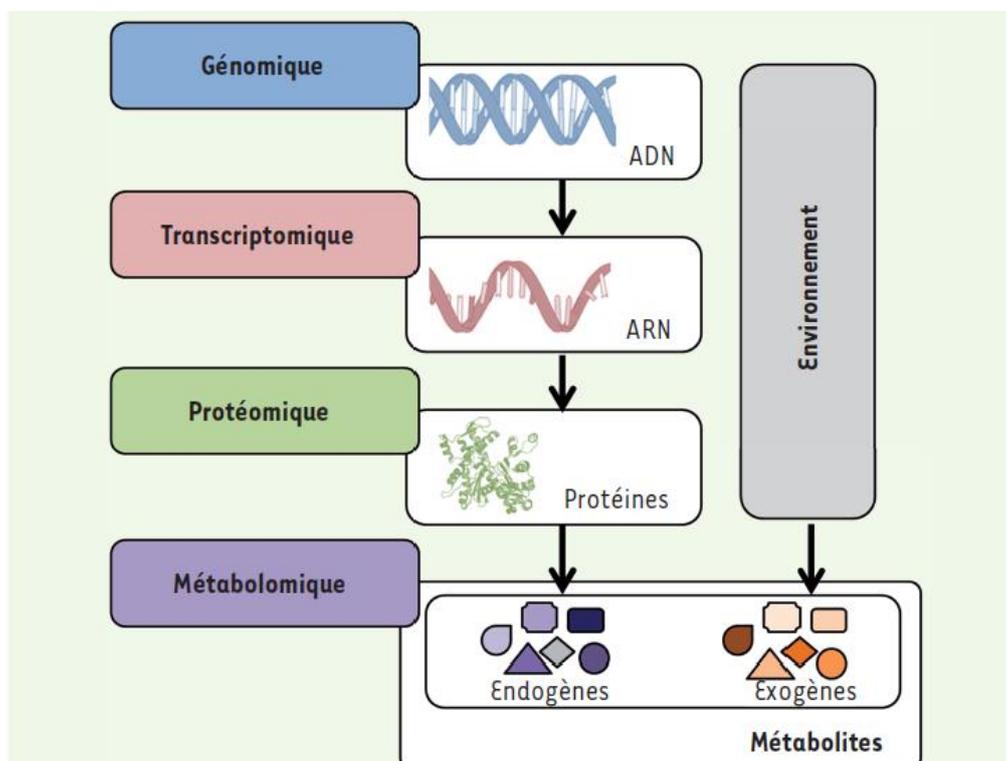


Figure 5 : La métabolomique dans la famille des « omiques » d'après (78)

Contrairement à l'ADN, à l'ARN et aux protéines, qui sont respectivement soumis à une régulation épigénétique, post-transcriptionnelle et post-traductionnelle, les métabolites sont la signature directe de l'activité biologique et sont donc mieux corrélés au phénotype d'un individu.

La métabolomique permet l'analyse simultanée d'un très grand nombre de molécules, issues de plusieurs voies métaboliques distinctes, permettant ainsi d'en souligner leurs dérégulations. Cette approche permet de définir un profil d'expression des métabolites endogènes et peut ainsi permettre de révéler une signature de l'ensemble des perturbations de l'organisme. La métabolomique peut être mise en œuvre afin d'identifier les métabolites et les voies métaboliques associées à des phénotypes particuliers, puis d'intégrer ces connaissances à des études biologiques fonctionnelles et mécanistiques. Elle trouve également tout son intérêt dans la découverte de biomarqueurs diagnostiques, pronostiques, ou prédictifs.

2. Approche ciblée et non ciblée

Deux principaux types d'approches métabolomiques sont envisageables pour permettre ces analyses (79). La première consiste en une approche de métabolomique dite ciblée, mesurant une liste prédéfinie de métabolites et qui se concentre sur une ou plusieurs voies métaboliques d'intérêt. Elle s'intéresse à des familles particulières de métabolites, telles que les glucides, les carnitines, les acides aminés, ou les lipides, par exemple. Cette approche, pouvant être quantitative ou non, permet d'estimer le taux de métabolites spécifiques de façon relative ou absolue dans l'échantillon et d'évaluer les perturbations de ces voies particulières d'intérêt. L'utilisation de cette approche intégrative nécessite de cibler lors du développement analytique les familles de métabolites recherchées. La seconde approche dite non ciblée est une technologie haut débit correspondant à l'analyse globale de toutes les molécules accessibles à l'analyse et présentes dans un échantillon à des concentrations mesurables, incluant des molécules non identifiées à l'heure actuelle. Cette approche permet un profilage métabolique global de l'échantillon. C'est une analyse holistique sans *a priori*, pour laquelle l'identification des métabolites est effectuée par comparaison avec des bases de données. L'identification des métabolites est l'étape limitante de la métabolomique non ciblée, un grand nombre de métabolites pouvant rester non identifiés alors qu'ils pourraient apporter des réponses sur les processus biologiques fondamentaux, ainsi

que sur leurs implications en pathologie et en thérapeutique.

Ces deux approches génèrent chacune un challenge analytique important (80). En effet, il s'agit de développer des méthodes permettant la détection d'un grand nombre de métabolites, d'une grande diversité physico-chimique (pKa, polarité, masse (81)) et présents à des concentrations très variables (du pM au mM (82)). De ce fait, à l'heure actuelle, il n'existe pas de méthode analytique « de référence » permettant l'analyse de l'ensemble du métabolome. De surcroît, le volume et la complexité des données obtenues peuvent également être des facteurs limitants dans l'utilisation de ces approches.

Les différentes étapes ou « workflow » d'une analyse de métabolomique sont successivement (Figure 6) : la préparation des échantillons, l'analyse des échantillons, le traitement des données à l'aide d'outils bioinformatiques, des analyses biostatistiques, l'identification des métabolites (en métabolomique non ciblée) et enfin l'interprétation biologique des résultats.

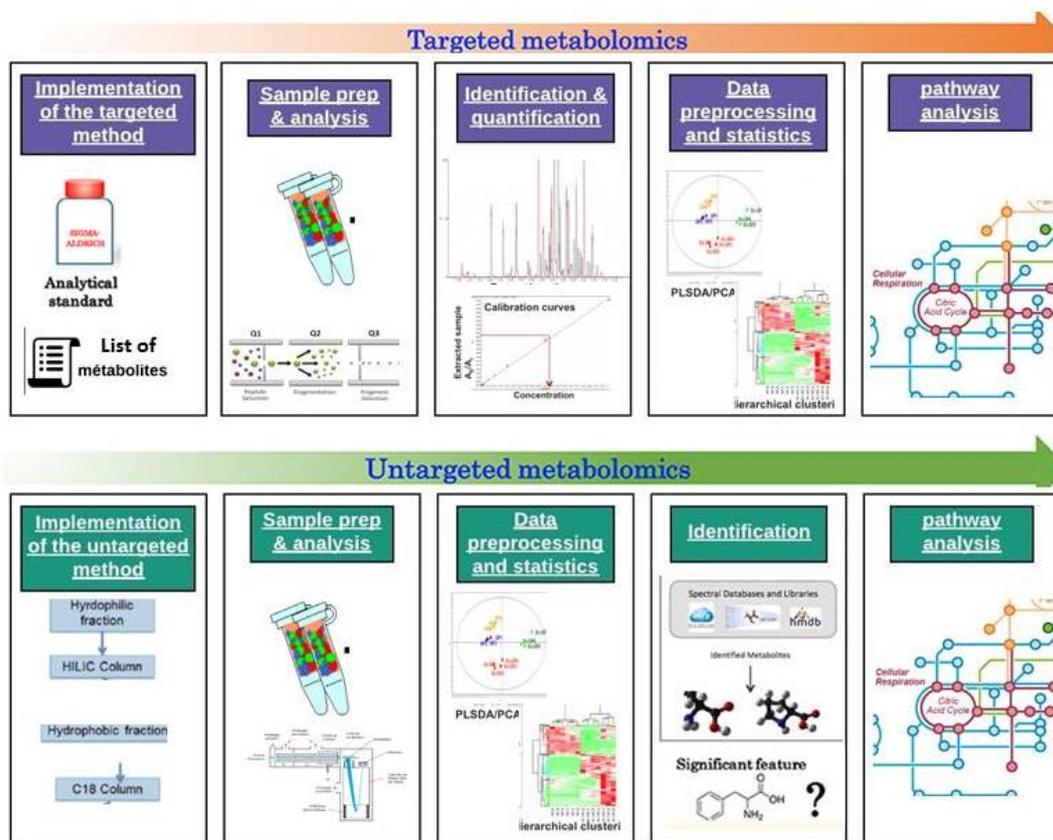


Figure 6 : Workflow d'une analyse de métabolomique ciblée et non ciblée

La standardisation des protocoles expérimentaux et analytiques ainsi que l'implémentation d'un contrôle qualité restent actuellement un challenge à relever, du fait du large éventail de conditions utilisables. Ce workflow complexe nécessite donc une pluridisciplinarité associant des compétences dans les domaines analytiques, bioinformatiques, biostatistiques et biologiques.

3. Techniques analytiques

Plusieurs techniques peuvent être mises en œuvre en métabolomique. Les métabolites peuvent être détectés essentiellement par la méthode de spectrométrie de masse (SM), souvent en couplage avec une méthode séparative (chromatographie en phase gazeuse ou liquide), ou par la résonance magnétique nucléaire (RMN) (83). Le Tableau 1 présente les principaux avantages et inconvénients de ces 2 techniques. Plus récemment, la spectroscopie infrarouge à transformée de Fourier (84), faisant partie des spectroscopies vibrationnelles, a également montré son intérêt en métabolomique.

Tableau 1 : Comparaison des approches analytiques utilisables en métabolomique

	Avantages	Inconvénients
Spectrométrie de masse	Faible volume d'échantillon Sensibilité Couplage à différentes méthodes séparatives : augmentation de la sélectivité	Préparation des échantillons Effet matrice Semi-quantitatif Fragmentation faiblement reproductible limitant l'utilisation des bases de données spectrales Ionisation différente selon la nature des composés
Résonance magnétique nucléaire	Préparation d'échantillon Analyse rapide Robuste Non destructif Haute reproductibilité Non sélective : classes variées de métabolites Quantitatif Elucidation structurale	Volume d'échantillon plus important Faible sensibilité Données manquantes Complexité des spectres
Spectroscopie infrarouge à transformée de Fourier	Faible volume d'échantillon Préparation des échantillons Vision globale Analyse rapide Coût Non sélective Reproductible Robuste	Faible sensibilité Identification et quantification

Finalement, chaque approche possède des spécificités techniques pouvant apporter des informations complémentaires et le choix technologique est dépendant des objectifs à atteindre (sensibilité, quantification, empreinte ou métabolites discriminants) et des ressources techniques disponibles dans les laboratoires et les unités de recherche. Il a été démontré que l'association de plusieurs techniques, favorisée par l'émergence de plateformes analytiques de métabolomique, permet d'élargir la couverture du métabolome (85). Le système utilisé pour ce projet est la chromatographie liquide d'ultra-haute performance couplée à un spectromètre de masse haute résolution (UPLC-HRMS), de type QTOF (quadripôle couplé à un détecteur à temps de vol ou *Time of Flight*). Il est composé de plusieurs modules comme illustré Figure 7.

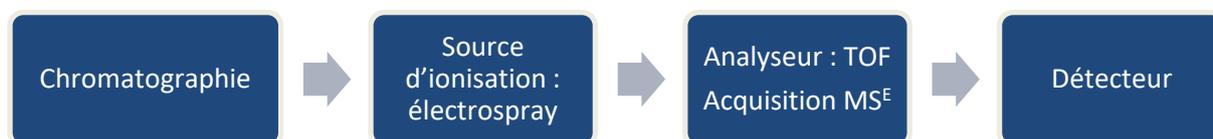


Figure 7 : Modules de l'UPLC-HRMS

3.1. Chromatographie liquide

La chromatographie liquide est une technique analytique permettant la séparation de composés en fonction de leurs propriétés physico-chimiques. Cette technique est basée sur les différences d'affinité entre la substance à analyser et deux phases, l'une stationnaire (la colonne), l'autre mobile (l'éluant). L'échantillon traverse la colonne poussée par la phase mobile. Les différences d'affinité des divers constituants du mélange pour ces deux phases vont entraîner des différences de vitesse de migration de ces composés au sein de la colonne, d'où une possibilité de séparation. Le temps de rétention (T_r) d'un composé est ainsi dépendant par ses interactions chimiques avec la phase mobile et la phase stationnaire choisies. Les conditions analytiques, telles que le type de colonne, ses dimensions, la composition de la phase mobile, la pression et la température, jouent un rôle déterminant dans la séparation des molécules. L'UPLC repose sur l'emploi d'une phase stationnaire composée de particules de moins de 2 μm , permettant d'augmenter le débit de la phase mobile et ainsi d'améliorer les performances analytiques en termes de durée d'analyse et de stabilité des T_r . Les molécules séparées par ce procédé peuvent ensuite être détectées, notamment grâce à la spectrométrie de masse.

3.2. Spectrométrie de masse

La spectrométrie de masse est une technique d'analyse physico-chimique permettant de détecter, d'identifier et de quantifier des molécules d'intérêt par mesure de leur rapport masse/charge (m/z). Son principe réside dans la transformation des molécules de leur état naturel en ions à l'état gazeux et la séparation en phase gazeuse de ces molécules chargées (ions) en fonction de ce rapport. De plus, la spectrométrie de masse permet de caractériser la structure chimique des molécules en les fragmentant. Un spectromètre de masse est composé d'une source d'ionisation, d'un analyseur qui trie les ions en fonction de leur rapport m/z , d'un détecteur qui comptabilise l'abondance des ions et d'un système informatique permettant de traiter et visualiser le signal.

3.2.1. Source d'ionisation

La source d'ionisation utilisée ici est l'électrospray, qui permet de faire passer des ions d'une phase liquide à une phase gazeuse, en utilisant le principe d'ionisation/désolvatation (Figure 8). La phase mobile contenant les molécules d'intérêt traverse un capillaire porté à un haut potentiel électrique. A la sortie du capillaire, un champ électrique va entraîner la formation de microgouttelettes chargées. Sous l'action d'une différence de potentiel, et à pression élevée, les ions vont être dirigés vers l'analyseur. Un gaz de désolvatation provoque l'évaporation de la phase mobile pour obtenir des ions libres chargés. Ce principe permet de ne pas dégrader les composés thermosensibles et d'ioniser des molécules de polarité et masse moléculaire variées. Il peut être utilisé en mode d'ionisation positif (ESI+) ou en mode d'ionisation négatif (ESI-), permettant ainsi d'étendre le champ de détection. Ces deux modes apportent des données complémentaires, puisque l'ionisation dépend des propriétés physico-chimiques de chaque molécule. Ainsi, certains métabolites pourront être détectés soit par l'un ou l'autre des modes d'ionisation, soit par les deux.

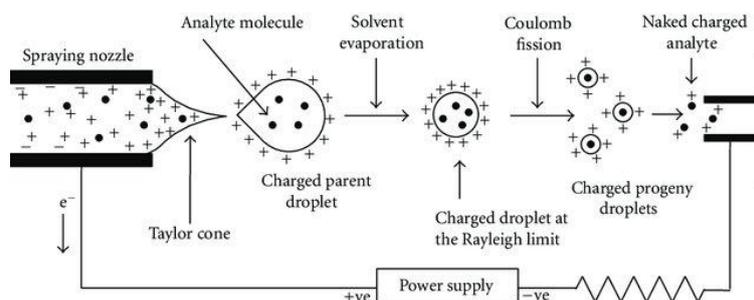


Figure 8 : Schéma de l'électrospray d'après (86)

3.2.2. Analyseur

Les molécules ionisées sont transférées vers l'analyseur afin d'être séparées en fonction de leur rapport m/z . L'analyseur utilisé est un spectromètre de haute résolution de type Q-TOF. Les ions précédemment formés sont accélérés grâce à une différence de potentiel dans un tube de vol. La vitesse de parcours dans ce tube varie en fonction du rapport m/z de chaque ion. Ainsi, un ion avec un faible rapport m/z arrive plus vite au détecteur.

L'analyseur de type TOF est dit de haute résolution. La résolution détermine la finesse des pics et donc la capacité à distinguer des pics proches. Un analyseur de haute résolution permet de déterminer la masse exacte des différents composés (ex : masse exacte de la kynurénine de formule $C_{10}H_{12}N_2O_3 = 208,0847$ Da) et de déduire leur composition élémentaire, *a contrario* des analyseurs de basse résolution qui permettent de déterminer une masse nominale des composés (ex : masse nominale de la kynurénine = 208,0 Da). Ce niveau de précision nécessite de prendre en compte de potentielles variations de mesure. L'équipement utilisé dans pour ce travail est calibré avec un mélange de composés de masses exactes connues (Major Mix IMS-QTOF calibration Kit, Waters®, Wilmslow, UK). De plus, l'injection d'une solution de référence appelée *lockmass*, dont le spectre est acquis à intervalles réguliers au cours de l'analyse, permet de corriger le signal tout au long de l'analyse et ainsi de s'affranchir de potentiels décalages de masse. À une masse exacte peut être associée plusieurs formules brutes, elles-mêmes pouvant correspondre à plusieurs structures chimiques et donc plusieurs composés différents.

3.2.3. Acquisition

Parmi les modes d'acquisition possibles, le mode balayage dit mode « fullscan » (MS^E) permet une acquisition sans *a priori*. Il est illustré Figure 9. Dans ce mode d'acquisition, un quadripôle ($MS1$) laisse indifféremment passer tous les ions présents vers une cellule de collision, où ils subissent ensemble une rampe d'accélération qui les fragmente. L'énergie appliquée dans la cellule de collision passe continuellement d'un profil de basse énergie à un profil de haute énergie. Les ions transférés vers l'analyseur à temps de vol ($MS2$) sont des ions parents (basse énergie) et des ions fils (haute énergie) permettant l'obtention de fragments spécifiques et leur spectre de masse tout au long du cycle d'analyse. Un algorithme permet la reconstruction des spectres de

fragmentation des ions fils et leur correspondance avec leur ion parent. Ceci permet de collecter les informations de masses exactes, d'ions précurseurs et de leurs fragments. Cette méthode de détection donne une mesure de la masse exacte de l'ion moléculaire de l'analyte détecté sous forme protoné ($[M+H]^+$, en ESI+), déprotoné ($[M-H]^-$, en ESI-) ou d'adduits ($[M+K]^+$, $[M+Na]^+$, $[M+Na-2H]^-$, $[M+NH_4]^+$...) et les spectres de masse associés.

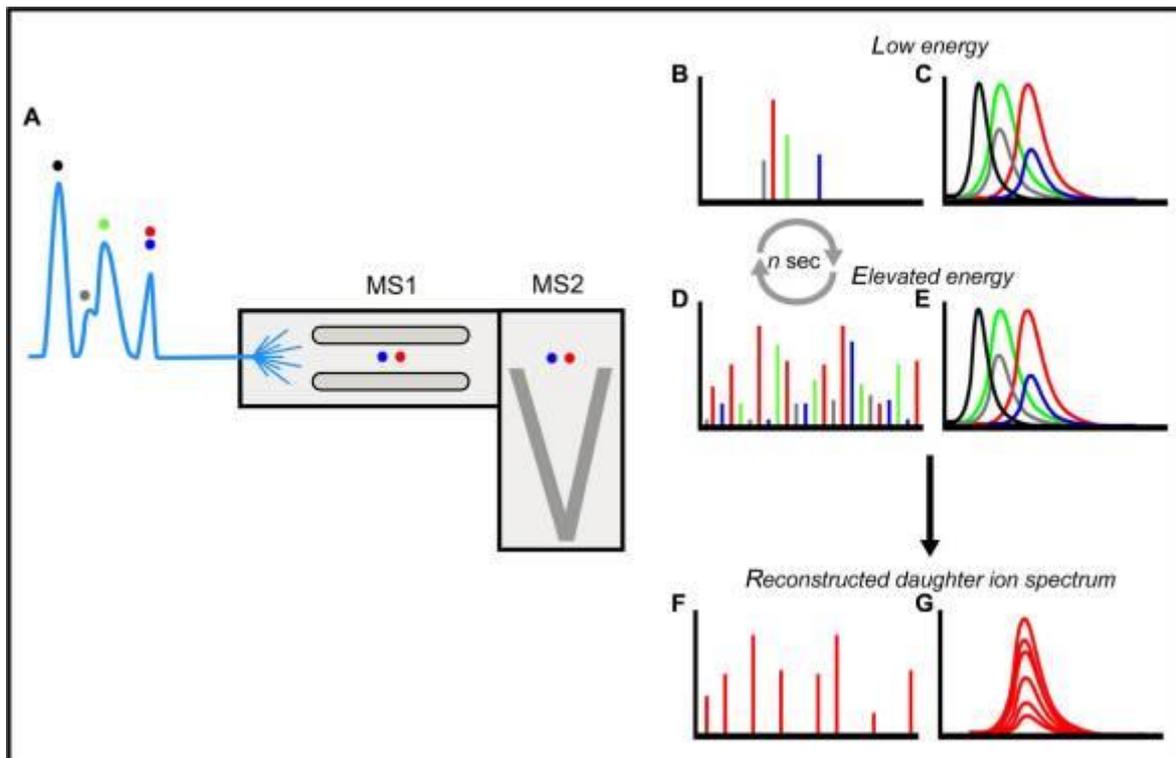


Figure 9 : Principe du mode d'acquisition MS^E d'après (87)

La spectrométrie de masse haute résolution apporte une spécificité et une sélectivité pour les rapports m/z . Néanmoins, dans le cas de matrices complexes, pouvant être composées d'ions ayant le même Tr et un m/z très proche, ou en cas de bruit de fond important, la seule combinaison du m/z et du Tr peut apparaître insuffisante pour une détection et une identification optimale des composés. Grâce à la mobilité ionique, les molécules peuvent être caractérisées par leur valeur de CCS, qui représente une dimension supplémentaire à la séparation des composés, basée sur la structure et la conformation 3D de la molécule (88).

3.3. La mobilité ionique

La mobilité ionique est une technique de séparation additionnelle des ions en phase gazeuse (89). Ainsi, les molécules ionisées au niveau de la source d'ionisation sont soumises à un champ électrique dans un courant de gaz. Les ions se déplacent en suivant le champ électrique à une vitesse qui dépendra de leur charge et de leur interaction avec le gaz, c'est-à-dire de leur masse, de leur taille et de leur forme, et sont donc séparés suivant leur mobilité ionique, ou encore appelée *drift time* (ms). Ce *drift time* est converti en une valeur de CCS ou *Collision Cross Section* (\AA^2). Plus une molécule est petite ou repliée, plus elle se déplacera rapidement dans le tube car elle subira moins de frictions, et inversement pour une molécule largement déployée. Pour deux ions ayant le même rapport m/z , l'ion avec une petite taille/conformation se déplace plus rapidement dans la cellule et sort en premier de la cellule de mobilité ionique. Cet ion a un *drift time* plus court, et donc une valeur de CCS correspondante plus basse. Il existe plusieurs technologies de mobilité ionique (90). Celle utilisée dans ce projet est une technologie « Travelling Wave Ion Mobility Spectrometry » ou « Spectrométrie de Mobilité Ionique à champ oscillant » (TWIMS, Figure 10). L'appareil est constitué d'un guide sur lequel une tension oscillante radiofréquence est appliquée. Une composante de tension continue est de plus appliquée successivement sur les électrodes du dispositif de manière à générer une « vague » de potentiel qui entraîne les ions le long de l'appareil à une vitesse qui dépend de leur mobilité. On obtient ainsi une dispersion temporelle des ions de mobilité différente (91).

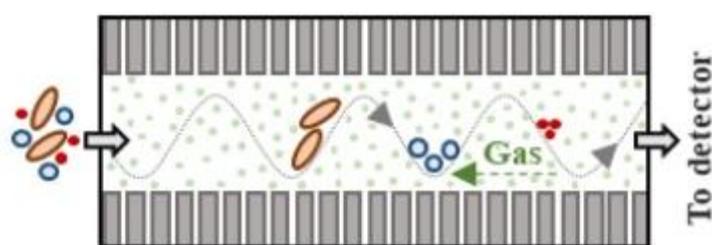


Figure 10 : Schéma de la séparation par mobilité ionique TWIMS d'après (92)

Il s'agit d'une technique de séparation pouvant être couplée à un spectromètre de masse. Le module est alors positionné après la chromatographie et avant le spectromètre de masse.

Les analyses de métabolomique peuvent être réalisées sur des échantillons complexes, dans le but de détecter et caractériser une grande diversité de molécules,

parfois en faibles concentrations. Dans ce contexte, l'ajout d'une dimension supplémentaire telle que la mobilité ionique permet théoriquement de générer des données plus claires et épurées (93), d'améliorer le rapport signal/bruit, la résolution et la séparation des métabolites (94). Il est important de noter que cette mobilité dépend de la structure et n'est pas affectée par l'équipement ou les facteurs expérimentaux (effets de matrice, variations de la composition de la phase mobile et des paramètres de chromatographie, des paramètres d'acquisition...), contrairement au Tr ou au spectre de fragmentation.

Il n'existe pas de relation simple entre le temps de parcours des ions et leur CCS. Pour obtenir une estimation des CCS, une calibration doit être réalisée à partir d'un mélange de composés de CCS connus (Major Mix IMS-QTOF calibration Kit, Waters®, Wilmslow, UK). La Figure 11 détaille la configuration de l'un des analyseurs utilisés dans ce projet, le Vion, commercialisé par la société Waters®.

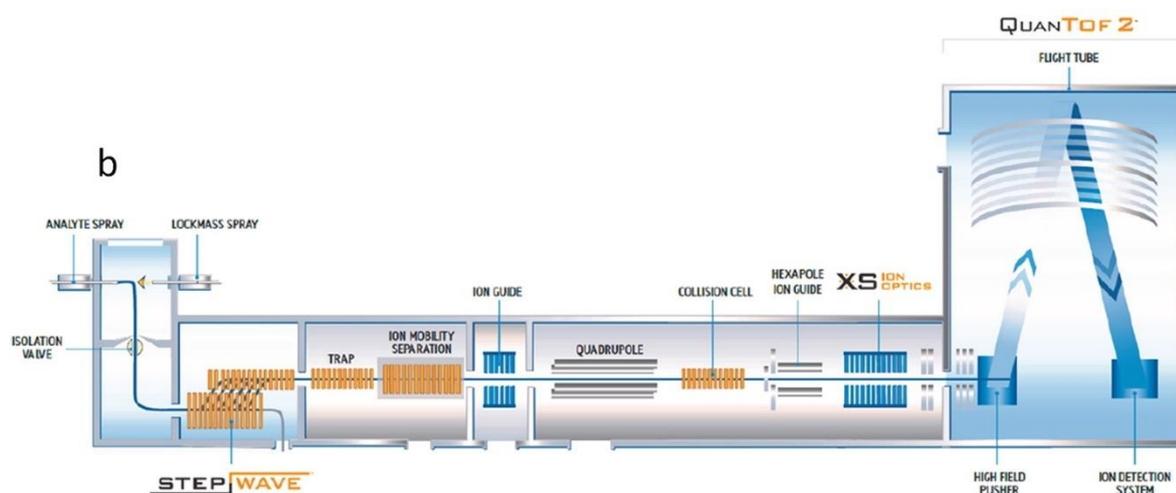


Figure 11 : Configuration de l'analyseur à temps de vol Vion (Waters®) d'après (95)

4. Sens biologique des données acquises

Les analyses de métabolomique réalisées avec les outils analytiques précédemment détaillés génèrent une quantité très importante de données. Elles doivent faire l'objet d'un traitement selon différentes étapes comme décrit dans la Figure 12. À partir de données instrumentales, les données extraites sont traitées par analyses chimiométriques pour apporter un sens biologique aux données et répondre aux questions de recherche posées.

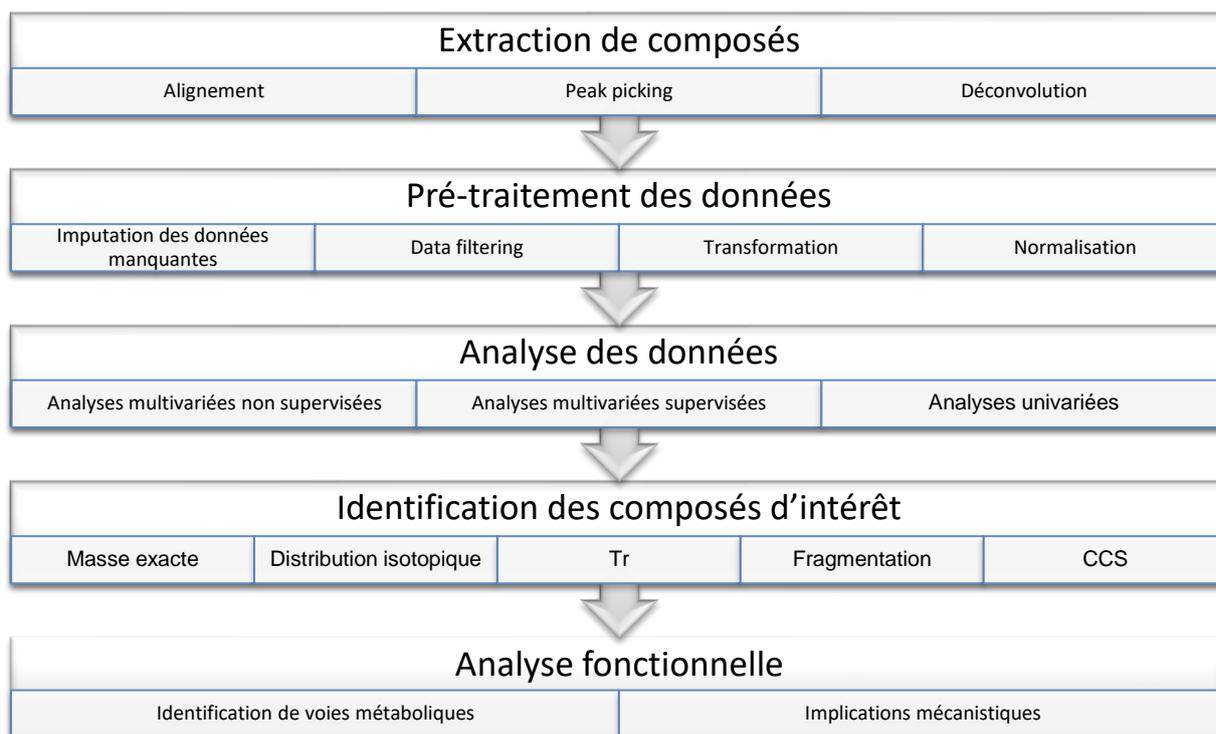


Figure 12 : Etapes du traitement de données en métabolomique non ciblée

4.1. Traitement des données

Les données brutes sont stockées dans des fichiers sous un format propriétaire accessible avec un logiciel spécifique du constructeur. Dans le cas de l'analyseur utilisé, ces données brutes ne peuvent pas être converties en un format accessible à d'autres outils de traitement.

4.1.1. Extraction des données

La première étape du traitement de données consiste en l'extraction de composés à partir de données instrumentales. La détection des pics est suivie de l'alignement des données en termes de Tr et de rapport m/z par rapport à un échantillon de référence. Ceci permet de corriger les déviations éventuelles survenant au cours d'une séquence d'analyse, dues par exemple aux variations de température, de pression, d'humidité, de pH de la phase mobile, de la matrice... Le « peak picking » est l'étape d'intégration des pics. Puisque plusieurs ions détectés peuvent être issus d'un même composé, la déconvolution est la recombinaison des ions pour les grouper en composés. Ceci est possible par association des données de Tr, de masse et d'adduits. Après ces étapes, les données peuvent être exportées sous forme de matrice de données exploitables

par d'autres logiciels, comme le logiciel R (96), pour réaliser un pré-traitement des données extraites.

4.1.2. Pré-traitement des données

Les jeux de données peuvent présenter des valeurs manquantes si l'abondance mesurée d'un composé est en-dessous de la limite de quantification, si le pic est masqué par le bruit d'autres pics co-élués, ou si le composé concerné est réellement absent de l'échantillon étudié. Les valeurs manquantes peuvent donc être déterminantes lors d'analyses différentielles, dans lesquelles il s'agit d'identifier les composés différentiellement exprimés entre deux conditions d'exposition. L'imputation consiste à compléter ces données manquantes en utilisant les informations des autres échantillons pour lesquels le pic est détecté. Plusieurs méthodes sont possibles par exemple en complétant les données par 1/5 de la valeur la plus basse mesurée pour ce composé (ou 1/5 de la médiane ou de la moyenne des valeurs), ou en utilisant des outils de prédiction des valeurs comme la méthode du « K Nearest Neighbors », ou encore du « Random Forest » (97).

Après l'imputation, une filtration des données a pour but d'éliminer les signaux dont les fluctuations sont dues à la variabilité analytique et non à la variabilité biologique des échantillons. Pour cela, un échantillon de contrôle de qualité (QC) est injecté à intervalle régulier tout au long de la séquence analytique (98). La variation des signaux d'une injection à l'autre de ce QC reflète la répétabilité analytique. La mesure est considérée comme fiable pour un métabolite si son coefficient de variation est inférieur à 20 % entre les QC (99,100). Ce seuil est défini pour éliminer au maximum les fluctuations d'origine analytique. D'autres seuils moins restrictifs sont parfois décrits dans la littérature (101,102).

Les données conservées après l'étape de filtration subissent alors une étape de transformation qui consiste à mettre les valeurs dans une échelle qui convient à leur distribution (103). Différentes méthodes de transformation sont possibles comme la transformation logarithmique, la transformation par la racine carrée ou encore par la racine cubique.

Pour finir, la normalisation a pour objectif de réduire les écarts systématiques entre les valeurs de différents échantillons afin de pouvoir les comparer (103). Différentes

méthodes de normalisation peuvent être utilisées, et notamment la méthode de normalisation par « cyclic LOESS », la standardisation, la normalisation pareto, la « robust scaling ».

Ainsi, il n'existe pas de méthode universelle de pré-traitement des données et plusieurs méthodes pourront être testées et évaluées pour pouvoir choisir la combinaison de méthodes la plus adaptée aux données à pré-traiter.

4.1.3. Analyses statistiques

Des analyses différentielles par analyses statistiques sont réalisées sur la matrice de données pré-traitée, dans l'objectif de faire ressortir des profils métaboliques pertinents, et ainsi répondre à la question biologique posée. Deux types d'analyses statistiques sont utilisables : les analyses univariées et multivariées.

Les analyses statistiques univariées sont des techniques classiquement utilisées en biologie. Elles permettent d'analyser une à une les variables explicatives (ex : un ion) en fonction d'une métadonnée (ex : exposition) sans tenir compte des interactions existant entre les variables. Ces analyses permettent de savoir s'il existe une différence d'expression d'un composé entre un groupe exposé et un groupe témoin, avec une probabilité d'erreur de 5% lorsque le seuil de significativité de la p-value est fixé à 0,05. Plusieurs analyses univariées sont possibles. Par exemple le t-test de Student peut être utilisé pour comparer entre deux groupes les moyennes de chaque composé. L'ANOVA peut également être utilisée pour comparer pour chaque composé les variances des échantillons appartenant à plusieurs groupes. Afin de prendre en considération le risque d'augmenter le nombre de faux-positifs par la multiplication des comparaisons, les résultats des tests statistiques doivent être corrigés. Plusieurs méthodes, plus ou moins conservatrices, existent telles que la correction par le « False Discovery Rate » (FDR) ou encore la correction de Bonferroni (104). Une des limites de l'approche univariée est l'absence de considération des corrélations entre les variables, qui ne sont pas toujours indépendantes.

Les analyses multivariées, plus utilisées pour les « Big Data », permettent l'étude simultanée de plusieurs variables. Une partie de l'analyse multivariée regroupe les techniques pour réduire la dimensionnalité des données afin de résumer et visualiser dans un espace réduit l'information biologiquement pertinente contenue dans le jeu de

données. L'analyse statistique comprend typiquement des méthodes non supervisées (ex : Analyse en Composante Principale ou ACP) et supervisées (ex : régression des moindres carrés partiels ou PLS). Les méthodes non-supervisées sont considérées comme des analyses statistiques descriptives, dont l'objectif est d'obtenir une vision globale des données sans *a priori*, afin de visualiser la répartition spontanée des échantillons dans un espace à deux dimensions. Les méthodes supervisées, dites explicatives et prédictives, ont pour objectif de relier un bloc de variables à expliquer Y (ex : le type d'exposition) à un bloc de variables explicatives X (ex : les abondances des métabolites).

La combinaison d'approches multivariées et univariées est fortement recommandée afin d'augmenter l'extraction d'informations pertinentes à partir de données métabolomiques. Une fois le pré-traitement des données et les analyses statistiques réalisées, l'identification des composés discriminants peut être réalisée.

4.2. Identification des composés détectés

La caractérisation des composés repose sur la masse exacte, le spectre de fragmentation, la distribution isotopique, le Tr si associé à une méthode séparative, et le CCS en cas de couplage à la mobilité ionique. L'identification des composés est basée sur la comparaison des données acquises avec celles de bases de données de référence, ou librairies. Un exemple est illustré par la Figure 13.

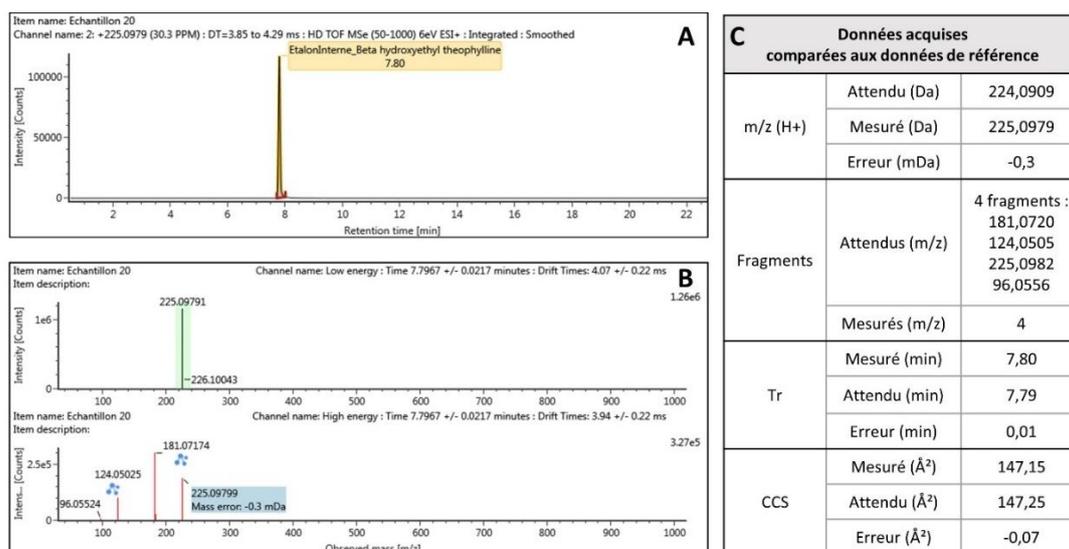


Figure 13 : Exemple de résultat obtenu lors de l'identification de la β -hydroxyéthyl-théophylline A : chromatogramme, B : spectres à haute et basse énergie, C : les données acquises sont confrontées à celles de la base de données « maison » de référence

Les librairies peuvent être établies au laboratoire à partir de standards (librairie « maison »), mais il existe également des bases de données externes de référence (Tableau 2) accessibles en ligne comme Human Metabolome Database (HMDB) (105), Metlin (106), MassBank (107), Kyoto Encyclopedia Genes and Genomes (KEGG) (108)..., qui sont continuellement mises à jour par la communauté scientifique.

Tableau 2 : Quelques-unes des bases de données les plus utilisées pour l'annotation des composés issues d'analyses LC-MS non ciblées, adapté de (109)

Nom	URL	Libre accès	MS/MS	Courte description
HMDB	www.hmdb.ca/	Oui	Oui	Base de métabolites humains
METLIN	metlin.scripps.edu/	Quelques fonctionnalités	Oui	Base de spectres de masse
MassBank	massbank.eu/	Oui	Oui	Base de spectres de masse
NIST	www.nist.gov/srd/nist-standard-reference-database-1a-v17	Non	Oui	Base de spectres de masse
MONA	mona.fiehnlab.ucdavis.edu/	Oui	Oui	Base de spectres de masse
GNPS	gnps.ucsd.edu/	Oui	Oui	Base de spectres de masse
LipidMaps	www.lipidmaps.org/	Oui	Oui	Base de lipidomique
KEGG	www.genome.jp/kegg/	Quelques fonctionnalités	Non	Base de voies métaboliques
MetaCyc	metacyc.org/	Quelques fonctionnalités	Non	Base de voies métaboliques
SMPDB	https://www.smpdb.ca/	Quelques fonctionnalités	Non	Base de voies métaboliques

La combinaison de plusieurs propriétés orthogonales augmente le niveau de confiance attribuée à une identification. Un consensus sur ces niveaux a été proposé par le groupe de travail “Chemical Analysis Working Group” de la “Metabolomics Standards Initiative (MSI)” (110). Le niveau 1 est le niveau de confiance le plus élevé, définissant une identification sûre et définitive d'un composé. Ce niveau est atteint lorsque deux propriétés orthogonales ou plus correspondent aux données de standards analysés dans les mêmes conditions analytiques. Il convient de noter que certains analytes requièrent une prudence particulière pour l'identification structurale sans ambiguïté, comme les isomères. L'annotation de niveau 2 est une annotation théorique d'un composé, et le niveau 3 d'une classe de composé. Ici, l'annotation est basée sur la comparaison des données avec des bibliothèques publiques sans confrontation à des standards dans les mêmes conditions analytiques. Ceci peut générer un nombre élevé d'erreurs d'annotation (faussement positives). Malheureusement, il n'existe pas de

librairie universelle. Le niveau de confiance le plus bas (niveau 4) rassemble les composés inconnus. Bien que non identifiés et non classifiés, ces métabolites peuvent néanmoins encore être différenciés.

4.3. Interprétation biologique

Après l'étape d'identification des composés, l'interprétation biologique des résultats est nécessaire pour resituer les métabolites au sein de réactions métabolites, de voies métaboliques, et du réseau métabolique qui est une visualisation de l'ensemble de ces voies. Cette étape d'interprétation biologique permet d'orienter les fonctions biologiques susceptibles d'être perturbées par le stimuli étudié, et donc de réaliser un lien biologique avec la question de recherche posée. Ceci est possible grâce la comparaison des métabolites d'intérêt avec des bases de données de voies métaboliques. Des bases de données telles que KEGG (Kyoto Encyclopedia of Genes and Genomes (108)) ou encore SMPDB (The Small Molecule Pathway Database (111)), contiennent des informations sur les voies métaboliques, les réactions et les enzymes les catalysant, et les liens entre eux. A titre d'exemple, la Figure 14 illustre l'ensemble des voies métaboliques de la base de données KEGG.

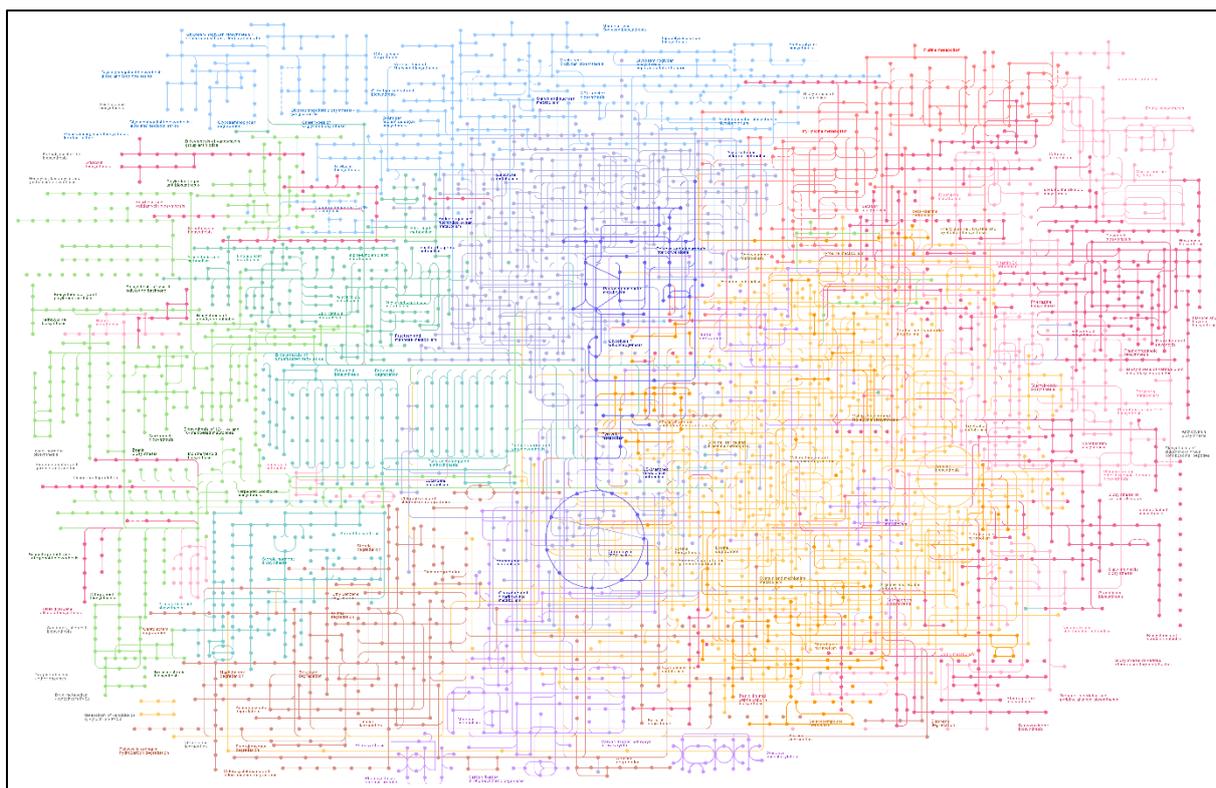


Figure 14 : Voies métaboliques de la base de données KEGG

Vu l'étendue du métabolome et des voies métaboliques déjà décrites, des outils statistiques permettent de traiter les données de métabolomique en confrontant les métabolites identifiés aux bases de données de voies métaboliques (112). L'analyse de surreprésentation (« over representation analysis », ou ORA) consiste à vérifier à partir de la liste des métabolites d'intérêt si le nombre de métabolites associés à une voie métabolique est dû au hasard ou non (113). Cette approche est utilisée lorsque les métabolites d'intérêt sont préalablement sélectionnés, elle ne nécessite donc qu'une liste d'identifiants pour être réalisée. Cette analyse considère seulement le nombre de métabolites par voie métabolique et n'intègre pas l'amplitude du changement de leur abondance entre les conditions testées. Quand ceci est souhaité, il est possible de réaliser une analyse d'enrichissement des voies (« metabolite set enrichment analysis » ou MSEA) qui prend en compte le « poids » de chaque métabolite (114). Cette approche ne nécessite pas de sélection préalable des métabolites, mais les concentrations sont requises. Ces approches apportent ainsi une hiérarchisation des voies en fonction de leur significativité. Il est à noter que les résultats sont fortement dépendants des bases de données utilisées qui s'avèrent parfois incomplètes. Les résultats sont confrontés à des bases de données spécifiques

d'une espèce définie, mais ne tiennent pas compte de la compartimentation cellulaire, pouvant conduire à des résultats incohérents par rapport au compartiment cellulaire étudié. De plus, un même métabolite peut appartenir à plusieurs voies métaboliques, entraînant parfois des biais dans les analyses statistiques. Une autre approche possible est de réaliser une interprétation biologique des données par rapport aux classes chimiques des composés identifiés (115). L'avantage est ici qu'une seule classe chimique est associée à chaque métabolite, et tous les métabolites possèdent une attribution de classe chimique. Cette méthodologie est complémentaire de celle par voies métaboliques car elle permet de palier à certaines difficultés liées au fait qu'à un métabolite peuvent être associés un, plusieurs, ou même aucune voie métabolique.

CHAPITRE III : LA MÉTABOLOMIQUE EN TOXICOLOGIE

1. Applications de la métabolomique

La métabolomique vise à évaluer de manière exhaustive les changements du métabolome induits par des facteurs endogènes et/ou exogènes, de cribler les métabolites significativement différents et ainsi d'identifier des biomarqueurs potentiels. Les applications actuelles de la métabolomique touchent différents domaines d'activités tels que la nutrition et l'agronomie (116), la biologie végétale (117), la recherche et le développement de nouveaux médicaments (118), ou encore la santé humaine (119). En toxicologie biologique et médico-légale, la métabolomique se place comme une approche novatrice 1) pour l'étude des profils métaboliques signant une exposition à un xénobiotique ou ses effets sur l'organisme 2) pour l'amélioration de l'interprétation *post-mortem* de situations médico-légales. Ces applications sont développées dans la revue bibliographique que j'ai pu réaliser sur le sujet dans le cadre de cette thèse (120).

Dans le domaine des expositions environnementales, les études de toxicité des aérocontaminants basées sur la métabolomique peuvent fournir de nouvelles informations sur les impacts de ces expositions sur la santé humaine et améliorer l'étude des mécanismes pathologiques associés (121). L'étude des conséquences métaboliques d'expositions environnementales reste un challenge qui nécessite la mise en œuvre de protocoles pertinents prenant en compte les éventuels facteurs confondants. Ceci est d'autant plus vrai en métabolomique, puisque le métabolome peut être influencé par tout type de stimuli, et pas seulement par l'exposition étudiée. Les expositions contrôlées sur cultures cellulaires ou modèles animaux permettent de réduire la variabilité interindividuelle.

2. Métabolomique et exposition aux produits alternatifs du tabac

Une série d'études a montré la faisabilité des analyses de métabolomique lors d'une exposition à la cigarette conventionnelle. Ces études se sont focalisées sur la mise en évidence de profils métabolomiques spécifiques d'exposition ou de sevrage tabagique (122–124), ou encore sur la recherche de dérégulations métaboliques pouvant apporter des hypothèses mécanistiques quant à la toxicité de la cigarette

conventionnelle, en termes d'addiction et de pathologies secondaires. Ces études ont été réalisées dans différentes matrices telles que des cultures cellulaires (125–127), ou dans des prélèvements tissulaires, du sang périphérique (128–130), des urines (131), ou encore du condensat d'air exhalé (132) chez l'animal ou chez l'Homme.

Rushing *et al.* ont réalisé une revue de la littérature traitant des effets du tabac sur le métabolisme endogène (133). Globalement, un grand nombre de voies métaboliques sont dérégulées chez les fumeurs par rapport aux non-fumeurs. Le métabolisme des acides aminés (ex : ceux qui participent à la production d'énergie et au métabolisme des folates et de la méthionine), le métabolisme énergétique (ex : glucides, acides gras, lipides), et le métabolisme oxydant (ex : Glutathion) et certaines vitamines (Ex : Vitamines A et B) étaient notamment dérégulés. Bien que de nombreuses études de métabolomique s'intéressent à l'exposition à la cigarette conventionnelle, très peu d'études de métabolomique abordent la question de l'exposition aux autres produits de consommation tels que les e-cig ou l'HTP.

2.1. Métabolomique et exposition à la cigarette électronique

Quelques études traitant d'analyses de métabolomique et de l'exposition à l'e-cig ont été publiées. Une revue de ces études, des conditions expérimentales mises en œuvre, et des principaux résultats est résumée dans le Tableau 3.

Une étude *in vitro* a comparé les profils métabolomiques de cellules épithéliales bronchiques humaines exposées directement à du e-liquide ou à un condensat de fumée de cigarette pendant différents temps d'exposition allant de 1h à 13h d'exposition (134). Cette étude a montré des altérations de 24 % (e-cig) et 35 % (tabac) du métabolome étudié après 1h d'exposition. Les fluctuations observées lors de l'exposition au e-liquide étaient biphasiques avec un premier maximum après 5h d'exposition, et un deuxième après 13h d'exposition. Certaines altérations métaboliques étaient communes entre les deux types d'exposition. Les voies métaboliques impliquées étaient notamment le métabolisme des acides aminés, le métabolisme énergétique, et la bêta-oxydation des acides gras. Globalement, ces résultats étaient en faveur d'un impact métabolomique partiellement comparable entre e-liquide et fumée de cigarette.

In vivo, Ren *et al.* (135) ont étudié l'effet de différentes durées d'expositions aiguës (de 1h à 8h) aux aérosols d'e-cig sur le métabolome sérique de souris sacrifiées deux semaines après les expositions. Le nombre de métabolites modulés était croissant avec la durée d'exposition, allant de 25 à 50 métabolites dérégulés par rapport aux contrôles non exposés. Les auteurs ont montré qu'ils appartiennent à 24 voies métaboliques, qui régulent principalement le métabolisme des acides aminés liés au cycle de Krebs. L'analyse fonctionnelle a permis d'identifier des pathologies en lien avec des variations de ces métabolites tels que le diabète de type 2, l'acidémie propionique, le défaut de transport des acides gras à longue chaîne ou encore le cancer du poumon. Les métabolites de ces empreintes étaient différents en fonction de la durée d'exposition (4 composés en commun uniquement après 1h, 2h, 4h, 8h d'exposition).

L'exposition chronique aux aérosols d'e-cig est également à l'origine d'empreintes métabolomiques spécifiques, impliquant notamment les lipides. Madison *et al.* (136) ont évalué l'impact de 4 mois d'exposition de souris aux aérosols d'e-cig, avec ou sans nicotine, sur le lipidome bronchique. Cette étude a montré une accumulation de phospholipides dans le compartiment cellulaire (macrophages) du lavage broncho-alvéolaire, et une perturbation de l'homéostasie phospholipidique de la partie acellulaire du liquide de lavage broncho-alvéolaire, et donc du surfactant pulmonaire. De plus, les altérations de l'homéostasie lipidique observées semblent indépendantes de la présence de nicotine dans les e-liquides testés, ce qui justifie des études plus approfondies sur les solvants utilisés dans les e-cig.

Chez l'Homme, l'étude du métabolome plasmatique et urinaire par RMN a également été réalisée dans le cas d'un court arrêt (5 jours) d'utilisation de l'e-cig avec ou sans nicotine, chez des utilisateurs réguliers de ce dispositif (137). Bien que le métabolome plasmatique ne soit pas significativement dérégulé, les auteurs ont pu mettre en évidence une signature urinaire de l'exposition à l'e-cig, ainsi qu'une réversion de ce profil vers un profil non vapoteur en cas d'arrêt du vapotage, avec ou sans nicotine. Ces données de métabolomique, associées à une amélioration des paramètres cardio-respiratoires (diminution de la fréquence cardiaque, de marqueurs de l'inflammation pulmonaire, amélioration de marqueurs fonctionnels respiratoires) suggèrent que l'e-cig altère la fonction respiratoire et que ces altérations pourraient être réversibles en cas d'arrêt d'utilisation.

De même, Hsiao *et al.* (138) ont mis en évidence des modifications des métabolites urinaires entre des vapoteurs, des fumeurs et des non-utilisateurs. 12 % des métabolites de ces signatures étaient communs entre les deux groupes d'exposés. Une augmentation plus importante d'acylcarnitines et acylglycines était observée chez les vapoteurs, suggérant une dérégulation de la peroxydation lipidique plus importante chez les vapoteurs que chez les fumeurs. Un biomarqueur spécifique du cancer ((5)-methyl-N(5)-formyl-2,5,6-triamino-4-hydroxypyrimidine ou Me-Fapy, un sous-produit de réparation ou de dégradation de l'adduit de l'ADN lié aux lésions de l'ADN par alkylation) était également augmenté dans le groupe de fumeurs. Dans le groupe de vapoteurs, cette augmentation n'était pas significative, même si une tendance à l'augmentation était également observée.

Pour une exposition chronique chez l'Homme, des sujets exposés pendant plus de 2 ans à l'e-cig présentaient une dérégulation des métabolites du cycle de Krebs, tandis que ceux exposés à la cigarette conventionnelle présentaient une altération des sphingolipides (139), par rapport au groupe non exposé. Les signatures métabolomiques observées, ici spécifiques de chaque type d'exposition, sont proposées par les auteurs comme de potentiels marqueurs de maladies métaboliques. Néanmoins, le petit effectif de sujets inclus dans cette étude chez l'Homme (e-cig n = 12, tabac n = 6, non exposés n = 6) nécessite une confirmation sur un nombre plus importante d'individus.

Pour finir, des altérations métaboliques ont également été mises en évidence dans d'autres matrices biologiques comme la salive (140). Ainsi, une augmentation de métabolites impliqués dans le métabolisme de l'acide arachidonique et dans la réponse immunitaire étaient en faveur d'une altération de voies spécifiques de maladies bucco-dentaires après exposition à l'e-cig, et notamment l'inflammation gingivale, connue pour précéder le développement de parodontite.

Une étude comparant différentes modalités d'exposition de souris a montré que le type d'appareil électronique, la composition du e-liquide, la durée d'exposition et le sexe entraînent des variabilités des profils métabolomiques (141). L'ajout d'aromatisants comme la vanilline (142), ou le maltol (143) est également mis en cause dans les altérations du métabolome.

Une étude réalisée par l'industrie du tabac s'est intéressée aux effets du passage (ou « switch ») de la cigarette conventionnelle à l'e-cig chez l'Homme (144). Des prélèvements d'urine et de plasma réalisés avant le « switch », et 5 jours après ce dernier, ont permis de mettre en évidence une signature métabolomique à la suite du « switch » vers l'e-cig, comprenant des modulations des métabolites de la nicotine, des xénobiotiques issus de la fumée de cigarette, ainsi que des métabolites issus du métabolisme des vitamines et du stress oxydant. Les métabolites de la nicotine et autres xénobiotiques étaient significativement réduits dans ces deux matrices. Une diminution des concentrations plasmatiques des métabolites du glutathion, ainsi qu'une augmentation des concentrations urinaires et plasmatiques de vitamines et antioxydants ont également été mises en évidence après le « switch ». Les auteurs suggèrent donc une réduction de l'exposition aux toxiques, associée à une potentielle amélioration au niveau du métabolisme des vitamines et des voies de défense antioxydante lors du passage à l'e-cig. Cette étude n'a néanmoins pas inclus de groupe contrôle non exposé, ne permettant pas d'évaluer la toxicité propre de l'e-cig, et les durées d'exposition étaient courtes (seulement 5 jours d'exposition à l'e-cig).

Tableau 3 : Etudes de métabolomique traitant de l'exposition à la cigarette électronique et leurs conclusions principales

Conditions expérimentales	Dispositif d'exposition	Modèle / Matrice	Effectif	Technique analytique	Métabolites/classes/voies métaboliques mises en évidence	Conclusions	Conflit d'intérêt avec l'industrie du tabac et du vapotage	Référence
Exposition aiguë à l'e-liquide ou au condensat de fumée de cigarette	e-liquide avec nicotine	Cellules épithéliales bronchiques humaines / Contenu intracellulaire	triplicat	LC-HRMS	Les expositions durant 1h, 2h, 5h, 7h, ou 13h ont affecté 392 métabolites pour le e-liquide et 569 métabolites pour le condensat de fumée. 138 étaient communs à au moins 1 temps. Altérations de 24 % (e-cig) et 35 % (tabac) du métabolome après 1h d'exposition. Pour l'e-cig, fluctuation biphase avec un premier maximum après 5h d'exposition, et un deuxième après 13h d'exposition. Certaines altérations métaboliques étaient communes entre les deux types d'expositions. Voies métaboliques impliquées : métabolisme des acides aminés, métabolisme énergétique (ADP) et bêta-oxydation des acides gras (phospholipides).	Modifications du métabolome après exposition au e-liquide, et de manière partiellement comparable au condensat de fumée de cigarette. Ne soutient pas le concept d'innocuité des e-cig.	NON	(134)
Exposition aiguë à l'e-cig avec nicotine	Non décrit	Souris / Sérum	40	LC-HRMS	2 semaines après l'exposition aiguë, variations métaboliques dépendantes de la durée d'exposition (26, 25, 55 ou 50 métabolites dérégulés après 1h, 2h, 4h ou 8h d'exposition). 24 voies métaboliques impactées, régulant principalement le métabolisme des acides aminés lié au cycle de Krebs. Les pathologies pouvant être liées à ces métabolites sont le diabète de type 2, l'acidémie propionique, le défaut de transport des acides gras à longue chaîne ou encore le cancer du poumon.	Augmentation des altérations métaboliques avec la durée d'exposition. Mise en évidence de marqueurs de pathologies métaboliques et cancéreuses.	NON	(135)
Exposition chronique à l'e-cig avec ou sans nicotine	e-cig de 2 ^{ème} génération	Souris / Lavage broncho-alvéolaire	9	LC-HRMS	Après 4 mois d'exposition, augmentation des phospholipides dans la partie cellulaire (macrophage) et acellulaire (surfactant) du liquide de lavage broncho-alvéolaire. Diminution concomitante des triglycérides du surfactant.	Altérations de l'homéostasie lipidique qui pourrait favoriser l'accumulation de phospholipides, de façon indépendante de la nicotine, justifiant des études plus approfondies sur les e-liquides.	NON	(136)
Arrêt du vapotage chez les utilisateurs réguliers d'e-cig avec ou sans nicotine	e-cig de 4 ^{ème} génération	Homme / Urine Plasma	30	RMN	Un court arrêt (5 jours) d'utilisation de l'e-cig (avec et sans nicotine) chez des utilisateurs réguliers impacte le métabolome urinaire mais pas plasmatique. Signature métabolomique urinaire permettant de différencier les vapoteurs de ceux ayant arrêté de vapoter : <ul style="list-style-type: none"> ➤ 3-hydroxyisovalérate et pyruvate plus élevés chez les vapoteurs avec nicotine ➤ Triméthylamine oxyde, hippurate et N-Phenylacetyl-glycine plus bas chez les sujets ayant arrêté le vapotage 	Chez les utilisateurs réguliers d'e-cig, l'arrêt du vapotage à court terme semble modifier le métabolome urinaire.	NON	(137)

Conditions expérimentales	Dispositif d'exposition	Modèle / Matrice	Effectif	Technique analytique	Métabolites/classes/voies métaboliques mises en évidence	Conclusions	Conflit d'intérêt avec l'industrie du tabac et du vapotage	Référence
Exposition chronique à l'e-cig ou au tabac	e-cig avec nicotine (non décrit)	Homme / Urine	117	LC-HRMS	Signature métabolique de 839 composés pour les fumeurs, et 396 composés pour les vapoteurs, dont 12 % en commun. Augmentation des acylcarnitines et acylglycines chez les vapoteurs, en faveur d'une dérégulation de la peroxydation lipidique. Un biomarqueur spécifique du cancer (Me-Fapy) augmenté dans le groupe fumeur. Non significatif chez les vapoteurs, mais tendance à l'augmentation.	Dérégulation du métabolome urinaire chez les fumeurs et les vapoteurs, avec augmentation du catabolisme lipidique chez les vapoteurs.	NON	(138)
Exposition chronique à l'e-cig ou au tabac	e-cig avec nicotine (non décrit)	Homme / Plasma	24	LC-HRMS	L'exposition à l'e-cig est à l'origine d'une dérégulation des métabolites du cycle de Krebs. L'exposition à la cigarette est à l'origine d'une altération des sphingolipides.	Mise en évidence de potentiels biomarqueurs de pathogenèse des maladies cardio-pulmonaires spécifiques de l'exposition à l'e-cig ou au tabac.	NON	(139)
Exposition à l'e-cig et effets sur la santé buccale	e-cig de 4 ^{ème} génération	Homme / Salive	30	LC-HRMS	Signature métabolique de 368 métabolites chez les exposés. Augmentation de prostaglandines, de leucotriènes en faveur d'un métabolisme accru de l'acide arachidonique, et dérégulation d'autres métabolites comme les gangliosides, les céramides, l'angiotensine en faveur d'une dérégulation de la réponse immunitaire. Voies métaboliques impliquées dans l'inflammation gingivale, qui précède le développement de la parodontite.	Mise en évidence de potentiels biomarqueurs prédictifs de risque de maladie bucco-dentaire.	NON	(140)
Exposition à différentes conditions de vapotage	e-cig de 2 ^{ème} , 3 ^{ème} et 4 ^{ème} génération	Souris / Plasma	6 réplicats	LC-HRMS	Variation des profils métaboliques en fonction du type d'appareil électronique, la composition du e-liquide, la durée d'exposition et le sexe. Signatures métaboliques associées aux maladies cardiovasculaires et à l'insuffisance rénale chronique.	Chaque génération d'e-cig et d'e-liquide est susceptible d'entraîner son propre effet sur la santé.	NON	(141)
Exposition à la vanilline utilisée comme aromatisant dans l'e-liquide	e-liquide sans nicotine	Cellules épithéliales bronchiques humaines / Contenu intracellulaire	triplicat	LC-HRMS	L'exposition avec la vanilline perturbe le métabolisme énergétique, celui des acides aminés, des antioxydants et des sphingolipides. Voies métaboliques associées à différentes pathologies respiratoires comme l'asthme, la fibrose idiopathique pulmonaire, ou le syndrome de détresse respiratoire aigüe.	Les aromatisants pourraient conduire le microenvironnement métabolique pulmonaire à un état plus pathogène.	NON	(142)
Exposition au maltol utilisé comme aromatisant dans l'e-liquide	e-cig de 3 ^{ème} génération	Cellules épithéliales bronchiques humaines / Contenu intracellulaire	triplicat	LC-HRMS	Perturbation du stress oxydant après exposition aux aérosols de e-cig, avec ou sans maltol. Perturbation du métabolisme des acides aminés spécifique après exposition avec maltol.		NON	(143)

Conditions expérimentales	Dispositif d'exposition	Modèle / Matrice	Effectif	Technique analytique	Métabolites/classes/voies métaboliques mises en évidence	Conclusions	Conflit d'intérêt avec l'industrie du tabac et du vapotage	Référence
Passage de la cigarette vers l'e-cig	e-cig de 1 ^{ème} génération	Homme / Urine Plasma	75	LC-HRMS	5 jours après un « switch » de cigarette vers e-cig : amélioration du métabolisme des vitamines et réduction du stress oxydant.	Réduction des biomarqueurs de potentielle toxicité « switch » de cigarette conventionnelle à e-cig.	OUI	(144)

2.2. Métabolomique et exposition au tabac chauffé

L'effet de l'exposition au tabac chauffé sur le métabolome a été jusqu'ici très peu étudié. Une revue de ces études, des conditions expérimentales mises en œuvre, et des principaux résultats, est résumée dans le Tableau 4.

Une étude *in vitro* réalisée par l'industrie du tabac a comparé les profils métabolomiques après exposition de cultures de cellules épithéliales gingivales humaines aux aérosols issus du tabac chauffé ou de la cigarette conventionnelle (145). L'exposition était de 28 minutes, répétée 3 jours. Le métabolome a été analysé 4 heures après la fin de l'exposition. Les auteurs ont ici souligné une perturbation du métabolome beaucoup moins importante lors d'une exposition au tabac chauffé (13 métabolites significativement perturbés) que lors d'une exposition à la cigarette conventionnelle (181 métabolites significativement perturbés) par rapport aux contrôles non exposés. En particulier, une altération du stress oxydant et de la voie du glutathion, non observée après exposition au tabac chauffé, a été montrée. De plus, un médiateur lipidique (le 15-Hydroxyeicosatetraenoic acid ou 15-HETE) était significativement augmenté après exposition au tabac conventionnel, mais pas après exposition au tabac chauffé. Ce médiateur serait associé à la régulation de la réponse immunitaire et inflammatoire, ainsi qu'au processus d'athérosclérose.

Une étude du lipidome du tissu pulmonaire, également réalisée par l'industrie du tabac, a comparé les profils lipidiques de souris exposées à la cigarette conventionnelle ou au tabac chauffé (146). Les auteurs ont mis en évidence d'importantes variations en lipides et protéines de surfactant, en eicosanoïdes, en céramides et sphingolipides, lors de l'exposition à la cigarette conventionnelle, qui seraient potentiellement associées à des effets négatifs sur le poumon. L'analyse des profils lipidomiques des souris exposées au tabac chauffé n'a montré des variations que très limitées, notamment en eicosanoïdes, par rapport à la cigarette conventionnelle. Pour les souris ayant subi un « switch » d'exposition de la cigarette conventionnelle vers l'HTP, une normalisation du profil lipidique étudié était observée, mettant en avant l'intérêt de substituer la consommation de cigarette conventionnelle par celle du tabac chauffé en termes de marqueurs lipidiques. Les mêmes auteurs ont plus récemment complété cette étude de lipidomique par des analyses de métabolomique (147) et ont montré des perturbations beaucoup moins importantes lors d'une exposition au tabac chauffé

que lors d'une exposition à la cigarette conventionnelle, notamment concernant des métabolites immunorégulateurs et du stress oxydant. Certains métabolites immunorégulateurs étaient toujours augmentés après le « switch » vers l'HTP, mais cette observation n'a pas été plus amplement discutée dans l'article.

A notre connaissance, aucune étude indépendante de l'industrie du tabac n'a, à ce jour, étudié les variations du métabolome survenant après une exposition au tabac chauffé. De plus, aucune étude n'a comparé les profils métabolomiques après exposition à l'e-cig et au tabac chauffé.

Tableau 4 : Etudes de métabolomique traitant de l'exposition au tabac chauffé et leurs conclusions principales

Conditions expérimentales	Dispositif	Modèle / Matrice	Effectif	Technique analytique	Métabolites/classes/voies métaboliques mises en évidence	Conclusions	Conflit d'intérêt avec l'industrie du tabac et du vapotage	Référence
Exposition au tabac chauffé par rapport à la cigarette	IQOS	Cellules épithéliales gingivales humaines / Contenu intracellulaire	5 réplicats	LC-HRMS	Perturbation du métabolome beaucoup moins importante lors d'une exposition au tabac chauffé (13 métabolites) que lors d'une exposition à la cigarette conventionnelle (181 métabolites). Perturbation du stress oxydant, de la voie du glutathion, et d'un médiateur lipidique (15-HETE) impliqués dans la réponse inflammatoire et l'athérosclérose après exposition au tabac mais pas au tabac chauffé.	L'exposition au tabac chauffé a un impact moindre sur les cellules épithéliales gingivales humaines par rapport à la cigarette conventionnelle.	OUI	(145)
Exposition chronique au tabac chauffé par rapport à la cigarette, et de l'arrêt du tabac ou du « switch » vers l'HTP	IQOS	Souris / Poumon	8 réplicats	LC-HRMS	Importantes variations en lipides et protéines de surfactant, en eicosanoïdes, en céramides et sphingolipides lors de l'exposition à la cigarette conventionnelle, potentiellement associées à des effets négatifs sur le poumon. Variations très limitées des profils lipidomiques (eicosanoïdes) après exposition au tabac chauffé par rapport à la cigarette conventionnelle. Normalisation du profil lipidique lors de l'arrêt du tabac ou lors du « switch » vers l'HTP.	Arrêt du tabac ou passage au tabac chauffé bénéfique au niveau pulmonaire.	OUI	(146)
Exposition chronique au tabac chauffé par rapport à la cigarette, et effet de l'arrêt du tabac ou du « switch » vers l'HTP	IQOS	Souris / Poumon	9	LC-HRMS	Perturbation du métabolome beaucoup moins importante lors d'une exposition au tabac chauffé que lors d'une exposition à la cigarette conventionnelle. Augmentation des métabolites jouant un rôle immunorégulateur (itaconate, polyamines, quinolininate) et des métabolites de la réponse au stress oxydant (voie de l'hème-biliverdine-bilirubine) lors de l'exposition à la cigarette, en faveur d'une induction de la réponse inflammatoire et du stress oxydant. Normalisation du profil métabolomique lors de l'arrêt du tabac. Certains métabolites immunorégulateurs (itaconate, polyamines) toujours augmentés après le « switch » vers l'HTP.	Arrêt du tabac bénéfique sur le profil métabolique du poumon.	OUI	(147)

CHAPITRE IV : OBJECTIFS DE LA THÈSE

L'ULR4483 a développé depuis 2014 des méthodes d'analyse des principaux composants chimiques des fumées de cigarette et des aérosols d'e-cig ou de tabac chauffé générés par une machine à fumer, ainsi que des méthodes d'exposition sur des modèles cellulaires (39,148). En complément des méthodes classiques d'étude de l'atteinte cellulaire et tissulaire (analyses histopathologiques, du stress oxydant, de l'inflammation, de la génotoxicité, des altérations épigénétiques...), l'utilisation d'approches globales telles que les « omiques » (métabolomique, transcriptomique, MIRnomique...) devrait permettre de mieux comprendre les mécanismes moléculaires et cellulaires de toxicité et d'identifier des marqueurs pertinents d'effets et/ou d'exposition.

L'objectif général de mes travaux est le développement et l'application d'une approche de métabolomique pour l'identification de marqueurs précoces de toxicité de ces aérocontaminants émergents.

Le premier objectif de ce travail est de développer une méthode de métabolomique. Les méthodes de chromatographie liquide et de spectrométrie de masse haute résolution seront développées et optimisées à partir de standards de métabolites. L'utilisation d'outils de *machine learning* permettra de mettre en place une base de données de référence de grand dimension, facilitant l'analyse des données de métabolomique.

Le deuxième objectif de ce projet consiste à analyser le métabolome de cellules épithéliales pulmonaires humaines immortalisées (BEAS-2B) cultivées à l'interface air-liquide et exposées à de l'air stérile (contrôle) ou aux émissions d'une cigarette de référence, d'HTP, ou d'e-cig en utilisant la machine à fumer Vitrocell. Puisque l'HTP et l'e-cig sont vendus par les industriels comme des alternatives à la cigarette conventionnelle, il est important de comparer la toxicité de ces trois dispositifs entre eux. L'analyse chimiométrique des données de métabolomique permettra de mettre en évidence des signatures métabolomiques pour chaque type d'exposition. L'identification des voies métaboliques dérégulées pourra offrir de nouvelles perspectives pour une meilleure compréhension des variations du métabolisme

cellulaire après une exposition à la cigarette, à l'e-cig ou au HTP et générer des hypothèses mécanistiques sur leurs effets toxiques.

Au total, ces recherches pourraient apporter de meilleures connaissances des mécanismes cellulaires et moléculaires de toxicité des produits alternatifs du tabac, et ainsi fournir des informations supplémentaires aux autorités pour la réglementation de ces produits ainsi que pour le développement de politiques de santé publique de réduction du tabagisme.

PARTIE II : TRAVAUX PERSONNELS

CHAPITRE I : DÉVELOPPEMENT ANALYTIQUE

1. Mise au point d'une méthode d'UPLC-HRMS

En métabolomique, le choix des conditions analytiques est d'une importance capitale, puisqu'il doit permettre la détection simultanée d'un grand nombre de métabolites d'une grande diversité en termes de propriétés physico-chimiques et de gamme de concentrations (149). La méthode de référence utilisée pour cette mise au point est décrite par Dunn *et al.* (100), et adaptée au parc analytique du laboratoire. Ainsi, le système utilisé pour initier ce projet est la chromatographie liquide d'ultra-haute performance couplée à un spectromètre de masse haute résolution, initialement sur un analyseur G2-XS-QTOF (Waters®).

La séparation chromatographique est réalisée par un système Acquity UPLC I-Class (Waters®) équipé d'une colonne de type phase inverse Acquity UPLC® HSS T3 (150 x 2,1 mm ; 1,8 µm, Waters®) maintenue dans une enceinte thermostatée à 50°C. Ce type de colonne possède un greffage permettant de mieux retenir les composés polaires que les colonnes C18 classiques, et est également plus rétentive pour les composés apolaires. Elle présente une meilleure compatibilité avec la phase mobile aqueuse, tolérant jusqu'à 100 % de phase aqueuse si nécessaire. La colonne HSS T3 présente donc un meilleur profil de rétention, en faisant la colonne de choix pour une analyse de métabolomique en phase inverse. Les dimensions de la colonne ont été fixées pour pouvoir optimiser la séparation et les Tr pour une durée d'analyse fixée à 30 minutes. La phase mobile, constituée d'un mélange de formiate d'ammonium 5 mM pH 3 (A) et d'acétonitrile avec 0,1 % d'acide formique (B), est utilisée en mode gradient (Tableau 5), le débit étant de 0,4 mL/min. La composition des phases mobiles a été fixée en fonction de la compatibilité avec la phase stationnaire, ainsi que les contraintes instrumentales.

Tableau 5 : Gradient d'élution initialement choisi

Temps (min)	Phase mobile A (%)	Phase mobile B (%)	Curve
0	97	3	7
4	97	3	7
25	3	97	7
27	3	97	7
28	97	3	7
30	97	3	7

Les conditions de spectrométrie de masse, quant à elles, ont été choisies pour détecter de façon spécifique un grand nombre de composés, avec la meilleure sensibilité possible (Tableau 6). L'ionisation est faite en modes électrospray positif et négatif. L'acquisition en mode MS^E sur analyseur Q-TOF a été choisie dans l'objectif de pouvoir acquérir une quantité importante de données concernant les métabolites mesurables dans les échantillons.

La gamme de balayage de 50 à 1 000 *m/z* choisie permet de minimiser le plus possible le bruit de fond afin d'optimiser la détection. Pour les molécules de bas poids moléculaire (autour de 100 *m/z*), cette gamme permet d'avoir suffisamment de fragments spécifiques détectés pour caractériser ces composés. Le « scan time » est la vitesse à laquelle les scans des spectres de masse sont réalisés. Plusieurs scans permettent de définir un pic chromatographique. Le nombre moyen de points par pic habituellement recherché est d'environ 20 points par pic. Une vitesse de scan trop faible peut entraîner un nombre de points insuffisant pour bien définir un pic chromatographique. A l'inverse, un nombre de points par pic trop important peut être à l'origine d'une diminution du ratio signal/bruit et une forme de pics mal définie. C'est pourquoi la valeur du « scan time » est définie à 0,25 seconde.

Tableau 6 : Conditions analytiques de spectrométrie de masse

Température de source	120 °C
Débit de gaz de nébulisation	50 L/h
Débit de gaz de désolvatation	1000 L/h
Température de désolvatation	600 °C
Tension de cône	20 V
Tension des capillaires (inlet et Lockmass)	0,5 kV (ESI +) / 2,5 kV (ESI -)
Gamme de masse	50 et 1000 <i>m/z</i>
Acquisition en mode MS ^E	Basse énergie : 6 eV Haute énergie : rampe de collision de 10 à 40 eV
Scan time	0,25 s
Lockmass - Leucine Enképhaline	Débit = 10 µL/min Concentration = 2 ng/µL Acquisition toutes les 15 s Scan time = 0,25 s

Le pilotage de l'appareil et le traitement des données, ou « processing », sont réalisés avec le logiciel UNIFI (version 1.8.2, Waters® MS Technologies, Manchester, UK). Le « processing » des pics est réalisé en 3D. La masse et le Tr sont le résultat de la projection de l'apex du pic dans le plan 2D *m/z*-Tr. La réponse est déterminée par le volume du pic en 3D.

Le choix de ces conditions analytiques va nécessiter certaines optimisations avant sa mise en œuvre dans le contexte biologique choisi. Différents paramètres, illustrés dans la Figure 15, permettent d'évaluer les performances analytiques testées lors des phases d'optimisation et notamment :

- le temps de rétention ;
- la hauteur du pic (H) ;
- la largeur du pic à sa base (width) : ce paramètre permet d'évaluer la finesse du pic chromatographique ;
- le nombre de points par pic ;
- la réponse : c'est la modification de la réponse de l'instrument en présence du composé, exprimée en volume du pic ;
- le bruit : c'est la fluctuation du signal du bruit de fond de l'instrument ;
- le rapport signal sur bruit : c'est le rapport du signal d'un composé et du signal de bruit de fond mesuré ;
- le spectre de fragmentation en basse énergie et haute énergie.

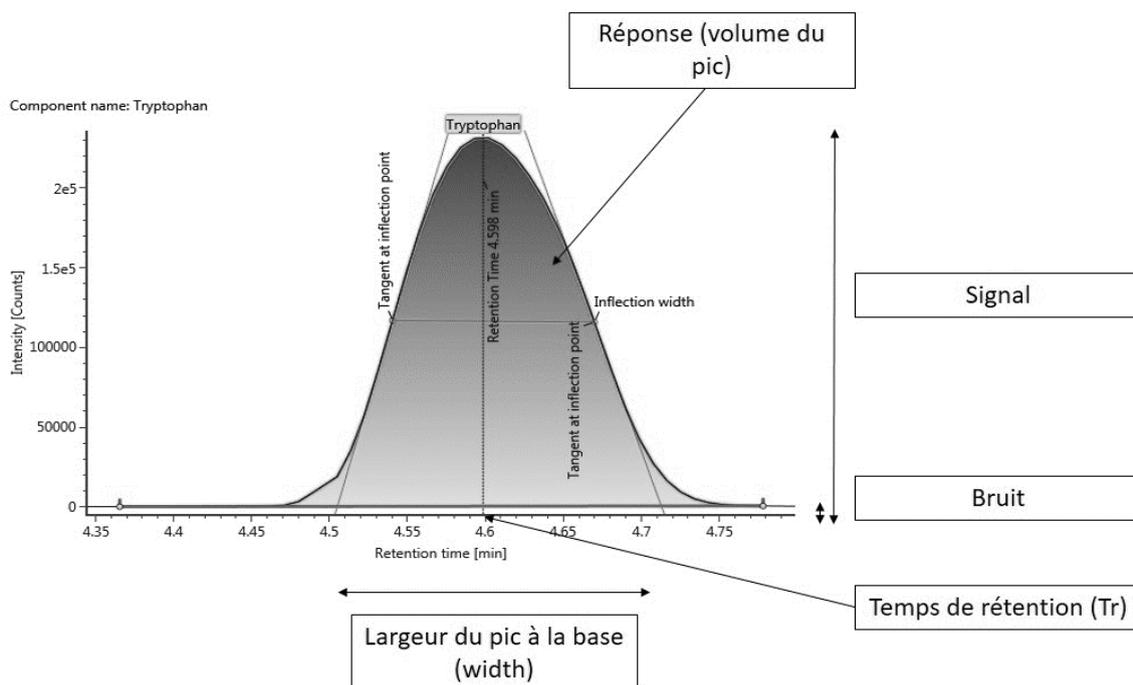


Figure 15 : Performances chromatographiques évaluées

La rétention, la forme des pics, l'intensité du signal mesuré, ainsi que le rapport signal/bruit doivent être optimaux pour que les données générées puissent être exploitables en métabolomique.

1.1. Méthode chromatographique

Les phases d'optimisation analytique des conditions chromatographiques ont été réalisées uniquement en mode d'ionisation positif, dans le but d'améliorer la séparation des composés, d'affiner la forme des pics chromatographiques, tout en gardant une rétention des composés polaires en début de gradient, et de raccourcir la durée totale de l'analyse. La méthode chromatographique finale optimisée sera appliquée au mode d'ionisation positif et négatif. Différentes variations du gradient de phase mobile ont été appliquées (Figure 16), permettant une optimisation des 4 principaux paramètres : a) la pente d'augmentation de la phase organique au cours du gradient, b) la durée du plateau initial de phase mobile en début de gradient, c) la proportion de phase aqueuse et de phase organique en début de gradient, et d) la durée de la pente (et donc la durée totale du gradient).

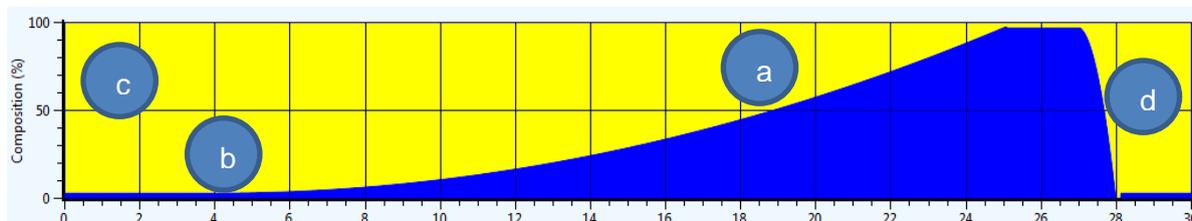


Figure 16 : Gradient de phase mobile initial et ses points d'optimisation

1.1.1. Matériel

Les solutions utilisées étaient les suivantes : acétonitrile (qualité UPLC-MS, Waters[®], Milford, MA, États-Unis), méthanol (qualité UPLC-MS, Waters[®]), eau purifiée Milli-Q (Millipore, Burlington, MA, États-Unis), acide formique (qualité UPLC-MS, Honeywell, Charlotte, NC, États-Unis), et formiate d'ammonium (qualité Reagent, Sigma-Aldrich, St. Louis, MI, États-Unis). L'optimisation analytique de la méthode chromatographique a été réalisée et évaluée à partir d'un échantillon composé de 10 standards. Sa composition a été choisie en fonction de la disponibilité de standards purs au laboratoire achetés chez Sigma Aldrich (St. Louis, MO, USA), de la distribution de leur rapport m/z , et du Tr de chaque standard dans les conditions analytiques initiales, pour couvrir au maximum la gamme temporelle de rétention (Tableau 7). Ces composés purs ont été mélangés pour former une solution méthanolique.

Tableau 7 : Composition de la solution de standards utilisée pour l'optimisation analytique

Composé	Tr (min)	Masse monoisotopique (Da)	Concentration (mg/L)
Arginine	0,86	174,1116	20
Créatinine	0,96	113,0589	20
3-hydroxy-kynurénine	1,15	224,0797	20
Nicotine	1,96	162,1157	20
Kynurénine	3,54	208,0847	20
Tryptophane	7,90	204,0898	20
β -hydroxyéthyl-théophylline	10,83	224,0909	1,6
Caféine	11,89	194,0803	20
Acide glycocholique	19,10	465,3090	2
Méthylclonazépam	20,02	329,0573	0,125

10 μ L de la solution de standards sont dilués dans 390 μ L d'un mélange de 97 % de phase aqueuse et 3 % de phase organique pour travailler dans les conditions identiques aux conditions initiales du gradient de phase mobile. Les échantillons sont stockés à -20°C jusqu'à analyse. Au cours de l'analyse, la température de l'échantillonneur automatique est fixée à 10 °C et le volume d'injection est de 10 μ L.

Parmi les 10 composés inclus dans l'échantillon de standards, les résultats de performances analytiques de 8 composés ont servi pour l'optimisation des conditions chromatographiques. En effet, les pics chromatographiques obtenus pour l'arginine et la créatinine présentaient une mauvaise intégration automatique par le logiciel UNIFI. Cette intégration n'est malheureusement pas modifiable dans le cas de l'intégration du pic en 3D comme dans ce travail. Dans ces conditions, ces deux composés ont été exclus de l'analyse des résultats de la phase d'optimisation analytique.

1.1.2. Type de pente

1.1.2.1. Méthode

La pente correspond à l'augmentation de la proportion de phase organique par rapport à la phase aqueuse au cours du gradient de phase mobile. La pente choisie pour le gradient de phase mobile de la méthode initiale commence après une période de 4 minutes de gradient isocratique composé de 97 % de phase aqueuse, et est une « curve 7 » du logiciel UNIFI, correspondant à une pente convexe.

Pour l'optimisation analytique, les pentes %B testées sont de type linéaire (curve 6) et concave (curve 5), comme illustré sur la Figure 17. L'influence de ces variations est attendue pour les composés dont le temps de rétention est compris dans la période d'ascension de la proportion de phase organique, c'est-à-dire après 4 minutes.

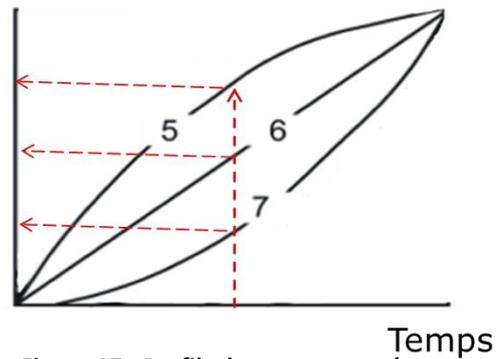


Figure 17 : Profils des pentes testés

La comparaison des différents profils de la Figure 17 fait supposer une diminution des Tr pour ces composés en utilisant une pente de type linéaire ou concave.

1.1.2.2. Résultats

Les composés éluant après 4 min sont influencés par le changement de pente du gradient de phase mobile, notamment leur Tr (Figure 18). Les composés sortant le plus tôt présentent un faible raccourcissement de leur Tr après application d'une pente linéaire ou concave. Les composés éluant le plus tard présentent un important raccourcissement de leur temps de rétention avec la pente linéaire et la pente concave.

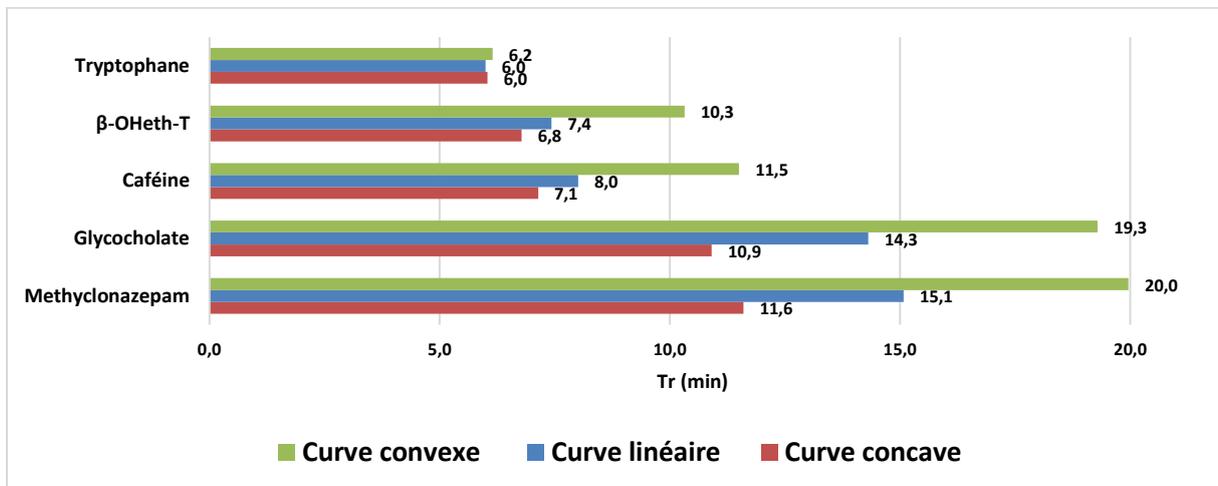


Figure 18 : Influence de la pente sur le Tr des composés éluant après 4 min

L'influence du changement de pente sur la largeur du pic à sa base, sur le nombre de points par pic, et sur la réponse (volume du pic) a été également évaluée. Ces données montrent une tendance à l'affinement des pics entre la pente convexe et les pentes linéaire et concave, de -17,5 % en moyenne pour la pente linéaire et de -17,8 % en moyenne pour la pente concave. La réponse a tendance à diminuer entre les conditions de pente convexe et les deux autres pentes testées d'en moyenne -22,3 % avec la pente linéaire et -31,9 % avec la pente concave.

Ainsi, la modification de la pente du gradient de phase mobile a montré des bénéfices en termes de raccourcissement des T_r , de réponse et d'affinement des pics, orientant le choix vers une pente linéaire.

1.1.3. Durée du plateau en début de gradient

1.1.3.1. Méthode

La méthode initiale commence par une période de 4 minutes de gradient isocratique composé de 97 % de phase aqueuse. Une variation de la durée initiale de ce plateau est testée en utilisant des durées de 3 minutes et 2 minutes de plateau. Dans ces conditions, les composés éluant après cette période pourraient être impactés.

1.1.3.2. Résultats

Globalement, les composés éluant avant 3 minutes présentent une faible modification de leur temps de rétention avec les deux conditions testées, et ceci dans la même proportion. Les composés sortant après 3 minutes présentent un raccourcissement de leur T_r de 0,7 minute en moyenne pour la condition 3 minutes, et de 1,5 minute en moyenne pour la condition 2 minutes. L'influence de la durée du plateau initial sur la largeur du pic à sa base (Figure 19), sur le nombre de points par pic et sur la réponse (aire du pic) a été également évaluée. Ces données montrent une tendance à l'affinement des pics en début de gradient, les autres composés n'étant pas affectés. L'intensité du signal (la réponse) n'est quant à elle pas affectée par ces variations.

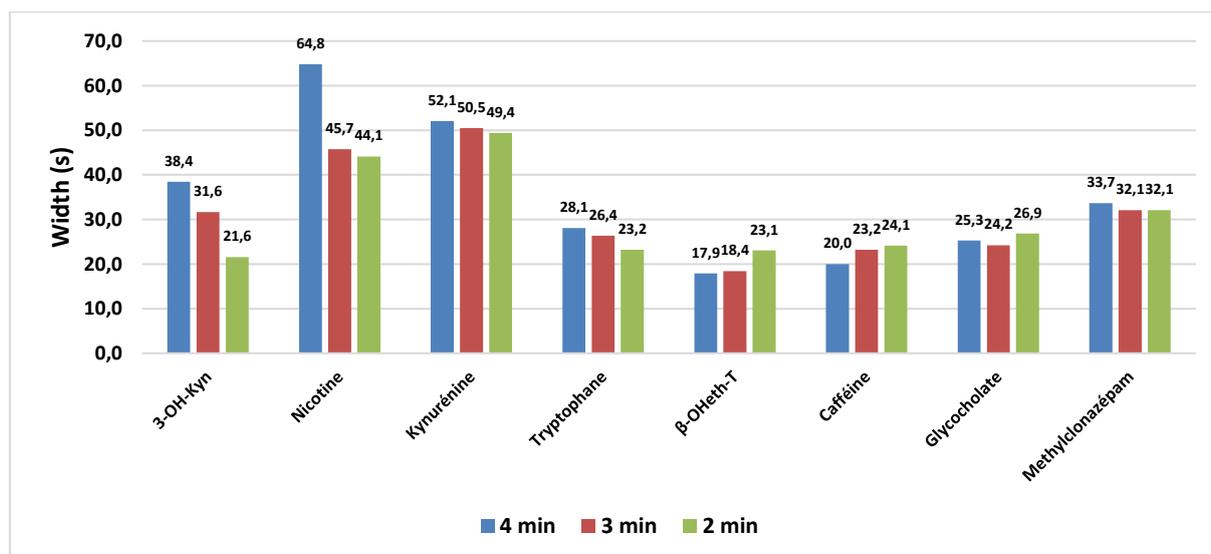


Figure 19 : Influence de la durée du plateau initial sur la largeur à la base des pics (width)

Ainsi, la modification de la durée du plateau en début de gradient affecte donc principalement la rétention des composés après le plateau, et la finesse des pics éluant au cours du plateau.

1.1.4. Composition de la phase mobile en début de gradient

1.1.4.1. Méthode

L'utilisation d'une colonne HSS T3 permet de travailler avec une composition initiale de phase aqueuse allant jusqu'à 100 %, ce qui pourrait avoir un effet bénéfique sur la rétention des molécules polaires. Ainsi, des conditions initiales à 99 % et à 100 % de phase aqueuse ont été testées. Dans ces conditions, tous les composés pourraient être impactés par ces modifications.

1.1.4.2. Résultats

Les deux conditions testées mettent en évidence une influence importante de la proportion de phase aqueuse dans la rétention de composés sortant en début de gradient. Les composés testés présentent un allongement de leur T_r , et cette tendance est plus importante pour les composés les plus polaires (Figure 20), c'est-à-dire pour les composés sortant en début de gradient.

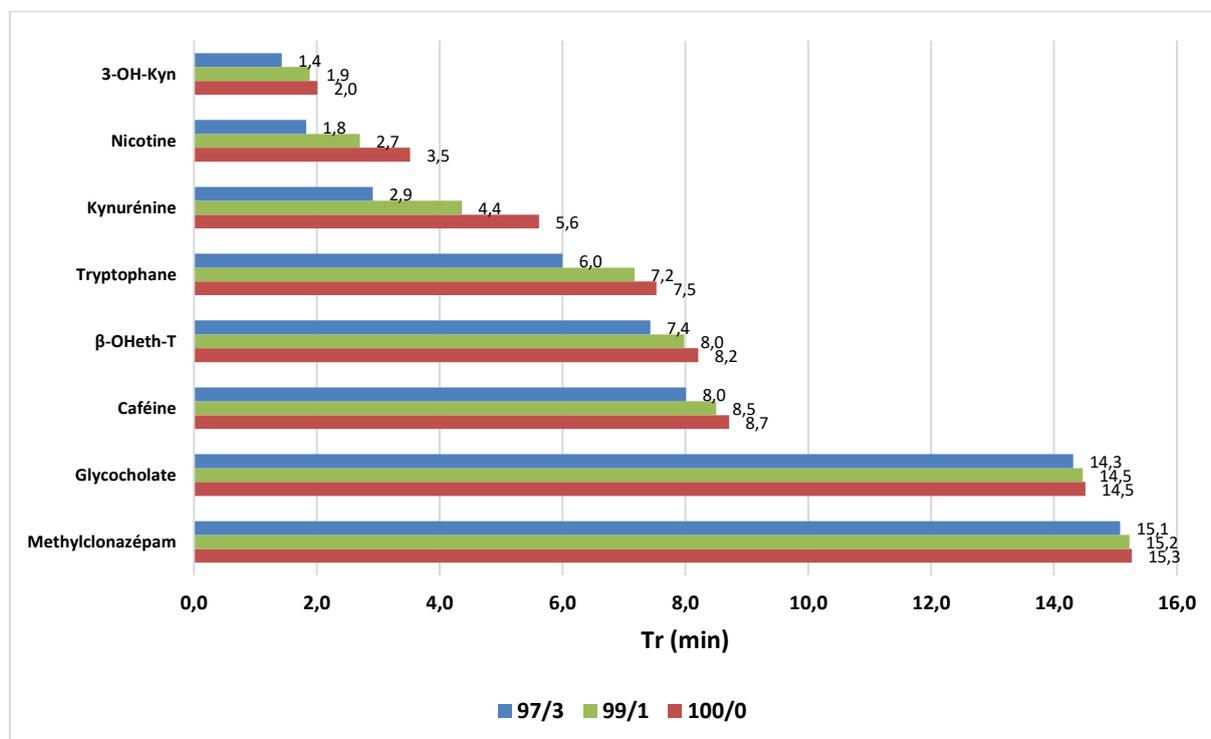


Figure 20 : Influence de la proportion de phase aqueuse/organique de la phase mobile sur le T_r

L'augmentation de la proportion de phase aqueuse entraîne également un élargissement des pics chromatographiques des composés sortant en début de gradient, pouvant aller jusqu'à un doublement de ce paramètre pour certains composés (Figure 21).

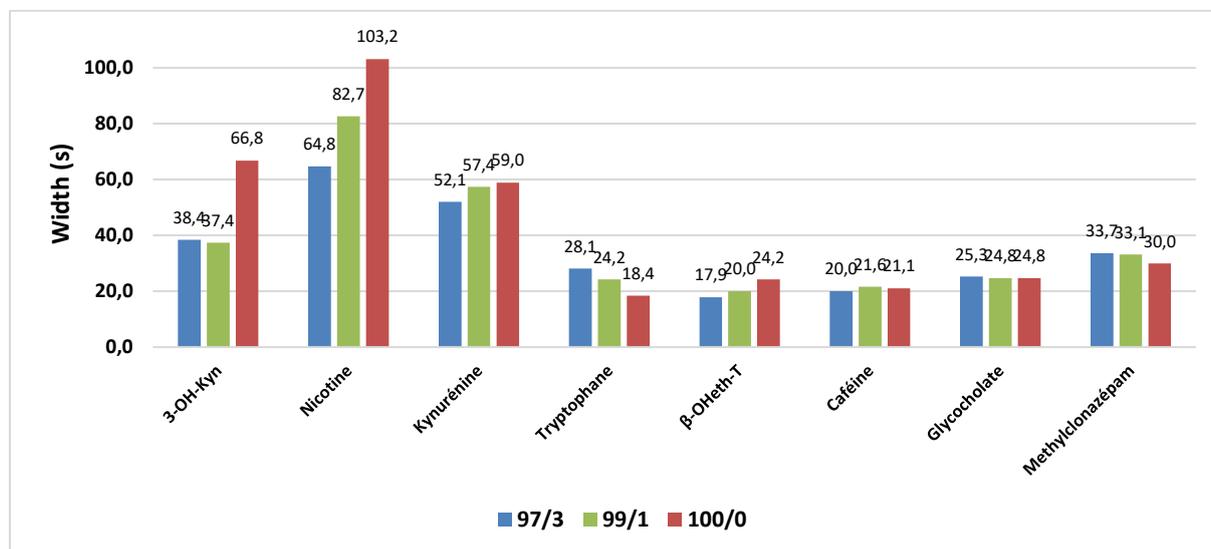


Figure 21 : Influence de la proportion de phase aqueuse/organique de la phase mobile sur la largeur à la base des pics (width)

Ainsi, en augmentant la proportion de phase aqueuse, les composés les plus polaires sont mieux retenus. Dans les conditions testées, certains composés présentent un élargissement de leur pic chromatographique lors de cette augmentation, allant ainsi plutôt vers une tendance à la détérioration des résultats.

1.1.5. Durée de la pente

1.1.5.1. Méthode

Dans les conditions chromatographiques initiales, la pente présente une durée totale de 21 minutes. Un raccourcissement de la durée de cette pente à 19 minutes ou à 13 minutes, a été testé. Les composés influencés par ces variations sont ceux dont le T_r est supérieur à 4 minutes.

1.1.5.2. Résultats

Le raccourcissement de la durée de la pente entraîne un fort écourtement des T_r des composés élués au cours de cette pente (Figure 22). Les composés élués avant le début de la pente ne sont pas impactés. Les composés impactés n'ont pas présenté de modification dans la largeur à la base de leur pic, ni dans la réponse générée.

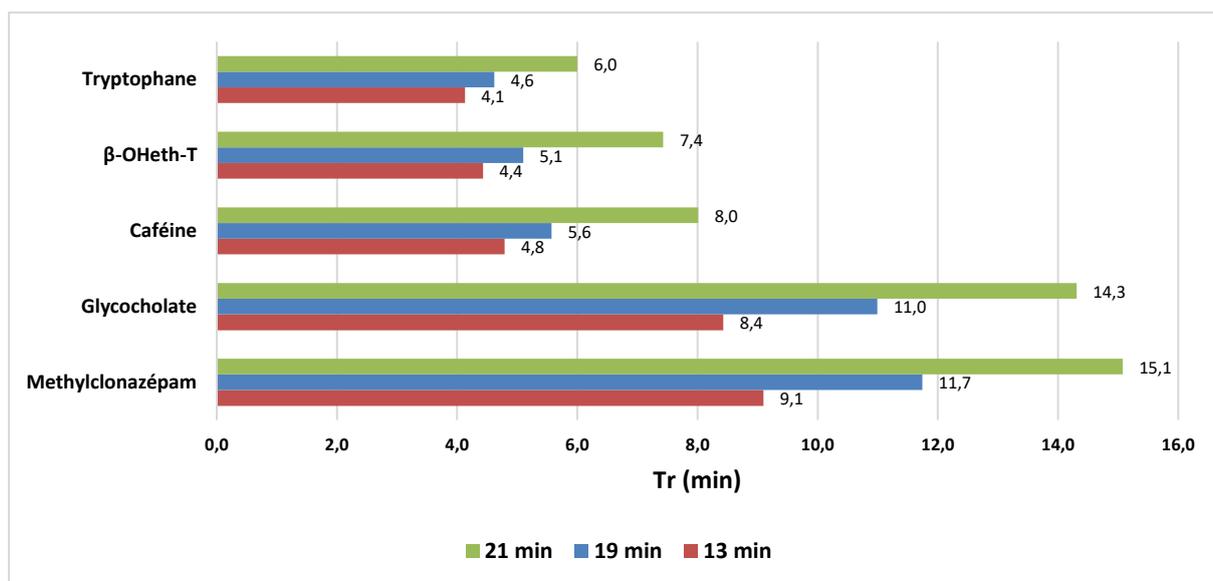


Figure 22 : Influence de la durée de la pente sur le Tr

La réduction de la durée de la pente entraîne donc une élution plus précoce des composés élués au cours de cette période, sans influence péjorative sur la forme des pics chromatographiques.

1.2. Méthode de spectrométrie de masse

Une évaluation de l'apport de la mobilité ionique pour la méthode développée a été réalisée en comparant les performances obtenues entre deux analyseurs : l'UPLC-QTOF (G2-XS-QTOF, Waters®) ayant servi à l'optimisation de la méthode chromatographique, et l'UPLC-IMS-QTOF (Vion Waters®) qui possède une cellule de mobilité ionique.

1.2.1. Matériel

La solution composée des 10 standards est analysée sur l'analyseur G2-XS-QTOF et le Vion. Les paramétrages de la méthode sont identiques pour les deux analyseurs et comme décrit précédemment, hormis la rampe de collision du Vion qui est augmentée (le gaz de collision utilisé est l'azote, la rampe de collision croît de 14 à 56 eV) par rapport à celle du G2-XS (le gaz de collision utilisé est l'argon, la rampe de collision croît de 10 à 40 eV).

1.2.2. Résultats

Tout d'abord, une comparaison des spectres de fragmentation en basse énergie et haute énergie est réalisée. Un exemple est donné Figure 23. Grâce à l'utilisation de la mobilité ionique, les spectres sont beaucoup plus épurés, ce qui présente un avantage en termes de spécificité.

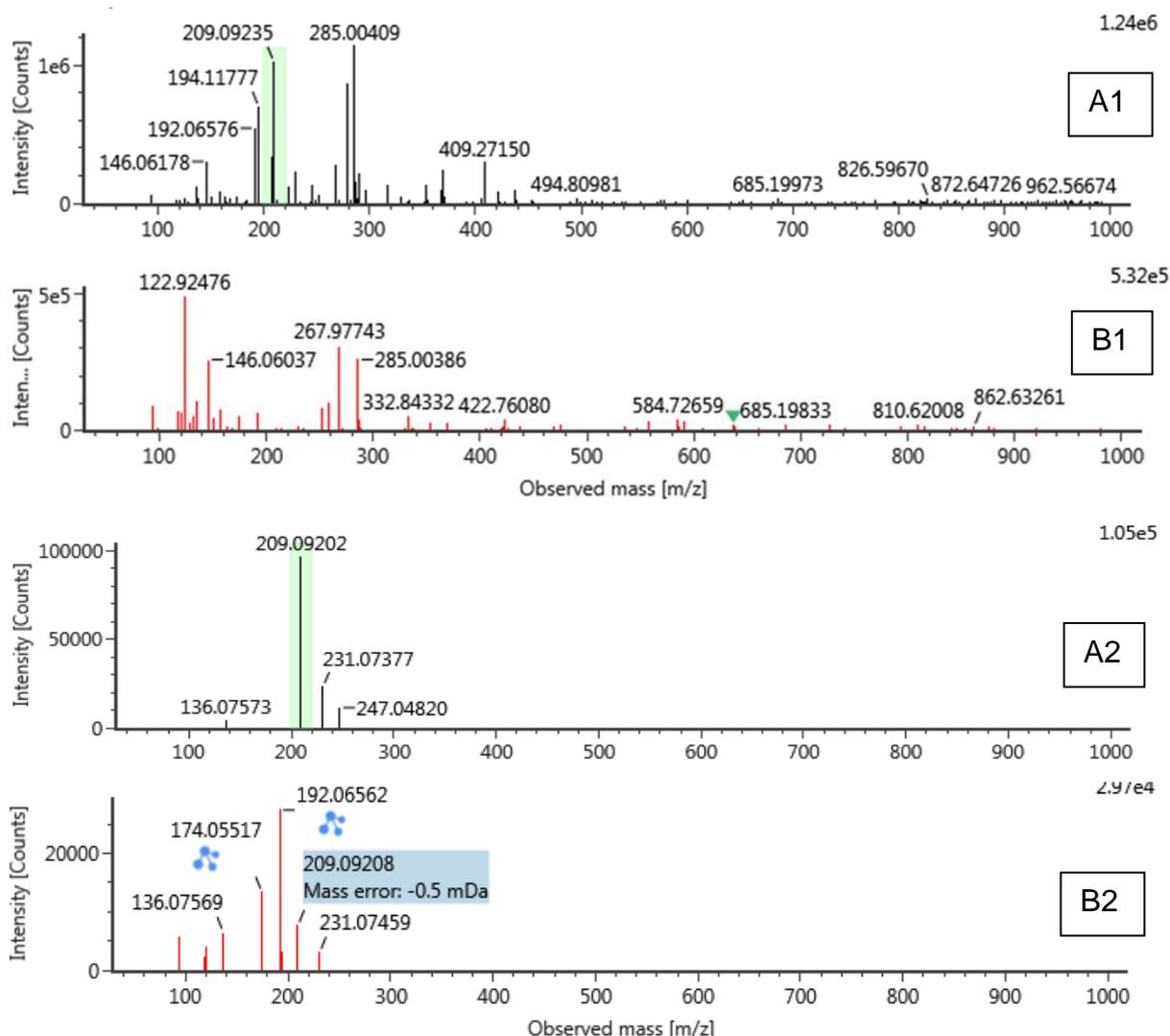


Figure 23 : Spectre de fragmentation en basse énergie (A) et haute énergie (B) de la kynurénine après analyse sur le G2-XS (A1 et B1) et le Vion (A2 et B2)

La comparaison des chromatogrammes met en évidence une meilleure délimitation des pics chromatographiques, facilitant ainsi son intégration, et à terme la reproductibilité des intégrations lors de l'analyse de plusieurs échantillons différents. Comme illustré dans la Figure 24, l'épaulement présent sur le pic d'arginine pour l'échantillon analysé avec le G2-XS peut générer des erreurs de « processing » et un

défaut de reproductibilité de l'intégration pour d'autres échantillons. Cet épaulement n'est pas observé en réalisant la même analyse sur le Vion.

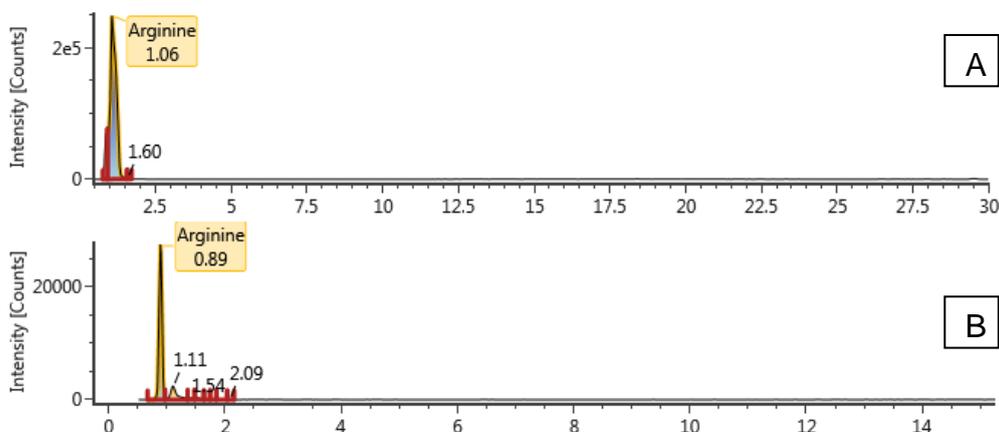


Figure 24 : Chromatogramme de l'arginine analysée avec le G2-XS (A) et le Vion (B)

Les différents paramètres illustrés dans la Figure 15, ont permis d'évaluer les performances analytiques testées sur les deux analyseurs. Les résultats sont décrits dans le Tableau 8. Les performances analytiques testées sont évaluées soit en différence (Vion - G2XS) pour le Tr, soit en pourcentage pour les autres critères ($[\text{Vion} - \text{G2XS}] / \text{G2XS} * 100$). Les variations positives sont surlignées en gris, tandis que les variations négatives présentent un fond blanc.

Tableau 8 : Comparaison des performances entre le Vion et le G2-XS

Composé	ΔTr (min)	ΔH (%)	ΔWidth (%)	$\Delta\text{Points par pic}$ (%)	$\Delta\text{Réponse}$ (%)	ΔBruit (%)	$\Delta\text{S/N}$ (%)
Arginine	-0,17	-89,1	-54,1	-57,5	-88,4	-60,4	-73,3
Créatinine	-0,08	-99,9	-43,9	-49,2	-99,7	-99,9	0,0
3-hydroxy-kynurénine	-0,12	-26,3	-69,2	-72,0	23,2	-82,4	317,2
Nicotine	-0,14	-96,9	-72,8	-74,4	-91,5	-99,3	362,0
Kynurénine	0,06	-79,7	-44,1	-50,0	-60,6	-37,8	-67,5
Tryptophane	0,25	-92,5	-29,5	-37,5	-79,9	-86,8	-42,9
β -hydroxyéthyl-théophylline	0,83	-42,7	-7,5	-16,4	38,8	-52,9	21,3
Caféine	0,78	-72,7	-11,2	-20,8	-9,3	-92,9	266,8
Ac glycocholique	0,73	-19,9	-30,8	-37,0	149,2	-86,7	523,0
Méthylclonazépam	0,76	-46,4	4,5	-6,7	-46,2	-75,0	106,5

Le T_r de chaque composé présente une faible variation, due à l'utilisation d'un module de chromatographie différent, et sans incidence sur la qualité des résultats. Les pics chromatographiques présentent globalement une diminution de leur largeur à la base avec le Vion, signe d'un affinement de ces derniers. Le nombre de points par pic diminue lors de l'utilisation du Vion, mais cette valeur reste systématiquement supérieure à 20. Ce constat est en faveur d'une amélioration, puisqu'un trop grand nombre de points par pic peut être à l'origine d'erreurs lors du « processing » des données entraînant des intégrations de pics erronés. La diminution de la réponse est en cohérence avec la diminution observée de la hauteur du pic associée à celle de la largeur à la base du pic. Bien que la réponse soit moins importante avec le Vion, la diminution importante du bruit, obtenue grâce à l'utilisation de la mobilité ionique, permet globalement une nette amélioration du rapport signal/bruit. Ainsi, l'association de la mobilité ionique à l'analyseur utilisé permet d'améliorer la qualité des chromatogrammes et des spectres de masse et de gagner en spécificité, mais néanmoins avec une baisse de la sensibilité.

1.3. Conditions analytiques finales choisies

Une comparaison de différents paramètres analytiques a pu être réalisée entre l'analyseur UPLC-IMS-QTOF (Vion Waters®) et UPLC-QTOF (G2-XS-QTOF, Waters®), en faveur de l'intégration du spectromètre de mobilité ionique.

L'optimisation de la méthode chromatographique a été réalisée en faisant varier quatre paramètres principaux. La modification de la pente du gradient de phase mobile a montré des bénéfices en termes de raccourcissement des T_r , de réponse et d'affinement des pics, orientant le choix vers une pente linéaire. De plus, la pente linéaire est largement plus universelle en comparaison à d'autres analyseurs, justifiant de surcroît ce choix dans l'objectif de publication de notre méthodologie de travail. La réduction de la durée de la pente a également été retenue, puisqu'elle entraîne une élution plus précoce des composés élués au cours de cette période, sans influence péjorative sur la forme des pics chromatographiques, dans l'objectif de raccourcir la durée totale de la chromatographie. Au regard des améliorations déjà obtenues grâce à ces modifications, la modification de la durée du plateau en début de gradient n'a pas été retenue. En revanche, en augmentant la proportion de phase aqueuse, les composés les plus polaires sont mieux retenus. Dans les conditions testées, certains

composés présentent un élargissement de leur pic chromatographique lors de cette augmentation, allant ainsi plutôt vers une tendance à la détérioration des résultats. La condition choisie est donc de commencer le gradient de phase mobile à 100 % de phase aqueuse sur une durée de plateau de 1 minute. Après 1 minute, l'apport progressif de phase organique sur 3 minutes permettra de pallier l'élargissement des pics chromatographiques. En fin de gradient, un plateau de 3 minutes de 99 % de phase organique permet d'éluer l'ensemble des composés contenus dans l'échantillon, avant un retour dans les conditions initiales de gradient pour rééquilibrer la colonne avant une nouvelle injection. Cette méthode chromatographique, optimisée pour les analyses en mode d'ionisation positif, sera identique pour la méthode ayant recours à l'ionisation négative.

Au total, la méthode optimisée retenue utilise donc le système UPLC-IMS-QTOF Vion (Waters®), dont les détails techniques sont décrits dans le Tableau 9 pour la méthode chromatographique, et le Tableau 10 pour la méthode de spectrométrie de masse. La séparation chromatographique est réalisée sur une colonne de type phase inverse Acquity UPLC® HSS T3 (150 × 2,1 mm ; 1,8 µm, Waters®) maintenue dans une enceinte thermostatée à 50°C. Le temps d'analyse est de 23 minutes. L'ionisation est faite en modes électrospray positif et négatif.

Tableau 9 : Gradient d'élution pour la méthode optimisée

Temps (min)	Débit (mL/min)	Phase mobile A (%) : Formiate d'ammonium 5 mM pH 3	Phase mobile B (%) : Acétonitrile avec 0,1 % d'acide formique	Curve
0	0,4	100	0	6
1		100	0	
2		99	1	
4		97	3	
17		1	99	
20		1	99	
20,5		100	0	
23		100	0	

Tableau 10 : Conditions analytiques de spectrométrie de masse pour la méthode optimisée

Température de source	120 °C
Gaz de nébulisation	50 L/h
Gaz de désolvatation	1000 L/h
Température de désolvatation	600 °C
Tension de cône	20 V
Tension des capillaires (inlet et Lockmass)	0,5 kV (ESI +) / 2,5 kV (ESI -)
Gamme de masse	50 et 1000 <i>m/z</i>
Acquisition en mode MS ^E	Basse énergie : 6 eV Haute énergie : rampe de collision de 14 à 56 eV
Scan time	0,25 s
Lockmass - Leucine Enkephaline	Débit = 10 µL/min Concentration = 2 ng/µL Acquisition toutes les 5 minutes « Scan time » = 0,25 s

2. Création d'une librairie interne de référence

Une méthode de séparation et de détection par LC-HRMS a été développée pour les analyses de métabolomique. En métabolomique, l'identification des composés détectés repose sur la confrontation des données expérimentales à des bases de données de métabolites. C'est pourquoi une librairie interne de référence est mise en place à partir de standards de métabolites.

2.1. Matériel et méthode

2.1.1. Échantillons

Les standards de métabolites endogènes (Sigma MSMLS library) ont été achetés chez Sigma-Aldrich. Cette librairie commerciale a été choisie pour la grande diversité chimique et fonctionnelle des métabolites qu'elle contient. Elle contient 634 standards échantillonnés dans sept plaques de 96 puits à raison de 5 µg de composé pur par puits, dont 37 doublons. Un tableau associé contenant des informations telles que le nom du métabolite, la formule moléculaire, le SMILES (Simplified Molecular Input Line Entry System représentation, qui correspond à la description de la structure d'un composé sous langage symbolique (150)) a été utilisé pour alimenter notre librairie interne (ou librairie « maison »). Des solutions stocks de ces standards ont été préparées selon les recommandations préconisées par le fournisseur (5 % de méthanol pour les plaques 1-5 et chloroforme:méthanol:1:1:0,3 pour les plaques 6-7) pour obtenir une concentration finale à 20 µg/mL. Les solutions utilisées étaient les suivantes : méthanol (qualité UPLC-MS, Waters®), eau purifiée Milli-Q (Millipore,

Burlington, MA, États-Unis), acide formique (qualité UPLC-MS, Honeywell, Charlotte, NC, États-Unis), et chloroforme (VWR Chemicals, Radnor, PA, USA). Les solutions stocks de standards de métabolites sont mélangées par 12 composés à une concentration finale de 1,6 µg/mL, puis injectées au volume de 15 µL pour l'acquisition des données spectrales, chromatographiques et de mobilité ionique de référence en ionisation positive et négative. Le système analytique et les conditions chromatographiques et de spectrométrie de masse utilisés sont ceux décrit précédemment. Les adduits pris en compte étaient [M+H]⁺, [M+K]⁺, [M+Na]⁺, [M-H]⁻, [M+Cl]⁻, [M+HCOO]⁻ et [M+CH₃OO]⁻. Pour chaque mode d'ionisation, les analyses sont réparties en 7 séquences comprenant chacun 8 échantillons.

2.1.2. Méthodologie d'indexation dans la librairie

L'analyse des données a été réalisée de manière semi-automatique à l'aide du logiciel UNIFI (version 1.9.4 Waters[®] MS Technologies, Manchester, UK), suivie d'une vérification manuelle. La base de données a été construite par l'indexation de la masse exacte, le temps de rétention, le spectre de fragmentation et la valeur de CCS de chaque composé. L'indexation de ces données dans la librairie de référence entraîne la création d'une « fiche métabolite » pour chacun des métabolites. Dans chacune d'elles, des données complémentaires recherchées dans les bases de données sont ajoutées, telles que les codes d'identification des métabolites (HMDB, KEGG, Chempider), des propriétés physico-chimiques (LogP, pKa, charge), ou encore les voies métaboliques d'intérêt dans lesquelles un métabolite est impliqué.

2.2. Résultats

L'analyse des échantillons de la librairie commerciale Sigma a permis de caractériser 266 métabolites en mode ESI + et 301 métabolites en ESI -. 187 métabolites sont communs entre les deux modes d'ionisation. Au total, ces deux librairies réunissent 380 métabolites différents. Les métabolites appartiennent à différentes classes chimiques comme des glucides, des acides carboxyliques, des lipides, des nucléotides et des composés organohétérocycliques (Figure 25), impliqués dans différentes voies métaboliques. Ces deux librairies internes de référence ainsi développées réunissent les paramètres d'identification de chaque métabolite, c'est-à-dire la masse exacte, le Tr, le spectre de fragmentation et la valeur de CCS.

Certains composés n'ont pas pu être indexés dans la librairie. Ceci peut être dû à une réponse trop faible, une masse hors de la gamme de masse choisie, un pic chromatographique inexploitable (forme du pic, dédoublement...).

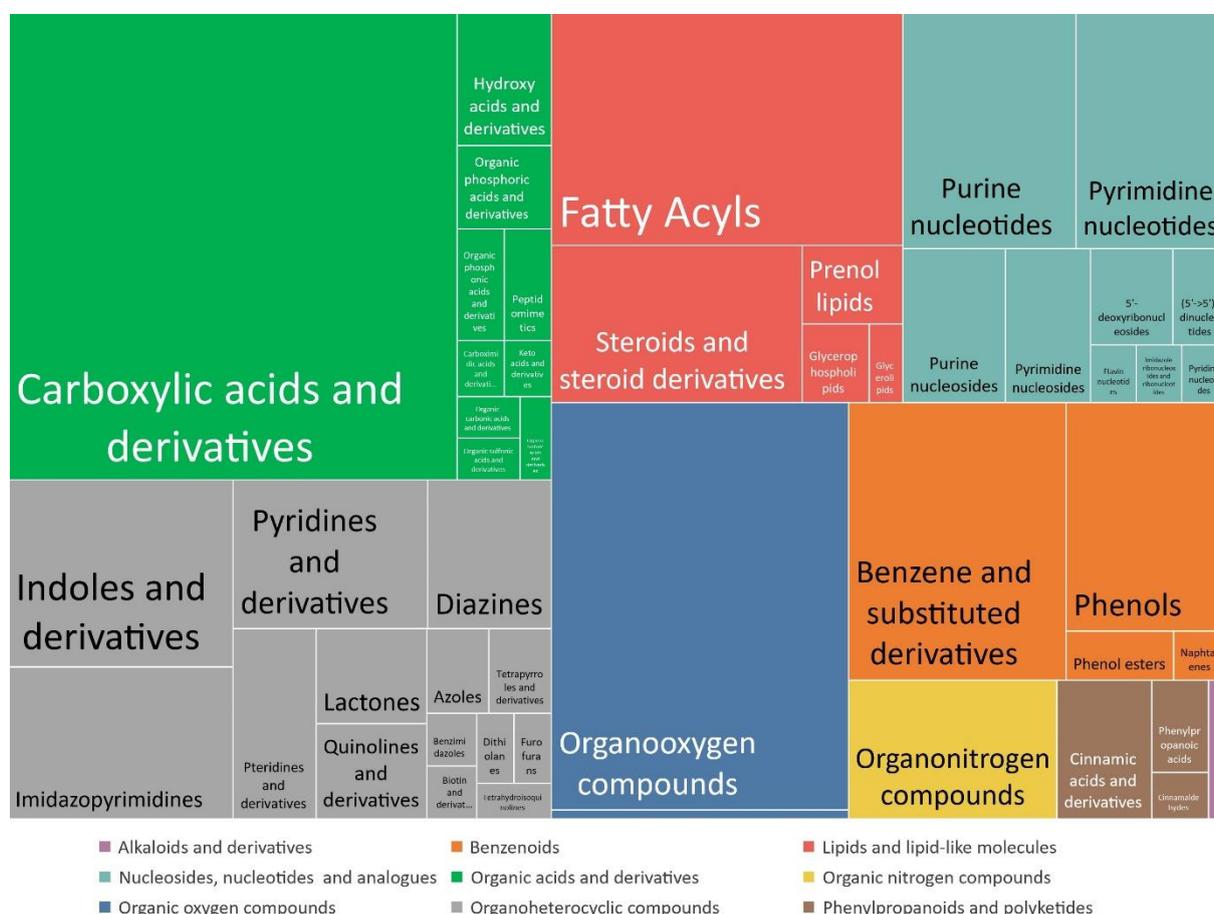


Figure 25 : Taxonomie chimique des métabolites inclus dans la librairie interne de référence et proportion de chaque superclasse et classe

3. Création d'une librairie de métabolomique non ciblée

Les données acquises en MS^E peuvent faire l'objet d'un retraitement non ciblé. Dans ce cas, l'étape d'identification des composés est réalisée en confrontant les données expérimentales à celles de bases de données. Leur confrontation aux données de la librairie interne permet une identification avec un niveau de confiance 1 selon le consensus de la "Metabolomics Standards Initiative (MSI)" (110). Néanmoins, l'identification est ici limitée par le nombre de métabolites présents dans la librairie interne de référence, et reste donc limitée en termes de fraction identifiable du métabolome. Il est donc nécessaire de confronter les données expérimentales à des bases de données en ligne, qui rassemblent généralement les données de masse

exacte et de fragmentation. Les données de Tr ne sont jamais disponibles dans ces bases de données, puisque spécifiques des conditions chromatographiques utilisées. Les données de CCS sont quant à elles rarement disponibles. Dans ce contexte, une identification sûre à partir des données de masse exacte et de fragmentation peut rapidement être un challenge, notamment lorsqu'aucun candidat n'est trouvé (non indexé dans les bases de données), ou lorsque plusieurs candidats sont proposés pour un même composé (faux positifs), augmentant le risque d'une identification erronée (151,152). C'est pourquoi l'utilisation d'autres propriétés orthogonales d'identification, comme le Tr ou le CCS, peuvent permettre une meilleure confiance dans l'identification des métabolites. L'utilisation d'outils d'intelligence artificielle, comme le *machine learning*, peut permettre de prédire ces données supplémentaires d'identification.

L'objectif de ce travail est donc de développer une base de données de prédiction des CCS et des Tr à partir des données de la librairie interne de référence précédemment développée. L'intérêt de cette approche est de pouvoir apporter des propriétés orthogonales supplémentaires qui seront utilisées lors de la confrontation des données expérimentales aux bases de données. Avec cette stratégie basée sur le *machine learning*, les valeurs de CCS et de Tr peuvent être combinées avec la masse monoisotopique, le spectre de fragmentation, et la similarité isotopique ce qui offre de nouvelles possibilités pour réduire le nombre de candidats proposés pour un même composé, accélérant ainsi le processus d'identification et limitant le risque d'erreurs. Par conséquent, cette approche permettra une meilleure interprétation biologique des données de métabolomique non ciblée.

Plusieurs études ont développé ou appliqué des approches de prédiction de CCS par *machine learning* (153). Des outils comme AllCCS (154), CCS Predictor (155), DeepCCS (156), MetCCS Predictor (157) ou LipidCCS predictor (158) permettent de créer des modèles de prédiction à partir de descripteurs moléculaires (159). Ce sont des informations numériques exprimant les propriétés structurales et physico-chimiques des métabolites (ex : polarité, LogP...) de manière exploitable par un modèle mathématique (160). *A contrario*, CSSBase est une interface web (<https://CCSbase.net>) qui donne accès à un modèle prédictif prêt à l'emploi, permettant la prédiction rapide des CCS directement à partir des structures SMILES. Cette plateforme peut donc être plus facilement mise en œuvre.

Il existe également des approches de prédiction de Tr par *machine learning*, y compris des modèles QSRR (quantitative structure retention relationship) (161–165). Dans ces modèles, les prédictions s'appuient sur les caractéristiques structurales et physico-chimiques des analytes, issues des SMILES, et des Tr acquis pour des composés standards. L'identification des relations entre la rétention chromatographique et les descripteurs moléculaires permet théoriquement de prédire le Tr pour tout métabolite disposant de descripteurs moléculaires (166). Des logiciels, tels que le QSRR Automator (167), permettent de faciliter la création de modèles de prédiction.

3.1. Matériel et méthode

3.1.1. Bases de données

Les données de CCS issues de la librairie interne de référence décrite précédemment sont utilisées pour valider les prédictions réalisées par *machine learning*. Les données de Tr issues de la librairie interne de référence décrite précédemment sont utilisées pour optimiser et entraîner un modèle de *machine learning*, et pour valider les prédictions réalisées par ce dernier. Les prédictions sont réalisées pour les composés de la base de données HMDB, base de données électronique d'accès gratuit contenant des informations détaillées sur les métabolites retrouvés dans le corps humain. La version 4.0 de cette base de données contient 114 000 entrées comprenant aussi bien des métabolites polaires que des lipides (105). Cette base de données peut être téléchargée, utilisée et adaptée pour générer une base de données d'identification interne plus large.

3.1.2. Méthodologie de prédiction des CCS

La plateforme CCSbase (<https://CCSbase.net>) (150) a développé et validé un modèle d'apprentissage automatique permettant de prédire le CCS des différents adduits d'un composé à partir de son SMILES. L'ensemble des composés de HMDB v4.0 a été soumis à ce modèle, permettant la prédiction des CCS des adduits [M+H]⁺, [M+K]⁺, [M+Na]⁺, [M+Na-2H]⁻, [M+NH₄]⁺, [M]⁺, [M-H]⁻ et [M]⁻. Une validation des résultats obtenus a été réalisée en comparant, par régression linéaire, les CCS prédits et les CCS observés pour différents adduits des standards analysés lors du développement de la librairie interne de référence. Le modèle est évalué par le coefficient de corrélation R² et l'erreur absolue moyenne entre les résultats de prédiction et les

résultats mesurés. La droite de meilleur ajustement a été calculée, avec un intervalle de +/- 2 écarts-types (ET). Les analyses statistiques et figures associées ont été réalisées à l'aide du logiciel R (96).

3.1.3. Méthodologie de prédiction des temps de rétention

Le modèle QSRR (167) permet une prédiction du Tr des composés de la base HMDB à partir des Tr des standards de composés analysés lors du développement de la librairie interne de référence. La méthodologie de prédiction est illustrée Figure 26. Les données sont séparées en un groupe d'apprentissage (training set) et un groupe de validation (validation set) du modèle. Dans un premier temps, des descripteurs moléculaires, établis à partir des SMILES, sont déterminés par le logiciel pour chaque composé. L'apprentissage du modèle est basé sur l'association des descripteurs moléculaires aux Tr mesurés à partir du sous-groupe d'apprentissage. Ainsi, l'algorithme de *machine learning* apprend la "règle" entre les descripteurs moléculaires les plus influents sur les performances du modèle et les valeurs de Tr expérimentales, dans le but d'établir des modèles de prédiction et sélectionner le meilleur modèle. Différentes méthodes comme la machine à vecteurs de support (Support vector regression, SVR) ou encore les forêts d'arbres décisionnels (Random Forest) peuvent être testées. Le modèle est évalué par le coefficient de corrélation R^2 entre les données de prédiction et celles d'apprentissage, et l'erreur absolue moyenne. Le modèle est ensuite validé par validation croisée (cross validation) à partir du sous-groupe de validation, rassemblant des données qui n'ont pas été utilisées dans le développement du modèle. La validation croisée est également évaluée avec les mêmes paramètres. Une fois qu'un modèle valide a été sélectionné, des prédictions de Tr ont été effectuées pour les métabolites de la base HMDB v4.0.

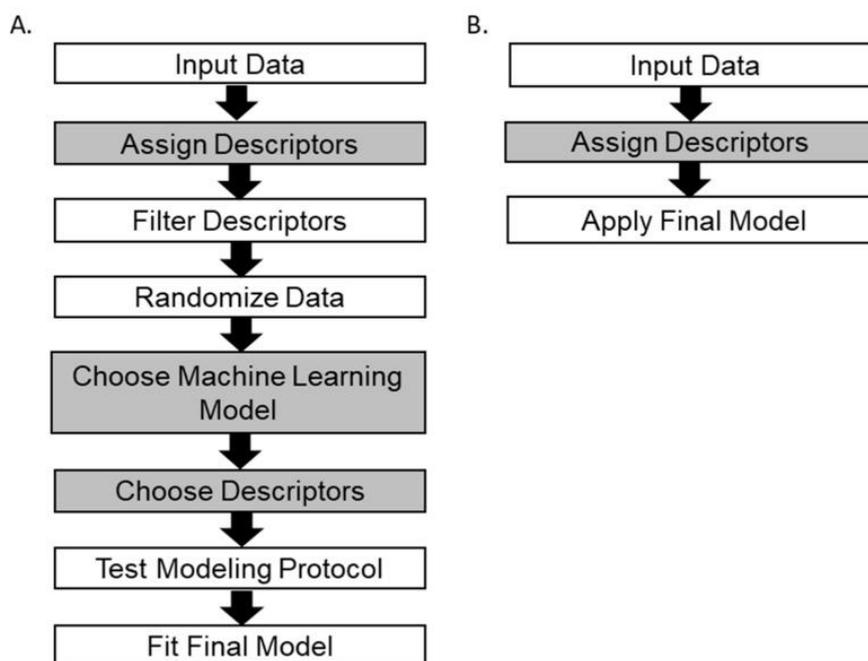


Figure 26 : Méthodologie de prédication de QSRR Automator (167)

A : génération du modèle, B : application du modèle généré pour obtenir des prédictions

3.2. Résultats

3.2.1. Prédiction des CCS

La prédiction des CCS a été réalisée sur près de 114 000 composés de la base de données HMDB v4.0, générant 916 104 valeurs de CCS en incluant tous les adduits. Les résultats ont été validés avec un ensemble de validation composé de 501 valeurs de CCS mesurées à partir de 297 standards dans les deux modes d'ionisation. La corrélation entre CCS prédits et CCS mesurés est illustrée Figure 27. La ligne bleue correspond au meilleur ajustement de la régression linéaire, tandis que les lignes en pointillé correspondent au meilleur ajustement ± 2 ET. Les valeurs aberrantes (« outliers ») ont été conservées en l'absence de toute preuve qu'elles étaient le résultat d'une erreur technique ou de prédiction. Le score R^2 de la régression linéaire est évalué à 0,938. L'erreur absolue moyenne a été calculée à $3,94 \text{ \AA}^2$ tandis que l'ET atteint $6,11 \text{ \AA}^2$, soit 3,36 %. Le CCS prédit est égal à $0,95 \times \text{CCS mesuré} + 7,92$. Le tableau de sortie des CCS qui résulte des prédictions nous a permis de construire notre base de données de référence interne prédite.

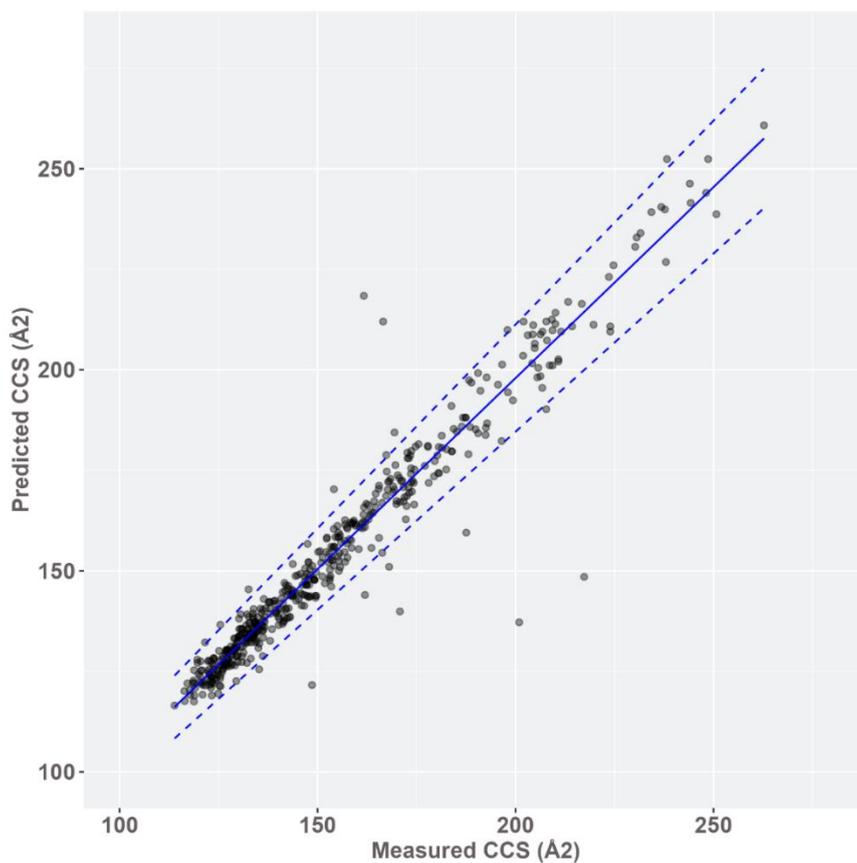


Figure 27 : CCS prédit (\AA^2) à l'aide de l'algorithme CCSBase en fonction du CCS mesuré (\AA^2)

3.2.2. Prédiction des temps de rétention

204 composés de la librairie interne de référence ont été sélectionnés pour le modèle QSRR. 114 ont été détectés à la fois dans ESI + et ESI -, tandis que 90 ont été détectés uniquement dans un mode d'ionisation (45 pour chaque mode d'ionisation). L'algorithme de SVR basé sur 113 descripteurs moléculaires a présenté les meilleures performances, avec un score R^2 de 0,999 et une erreur absolue moyenne de 0,10 min pour l'ensemble d'entraînement (Figure 28).

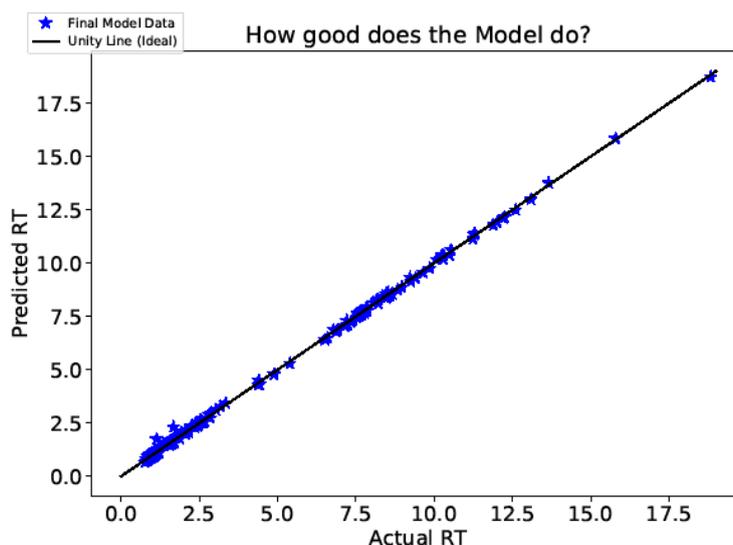


Figure 28 : Tr (min) prédit à l'aide de l'algorithme QSRR Automator en fonction du Tr (min) mesuré
Cette figure a été générée par QSRR Automator à partir de toutes les données disponibles pour la
construction du modèle (y compris les données d'apprentissage)

Le set de validation testé par validation croisée (n = 10) a validé le modèle avec les performances suivantes : moyenne du score R^2 du SVR égale à 0,898, erreur absolue moyenne égale à 0,81 min, écart-type de l'erreur absolue moyenne égal à 0,15 min. L'ensemble des performances du modèle établi est résumé dans le Tableau 11. Les prédictions de Tr ont ensuite été effectuées pour près de 114 000 métabolites de la base HMDB v4.0. Le tableau de sortie qui résulte des prédictions du modèle QSRR nous a permis de construire notre base de données de référence interne prédite.

Tableau 11 : Performances du modèle SVR utilisant l'algorithme QSRR Automator

Scoring Metric	Value(s)
Final Model R^2 Score	0.999
Final Model Mean Absolute Error	0.100
Cross Validation R^2 Values	[0.917 0.938 0.849 0.890 0.884 0.843 0.909 0.872 0.949 0.926]
Mean of Cross Validation R^2 Values	0.898
Median of Cross Validation R^2 Values	0.900
Standard Deviation of Cross Validation R^2 Values	0.034
Cross Validation Mean Absolute Error Values	[0.933, 0.649, 0.761, 0.948, 1.035, 1.059, 0.666, 0.752, 0.661, 0.676]
Mean of Cross Validation Mean Absolute Error Values	0.814
Median of Cross Validation Mean Absolute Error Values	0.756
Standard Deviation of Cross Validation Mean Absolute Error Values	0.154

4. Discussion

La LC-HRMS est une approche analytique centrale largement utilisée en métabolomique (90). Nous avons utilisé un UPLC-IMS-QTOF (Vion, Waters®) pour développer une méthode analytique adaptée à l'analyse du métabolome. La séparation est réalisée sur une colonne Acquity UPLC HSS T3 qui possède une bonne rétention des composés polaires et une compatibilité avec la phase mobile aqueuse plus importante par rapport aux phases stationnaires plus classiques. Elle peut être utilisée pour la rétention d'analytes mi-polaires à apolaires. Le gradient de phase mobile a été optimisé pour obtenir une bonne séparation des composés, affiner la forme des pics chromatographiques, tout en gardant une rétention des composés polaires en début de gradient. La durée du gradient a été également optimisée, pour une durée totale de 23 minutes, plus compatible avec l'analyse de séries comprenant de nombreux échantillons. Une approche de métabolomique telle que développée ici reste limitée en termes de fraction du métabolome détectable. Seule une combinaison de plusieurs méthodes pourrait permettre une meilleure représentativité du métabolome (149). Concernant la détection des composés très polaires, l'idéal serait le développement d'une méthode complémentaire, qui mettrait en œuvre une séparation chromatographique sur colonne HILIC (chromatographie liquide à interaction hydrophile, qui utilise une phase stationnaire à greffage hydrophile, plus adaptée pour séparer les composés polaires). Néanmoins, ce type de méthode présente certains inconvénients. En particulier, elle nécessite une phase d'équilibration de la colonne plus importante pour avoir une bonne reproductibilité des T_r et une forme des pics optimale, en une durée d'analyse potentiellement plus longue.

Nous avons créé une librairie interne de référence de 380 métabolites à l'aide d'une série de standards. Elle permettra une identification atteignant le niveau de confiance le plus élevé (niveau 1 selon le MSI (110)). La librairie commerciale Sigma utilisée contient certains métabolites polaires qui ne peuvent pas être retenus dans les conditions analytiques choisies, à l'origine de l'absence de détection de certains métabolites (149). De plus, avec environ 114 000 métabolites répertoriés dans la base de données HMDB, seule une petite fraction a pu être couverte par les standards. Par conséquent, l'identification des composés dans le cadre d'une approche non ciblée de métabolomique reste un challenge.

En prédisant le Tr et le CCS, la librairie de prédictions générée peut potentiellement être un atout pour l'identification des composés en métabolomique non ciblée, apportant une confiance d'identification de niveau 2 ou 3 (110). Ainsi, à partir des résultats expérimentaux, nous avons utilisé deux modèles de prédiction déjà publiés (CCSBase et QSRR Automator) pour prédire les CCS et Tr d'une base de données de grande dimension. Dans le processus d'identification des métabolites, le Tr et le CCS se positionnent comme des paramètres additionnels d'identification, fournissant des informations supplémentaires provenant de la séparation chromatographique et de la mobilité ionique. Augmenter le nombre de critères d'identification apparaît comme pertinent pour améliorer la confiance accordée à l'identification des métabolites, et limiter le nombre de mauvaises identifications.

Pour la prédiction des CCS, CCSBase utilise un modèle de *machine learning* entraîné par ses auteurs à partir d'une combinaison de plusieurs bases de données, permettant de couvrir une grande variété de composés aux propriétés physico-chimiques variées, en améliorant ainsi la qualité des prédictions réalisées. Dans notre étude, la prédiction des CCS a été validée en comparaison à nos données expérimentales avec un faible biais et un R^2 élevé de 0,938. Lors de l'identification d'un composé, les données acquises seront comparées aux données prédites avec une tolérance d'erreur de CCS d'une valeur correspondant à l'erreur moyenne +/- 2 écarts-types, c'est-à-dire un maximum de $16,2 \text{ \AA}^2$. Ce seuil permet ainsi de limiter le nombre de propositions d'identification faussement attribuées. Cet intervalle de tolérance est relativement important par rapport à d'autres travaux publiés par ailleurs, présentant une erreur relative médiane de l'ordre de 3 à 5 % (154–158). Néanmoins, limiter les erreurs d'identification avec la tolérance définie dans notre étude reste très pertinent si l'on considère le nombre important d'identifications possibles lorsque l'on n'utilise pas le CCS comme donnée orthogonale d'identification supplémentaire. De plus, les prédictions ont été réalisées pour les ions protonés et déprotonés ainsi que pour les adduits, chacun ayant le même Tr mais un CCS différent (168). La présence d'adduits pour un même composé est fréquente lors de l'analyse d'échantillons biologiques hétérogènes (169), et apporte une information précieuse puisqu'elle peut permettre une double validation de l'identification.

Pour la prédiction des Tr, la méthodologie utilisée était différente puisque nous avons réalisé les étapes d'apprentissage et de validation du modèle de *machine learning*. En

effet, les données expérimentales de Tr du sous-groupe d'apprentissage ont permis d'entraîner le modèle, tandis que celles du second sous-groupe indépendant ont permis de valider le modèle. Le modèle final obtenu présentait une faible erreur de prédiction. Lors de l'identification d'un composé, les données acquises seront comparées aux données prédites avec une tolérance d'erreur de Tr d'une valeur correspondant à l'erreur moyenne +/- 2 ET, c'est-à-dire un maximum de 1,11 min. Ce seuil permet ainsi de limiter le nombre de propositions d'identification faussement attribuées. Les erreurs de prédiction peuvent être dues à un biais du logiciel de prédiction, à une erreur aléatoire dans les données utilisées ou à des erreurs dans l'attribution d'une identification des standards. Naylor *et al.* ont décrit les performances initiales de QSRR Automator sur différentes colonnes chromatographiques. Ils ont montré que les erreurs de prédiction étaient inférieures à 1 minute pour la majorité des prédictions, et inférieures à 2 minutes pour la quasi-totalité des prédictions (167). Les performances de QSRR sont ici comparables à celles des méthodes publiées antérieurement (164,170). Pour les méthodes de relativement courte durée d'analyse comme la nôtre, de nombreux métabolites ont des Tr très proches, comme par exemple des isomères dont la différence de Tr est inférieure à 10 s. Notre base de données de prédiction ne permet pas de faire la distinction entre ces composés, mais est adaptée pour distinguer des composés bien séparés mais de même masse, permettant de réduire les identifications faussement positives, et ainsi faciliter l'interprétation biologique des résultats. Même si le modèle généré est basé sur des composés aux propriétés physico-chimiques variées, séparés et identifiés avec une méthode optimisée, une attention particulière doit être accordée pour éviter les surinterprétations, notamment concernant : (i) les composés non retenus dans la colonne, (ii) les composés retenus après l'intervalle de Tr mesuré pour les composés du sous-groupe d'apprentissage et (iii) les composés dont les propriétés physico-chimiques diffèrent de celles des composés du sous-groupe d'apprentissage. Par exemple, la base de données de prédiction a été générée à partir d'une grande variété de standards dédiés aux analyses de métabolomique, mais n'incluait pas les composés complexes de haut poids moléculaire. Les prédictions pour ces métabolites doivent donc être exclues.

Il convient de mentionner certaines limites de la présente étude. Le nombre de standards est relativement faible, ce qui se traduit par un nombre limité de données

expérimentales, susceptibles d'influencer les performances des modèles. Toutefois, cette limitation est contrebalancée par la qualité des données, puisque nous avons inclus des données provenant d'authentiques standards, présentant une identification avec le plus haut degré de confiance possible. De plus, la base de données de Tr prédits est strictement liée aux conditions chromatographiques choisies ; elle ne peut donc pas être utilisée en cas de modification des conditions séparatives.

La plupart des modèles de prédiction précédemment publiés traitent de l'une ou l'autre des propriétés prédites (CCS ou Tr), et rarement de la prédiction de plusieurs dimensions de données (171–174). Il est intéressant de noter que ces études se focalisent sur la prédiction de composés lipidiques ou d'autres composés exogènes, tandis que notre travail traite des métabolites humains de façon plus globale. L'utilisation de critères d'identification supplémentaires tels que le CCS et le Tr permet ainsi de réduire de manière significative les identifications faussement attribuées, permettant de réduire le champ de recherche de l'identité du composé et d'améliorer la précision de l'identification.

Pour conclure, une méthode de séparation et de détection par LC-HRMS a été développée pour les analyses de métabolomique. En parallèle, une librairie de référence basée sur les composés de la base de données HMDB a été mise en place et rassemble les données nécessaires à l'identification de chaque composé, c'est-à-dire sa masse exacte, son temps de rétention, et son CCS. Cette base de données de CCS et Tr prédits sera utilisée comme référence pour faciliter l'identification des composés d'intérêt de notre étude toxicologique sur le tabac, les e-cig et l'HTP. Ces résultats ont fait l'objet d'une publication disponible à la fin de ce manuscrit (175). Un exemple d'application en toxicologie illustre l'intérêt d'utiliser cette base de données interne de grande dimension pour réduire le nombre d'identifications faussement positives en métabolomique non ciblée.

CHAPITRE II : APPLICATION BIOLOGIQUE

À la suite du développement analytique réalisé, l'approche de métabolomique est appliquée dans l'objectif d'identifier de potentiels marqueurs précoces de toxicité des émissions de tabac, de tabac chauffé et de cigarette électronique. Pour cela, nous avons réalisé l'analyse du métabolome de BEAS-2B cultivées à l'interface air-liquide et exposées à de l'air stérile (contrôle) ou aux émissions d'une cigarette de référence, d'HTP, ou d'e-cig en utilisant la machine à fumer Vitrocell. L'interprétation des résultats est réalisée par comparaison des profils métabolomiques spécifiques de chaque type d'exposition.

1. Matériel et méthode

1.1. Cultures cellulaires et échantillonnage

1.1.1. Cultures cellulaires

L'étude de la toxicité des aérocontaminants est réalisée *in vitro* sur des cellules BEAS-2B qui sont des cellules immortalisées d'origine épithéliale bronchique humaine. Cette lignée cellulaire est considérée comme étant un modèle *in vitro* pertinent pour l'étude de la toxicité des substances inhalables et a déjà été précédemment utilisée au sein de l'ULR4483 pour l'étude de certains effets toxiques de la cigarette conventionnelle, de la e-cig ou des HTP.

La lignée BEAS-2B estensemencée dans des flasques (CellBIND de 75 cm²) contenant du milieu de culture défini LHC9 et supplémenté d'antibiotiques (pénicilline/streptomycine 1 %). Les flasques sont placées dans un incubateur (T° = 37°C, CO₂ = 5 % et humidité = 85 %). A 80-90 % de confluence, les cellules sont réensemencées sur des inserts Transwell® PET 0,4 µm (75 000 cellules/insert) suspendus dans des plaques 6 puits. A 80 % de confluence des cellules, le milieu de culture recouvrant l'insert est retiré afin de placer les cellules en interface air-liquide ; les inserts sont ensuite placés dans le module d'exposition de la machine à fumer VC1 de Vitrocell® (39,148).

1.1.2. Méthode d'exposition

1.1.2.1. Machine à fumer

La machine à fumer VC1 de Vitrocell® est composée de deux parties (Figure 29). La première partie, permettant la génération des aérosols ou des fumées, est composée d'une pompe à piston dont le paramétrage du volume, de la durée d'aspiration/expiration et de la fréquence d'activation permet de mimer une bouffée. La seconde partie est un module d'exposition des cultures cellulaires à l'interface air-liquide permettant un contact direct des cellules aux aérosols ou fumées générés. Ce module, relié à la première partie par des tuyaux et un cylindre de dilution, est constitué de 3 puits pouvant contenir les inserts issus des cultures cellulaires, qui subissent alors une exposition identique dans chaque puits. Une aspiration active après les inserts permet d'assurer une circulation des fumées ou des aérosols en évitant leur stagnation, reflétant mieux le modèle respiratoire.

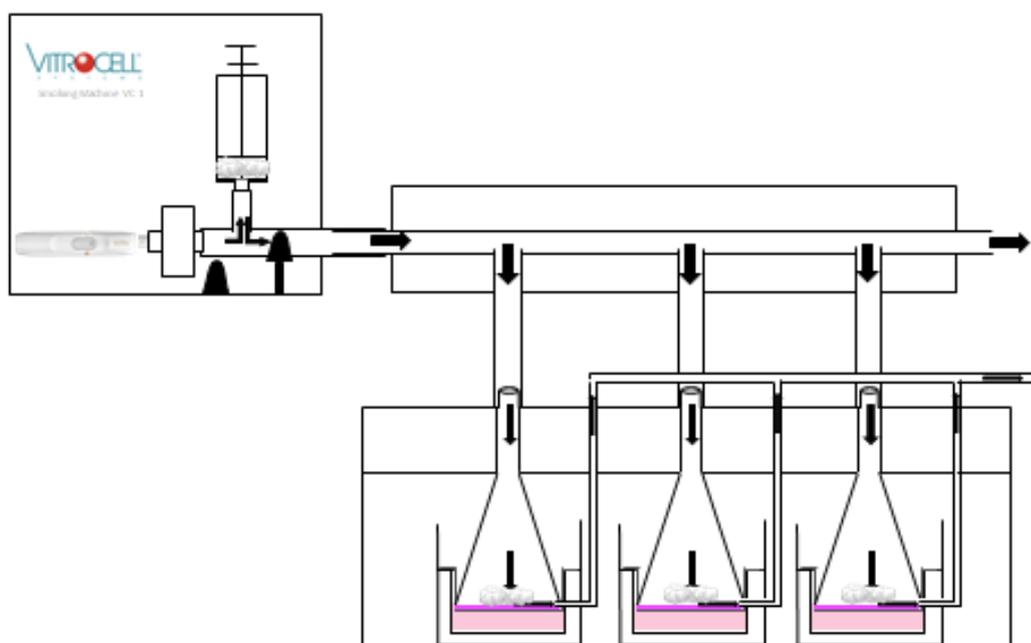


Figure 29 : Dispositif de génération de fumées et d'aérosols (Vitrocell®) et d'exposition des cultures cellulaires

1.1.2.2. *Dispositifs d'exposition*

La fumée de cigarette est générée à partir de cigarettes 3R4F (Lexington, KY, USA). Ces cigarettes sont une référence largement utilisée pour les analyses toxicologiques *in vitro* et *in vivo* car elles sont préparées avec un minimum de variabilité d'une cigarette à l'autre permettant de reproduire et de comparer dans le temps et d'une étude à l'autre les résultats obtenus.

Les aérosols d'e-cig sont générés à partir de modèle ModBox de la marque NHOSS® (Innova, Bondues, France). Il s'agit d'un modèle de 3^{ème} génération dont la puissance et la résistance peuvent être modulées. Pour ce projet, une faible puissance (MB-18W) et une puissance moyenne (MB-30W) ont été testées avec une résistance de 0,5 Ohm. L'objectif d'utiliser ces deux conditions expérimentales est de mettre en évidence un potentiel impact de la puissance de réglage de l'e-cig sur la composition chimique des aérosols et, par voie de conséquence, sur la nature et/ou l'intensité des effets toxicologiques des aérosols. L'e-liquide testé est également de la marque NHOSS® (Innova, Bondues, France). La composition mentionnée sur le produit est de 65 % de propylène glycol et de 35 % de glycérine, complétés d'arômes alimentaires pour lui donner un goût « tabac blond » et de nicotine à la concentration de 16 mg/mL. Pour préchauffer la résistance, le dispositif est activé par appui sur un bouton poussoir, initié 1 s avant le début de la bouffée.

Les aérosols d'HTP sont générés à partir du modèle IQOS 2.4 (PMI, Neuchâtel, Suisse). Les sticks de tabac, également commercialisés par PMI sont de type « ambre », décrits comme donnant un goût proche des cigarettes Marlboro Red. Un stick permet de générer 12 bouffées après activation d'un bouton poussoir. Comme recommandé par le fabricant, le dispositif est nettoyé tous les 20 sticks.

1.1.2.3. *Choix des durées d'exposition*

Les durées d'exposition ont été choisies sur la base des résultats d'études de cytotoxicité précédemment menées au laboratoire (176). L'objectif était de définir les durées d'exposition pour lesquelles une sub-toxicité était observée, c'est-à-dire moins de 20 % de cytotoxicité, afin de comparer les effets des différents aérosols ou fumées entraînant une cytotoxicité similaire. Le profil de bouffée était standardisé selon le profil HCl (Health Canada Intense smoking regime) qui correspond à un volume de bouffée de 55 mL, un temps d'aspiration de 2 s, toutes les 30 s (177). La viabilité cellulaire,

évaluée par la mesure de l'ATP intracellulaire, a été comparée à différents temps d'exposition aux aérosols d'e-cig, d'HTP ou à la fumée de cigarette de référence 3R4F (176). Une sub-toxicité était observée jusqu'à une exposition de 60 min (ou 120 bouffées) pour l'e-cig et l'HTP, et seulement 2 min (ou 4 bouffées) pour la 3R4F, permettant d'établir le choix des durées d'exposition pour ce projet (Tableau 12). Deux durées d'exposition, désignées par « T1 » et « T2 » dans la suite du manuscrit, sont choisies dans l'objectif de mettre en évidence un potentiel impact de la durée d'exposition sur la nature et/ou l'intensité des effets toxicologiques des émissions étudiées. De plus, des inserts sont laissés dans l'incubateur sans exposition et servent de contrôle négatif (désignés par « T0 »). Des tests préalablement réalisés n'avaient en effet pas objectivé de différence significative sur la viabilité cellulaire entre des cellules exposées à de l'air stérile ou des cellules laissées dans l'incubateur sans exposition (39).

Tableau 12 : Protocole d'exposition

	Contrôle (T0)	Durée d'exposition 1 (T1)	Durée d'exposition 2 (T2)
3R4F	0 min	1 min	2 min
MB-18W	0 min	30 min	60 min
MB-30W	0 min	30 min	60 min
HTP	0 min	30 min	60 min

Chaque exposition a été répétée sur 4 cultures cellulaires différentes, générant un total de 48 échantillons.

1.1.3. Extraction des métabolites

Vingt-quatre heures après les expositions, le milieu de culture est récupéré et conservé à -80 °C. L'échantillon sur insert est quant à lui préparé pour permettre une extraction optimale des métabolites en vue des analyses de métabolomique. Après rinçage au PBS (tampon phosphate salin), les plaques sont incubées avec 1 mL de solution de « quenching » composées de 80 % de méthanol pendant 20 min à -20°C. La suite de la préparation des échantillons est réalisée sur glace. Les inserts sont grattés pour récupérer les cellules, et le lysat est transféré dans un tube Eppendorf avant centrifugation 5 minutes à 14 000 g à + 4°C. Le surnageant est transféré dans un tube Eppendorf, tandis que le précipité est de nouveau nettoyé avec 500 µL de la solution d'extraction et centrifugé selon les mêmes conditions. L'extrait total obtenu est

finalement séché sous speedvac sans chauffage et stocké à -80 °C jusqu'aux analyses. Juste avant injection, l'extrait sec est remis en solution par 100 µL du solvant de reprise composé d'eau à 10 % de méthanol contenant les standards internes : la β-hydroxyethyl-theophylline (Sigma-Aldrich) à 1,6 mg/L et le méthylclonazépam (Roche) à 0,125 mg/L, qui seront détectés en mode d'ionisation positif, et le phénobarbital-D5 (Sigma-Aldrich) à 1 mg/L qui sera détecté en mode d'ionisation négatif.

1.2. Mesure de la cytotoxicité

La lactate déshydrogénase (LDH) est une enzyme cytoplasmique stable présente dans toutes les cellules et rapidement libérée dans le milieu de culture cellulaire après lyse de la membrane plasmique en cas de mort cellulaire par nécrose ou apoptose tardive. L'activité LDH est déterminée à l'aide d'un kit commercial de mesure de la cytotoxicité (Cytotoxicity Detection Kit PLUS, Roche Diagnostics GmbH, Mannheim, Allemagne), faisant intervenir une réaction enzymatique couplée (Figure 30) : le NAD⁺ est réduit en NADH/H⁺ par la conversion catalysée par la LDH du lactate en pyruvate. Dans la deuxième étape, le catalyseur (diaphorase) transfère H/H⁺ de NADH/H⁺ au sel de tétrazolium qui est réduit en formazan. L'augmentation de la quantité de formazan produite dans le surnageant de culture est directement liée à l'augmentation du nombre de cellules lysées. Le colorant formazan est soluble dans l'eau et peut être détecté par spectrophotométrie.

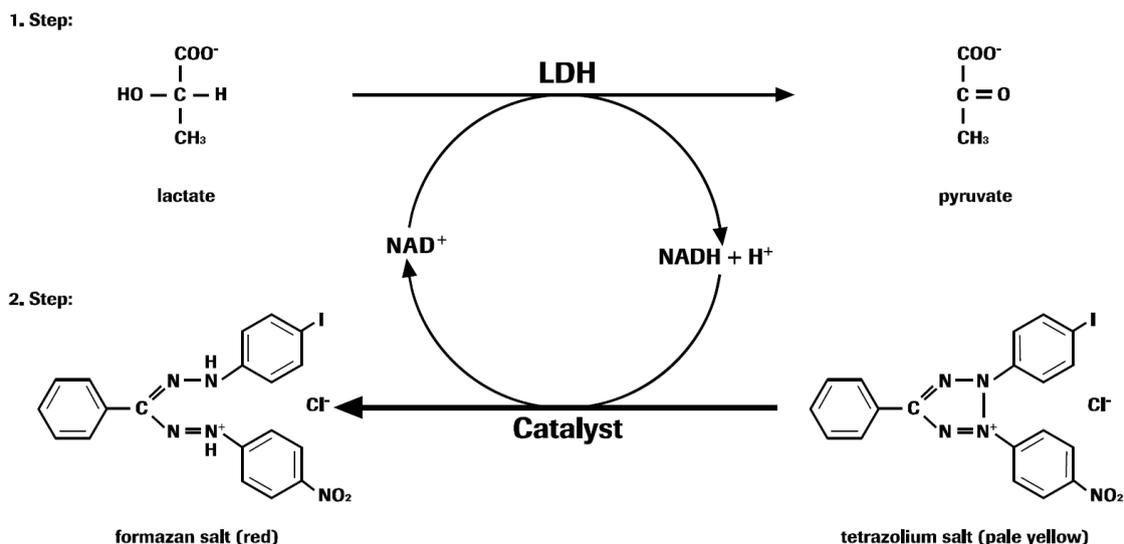


Figure 30 : Méthode de mesure de l'activité de la LDH

L'absorbance finale est définie comme la différence entre l'absorbance mesurée à 492 nm et celle mesurée à 690 nm (longueur d'onde de référence). L'inclusion d'un

échantillon de contrôle bas (milieu de culture de cellules non exposées), d'un contrôle haut (milieu de culture d'un échantillon totalement lysé au triton), et d'un blanc (milieu de culture LHC-9) permet ensuite de calculer pour chaque échantillon un pourcentage de cytotoxicité, selon la formule suivante :

$$\text{cytotoxicité (\%)} = \frac{\text{exp. value} - \text{low control}}{\text{high control} - \text{low control}} \times 100$$

100 µL d'échantillons de milieu de culture, de contrôle ou de blanc sont déposés dans chaque puit d'une plaque 96 puits. 100 µL de solution de réaction sont ajoutés à chaque puit pré-rempli, avant incubation 30 minutes à température ambiante. 50 µl de solution stop sont alors ajoutés dans chaque puit pour arrêter la réaction. Après mélange pendant 10 s, les absorbances sont mesurées à 492 et 690 nm. Ces analyses sont réalisées en triplicats techniques.

A partir des résultats obtenus, les valeurs médianes de la différence d'absorbance ($A_{492} - A_{690}$) permettent de calculer la cytotoxicité médiane mesurée dans chaque échantillon. Une cytotoxicité inférieure à 20 % permet de confirmer que les durées d'exposition choisies sont sub-toxiques.

Pour compléter ces conclusions, un test de Kruskal-Wallis, équivalent non paramétrique du test ANOVA, permet d'évaluer s'il existe des différences significatives entre les groupes 3R4F, HTP, MB-18W et MB-30W, au seuil de significativité de p-value de 0,05. En cas de résultat significatif, un test de Wilcox, équivalent non paramétrique du test de Student, permet de mettre en évidence les différences significatives entre échantillons (T0 vs T1 ou T0 vs T2 ou T1 vs T2), au seuil de significativité de p-value de 0,05. Ces analyses statistiques sont réalisées avec le logiciel R (96).

1.3. Analyses de métabolomique

1.3.1. Analyse des échantillons

Les analyses de métabolomique sont réalisées sur l'analyseur UPLC-IMS-QTOF (Vion Waters®) selon la méthode décrite dans le chapitre I. L'analyseur est piloté par le logiciel UNIFI (Waters®). Les 48 échantillons préparés sont déposés dans l'injecteur maintenu à + 10°C. Le volume d'injection est fixé à 10 µL d'échantillon. Tous les échantillons sont traités au cours d'un premier batch en ESI+ et d'un second batch en ESI-. La séquence analytique est composée de différents échantillons et notamment de :

- Solution commerciale de contrôle, ou « test mix » composée des standards suivants : caféine (ESI+), leucine-enképhaline (ESI+/ESI-), réserpine, sulfadiméthoxine (ESI+/ESI-), sulfaguanidine (ESI+/ESI-), terfénaire (ESI+), Val-Tyr-Val (ESI+/ESI-), vérapamil (ESI+), paracétamol (ESI+). Elle permet d'évaluer les performances de la plateforme analytique, c'est à dire l'état de la colonne et la sensibilité de la spectrométrie de masse. Cette solution utilisée en routine est donc bien caractérisée. Les composés de cette solution commerciale présentent une large gamme de ratios m/z et présentent des Tr différents. Les performances analytiques seront donc rapidement vérifiées ;
- Echantillon de conditionnement, composé d'un mélange des échantillons d'intérêt dilués, pour conditionner la colonne chromatographique avec la matrice d'intérêt ;
- Echantillon de contrôle de qualité (QC), composé d'un mélange des échantillons d'intérêt. Ce QC est qualitativement et quantitativement représentatif des échantillons biologiques étudiés. Il est injecté à intervalles réguliers (tous les 7 ou 8 échantillons environ) au cours du batch, afin de tester la stabilité analytique au sein d'une série et corriger le signal en cas de fluctuation de réponse (pré-traitement des données). Ce QC est également dilué au 1/3 (QCdil) et injectés à intervalles (tous les 15 échantillons environ) ;
- Echantillons d'intérêt, numérotés E1 à E48, dans un ordre d'injection randomisé, pour éviter une distribution non équilibrée ou des clusters et donc des distributions/séparations inégales entre les groupes.

Les différents types d'échantillons injectés permettent de mettre en place plusieurs niveaux de maîtrise et de contrôle de la qualité des analyses (109). L'ordre d'analyse de ces échantillons est décrit à la Tableau 13.

Tableau 13 : Ordre d'analyse des différents types d'échantillons d'après (109)

Ordre de passage des échantillons	Echantillon analysé
1	Test Mix
2	5 échantillons de conditionnement
3	Contrôle de qualité (QC)
4	Contrôle de qualité dilué (QCdil)
5	7 échantillons randomisés
6	Contrôle de qualité (QC)
7	7 échantillons randomisés
8	Contrôle de qualité (QC)
9	Contrôle de qualité dilué (QCdil)
10	7 échantillons randomisés
...	...
12	Contrôle de qualité (QC)
13	Test Mix

1.3.2. Pré-traitement des données en non ciblé

Les données acquises en ESI+ et en ESI- font l'objet d'un retraitement séparément et selon le workflow décrit dans la Figure 12. Il n'existe pas de méthode universelle de traitement des données, nécessitant d'en tester plusieurs. L'analyse des performances de chaque type de traitement testé a permis le choix d'une méthode finale adaptée aux données acquises.

Les données sont exportées depuis le logiciel d'acquisition UNIFI (Waters®) vers le logiciel de traitement de données Progenesis (Waters®) pour être soumises à l'étape d'extraction des composés. Les données subissent les étapes d'alignement, de « peak picking » et de déconvolution. Une fois les données extraites, elles sont exportées sous forme de matrice de données dans le logiciel R.

L'imputation des données manquantes a fait appel à une méthode en deux étapes : 1) retrait des composés dont plus de 50 % des valeurs sont manquantes et 2) pour les composés conservés, les valeurs manquantes sont remplacées par 1/5 de la valeur la plus basse (LOQ). Les données sont ensuite filtrées en fonction du coefficient de variation de chaque variable au cours de l'analyse, calculé selon la méthode de l'écart relatif à la médiane (Relative Standard Deviation) :

$$RSD = \frac{\text{Median Absolute Deviation}}{\text{Median}} * 100$$

Avec la median absolute deviation (MAD, écart absolu médian) :

$$MAD = \text{médiane}(|X_i - \text{médiane}X_i|)$$

La mesure d'intensité est considérée comme fiable pour un métabolite si ce coefficient est inférieur à 20 % entre les QC (99,100). Ce seuil est défini pour éliminer au maximum les fluctuations d'origine analytique. Les données subissent enfin une transformation logarithmique et une normalisation de type « cyclic LOESS » (178).

L'évaluation de l'efficacité du pré-traitement des données est réalisée par détermination du coefficient de corrélation intra-classe (ICC) calculé à partir des données de tous les QC (non dilués et dilués). Le type de modèle utilisé est le modèle « two way random effects, absolute agreement, single rater/measurement » décrit par Koo *et al.* (179). Un ICC inférieur à 0,5, entre 0,5 et 0,75, entre 0,75 et 0,9, ou supérieur à 0,90 indique respectivement une corrélation faible, modérée, bonne, ou excellente. Enfin, une analyse en composante principale (ACP) est réalisée. La carte factorielle du poids des individus issue de l'ACP, encore appelée « score plot », permet une représentation visuelle de ces données, montrant les variations entre les échantillons selon les nouvelles composantes.

Les données issues des analyses en ESI+ et en ESI- pré-traitées indépendamment sont finalement associées pour réaliser les analyses statistiques sur une seule matrice de données.

1.3.3. Statistiques

La matrice de données est traitée par analyse multivariée supervisée PLS-DA, qui a pour objectif de relier un bloc de variables à expliquer Y (le type d'exposition) à un bloc de variables explicatives X (les abondances des métabolites). La première étape de la construction du modèle est de définir le nombre de composantes orthogonales du modèle par validation croisée (cross-validation). La qualité des modèles ainsi construits peut être évaluée par le Q²_Y (cum, capacité à prédire la variable Y) et R²_Y (validation croisée, pourcentage de la variabilité qui peut être expliquée par la variable Y). Le R²_X correspond quant à lui à la proportion de la variabilité représentée par le modèle. Le nombre d'échantillons analysés étant nettement inférieur au nombre de variables exploitées, il est important de vérifier que la discrimination observée n'est

pas liée à des variables non significatives. Des tests de permutation (nombre de permutations = 1000) ont alors été effectués pour évaluer la significativité des critères diagnostiques Q²Y et R²Y. On obtient ainsi des p-values, pR²Y et pQ², qui sont des indicateurs de performance du modèle, dont les seuils sont fixés à 0,05. En cas de validité du modèle, les variables issues de la PLS-DA sont classées par ordre croissant selon le critère VIP (Variable Importance in the Prediction) qui prend en compte l'importance de chaque variable dans la création des composantes et dans la prédiction des groupes. Les métabolites présentant les meilleurs scores VIP sont ensuite sélectionnés pour subir une analyse multivariée non supervisée de heatmap. Il s'agit d'une technique de visualisation graphique de données statistiques qui fait correspondre à l'intensité d'une grandeur variable un nuancier de couleurs sur une matrice à deux dimensions.

Les résultats obtenus par analyses multivariées sont confirmés par des analyses univariées de la matrice de données. L'ANOVA permet de comparer les variances des échantillons appartenant aux 4 groupes d'exposition. Pour chaque métabolite, ce test permet de comparer les résultats pour les différents temps d'exposition et ainsi d'évaluer les différences intra-groupes. Le test est significatif si au moins un des temps d'exposition est différent, au seuil de significativité de 0,05 après correction de la p-value par le FDR. Pour les métabolites montrant une différence significative, un test de Student est réalisé pour comparer les effets aux différents temps d'exposition : T0 vs T1 et T0 vs T2. Le seuil de significativité est fixé à 0,05 après correction de la p-value par le FDR. Les métabolites discriminants pourront alors être identifiés.

1.3.4. Sens biologique des données acquises

1.3.4.1. *Identification*

Les paramètres d'identification d'un composé sont sa masse exacte, son Tr, sa similarité isotopique, sa fragmentation et son CCS acquis expérimentalement. L'étape d'identification des composés discriminants est réalisée en confrontant les données acquises à celles de la librairie interne de référence (établie à partir de 380 authentiques standards) avec le logiciel Unifi (Waters®), ou à celles de bases de données avec le logiciel Progenesis (Waters®). Dans ce deuxième cas, chaque composé peut être associé à une ou plusieurs identifications. Chaque identification est

caractérisée par un score de correspondance. Chacun des 5 paramètres d'identification contribue à 20 % du score total. Les identifications présentant le plus haut score de correspondance ont été conservées. A chaque identification d'un métabolite est associé un identifiant HMDB, ainsi qu'un SMILES.

Les bases de données utilisées, par ordre de confiance d'annotation, étaient 1) librairie interne de référence 2) base de données large prédite 3) Metabolic profiling CCS library (librairie fournie par Waters®). Les critères d'identification pour chaque base de données sont détaillés dans le Tableau 14. Dans le deuxième cas impliquant la base de données prédite, la méthode de recherche de fragmentation a été réalisée à partir d'une fragmentation théorique par le logiciel Progenesis (Waters®). Dans quelques cas, les spectres ont été confrontés manuellement à la base de données spectrales MassBank (180) pour élargir les possibilités d'identification.

Tableau 14 : Critère d'identification pour chaque base de données

Correspondance avec la base de données	Niveau d'identification MSI	Seuils d'acceptation		
		$\Delta m/z$	ΔCCS	ΔTr
Librairie interne de référence	1	3 mDa	10 %	2 min
Base de données large prédite	2 ou 3	5 mDa	16 A ²	1.1 min
Metabolic profiling CCS library	2 ou 3	5 mDa	16 A ²	-
Aucune correspondance	4	-	-	-

L'identification attribuée à chaque composé identifié, son numéro HMDB et son SMILES sont exportés dans un tableau d'identification.

1.3.4.2. *Interprétation biologique*

A chaque composé identifié est associé un SMILES, permettant de leur attribuer une classe chimique grâce à l'outil Classyfire (181). Une analyse des voies métaboliques associées à chaque identification a été réalisée grâce au logiciel MetaboAnalyst v5.0, par analyse de surreprésentation. Les identifiants HMDB des métabolites d'intérêt ont été importés pour assurer la standardisation des identifications. Les identifications sont confrontées à la librairie des voies *Homo sapiens* de la base de données KEGG. Les paramètres d'analyse utilisés sont le test hypergéométrique pour l'analyse de la surreprésentation et centralité relative entre les voies pour l'analyse de la topologie des voies (qui calcule l'impact des voies). Le seuil de significativité est fixé à une p-value < 0,05 associée à un FDR < 0,25.

2. Résultats

2.1. Mesure de la cytotoxicité

Les résultats de mesure de la cytotoxicité par le test LDH pour chaque dispositif et chaque temps d'exposition sont présentés dans le Tableau 15.

Tableau 15 : Cytotoxicité (%) en fonction du dispositif testé et de la durée d'exposition

	MB-18W			MB-30W			HTP			3R4F		
	T0	T1	T2	T0	T1	T2	T0	T1	T2	T0	T1	T2
Moyenne (%)	-0,8	0,2	-0,2	0,1	0,8	0,2	0,4	7,1	11,6	0,2	0,9	2,5
Médiane (%)	-0,8	-0,4	-0,7	0,2	1,1	0,3	0,3	8,2	13,2	0,1	0,9	1,7
Ecart-type (%)	0,1	1,7	1,3	0,5	1,2	0,7	0,7	4,9	7,5	0,4	0,4	2,5

Ces résultats confirment une très faible cytotoxicité pour les dispositifs MB-18W, MB-30W et 3R4F dans les conditions expérimentales utilisées. Pour l'HTP, la cytotoxicité mesurée atteint une valeur moyenne de 11,6 % ($\pm 7,5$ %) après 60 min d'exposition (T2). Ces résultats sont en faveur d'une cytotoxicité inférieure à 20 % pour toutes les conditions testées, permettant de confirmer que les durées d'exposition choisies sont sub-toxiques.

Le test de Kruskal-Wallis, utilisé pour évaluer s'il existe des différences entre les groupes 3R4F, HTP, MB-18W et MB-30W, met en évidence une différence significative du groupe HTP par rapport aux autres groupes (p -value = 0,03). Les autres groupes présentent une p -value non significative (3R4F p -value = 0,08 ; MB-18W p -value = 0,43 ; MB-30W p -value = 0,58). Le test de Wilcoxon réalisé sur les résultats obtenus pour le groupe HTP met en évidence une différence significative entre les échantillons du groupe contrôle vs T2 (p -value = 0,02). En revanche, il n'existe pas de différence significative entre les échantillons du groupe contrôle vs T1 (p -value = 0,057) et entre les groupes T1 vs T2 (p -value = 0,48).

En conclusion, les conditions testées sont bien sub-toxiques, et l'exposition aux aérosols d'HTP pour une durée de 60 min entraîne une cytotoxicité plus importante qu'en l'absence d'exposition. Se placer à des cytotoxicités basses permet de s'assurer que les différences observées dans les abondances de métabolites seront bien le reflet des variabilités biologiques dues aux expositions, et non pas à des différences de quantités de cellules vivantes au départ.

2.2. Métabolomique

2.2.1. Pré-traitement des données de métabolomique

L'extraction des données est réalisée avec le logiciel Progenesis (Waters®) selon les étapes successives d'alignement, de « peak picking » et de déconvolution. Elle génère une matrice de données composée de 5130 composés en mode ESI+ et 5037 composés en mode ESI-. Les deux matrices extraites sont pré-traitées selon les étapes successives d'imputation des données manquantes, de filtration, de transformation et de normalisation comme décrites précédemment. Une comparaison des données avant et après pré-traitement permet d'évaluer la qualité de ces étapes.

2.2.1.1. *Pré-traitement des données en mode d'ionisation positif*

Le pré-traitement des données en mode d'ionisation positif a permis de générer une matrice de données finale de 2398 composés.

La corrélation intra-classe (ICC) calculée à partir des données de tous les QC est évaluée à 0,778 [0,771 ; 0,785] avant pré-traitement, et atteint 0,848 [0,840 ; 0,856] après pré-traitement. Cette ICC élevée (plus proche de 1) indique une grande similarité entre les valeurs de tous les QC après pré-traitement. Puisque les données des QC résultent de l'injection d'un même composé à plusieurs reprises, ces résultats sont en faveur d'une amélioration des données après pré-traitement.

L'abondance totale des composés pour chaque échantillon est comparée entre les données avant et après pré-traitement. Cette comparaison est illustrée par la Figure 31. Cette figure permet de voir que les abondances totales des échantillons de QC dilués (colorés en bleu clair) sont largement inférieures aux autres échantillons. Ce résultat est attendu, reflétant la dilution de ces échantillons. En revanche de manière inattendue, deux échantillons nommés E16 (groupe MB-30W, exposition 30 min) et E19 (groupe MB-30W exposition 60 min) présentent également des abondances totales largement inférieures aux autres échantillons. Ces anomalies sont corrigées par le pré-traitement des données.

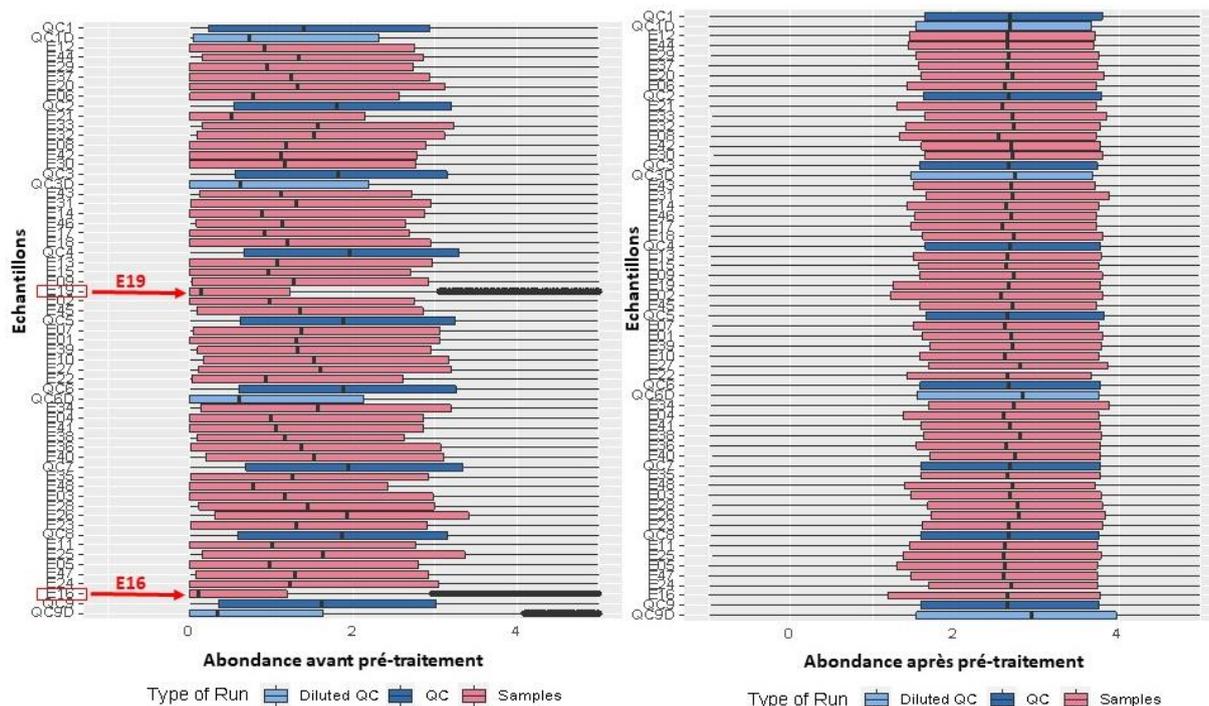


Figure 31 : Abondance des composés pour chaque échantillon pour les données acquises en mode d’ionisation positif avant vs après pré-traitement des données

Une attention particulière est apportée à l’effet du pré-traitement des données sur l’abondance des deux étalons internes ajoutés lors de la préparation des échantillons et détectables en ESI+. La comparaison entre les données avant et après pré-traitement est illustrée par la Figure 32. Cette figure permet de mettre en évidence une diminution du signal au cours de la séquence avant pré-traitement des données, phénomène bien connu en métabolomique. De plus, les abondances de ces étalons internes pour les échantillons E16 et E19 sont fortement diminuées avant pré-traitement. Cette observation, en plus de celle réalisée avec la Figure 31, est en faveur d’un problème technique lors de l’injection de ces deux échantillons, probablement à l’origine d’un volume d’injection inférieur aux autres échantillons. Ce problème technique est corrigé pour les étalons internes par le pré-traitement des données.

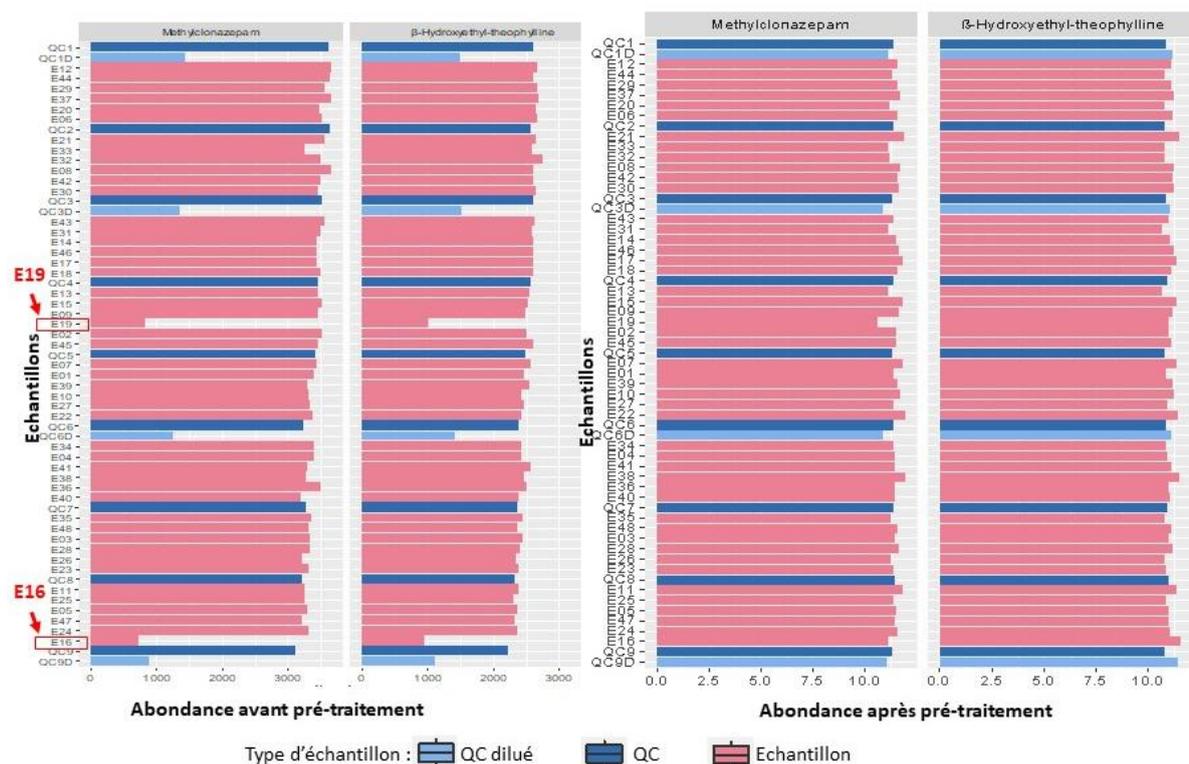


Figure 32 : Abondance des deux étalons internes pour les données acquises en mode d'ionisation positif avant vs après pré-traitement des données

Les données analytiques obtenues avec l'ensemble des échantillons sont également sujettes à des analyses multivariées d'ACP (Figure 33). Ici après pré-traitement, les données des échantillons sont plutôt réparties aléatoirement, signifiant que la variabilité de notre jeu de données biologiques est très importante. En revanche, les QC sont regroupés ensemble dans l'espace des composantes principales, ce qui signifie que les différences relatives entre les injections répétées, représentant les variations du système analytique, sont plus petites que les différences biologiques entre les échantillons. Au total, ces résultats indiquent que la variabilité analytique est faible et confortent la robustesse de la méthode. Avec cette visualisation en ACP, les échantillons E16 et E19 restent éloignés des autres échantillons après pré-traitement. Il est donc décidé de les considérer comme des « outliers » et de les retirer pour la suite des analyses de données.

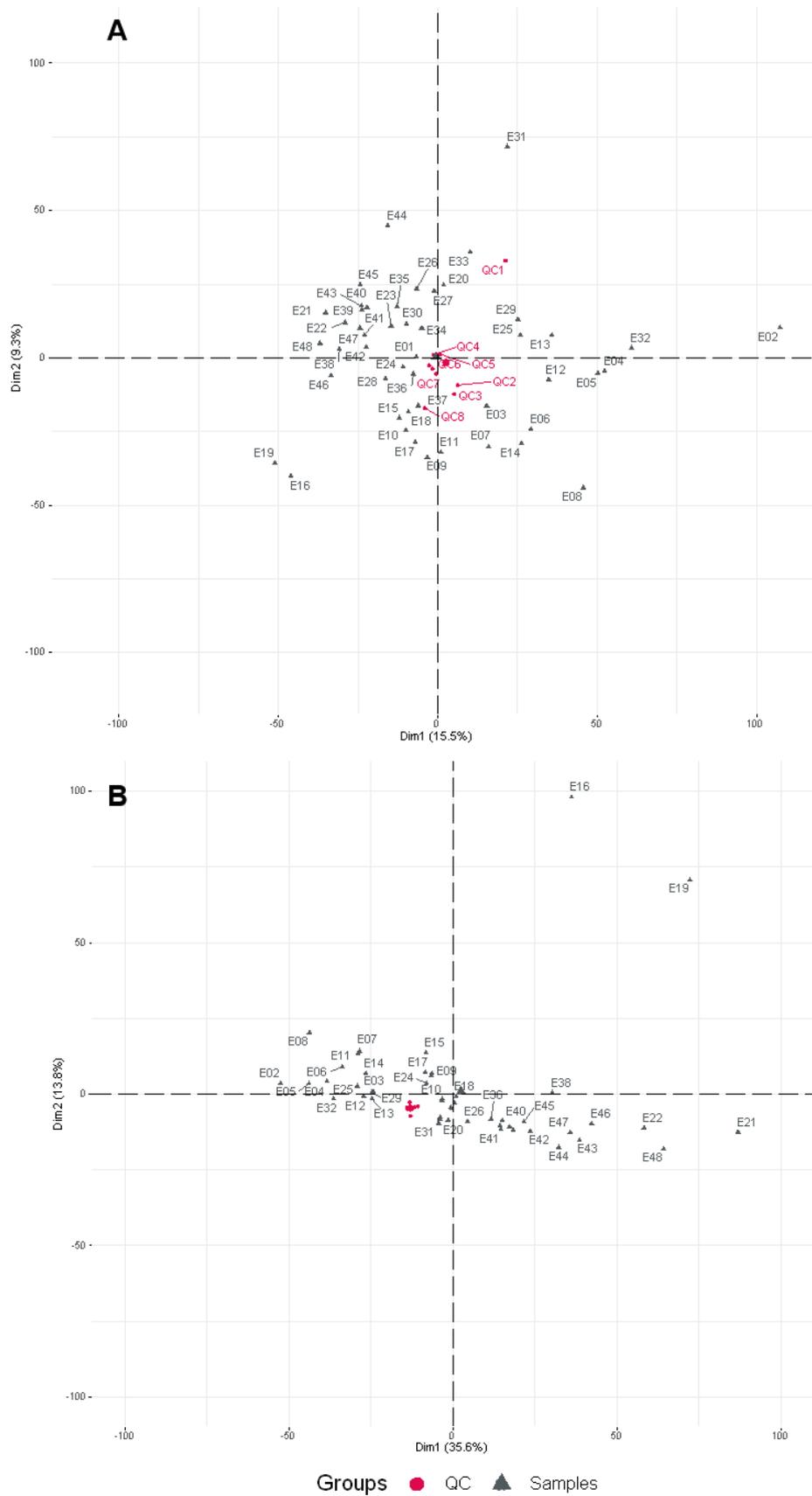


Figure 33 : Score plots des analyses ACP pour les données acquises en mode d'ionisation positif Echantillons (en gris) vs QC (en rouge) ; (A) avant pré-traitement ; (B) après pré-traitement

2.2.1.2. Pré-traitement des données en mode d'ionisation négatif

Le pré-traitement des données en mode d'ionisation négatif a permis de générer une matrice de données finale de 1193 composés.

L'ICC calculée à partir des données de tous les QC est évaluée à 0,716 [0,708 ; 0,725] avant pré-traitement, et atteint 0,851 [0,840 ; 0,861] après pré-traitement. Cette ICC élevée (plus proche de 1) indique une grande similarité entre les valeurs de tous les QC après pré-traitement. Ces résultats sont en faveur d'une amélioration des données après pré-traitement.

L'abondance totale des composés pour chaque échantillon est comparée entre les données avant et après pré-traitement. Cette comparaison est illustrée par la Figure 34. Avant pré-traitement, les abondances sont relativement variables. Ceci est corrigé par le pré-traitement des données.

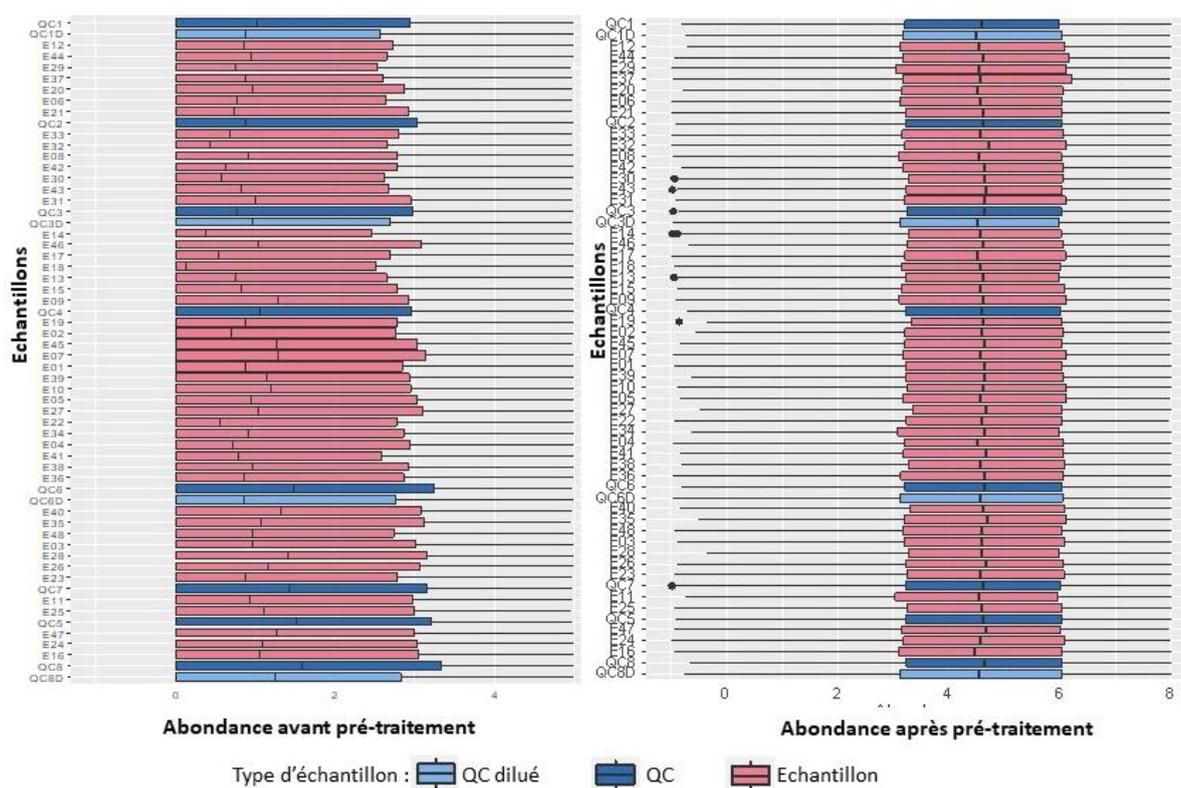


Figure 34 : Abondance des composés pour chaque échantillon pour les données acquises en mode d'ionisation négatif avant vs après pré-traitement

Le suivi de l'étalon interne utilisé pour le mode ESI- permet de mettre en évidence une diminution de l'abondance pour les 4 premiers échantillons injectés (E12, E44, E29, E37) (Figure 35). Ce résultat était attendu puisque le volume d'injection programmé pour ces échantillons était de 5 μ L au lieu de 10 μ L. Ce problème technique est néanmoins corrigé par le pré-traitement des données.

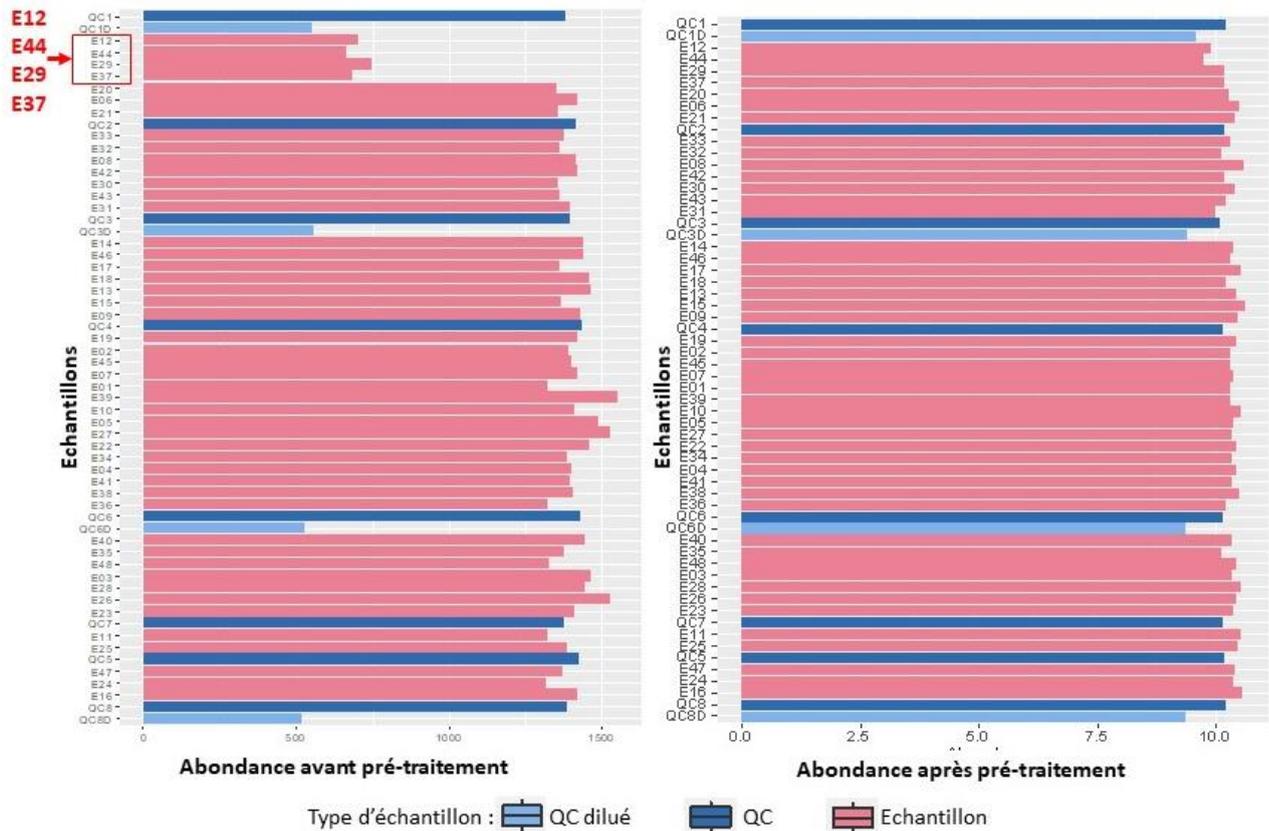


Figure 35 : Abondance de l'étalon interne pour les données acquises en mode d'ionisation négatif avant vs après pré-traitement

Comme pour les résultats en ESI+, l'ACP pour les données en mode d'ionisation négatif (Figure 36) montre la variabilité importante de notre jeu de données biologiques (échantillons dispersés) et une variabilité analytique faible (QC rassemblés). Ces données confortent la robustesse de la méthode.

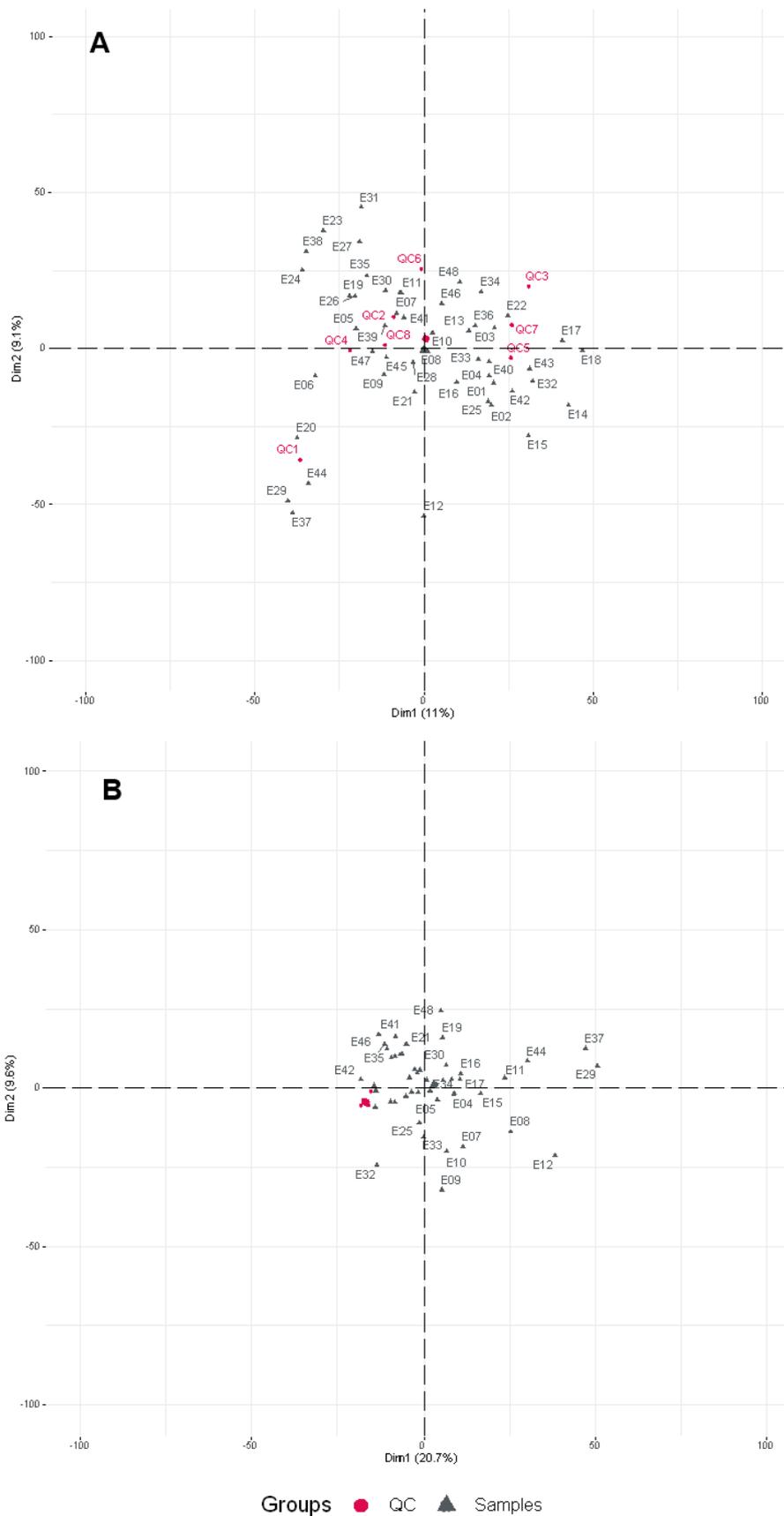
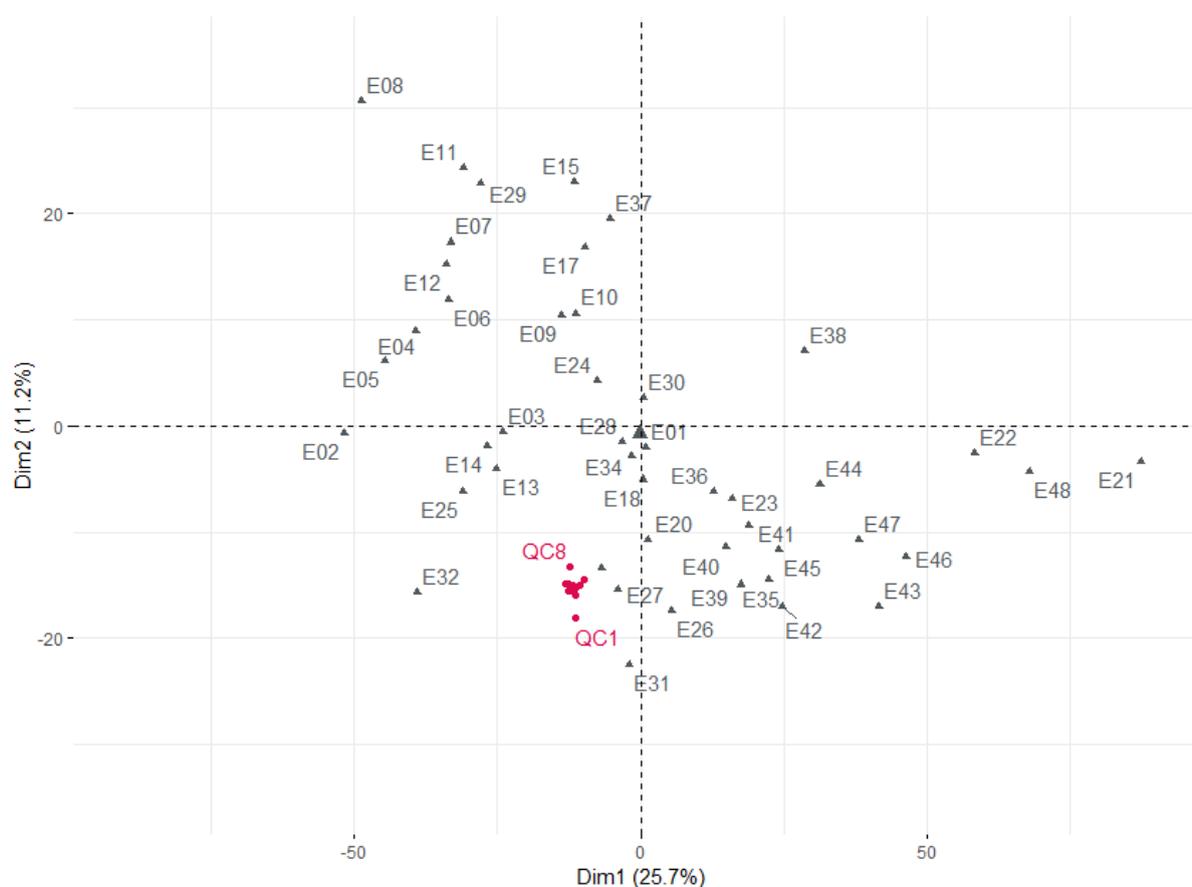


Figure 36 : Score plots des analyses ACP pour les données acquises en mode d'ionisation négatif Echantillons (en gris) vs QC (en rouge) ; (A) avant pré-traitement ; (B) après pré-traitement

2.2.1.3. Association des données des deux modes d'ionisation

Après association des données normalisées issues de l'acquisition en ESI+ et ESI-, la matrice de données finale est constituée de 46 échantillons (les échantillons E16 et E19 considérés comme « outliers » ont été retirés du jeu de données) et 3591 composés.

L'ACP de l'ensemble des données pré-traitée (Figure 37) montre la variabilité importante de notre jeu de données biologiques (échantillons dispersés), ainsi qu'une variabilité analytique faible (QC rassemblés).



**Figure 37 : Score plot de l'analyse ACP associant les données issues de l'acquisition en mode d'ionisation positif et négatif
Echantillons (en gris) vs QC (en rouge) après pré-traitement**

2.2.1.4. Statistiques

La PLS-DA est ici utilisée pour discriminer les groupes exposés, c'est-à-dire HTP (T1+T2 soit 8 échantillons), 3R4F (T1+T2 soit 8 échantillons), MB-18W (T1+T2 soit 8 échantillons) et MW-30W (T1+T2 soit 6 échantillons). Le modèle optimal est construit sur 7 composantes. La Figure 38 correspondant à l'ensemble des échantillons exposés montre qu'une proportion importante de la variabilité ($R^2X = 0,659$) est représentée par la PLS-DA, et a permis une bonne discrimination des échantillons en fonction de leur exposition. Le pourcentage de variabilité (R^2Y) expliqué par le modèle est de 96,1 %. Le pouvoir prédictif (Q^2Y) du modèle est de 73,8 %. Ces résultats montrent que le modèle permet bien de prédire l'appartenance d'un échantillon exposé à un groupe donné.

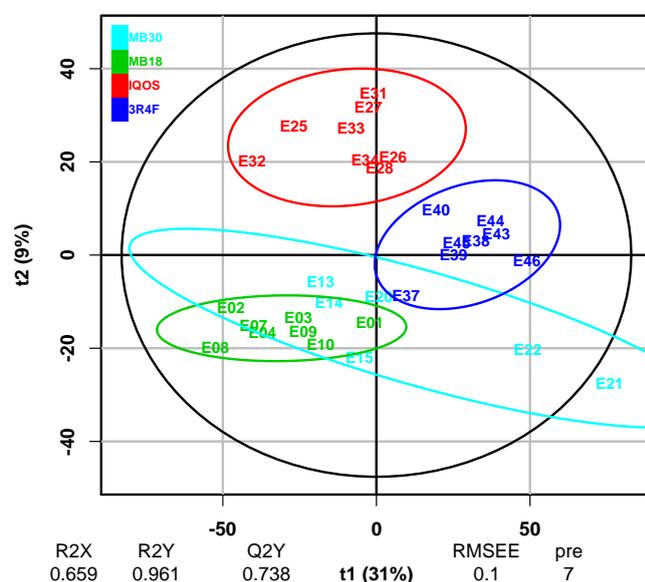


Figure 38 : Score plot de l'analyse PLS-DA pour les groupes d'échantillons exposés (T1 et T2) des dispositifs MB-18W, MB-30W, IQOS et 3R4F

Le test du modèle par méthode de permutation permet d'obtenir les performances $pQ^2 = 0,001$ et $pR^2 = 0,001$, qui valident le modèle de discrimination des groupes d'exposition. Les variables pour lesquelles le critère VIP était supérieur à 1,5, correspondant à 180 variables, ont été sélectionnées pour construire le Heatmap (Figure 39). Parmi elles, 27 variables ont un score VIP compris entre 1,8 et 2. Quatre variables ont un score VIP supérieur à 2. 1174 autres variables ont un score VIP compris entre 1 et 1,5. Les lignes représentent les échantillons tandis que les colonnes représentent les métabolites. Les métabolites augmentant sont colorés en rouge, tandis que ceux diminuant sont colorés en bleu. L'intensité de chaque couleur correspond à l'amplitude de la différence par rapport à la valeur moyenne.

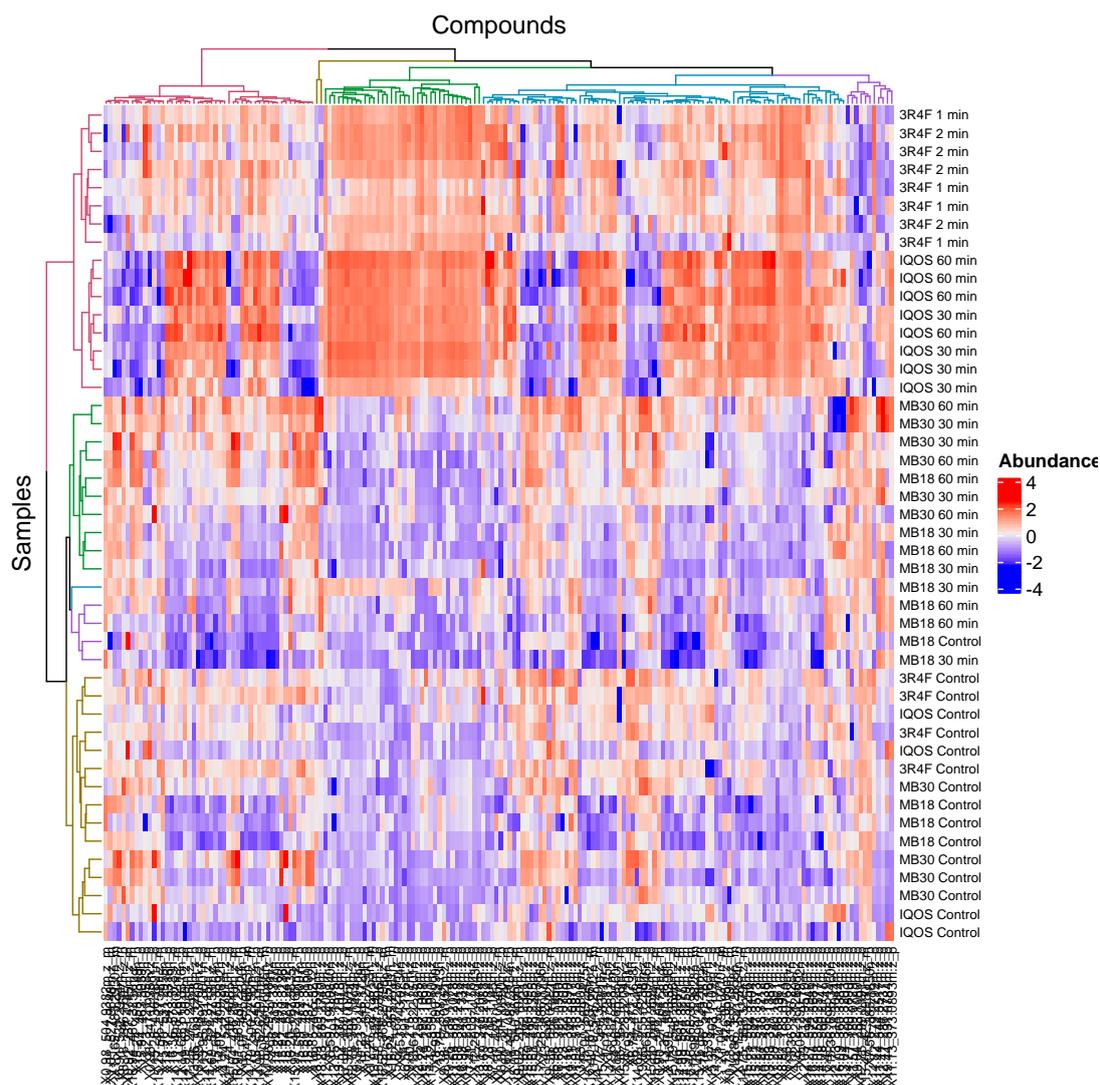


Figure 39 : Analyse heatmap des métabolites présentant un score VIP > 1,5

On peut ici observer une classification des échantillons en fonction de deux bras. Le premier réunit les échantillons exposés aux produits du tabac. Parmi eux, ceux exposés à l'HTP sont séparés de ceux exposés à la 3R4F. Le deuxième réunit les échantillons exposés à l'e-cig ou non exposés. Parmi eux, ceux exposés à l'e-cig sont assez bien discriminés de ceux non exposés. En revanche, les échantillons exposés à l'e-cig 18W ou 30W ne sont pas séparés.

Les résultats des analyses univariées d'ANOVA pour une p-value corrigée < 0,05 ont permis de mettre en évidence 214 composés significativement modulés, participant à la discrimination intra-groupe. Leur répartition entre les 4 groupes d'exposition est détaillée Figure 40 sous forme de diagramme de Venn. Il indique que 84 % des composés dérégulés à la suite de l'exposition aux fumées de la 3R4F le sont aussi avec les aérosols de tabac chauffé. Cela correspond à 43 composés en commun.

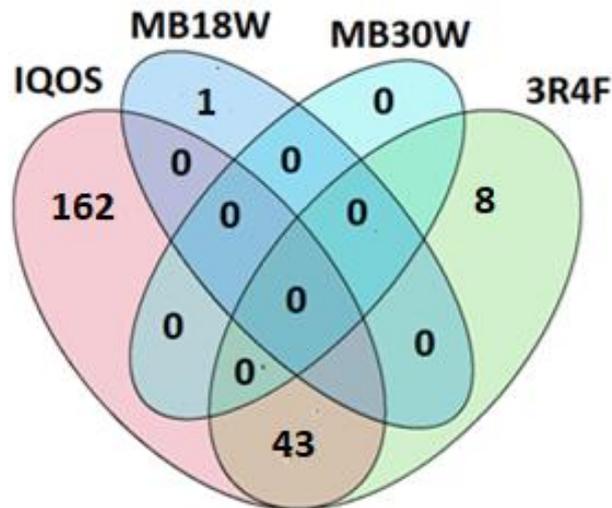


Figure 40 : Répartition des métabolites significatifs au test ANOVA parmi les différents types d'exposition

Ces résultats sont comparés à ceux obtenus en PLS-DA (Tableau 16). Les composés communs entre les deux tests soulignent une variabilité inter- et intra-groupes. Les composés significatifs en ANOVA, mais présentant un score VIP < 1,5 en PLS-DA, indiquent une variabilité intra-groupes, sans variabilité inter-groupes. Finalement, les composés considérés d'intérêt en PLS-DA, mais non significatifs en ANOVA, soulignent une variabilité inter-groupes, sans variabilité intra-groupes.

Tableau 16 : Principe des analyses de PLS-DA et ANOVA et comparaison de leurs résultats

	PLS-DA	ANOVA
Type d'analyse	Analyse multivariée supervisée	Analyse univariée
Objectif de l'analyse	Discrimination des groupes exposés : révélation des variables les plus discriminantes	Analyse de variance : pour chaque exposition, teste la différence entre les temps d'exposition pour une variable
Conception de l'analyse (groupes)	<ul style="list-style-type: none"> (T1+T2)_{HTP} vs (T1+T2)_{3R4F} vs (T1+T2)_{MB-18W} vs (T1+T2)_{MB-30W} 	<ul style="list-style-type: none"> T0_{HTP} vs T1_{HTP} vs T2_{HTP} T0_{3R4F} vs T1_{3R4F} vs T2_{3R4F} T0_{MB-30W} vs T1_{MB-30W} vs T2_{MB-30W} T0_{MB-18W} vs T1_{MB-18W} vs T2_{MB-18W}
Variables de l'analyse	Abondance de tous les métabolites dans une même analyse	Abondance d'un métabolite par analyse. Multi-test avec ajustement FDR des p-values
Critère d'interprétation	R2, Q2, pR2Y, pQ2 : qualité du modèle Score VIP : liste des variables les plus discriminantes des groupes	p-value ajustée : significativité de la différence entre les temps d'exposition pour une variable
Nombre de métabolites révélés	180 (Score VIP ≥ 1,5)	214 (p-value ajustée ≤ 0.05)
Nombre de métabolites communs aux 2 analyses	95	
Nombre de métabolites non communs aux 2 analyses	85	119

Pour les variables d'intérêt (résultats significatifs) de l'ANOVA, des t-tests sont réalisés. Pour le dispositif HTP, 198 composés présentent une différence significative entre T0 (contrôle non-exposé) et T1 (30 min d'exposition). 204 métabolites présentent une différence significative entre T0 et T2 (60 min d'exposition), dont 197 communs entre les deux durées d'exposition. Pour le dispositif 3R4F, 46 métabolites présentent une différence significative entre T0 et T1 (1 min d'exposition). 51 métabolites présentent une différence significative entre T0 et T2 (2 min d'exposition), dont 46 communs entre les deux durées d'exposition.

2.2.2. Sens biologique des données acquises

2.2.2.1. Identification

Sur les 214 composés discriminants, une identification a pu être attribuée pour 84 composés, les autres restant non identifiés. Le Tableau 17 liste les métabolites identifiés pour chaque mode d'ionisation et détaille le fold-change (exprimé en $\log_2(\text{FC})$) entre T0 et T1 (FC1) ou T0 et T2 (FC2). Un dégradé de couleur permet d'évaluer l'intensité des variations (du bleu au rouge). Le niveau de confiance des identifications de chaque métabolite (colonne « ID ») est chiffré selon le référentiel MSI décrit en introduction (110). En complément, les données de p-value de l'ANOVA, de FC et la score VIP de l'ensemble des composés discriminants (identifiés ou non) sont présentées en annexe.

Certains composés ont été identifiés comme des composés exogènes. La nicotine est significativement augmentée après exposition aux aérosols d'HTP (FC1 = 16,6 ; FC2 = 17,3) ou à la fumée de 3R4F (FC1 = 12,0 ; FC2 = 14,3). A noter que pour ce composé, la p-value de l'ANOVA après exposition aux aérosols d'e-cig 30 W était proche de la significativité, avec une p-value à 0,08 (Figure 41).

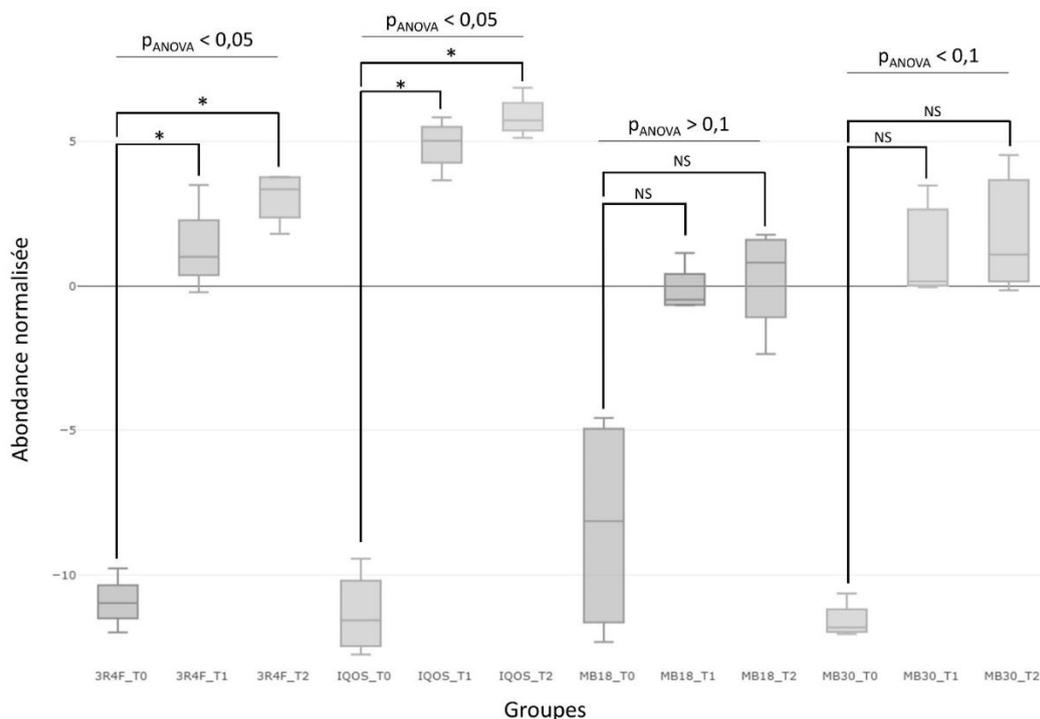


Figure 41 : Box plot de la nicotine pour les différents temps et dispositifs d'exposition
 p_{ANOVA} (corrige par le FDR) ; Test de Student : * $p < 0,05$; NS : Non significatif

Les niveaux de la plupart des autres composés exogènes augmentent par rapport aux contrôles non exposés, et d'autant plus lorsque le temps d'exposition est allongé. C'est le cas du 1-Naphthylamine après exposition aux aérosols d'HTP (FC1 = 15,4 ; FC2 = 15,6) ou à la fumée de 3R4F (FC1 = 10,0 ; FC2 = 12,0), du 3-Méthylindole après exposition aux aérosols d'HTP (FC1 = 4,8 ; FC2 = 6,6), et de la scopoletine après exposition aux aérosols d'HTP (FC1 = 13,1 ; FC2 = 13,5) ou à la fumée de 3R4F (FC1 = 7,0 ; FC2 = 9,5). En ce qui concerne la famille des β -carbolines, la β -carboline était augmentée après exposition aux aérosols d'HTP (FC1 = 9,5 ; FC2 = 10,3) ou à la fumée de 3R4F (FC1 = 7,6 ; FC2 = 10,3), ainsi que l'harman après exposition aux aérosols d'HTP (FC1 = 6,7 ; FC2 = 7,0) ou à la fumée de 3R4F (FC1 = 6,4 ; FC2 = 7,5). Un composé de la classe des pyrroline, le rollipyrrole, est également augmenté après exposition aux aérosols d'HTP (FC1 = 11,8 ; FC2 = 10,3) ou à la fumée de 3R4F (FC1 = 10,9 ; FC2 = 11,5). Un seul composé exogène diminue par rapport au contrôle non exposé, la theaspirane, après exposition aux aérosols d'HTP (FC1 = -2,4 ; FC2 = -3,3) ou à la fumée de 3R4F (FC1 = -2,0 ; FC2 = -2,6).

Les autres composés sont attribués au métabolisme cellulaire.

Tableau 17 : Métabolites identifiés et FC associés pour chaque exposition

NS : Non significatif

ESI	Pic	ID	Nom du composé	IQOS		3R4F		MB-18W		MB-30W	
				Log2(FC) T0 vs T1	Log2(FC) T0 vs T2						
NEG	0.93_565.0441m/z	2	Uridine diphosphate glucose	-1,4	-1,8	NS	NS	NS	NS	NS	NS
NEG	0.97_607.0776n	2	Uridine diphosphate -N-acetylglucosamine	-1,1	-1,2	NS	NS	NS	NS	NS	NS
NEG	0.97_628.0517m/z	2	Uridine diphosphate -N-acetylglucosamine	-2,3	-2,0	NS	NS	NS	NS	NS	NS
NEG	1.00_482.9586m/z	1	Uridine triphosphate	-0,8	-1,2	NS	NS	NS	NS	NS	NS
NEG	1.00_506.9926n	2	dGTP	-0,6	-1,0	NS	NS	NS	NS	NS	NS
NEG	1.04_191.0547m/z	1	Quinic acid	2,8	3,1	NS	NS	NS	NS	NS	NS
NEG	1.08_427.0267n	1	Adenosine 3',5'-diphosphate	1,2	1,5	NS	NS	NS	NS	NS	NS
NEG	1.08_604.0656m/z	1	Guanosine diphosphate mannose	-1,3	-1,5	NS	NS	NS	NS	NS	NS
NEG	1.09_429.0553m/z	2	CMP-2-aminoethylphosphonate	-0,7	-1,3	NS	NS	NS	NS	NS	NS
NEG	1.59_742.0631m/z	1	NADP	-0,4	-0,7	NS	NS	NS	NS	NS	NS
NEG	1.60_347.0374m/z	1	Inosine 5'-monophosphate	3,7	3,7	NS	NS	NS	NS	NS	NS
NEG	1.64_148.0420m/z	1	Methionine	-1,2	-1,3	NS	NS	NS	NS	NS	NS
NEG	2.35_321.0676n	2	Beta-Citryl-L-glutamic acid	-1,0	-1,3	NS	NS	NS	NS	NS	NS
NEG	2.96_612.1481n	2	Oxidized glutathione	NS							
NEG	14.11_498.2602m/z	2	LysoPE(20:5)	2,4	2,7	NS	NS	NS	NS	NS	NS
NEG	14.98_526.2911m/z	2	LysoPE(22:5)	2,5	2,9	NS	NS	NS	NS	NS	NS
NEG	15.47_506.3213m/z	2	LysoPE(20:1)	0,9	1,4	NS	NS	NS	NS	NS	NS
NEG	15.69_506.3213m/z	2	LysoPE(20:1)	1,0	1,9	NS	NS	NS	NS	NS	NS

ESI	Pic	ID	Nom du composé	IQOS		3R4F		MB-18W		MB-30W	
				Log2(FC) T0 vs T1	Log2(FC) T0 vs T2						
POS	0.96_404.0019n	2	Uridine 5'-diphosphate	-1,4	-1,3	NS	NS	NS	NS	NS	NS
POS	0.98_489.1138m/z	1	Cytidine 5' diphosphocholine	5,6	7,0	NS	NS	NS	NS	NS	NS
POS	0.99_506.9954n	2	dGTP	-0,7	-1,1	NS	NS	NS	NS	NS	NS
POS	1.07_192.0265n	1	Isocitric Acid	-1,0	-1,4	NS	NS	NS	NS	NS	NS
POS	1.07_321.0689n	2	Beta-Citryl-L-glutamic acid	-1,3	-1,5	NS	NS	NS	NS	NS	NS
POS	1.07_612.1507n	2	Oxidized glutathione	NS	NS	NS	1,4	NS	NS	NS	NS
POS	1.11_250.0931m/z	2	Deoxyguanosine	-1,4	-1,6	NS	NS	NS	NS	NS	NS
POS	2.01_227.0902n	1	Deoxycytidine	-0,5	-1,0	NS	NS	NS	NS	NS	NS
POS	2.29_321.0690n	2	Beta-Citryl-L-glutamic acid	-1,3	-1,6	NS	NS	NS	NS	NS	NS
POS	2.98_612.1497n	2	Oxidized glutathione	NS	0,9	NS	1,7	NS	NS	NS	NS
POS	3.63_132.0806m/z	2	3-Methylindole	4,8	6,6	NS	NS	NS	NS	NS	NS
POS	3.82_229.1781n	2	N1,N8-Diacetylspermidine	4,6	6,9	NS	2,8	NS	NS	NS	NS
POS	3.83_230.1859m/z	2	N1,N8-Diacetylspermidine	3,2	4,9	NS	NS	NS	NS	NS	NS
POS	4.22_221.1280m/z	2	3-Indolebutyric acid	8,7	9,7	NS	6,3	NS	NS	NS	NS
POS	4.53_143.0734n	2	1-Naphthylamine	15,4	15,6	NS	10,0	NS	NS	NS	NS
POS	5.62_163.1228m/z	1	Nicotine	16,6	17,3	NS	12,0	NS	NS	NS	NS
POS	6.54_191.1176m/z	2	N-Methylserotonin	10,6	11,3	NS	6,6	NS	NS	NS	NS
POS	6.59_190.0840n	2	3-Hydroxyisubutyric acid	6,7	6,8	NS	NS	NS	NS	NS	NS
POS	7.26_187.0629n	2	Indoleacrylic acid	-0,9	-0,9	NS	NS	NS	NS	NS	NS
POS	8.18_246.1697m/z	2	Isovalerylcarnitine	-1,0	-1,4	NS	NS	NS	NS	NS	NS
POS	8.53_169.0760m/z	2	Beta-carboline	9,5	10,3	NS	7,6	NS	NS	NS	NS
POS	8.82_183.0914m/z	2	Harman	6,7	7,0	NS	6,4	NS	NS	NS	NS
POS	9.00_215.1177m/z	2	Melatonin	11,1	12,4	NS	NS	NS	NS	NS	NS
POS	9.22_260.1851m/z	2	Hexanoylcarnitine	-1,1	-1,0	NS	NS	NS	NS	NS	NS
POS	9.34_193.0494m/z	2	Scopoleptin	13,1	13,5	NS	7,0	NS	NS	NS	NS
POS	11.43_236.2004m/z	2	Theaspirane	-2,4	-3,3	NS	-2,0	NS	NS	NS	NS
POS	12.54_262.1781n	2	Glycerol 5-hydroxydecanoate	NS	1,7	NS	NS	NS	NS	NS	NS
POS	12.78_391.1874m/z	3	Eicosanoid	11,3	11,8	NS	6,7	NS	NS	NS	NS
POS	12.85_253.1335m/z	2	Rollipyrrole	11,8	10,3	NS	10,9	NS	NS	NS	NS
POS	12.93_377.2081m/z	3	Eicosanoid	11,1	11,0	NS	10,1	NS	NS	NS	NS
POS	13.03_391.1889m/z	3	Eicosanoid	5,4	6,4	NS	2,1	NS	NS	NS	NS
POS	13.15_391.1875m/z	3	Eicosanoid	9,2	9,7	NS	NS	NS	NS	NS	NS
POS	13.17_377.2074m/z	3	Eicosanoid	12,2	11,0	NS	NS	NS	NS	NS	NS

ESI	Pic	ID	Nom du composé	IQOS		3R4F		MB-18W		MB-30W	
				Log2(FC) T0 vs T1	Log2(FC) T0 vs T2						
POS	13.36_336.2287n	3	Eicosanoid	9,6	10,5	6,4	8,3	NS	NS	NS	NS
POS	13.37_465.2850n	2	LysoPC(14:1)	2,2	2,8	NS	NS	NS	NS	NS	NS
POS	13.52_377.2078m/z	3	Eicosanoid	11,4	11,5	6,6	9,3	NS	NS	NS	NS
POS	13.61_583.2540m/z	2	Biliverdin	NS	NS	1,6	2,1	NS	NS	NS	NS
POS	13.86_467.3004n	2	LysoPC(14:0:0:0)	1,3	1,6	NS	NS	NS	NS	NS	NS
POS	13.97_359.1982m/z	3	Eicosanoid	9,4	9,2	3,2	5,1	NS	NS	NS	NS
POS	14.07_499.2692n	2	LysoPE(0:0:20:5)	2,4	3,1	NS	NS	NS	NS	NS	NS
POS	14.12_541.3164n	2	LysoPC(20:5)	5,6	7,2	NS	NS	NS	NS	NS	NS
POS	14.18_373.1770m/z	3	Eicosanoid	14,9	14,0	NS	NS	NS	NS	NS	NS
POS	14.24_493.3160n	2	LysoPC(16:1:0:0)	1,2	1,7	NS	NS	NS	NS	NS	NS
POS	14.27_518.3212m/z	2	LysoPC(16:0)	2,5	3,6	NS	NS	NS	NS	NS	NS
POS	14.42_359.1980m/z	3	Eicosanoid	17,1	16,1	10,5	11,5	NS	NS	NS	NS
POS	14.66_322.2497n	3	Eicosanoid	9,4	8,8	NS	NS	NS	NS	NS	NS
POS	14.68_519.3316n	2	LysoPC(18:2)	2,1	2,9	NS	NS	NS	NS	NS	NS
POS	14.72_543.3316n	2	LysoPC(20:4)	3,0	4,3	NS	NS	NS	NS	NS	NS
POS	14.83_508.3388m/z	2	LysoPC(17:1)	NS	2,6	NS	NS	NS	NS	NS	NS
POS	14.85_510.3983m/z	2	MG(10:0:0:0:0)	-3,3	-7,4	NS	NS	NS	NS	NS	NS
POS	14.96_519.3317n	2	LysoPC(18:2)	1,0	1,6	NS	NS	NS	NS	NS	NS
POS	15.00_569.3475n	2	LysoPC(22:5)	4,7	6,6	NS	NS	NS	NS	NS	NS
POS	15.07_495.3320n	2	LysoPC(16:0)	0,8	1,2	NS	NS	NS	NS	NS	NS
POS	15.21_584.3104m/z	2	LysoPC(20:3)	2,7	3,5	NS	NS	NS	NS	NS	NS
POS	15.22_547.3573m/z	2	1-Stearyl glycerophosphocholine	3,1	4,4	NS	NS	NS	NS	NS	NS
POS	15.24_453.2849n	2	LysoPE(16:0:0:0)	-1,8	-2,0	NS	NS	NS	NS	NS	NS
POS	15.27_318.2190n	3	Eicosanoid	11,2	11,2	4,6	7,3	NS	NS	NS	NS
POS	15.43_521.3476n	2	LysoPC(18:1)	0,9	1,3	NS	NS	NS	NS	NS	NS
POS	15.43_545.3462n	2	LysoPC(20:3)	2,0	2,7	NS	NS	NS	NS	NS	NS
POS	15.65_502.3256m/z	2	LysoPC(P-16:0)	4,0	4,5	NS	NS	NS	NS	NS	NS
POS	15.66_479.3361n	2	LysoPC(P-16:0)	3,1	3,7	NS	NS	NS	NS	NS	NS
POS	15.79_548.3696m/z	2	LysoPC(20:2)	2,5	3,2	NS	NS	NS	NS	NS	NS
POS	15.96_547.3629n	2	LysoPC(20:2)	2,6	3,5	NS	NS	NS	NS	NS	NS
POS	15.99_548.3700m/z	2	LysoPC(20:2)	1,8	2,7	NS	NS	NS	NS	NS	NS
POS	16.80_549.3785n	2	LysoPC(20:1)	1,5	2,2	NS	NS	NS	NS	NS	NS
POS	16.90_508.3750m/z	2	LysoPC(P-18:0)	5,4	6,0	NS	NS	NS	NS	NS	NS

2.2.2.2. Interprétation biologique

Parmi les 51 composés discriminants pour l'exposition à la 3R4F, 20 ont pu être identifiés. Sept composés sont classés comme exogènes, tandis que 8 composés sont des lipides appartenant à la classe des eicosanoïdes. Dans ce contexte, peu de composés peuvent faire l'objet d'une analyse de surreprésentation, c'est pourquoi l'interprétation biologique des résultats en réponse à l'exposition à la 3R4F sera réalisée plutôt métabolite par métabolite, en regard de la littérature existante.

Concernant les 205 composés discriminants pour l'exposition à l'HTP, la taxonomie chimique des métabolites identifiés et la proportion de chaque superclasse et classe est représentée en Figure 42. Parmi les composés identifiés, 38 appartiennent à la superclasse des lipides.

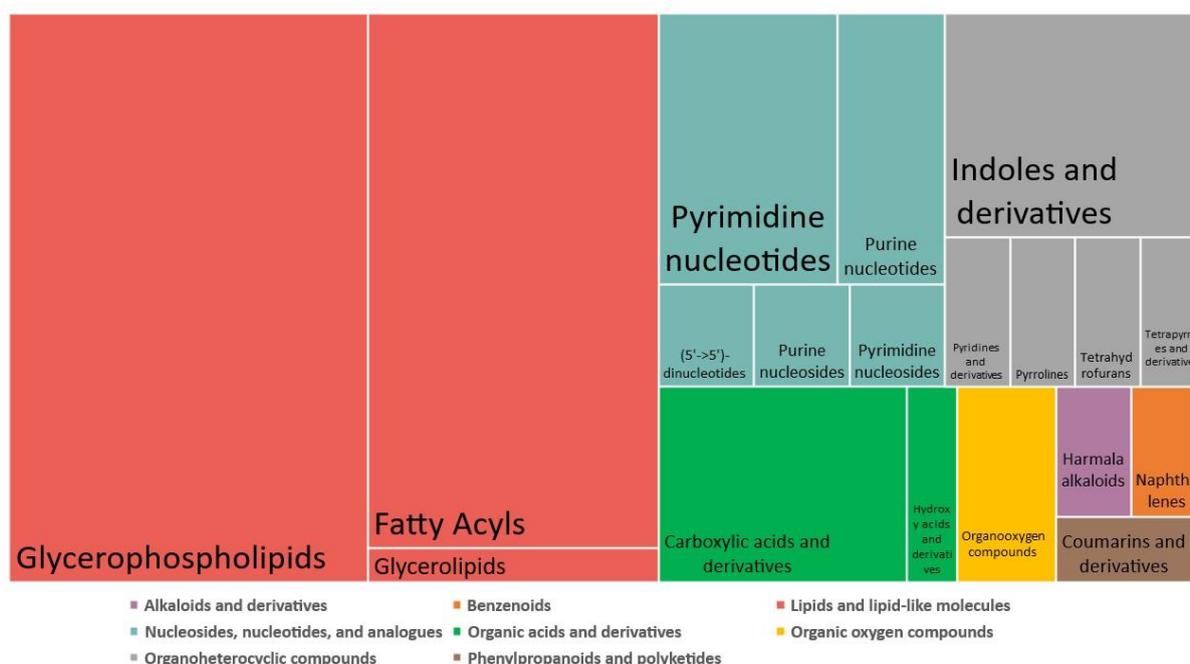


Figure 42 : Taxonomie chimique des métabolites identifiés pour l'exposition HTP et proportion de chaque superclasse et classe

Une analyse des voies métaboliques et fonctions biologiques dérégulées par les émissions d'HTP a été réalisée sur MetaboAnalyst v5.0. Les résultats représentés sous forme de diagramme de dispersion sont illustrés Figure 43. Cette représentation illustre les voies métaboliques identifiées par l'analyse de surreprésentation, projetées en fonction de la p-value et de la valeur d'impact de chacune de ces voies, évaluée par l'analyse topologique. Les couleurs plus foncées indiquent des changements plus significatifs des métabolites dans la voie correspondante.

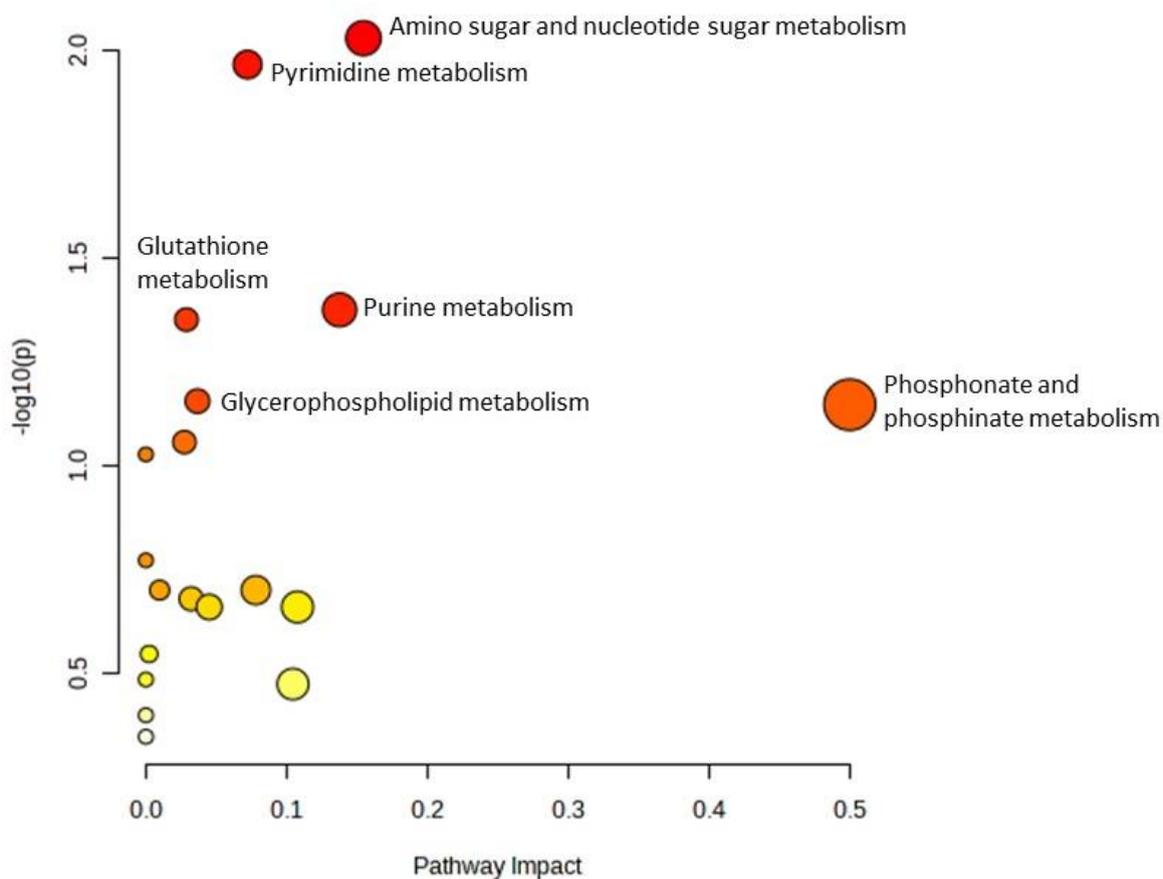


Figure 43 : Diagramme de dispersion des voies métaboliques générées sur le logiciel MetaboAnalyst pour l'exposition HTP

Les 10 premières voies métaboliques identifiées et les résultats associés sont listés dans le Tableau 18.

Tableau 18 : Top 10 des voies métaboliques de la base KEGG obtenues par l'analyse fonctionnelle (MetaboAnalyst) des composés identifiés

Voies métaboliques	Nombre total de métabolites	Hits	p-value	FDR	Impacte
Amino sugar and nucleotide sugar metabolism	37	3	0,009	0,45	0,15
Pyrimidine metabolism	39	3	0,01	0,45	0,07
Purine metabolism	65	3	0,04	0,93	0,13
Glutathione metabolism	28	2	0,04	0,93	0,03
Glycerophospholipid metabolism	36	2	0,07	0,98	0,04
Phosphonate and phosphinate metabolism	6	1	0,07	0,98	0,50
Tryptophan metabolism	41	2	0,08	0,98	0,03
Ascorbate and aldarate metabolism	8	1	0,09	0,98	0
Nicotinate and nicotinamide metabolism	15	1	0,19	1	0
Starch and sucrose metabolism	18	1	0,20	1	0,01

Aucun des voies métaboliques identifiées dans la base de données KEGG n'est significative (p -value < 0,05 et FDR < 0,25), mais ces résultats permettent d'identifier des voies métaboliques potentiellement dérégulées après exposition à l'HTP.

Certains métabolites identifiés participent au métabolisme des osamines et des nucléotides-oses. Une diminution de l'UDP-N-acetyl-alpha-D-glucosamine (FC1 = -2,3 ; FC2 = -2,0), de l'UDP-glucose (FC1 = -1,4 ; FC2 = -1,8) et du GDP-mannose (FC1 = -1,3 ; FC2 = -1,5) était observée. D'autres métabolites participent au métabolisme des nucléotides (purines et des pyrimidines). Après exposition aux aérosols d'HTP, le métabolisme des purines a été principalement affecté par une augmentation de l'inosine 5'-monophosphate (FC1 = 3,7 ; FC2 = 3,7) et une diminution de la déoxyguanosine (FC1 = -1,4 ; FC2 = -1,6), et du dGTP (FC1 = -0,7 ; FC2 = -1,1). En ce qui concerne le métabolisme des pyrimidines l'UDP (FC1 = -1,4 ; FC2 = -1,3), l'UTP (FC1 = -0,8 ; FC2 = -1,2) et la déoxycytidine (FC1 = -0,5 ; FC2 = -1,0) ont été identifiés comme des métabolites modulés à la baisse. L'exposition aux aérosols d'HTP impacte également des métabolites du métabolisme des glycérophospholipides avec une augmentation du LysoPC(16:0) (FC1 = 2,5 ; FC2 = 3,6), et de la CDP-choline (FC1 = 5,6 ; FC2 = 7,0).

Enfin, nous avons observé une modulation à la hausse d'un biomarqueur du stress oxydant, le glutathion oxydé, après 60 min d'exposition (FC2 = 0,9) et une tendance à la significativité après 30 min d'exposition, (p -value = 0,056 ; FC1 = 0,48). L'augmentation du glutathion oxydé est associée à une diminution du NADP (FC1 = -0,4 ; FC2 = -0,7), tous deux impliqués dans le métabolisme du glutathion.

3. Discussion

Les profils d'exposition utilisés ont été définis dans des travaux préalables, sur la base de la viabilité cellulaire mesurée par dosage de l'ATP intracellulaire. En effet, des durées d'exposition pour lesquelles une sub-toxicité était observée, c'est-à-dire moins de 20 % de cytotoxicité, avaient été retenues. Ces tests préliminaires avaient montré une viabilité cellulaire moyenne à T2 de 83,6 %, 83 %, 85 % et 91,3 % pour les cellules exposées respectivement aux dispositifs 3R4F, HTP, MB-18W, et MB-30W. La cytotoxicité évaluée par la mesure de la LDH dans les milieux de culture des échantillons inclus dans cette étude est cohérente avec ces données préliminaires puisqu'elle atteint au maximum 18,1 % de cytotoxicité (pour le dispositif HTP) par rapport à un échantillon non exposé. Il est à noter que ce test tient uniquement compte de la mort cellulaire par nécrose. Se placer à des cytotoxicités basses et comparables entre les différents dispositifs permet de s'assurer que les différences observées dans les abondances de métabolites seront bien le reflet des variabilités biologiques dues aux expositions, et non pas à des différences de quantités de cellules vivantes au départ. Ces résultats ont orienté le choix du type de normalisation à appliquer. En effet, il en existe deux principaux qui sont 1) la normalisation pré-acquisition, consistant à adapter la quantité d'échantillon pour normaliser les données avant leur acquisition, lorsque d'importantes variations quantitatives sont mises en évidence et 2) la normalisation post-acquisition, consistant à normaliser les données une fois acquise (182). Vu les résultats de cytotoxicité observés, il semble pertinent d'appliquer une méthode de normalisation post-acquisition.

Après acquisition des données par LC-HRMS, le logiciel Progenesis (Waters®) a permis une extraction exhaustive des pics à partir des acquisitions réalisées en mode d'ionisation ESI+ et ESI-. Les étapes du pré-traitement des données ont permis de sélectionner les pics correspondant à des composés, en éliminant les interférences analytiques (bruit de fond). Deux échantillons « outliers » ont pu être identifiés, et ont donc été retirés du jeu de données analysé par la suite. Ils appartenaient tous deux au groupe exposé à l'e-cig à puissance 30 W, l'un pour une durée de 30 min (T1) et l'autre pour une durée 60 min (T2).

Les données ont ensuite été analysées en utilisant des approches multivariées et univariées. La PLS-DA a permis de discriminer les quatre groupes exposés. L'ANOVA

couplée au t-test a quant à elle permis de mettre en évidence des empreintes métabolomiques permettant de distinguer les échantillons exposés des échantillons contrôles pour les groupes HTP et 3R4F. Les résultats ont été systématiquement corrigés par le FDR, méthode d'intérêt pour limiter les fausses associations de variables (faux positifs), mais assez stringente et donc pouvant entraîner un risque de rejet de vraies associations (faux négatifs). Le nombre de composés significativement dérégulés ainsi que l'intensité de ces modulations (FC) augmentent de manière limitée avec la durée d'exposition, suggérant qu'il existe peu d'effet temps-dépendant. Une importante proportion des composés discriminants sont communs entre les deux expositions aux produits du tabac. En effet, 84 % des composés discriminants après exposition à la cigarette 3R4F le sont également après exposition au tabac chauffé. Le plus grand nombre de métabolites discriminant l'exposition au tabac chauffé pourrait s'expliquer par des effets plus importants avec ce dispositif, mais plus vraisemblablement par la différence de durée d'exposition entre les aérosols de tabac chauffé (30 min pour T1 et 60 min pour T2) et la cigarette de référence (1 min pour T1 et 2 min pour T2). En revanche, alors même que les durées d'exposition à l'e-cig étaient les mêmes que pour l'HTP, on observe très peu d'effet sur le métabolome des échantillons exposés aux émissions d'e-cig. En effet, 1 seul métabolite était dérégulé de façon significative pour l'exposition à une puissance de 18 W à T1 et T2, tandis qu'aucun composé n'était modulé significativement pour la puissance la plus forte testée (30 W). Ce composé, non identifié, n'était pas significativement modulé avec les autres conditions d'exposition. Ceci suggère qu'il y a peu d'altérations métaboliques mesurables dans nos conditions expérimentales après l'exposition à l'e-cig. Au total, les résultats suggèrent que l'exposition des cellules aux produits du tabac, c'est-à-dire 3R4F et HTP, semble avoir plus d'impact au niveau métabolomique que l'e-cig. Notons que si des dérégulations similaires du métabolome sont observées dans les cellules exposées au tabac chauffé ou à la cigarette, elles apparaissent après des durées d'exposition beaucoup plus courtes pour la cigarette (< 2 min d'exposition) que pour l'HTP (30 à 60 min d'exposition).

L'étape d'identification des composés est ensuite une étape indispensable pour l'interprétation biologique de ces empreintes métabolomiques. Elle est réalisée par comparaison des données chromatographiques et spectrales acquises à des bases de données de référence. Dans ce travail, la méthodologie d'identification utilisée a fait

appel à 3 bases de référence : 1) une base de données interne établie à partir de standards, 2) une base de données plus large basée sur la base HMDB et contenant des données prédites, ainsi que 3) une base de données commerciale. Avec cette méthodologie, une proposition d'identification a pu être attribuée à près de 40 % des composés discriminants (soit 85 composés). Une identification sûre (niveau 1) a pu être attribuée à 11 composés. Les autres identifications ont un degré de confiance de niveau 2 ou 3. Certains composés sont d'origine exogène, tandis que d'autres sont attribués au métabolisme cellulaire.

3.1. Composés exogènes

Parmi les composés identifiés, 8 sont des composés exogènes dont 7 varient de manière similaire lors de l'exposition à la fumée de 3R4F ou aux aérosols d'HTP. Un autre composé, le 3-méthylindole, n'est augmenté que lors de l'exposition aux aérosols d'HTP.

La nicotine augmente significativement après exposition à la fumée de 3R4F et aux aérosols d'HTP par rapport aux contrôles non exposés. En revanche, sa variation n'est pas significative pour le groupe exposé à l'e-cig 18 W, mais proche de la significativité pour l'e-cig puissance 30 W (p-value = 0,08). Ces résultats, non attendus pour l'e-cig, peuvent être expliqués par l'importante dispersion des valeurs de nicotine pour les échantillons de contrôle de l'exposition à la MB-18W. Pour l'exposition aux aérosols de MB-30W, 2 échantillons « outliers » ont été retirés du jeu de données, l'un pour le temps d'exposition 30 min et l'autre pour le temps d'exposition 60 min, entraînant une perte de puissance statistique et une p-value non significative mais inférieure à 0,01. Des travaux préalables au sein de l'ULR4483 ont comparé la quantité de nicotine dans les aérosols (et non pas dans les cultures cellulaires) générés par l'HTP ou l'e-cig réglée aux mêmes puissances que notre étude en comparaison avec la 3R4F. Les résultats obtenus ont montré que la quantité de nicotine émise est très variable entre les dispositifs testés avec un profil de bouffée identique. En effet, l'HTP émet environ 30 % de nicotine en moins que la cigarette (63 µg/bouffée contre 95 µg/bouffée). Quant à l'e-cig, elle émet 60 µg de nicotine par bouffée lorsqu'elle est réglée à 18 W et 137 µg/bouffée lorsqu'elle est réglée à 30 W (18).

La 1-naphthylamine, augmentée dans les deux conditions d'exposition aux produits du

tabac, est un composé de la classe des naphthalènes qui sont des amines aromatiques. Ce composé est largement décrit comme composé de la fumée de cigarette conventionnelle (183,184), ainsi que comme marqueur d'exposition à la fumée de cigarette, par exemple dans la matrice urinaire (185).

Le 3-méthylindole, quant à lui augmenté après exposition aux aérosols d'HTP mais pas après exposition à la fumée de 3R4F (potentiellement due à la courte durée d'exposition pour la 3R4F), est un composé de la classe des indoles et dérivés. Il est formé par la pyrolyse du tryptophane lors de la combustion du tabac. Ce résultat pourrait être en faveur d'un phénomène de pyrolyse avec le dispositif HTP. Les résultats publiés par Vivarell *et al.* indiquent que les émissions d'HTP contiennent des sous-produits de pyrolyse et de dégradation thermique identiques à la fumée de cigarette conventionnelle, et qu'ils provoqueraient des lésions pulmonaires graves et augmenteraient le risque de cancer sur modèle animal (186). Le 3-méthylindole a été décrit comme cytotoxique sur les lignées BEAS-2B après bio-activation (187) et mutagène sur des modèles de microsomes pulmonaires humains, soutenant l'hypothèse d'un probable effet cancérigène pulmonaire chez l'homme (188). Sur cellules NHBE, l'exposition au 3-méthylindole a également provoqué des lésions et des mutations importantes de l'ADN, sans déclencher de défenses apoptotiques, ce qui renforce l'hypothèse selon laquelle ce composé inhalé de la fumée de cigarette pourraient être un agent cancérigène sélectif du poumon (189).

La β -carboline (appelée également norharman) et l'harman sont deux β -carboline, dont la source majeure décrite est la fumée de tabac (190). Elles sont augmentées après exposition aux aérosols d'HTP et à la fumée de 3R4F. Elles possèdent des propriétés inhibitrices des monoamines oxydases. A ce titre, leur implication dans les effets comportementaux liés à l'addiction a été plus précisément étudiée, sans mettre en évidence de relation avec cet effet (191). Elles ont montré des effets comportementaux, notamment des effets cognitifs (192), anxiolytiques (193) et antidépresseurs (194), qui sont attribués à cette famille d'alcaloïdes.

La scopoletine (encore appelé 6-méthoxy-7-hydroxycoumarin), également augmentée suite aux expositions aux produits du tabac, appartient à la classe des coumarines et dérivés et est un polyphénol également décrit dans la composition du tabac (195). Il est considéré comme un produit chimique dérivé de la plante, et non comme un produit

de combustion.

Certains autres composés (rollipyrrole, theaspirane) sont très peu décrits et demandent des investigations et confirmations supplémentaires.

Pour conclure, ces composés peuvent ainsi être considérés comme des marqueurs d'exposition aux produits du tabac, aussi bien issus des émissions de cigarette que d'HTP. Alors que l'industrie du tabac décrit les produits du tabac chauffé comme à faible risque pour la santé des usagers, ces résultats sont en faveur d'une toxicité néanmoins existante, puisque certains composés cancérigènes sont identifiés. Ainsi, la toxicité de l'HTP ne doit pas être sous-estimée.

Les autres composés identifiés sont des composés endogènes, les plaçant comme marqueurs d'effets.

3.2. Composés endogènes

3.2.1. Stress oxydant

L'exposition aux produits du tabac est associée à une réaction de stress cellulaire pulmonaire, et notamment une réponse au stress oxydant causée directement par l'exposition à certains composés chimiques, ou induite par la génération d'espèces réactives de l'oxygène (ERO). Nous avons ainsi observé des perturbations du métabolisme du glutathion, notamment avec une augmentation du glutathion oxydé après les plus longues durées d'exposition (T2), aussi bien pour l'exposition aux émissions d'HTP que de 3R4F. Ces résultats sont en faveur de la génération de stress oxydant à la suite de l'exposition aux produits du tabac. Ils sont cohérents avec les études précédemment réalisées au sein de l'ULR4483, ayant montré une augmentation du rapport glutathion oxydé sur glutathion réduit après exposition à la cigarette et au tabac chauffé (18). Une autre étude, également menée par l'ULR4483, a montré que l'HTP et la 3R4F induisaient l'activation du facteur de transcription Nrf2 et l'expression de ses gènes cibles, l'hème oxygénase 1 (HO1) et la NAD(P)H-Quinone Déhydrogénase 1 (NQO1), démontrant une réponse antioxydante après une exposition aux produits du tabac. Dans notre étude de métabolomique, une augmentation de la biliverdine a été constatée après les deux durées d'exposition aux émissions de 3R4F. La biliverdine est un métabolite de l'hème, produit sous l'action

d'une enzyme cytoprotectrice, l'hème oxygénase (HO) (196,197). La biliverdine est ensuite rapidement transformée en bilirubine par la biliverdine réductase. La biliverdine peut être régénérée à partir de la bilirubine par des réactions avec les ERO. Différentes propriétés ont été reconnues à la biliverdine et à la bilirubine, notamment des propriétés antioxydantes en réponse au stress oxydant. L'étude de Titz *et al.* (147) réalisée sur modèle murin, analysant le métabolome du tissu pulmonaire après exposition à la cigarette conventionnelle ou à l'HTP, a également mis en évidence des dérégulations du métabolisme de la bilirubine dans le groupe exposé à la cigarette, mais pas dans le groupe exposé à l'HTP. La bilirubine et son enzyme de synthèse, la biliverdine réductase présentaient une augmentation d'expression. La biliverdine présentait une tendance à l'augmentation, tandis que l'abondance des hèmes oxygénases était augmentée par l'exposition à la 3R4F. Notre étude a également mis en évidence des modulations du métabolisme des nucléotides (purines et pyrimidines) après exposition à l'HTP (mais pas pour la 3R4F). Le métabolisme des purines, a été rapporté comme producteur d'ERO par la voie de la xanthine oxydase (198). En outre, une dérégulation du métabolisme des nucléotides a déjà été décrite chez les fumeurs de cigarette, et a été décrite pour son rôle dans le développement de cancer (199,200). Nous avons constaté une diminution de la méthionine après exposition à l'HTP. Le stress oxydant peut induire l'oxydation de la méthionine (201), à l'origine d'une diminution de cette dernière. Une augmentation des produits de cette oxydation (méthionine sulfoxide et N-acétyl-méthionine-sulfoxide) a été décrite après une exposition de cultures de cellules épithéliales gingivales humaines exposées pendant 28 minutes sur 3 jours à la fumée de cigarette conventionnelle, mais pas aux aérosols d'HTP (145). Une diminution de la méthionine plasmatique a également été observée chez les utilisateurs de cigarette par rapport aux vapoteurs ou aux individus non exposés (139).

3.2.2. Métabolisme énergétique

L'augmentation de la demande d'énergie est une réponse attendue pour toutes les cellules soumises à un stress. Plusieurs métabolites signalent une altération du métabolisme énergétique après exposition aux produits du tabac. Le niveau d'ADP était augmenté après exposition aux aérosols d'HTP (mais pas pour la 3R4F). Un intermédiaire du cycle de Krebs, l'isocitrate, était diminué pour les cellules exposées à

l'HTP par rapport aux contrôles (mais pas pour la 3R4F). Cette voie métabolique est au centre du métabolisme énergétique cellulaire qui fournit l'énergie nécessaire à la respiration cellulaire. Il est bien connu que la fumée de cigarette inhibe la fonction mitochondriale respiratoire et dérégule le cycle de Krebs (202). L'isocitrate, plus précisément, a également déjà été décrit comme diminuée après des expositions répétées de culture de tissu bronchique 3D à la fumée de cigarette (125). Notre analyse a également montré une diminution significative de l'isovalérylcarnitine, carnitine à chaîne courte, et une diminution d'un acylcarnitine à chaîne intermédiaire, à savoir l'hexanoylcarnitine, après exposition aux émissions d'HTP. Les acylcarnitines sont des composés organiques contenant un acide gras. Ils jouent un rôle central dans le transport des acides gras à travers la membrane interne de la mitochondrie pendant la bêta-oxydation, voie métabolique de dégradation des acides gras pour produire notamment de l'acetyl-CoA qui va alimenter le cycle de Krebs. Leur perturbation confirme le dysfonctionnement mitochondrial et la perturbation de la bêta-oxydation des acides gras après exposition aux produits du tabac, qui peut refléter une demande énergétique intracellulaire importante.

3.2.3. Métabolisme lipidique

D'autres lipides que les carnitines étaient également perturbés. En effet, dès l'analyse taxonomique des métabolites identifiés (Figure 42), nous avons pu objectiver les effets de l'exposition aux aérosols d'HTP ou de 3R4F sur le métabolisme lipidique, avec des variations significatives de plusieurs classes de lipides telles que les glycérophospholipides (glycérophosphocholines (LysoPC) ou glycéroéthanolamines (LysoPE)) ou encore les acides gras (carnitines, eicosanoïdes). La majorité de ces espèces lipidiques significativement affectées étaient augmentées, avec plutôt une augmentation entre T1 et T2, en faveur d'une accumulation lipidique. Pour les deux produits du tabac, plusieurs pics attribués à la classe des eicosanoïdes étaient significativement augmentés après exposition. Les eicosanoïdes, dérivés de l'acide arachidonique sous l'action d'enzymes lipooxygénase et cyclooxygénase, sont des médiateurs de signalisation de la réponse inflammatoire. Une augmentation de l'activité de ces enzymes a déjà été décrite chez les fumeurs (203), et associée à l'inflammation et des pathologies pulmonaires (204) ou les cancers (205). Pour l'HTP, une augmentation de l'activité de la cyclooxygénase a également déjà été décrite

(186). Ainsi, l'exposition aux émissions de tabac et de tabac chauffé pourrait affecter la régulation de la réponse inflammatoire des cellules bronchiques. Il est à noter que la majorité des études analysant les eicosanoïdes fait appel à des techniques analytiques recourant à un mode d'ionisation négatif, tandis que les identifications ont été ici réalisées pour des composés détectés en mode d'ionisation positif. Dans ce contexte et compte tenu du risque d'identifications erronées, une confirmation de ces identifications sera indispensable pour confirmer ces conclusions. Dans notre étude, d'autres classes de lipides étaient également augmentées après exposition aux aérosols d'HTP. Il s'agit notamment de lysoglycérophospholipides de la sous-classe des glycérophosphocholines (lysophosphatidylcholines, LysoPC) et des glycérophosphoéthanolamines (lysophosphatidyléthanolamines, LysoPE). Les LysoPC sont produits par le clivage de la phosphatidylcholine (PC) par la phospholipase A2 (PLA2), réaction formant également des acides gras libres, par exemple l'acide arachidonique, qui est le précurseur des eicosanoïdes (206). Cette réaction est également possible sous l'effet d'ERO. Une augmentation de ces lipides est donc également en faveur d'un état inflammatoire. Les LysoPE sont également produits par le clivage de la phosphatidyléthanolamine (PE) par la PLA2. Des travaux précédents de l'ULR4483 avaient également démontré la mise en place d'une inflammation après exposition à l'HTP ou à l'e-cig par le dosage de dix médiateurs de l'inflammation 24 heures après une exposition des cellules BEAS-2B aux différentes émissions. Les résultats avaient montré une augmentation de la cytokine pro-inflammatoire IL-6 après une exposition à l'e-cig (18 et 30 W) et à l'HTP. De plus, une induction de l'IL-8 a également été mise en évidence après exposition à l'HTP (18).

Les études menées par l'industrie du tabac comparant le lipidome de tissu bronchique murin après exposition au tabac et au tabac chauffé avaient montré des dérégulations des lipides, mais beaucoup plus marquées pour le tabac que pour l'HTP (effets très limités pour HTP), leur permettant de conclure à un effet moindre de l'HTP par rapport à la cigarette conventionnelle (146,147). Vu nos résultats, il ne faut pas négliger l'impact de l'HTP sur ces voies métaboliques.

3.2.4. Autres métabolites

Outre les perturbations des voies métaboliques susmentionnées, d'autres perturbations ont été observées. La diacétylspermidine était augmentée à la suite des

expositions aux émissions de 3R4F et d'HTP. Ce composé est impliqué dans le métabolisme des polyamines, notamment intermédiaire du catabolisme de la spermidine en putrescine par les enzymes spermidine/spermine-N(1)-acetyltransferase-1 et spermine oxydase. Une augmentation des polyamines a été observée dans des pathologies respiratoires telles que l'asthme (207).

Certains autres composés sont peu décrits et demandent des investigations supplémentaires.

3.3. Conclusions

Dans la littérature, très peu d'études de métabolomique traitent de l'effet de l'exposition aiguë aux aérosols d'HTP en comparaison à la cigarette conventionnelle ou à des contrôles non exposés, et aucune n'inclus de comparaison avec des expositions aux aérosols d'e-cig, soulignant l'originalité de ce travail. Seulement trois études ont été identifiées lors de notre revue de littérature et elles étaient toutes réalisées par l'industrie du tabac, donc susceptibles d'être biaisées en raison de conflits d'intérêts. Ce travail par analyses de métabolomique offre une méthodologie permettant de « découvrir » les voies biologiques potentiellement affectées par l'exposition à l'HTP. La plupart des composés et voies impliqués tels que la réponse au stress oxydant, le métabolisme des lipides ou le métabolisme énergétique ont également déjà été décrits comme altérés après une exposition à la cigarette conventionnelle. Dans notre étude, l'exposition à la fumée de cigarette a dérégulé moins de métabolites que l'exposition aux aérosols d'HTP, mais la plupart de ces métabolites étaient communs avec ceux dérégulés après l'exposition à l'HTP. Ceci pourrait être expliqué par la différence de temps d'exposition, de 30 et 60 min pour l'HTP, contre 1 et 2 min pour la cigarette conventionnelle. Certains composés avaient été identifiés dans des maladies associées au tabagisme, telles que l'asthme, la BPCO et le cancer du poumon. Ainsi, nos résultats décrivent la possibilité que les métabolites dérégulés après les expositions à l'HTP soient associés à la pathogenèse des maladies pulmonaires. En revanche peu d'altérations métaboliques étaient mesurables dans nos conditions expérimentales après exposition à l'e-cig. La littérature fait néanmoins état d'altérations métaboliques possibles. Cette différence pourrait être expliquée par des conditions expérimentales non comparables, ou si des expositions aiguës ne suffisent pas pour que les cellules développent des modifications métabolomiques suffisantes

pour être détectées avec la méthode mise en œuvre. Des expositions réitérées sur plusieurs jours pourront être envisagées avec ces modèles *in vitro* ou en utilisant des modèles animaux afin de mettre en évidence les dérégulations métaboliques dues aux expositions à l'HTP ou à l'e-cig sur plus long-terme.

Cette étude représente donc un travail de « découverte » pour une évaluation plus complète en ciblant plus précisément les voies dérégulées.

3.4. Limites

Il est important de souligner certaines limites de cette analyse fonctionnelle et notamment concernant l'étape d'identification. En effet, une masse exacte peut être associée à plusieurs formules brutes, elles-mêmes pouvant correspondre à plusieurs structures chimiques, entraînant un nombre de possibilités d'identifications important pour un même signal. La multiplication du nombre de paramètres d'identification complémentaires à la masse exacte permet de réduire le risque d'erreur d'identification, qui reste néanmoins persistant. Concernant plus particulièrement la confrontation spectrale entre données acquises et données de référence, l'interprétation des spectres de masse est compliquée par le fait que les ions produits et leurs rapports observés sont dépendants de l'analyseur et des paramétrages établis pour la fragmentation des composés (208). Ceci est d'ailleurs la principale limite à la construction d'une base de données universelle pour la communauté travaillant en métabolomique. Au-delà de ces difficultés et d'un point de vue technique, il est possible que certains composés présentent une fragmentation non souhaitée dans la source d'ionisation. Ceci peut amener à la production d'ions qui ont la même masse ou la même structure que d'autres métabolites, et donc à de potentielles erreurs d'identification. Il est également possible que plusieurs espèces ioniques différentes (par exemple, espèces protonées et adduites, ions de fragments et pics isotopiques) soient détectées pour un même métabolite. Ces signaux redondants peuvent rendre difficile l'attribution correcte de ces différentes espèces pour un même métabolite (209). Dans d'autres cas, certains composés détectés avec une faible réponse de l'analyseur peuvent présenter une faible fragmentation, voir pas de fragmentation mesurable, empêchant la confrontation des données acquises aux bases de données (210).

Les données acquises sont comparées à des bases de données qui sont régulièrement incrémentées par la communauté scientifique, permettant le développement de bases de données toujours plus complètes. Certaines sont spécifiques d'un organisme comme par exemple HMDB pour l'Homme et d'autres spécifiques de classes de composés comme par exemple Lipidmaps pour les lipides. Dans la continuité de ce projet, la confrontation des données acquises à d'autres bases de données que celles déjà utilisées pourrait potentiellement améliorer l'identification des composés non identifiés. Cette démarche pourrait également permettre de confirmer les résultats déjà obtenus.

Finalement, pour s'assurer de l'exactitude de l'identification des composés de niveau d'identification 2 ou 3, il est nécessaire d'injecter dans les mêmes conditions analytiques un standard commercial du composé, pour que les données d'identification soient identiques. Ceci n'est pas toujours possible, notamment en termes de coût et de disponibilité commerciale de ces standards.

Une autre limite principale de cette étape d'identification est l'impossibilité d'identifier une part importante des composés discriminants. Ceci peut s'expliquer 1) si le métabolite ne figure pas encore dans les bases de données, 2) si les caractéristiques d'identification des composés ne sont pas disponibles dans les bases de données, comme par exemple certaines caractéristiques d'adduits, ou 3) s'il ne s'agit pas d'un métabolite (208).

Une autre approche d'identification possible, permettant une annotation de niveau 3 des composés c'est-à-dire en lui attribuant une classe plutôt qu'une identification précise, est l'approche par réseaux moléculaires (211). Il s'agit d'une approche bioinformatique capable d'organiser et de visualiser des données spectrales sous la forme de carte de similarité spectrale, mettant en lumière l'existence de groupes spectraux ainsi que leurs degrés de similarité. Une autre alternative possible nécessiterait de détecter, isoler et purifier chaque composé et de réaliser une analyse de RMN pour permettre une élucidation structurale de chaque composé (212).

C'est pour toutes ces raisons que la métabolomique non ciblée reste un outil de « découverte » de marqueurs. Une confirmation des résultats par méthodes spécifiques est nécessaire.

CONCLUSIONS GÉNÉRALES

CONCLUSIONS GÉNÉRALES

Alors que l'OMS prône le sevrage tabagique, l'arrivée sur le marché de la cigarette électronique puis plus récemment du tabac chauffé pose question sur l'innocuité de ces dispositifs. L'évaluation des effets du tabac chauffé et du vapotage sur la santé peut se faire à plusieurs niveaux. L'approche métabolomique a ici été retenue pour sa capacité à mesurer et comprendre les variations des métabolites, produits finaux des processus de régulation cellulaire, étroitement liées à la réponse biologique de l'organisme à la suite des expositions. Cette approche a permis à la fois d'étudier l'expression de métabolites exogènes, signature d'exposition, et de métabolites endogènes, signature d'effets.

Le premier point clé de ce travail de thèse était de développer une méthode de métabolomique par spectrométrie de masse haute résolution. Nous avons utilisé un UPLC-IMS-QTOF (Vion, Waters®) pour développer une méthode analytique adaptée à l'analyse du métabolome. Les avantages sont l'acquisition des données en MS^E et l'utilisation de la mobilité ionique comme critère orthogonal supplémentaire utile pour l'identification des composés détectés. Il est à noter qu'une approche de métabolomique telle que développée ici reste limitée en termes de fraction du métabolome détectable, puisque son étendue est évaluée à plusieurs dizaines de milliers de métabolites. Seule une combinaison de plusieurs méthodes pourrait permettre une meilleure représentativité du métabolome. Nous avons ensuite créé une librairie interne de référence à l'aide d'une série de standards. À partir de celle-ci, nous avons construit une base de données prédite de grande dimension, rassemblant les données d'identification prédites pour les composés de la base de données HMDB, soit près de 114 000 composés.

Le deuxième axe de ce travail de thèse était d'utiliser cette approche pour mieux caractériser les mécanismes de toxicité des nouveaux produits du tabac ou du vapotage. La métabolomique nécessite la coordination de plusieurs disciplines telles que la biologie, les statistiques, la bioinformatique et la chimie analytique.

Pour cette étude, le choix du modèle d'exposition s'est porté sur un modèle *in vitro*. Puisque le métabolome peut être influencé par tout type de stimuli, et pas seulement par l'exposition étudiée, le modèle cellulaire permet de s'affranchir d'un grand nombre

de facteurs potentiellement confondants qui pourraient affecter le métabolome lors d'études *in vivo*. Ainsi, des cellules épithéliales bronchiques humaines (lignée BEAS-2B) cultivées à l'interface air-liquide ont été exposées à de l'air stérile (contrôle) ou aux émissions d'une cigarette de référence, de tabac chauffé, ou d'e-cig. Deux durées d'exposition (T1 ou T2, 1 min ou 2 min pour la cigarette, 30 min ou 60 min pour l'HTP et l'e-cig) ont été choisies dans l'objectif de mettre en évidence un potentiel impact de la durée d'exposition sur la nature et/ou l'intensité des effets toxicologiques des émissions étudiées.

L'exposition de cellules épithéliales bronchiques humaines aux émissions des différents dispositifs a permis de mettre en évidence que les émissions de tabac chauffé et de cigarette induisaient des dérégulations du métabolome par rapport aux cellules non exposées. La plupart des composés dérégulés à la suite de l'exposition au tabac étaient également dérégulés après exposition au tabac chauffé, dès la première durée d'exposition. La signature métabolomique de l'exposition au tabac chauffé comprenait également d'autres composés non modifiés pour le tabac dans nos conditions expérimentales. Ce constat peut s'expliquer d'abord par la différence de composition des émissions d'HTP et de cigarette conventionnelle : en effet, une étude (indépendante de l'industrie du tabac) a montré que plus de 20 composés dangereux ou potentiellement dangereux sont retrouvés à des niveaux plus importants dans les aérosols d'HTP que dans la fumée de cigarette (59). De plus, les différences observées en termes de dérégulations métaboliques peuvent également être expliquées par la grande différence de temps d'exposition utilisés dans cette étude entre les deux produits du tabac. En effet, les durées d'exposition (1 et 2 min pour la cigarette vs 30 et 60 min pour l'HTP) ont été choisies sur la base de doses sub-toxiques comparables (> 80 % de viabilité cellulaire) pour s'assurer que les différences observées sont bien le reflet des variabilités biologiques dues aux expositions, et non pas à des différences de quantités de cellules vivantes au départ. Ainsi, la toxicité beaucoup plus importante de la fumée de cigarette ne permet pas de réaliser des expositions plus longues.

Les effets de l'e-cig ont également été étudiés dans ce projet mais très peu de dérégulations ont été mesurées (1 seul composé impacté par ces expositions), quels que soient la puissance et le temps d'exposition étudié. Il semble probable que des expositions aiguës ne suffisent pas pour que les cellules développent des

modifications métaboliques suffisantes pour être détectées avec la méthode mise en œuvre. Puisque les temps d'exposition sont identiques entre tabac chauffé et e-cig, il est possible de conclure que le tabac chauffé entraîne des dérégulations plus importantes du métabolome et donc que le tabac chauffé est plus toxique que l'e-cig dans ce modèle.

L'un des atouts de l'approche métabolomique est sa capacité à générer des hypothèses sur les voies de signalisation perturbées en fonction des différents métabolites dérégulés. L'identification de ces composés a pu montrer la présence de composés exogènes communs après des expositions à la cigarette ou au tabac chauffé, pour la plupart déjà largement décrits dans la fumée de tabac. Certains d'entre eux sont classés comme cancérigènes. Même présents à des niveaux plus faibles dans les émissions de tabac chauffé, ils peuvent provoquer des dommages de l'ADN. Des expositions répétées pourraient donc amener à des accumulations de mutations et ainsi participer au processus de cancérogenèse au même titre que la cigarette conventionnelle. Les composés endogènes des signatures métabolomiques mises en évidence ont été resitués dans leur contexte biologique par une analyse de voies métaboliques. Ainsi, le tabac chauffé dérégule des voies métaboliques impliquées notamment dans les processus de stress oxydant, le métabolisme énergétique, ou encore le métabolisme lipidique, reflétant la réaction de stress cellulaire à la suite de l'exposition aux aérosols de tabac chauffé. Pour finir, certains composés restent non identifiés. D'autres méthodes aidant à l'identification de ces métabolites, comme l'approche par réseaux moléculaires, pourraient être mises en œuvre dans la suite de ce projet.

La méthode de métabolomique mise en place dans cette étude est basée sur l'UPLC-HRMS, une méthode haut débit, apportant une vision globale des dérégulations à la suite des expositions. Ces résultats permettent de mieux comprendre les mécanismes moléculaires et cellulaires de toxicité des aérocontaminants d'intérêt et d'identifier des biomarqueurs pertinents d'effets et/ou d'exposition à ces aérocontaminants. Cette approche se positionne comme un outil de « découverte » nécessitant une validation ciblée des biomarqueurs potentiels et ouvre de nouvelles perspectives. Les résultats de ce travail de thèse suggèrent que l'exposition des cellules aux produits du tabac, c'est-à-dire 3R4F et HTP, a plus d'impact au niveau métabolomique que l'e-cig. Ceci est cohérent avec les résultats préliminaires obtenus pour les analyses de l'expression

génomique (transcriptomique), réalisées sur des échantillons exposés au même moment que ceux inclus pour les analyses de métabolomique. Le tabac chauffé induisait des altérations transcriptomiques très semblables à celles observées avec la cigarette. L'analyse fonctionnelle des gènes modulés a mis en évidence plusieurs voies d'intérêt. Ainsi, le tabac chauffé active de nombreuses voies impliquées dans les processus de cancérogenèse, avec en particulier une altération possible de la fonction mitochondriale. Intégrer les données de plusieurs niveaux biologiques, c'est-à-dire transcriptomique et métabolomique, pourrait permettre de mieux comprendre les mécanismes moléculaires et cellulaires de toxicité des produits du tabac.

Il sera important de valider ces résultats dans des modèles de cultures primaires cultivées en interface air-liquide, qui permettent un meilleur niveau de différenciation cellulaire et sont donc plus représentatives de l'épithélium bronchique humain que la lignée BEAS-2B. Des expositions répétées sur plusieurs jours pourront également être envisagées avec ces modèles *in vitro* ou en utilisant des modèles animaux afin de mettre en évidence les dérégulations métaboliques du HTP et de l'e-cig sur le long-terme.

En conclusion, les résultats issus de ce projet apportent des éléments novateurs sur la toxicité aiguë des émissions de tabac chauffé, en comparaison à la fumée de cigarette conventionnelle et aux aérosols d'e-cig. Ils démontrent qu'il ne faut pas sous-estimer la toxicité de ces nouveaux dispositifs. Ces données devraient participer au débat sur le rapport bénéfices/risques de ces produits et faciliter leur positionnement dans les campagnes de lutte contre le tabagisme.

RÉFÉRENCES

RÉFÉRENCES

1. Reitsma MB, Kendrick PJ, Ababneh E, Abbafati C, Abbasi-Kangevari M, Abdoli A, et al. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *The Lancet*. 19 juin 2021;397(10292):2337-60.
2. Pasquereau A, Andler R, Guignard R, Gautier A, Soullier N, Richard JB, et al. Prévalence nationale et régionale du tabagisme en France en 2021 parmi les 18-75 ans, d'après le Baromètre de Santé publique France. *Bull Épidémiologique Hebd*. 2022;(26):470-80.
3. European Commission, Directorate-General for Communication, Directorate-General for Health and Food Safety. Attitudes of Europeans towards tobacco and electronic cigarettes: report. Eur Comm [Internet]. 2021 [cité 6 déc 2022]; Disponible sur: <https://europa.eu/eurobarometer/surveys/detail/2240>
4. Organisation mondiale de la santé. Rapport de l'OMS sur l'épidémie mondiale de tabagisme, 2021 : les produits nouveaux et émergents [Internet]. [cité 21 déc 2022]. Disponible sur: <https://www.who.int/fr/publications-detail/9789240032095>
5. Jha P. The hazards of smoking and the benefits of cessation: A critical summation of the epidemiological evidence in high-income countries. Franco E, éditeur. *eLife*. 24 mars 2020;9:e49979.
6. Bonaldi C, Boussac M, Nguyen-Thanh V. Estimation du nombre de décès attribuables au tabagisme, en France de 2000 à 2015. *Bull Épidémiologique Hebd*. 28 mai 2019;15:278-84.
7. Products C for T. Harmful and Potentially Harmful Constituents in Tobacco Products and Tobacco Smoke: Established List. FDA [Internet]. 21 juill 2020 [cité 11 avr 2023]; Disponible sur: <https://www.fda.gov/tobacco-products/rules-regulations-and-guidance/harmful-and-potentially-harmful-constituents-tobacco-products-and-tobacco-smoke-established-list>
8. Rodgman A, Perfetti TA. *The Chemical Components of Tobacco and Tobacco Smoke*. 2^e éd. Boca Raton: CRC Press; 2013. 2332 p.
9. Talhout R, Schulz T, Florek E, van Benthem J, Wester P, Opperhuizen A. Hazardous compounds in tobacco smoke. *Int J Environ Res Public Health*. févr 2011;8(2):613-28.
10. United States Public Health Service Office of the Surgeon General, National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. *Smoking Cessation: A Report of the Surgeon General*. Washington (DC): US Department of Health and Human Services; 2020.
11. Organisation mondiale de la santé. WHO global report on trends in prevalence of tobacco use 2000-2025, third edition [Internet]. World Health Organization; 2019 [cité 22 déc 2022]. Disponible sur: <https://www.who.int/publications-detail->

redirect/who-global-report-on-trends-in-prevalence-of-tobacco-use-2000-2025-third-edition

12. Jerzyński T, Stimson GV, Shapiro H, Król G. Estimation of the global number of e-cigarette users in 2020. *Harm Reduct J.* 23 oct 2021;18(1):109.
13. Pasquereau A, Quatremère G, Guignard R, Andler R, Verrier F, Pourchez J, et al. Baromètre de Santé publique France 2017. Usage de la cigarette électronique, tabagisme et opinions des 18-75 ans. *Saint-Maurice Santé Publique Fr.* 2019;17.
14. Goniewicz ML, Knysak J, Gawron M, Kosmider L, Sobczak A, Kurek J, et al. Levels of selected carcinogens and toxicants in vapour from electronic cigarettes. *Tob Control.* mars 2014;23(2):133-9.
15. Belushkin M, Tabin Djoko D, Esposito M, Korneliou A, Jeannet C, Lazzerini M, et al. Selected Harmful and Potentially Harmful Constituents Levels in Commercial e-Cigarettes. *Chem Res Toxicol.* 17 févr 2020;33(2):657-68.
16. Sleiman M, Logue JM, Montesinos VN, Russell ML, Litter MI, Gundel LA, et al. Emissions from Electronic Cigarettes: Key Parameters Affecting the Release of Harmful Chemicals. *Environ Sci Technol.* 6 sept 2016;50(17):9644-51.
17. Beauval N, Antherieu S, Soyez M, Gengler N, Grova N, Howsam M, et al. Chemical Evaluation of Electronic Cigarettes: Multicomponent Analysis of Liquid Refills and their Corresponding Aerosols. *J Anal Toxicol.* 1 oct 2017;41(8):670-8.
18. Dusautoir R, Zarcone G, Verrièle M, Garçon G, Fronval I, Beauval N, et al. Comparison of the chemical composition of aerosols from heated tobacco products, electronic cigarettes and tobacco cigarettes and their toxic impacts on the human bronchial epithelial BEAS-2B cells. *J Hazard Mater.* 5 janv 2021;401:123417.
19. Hartmann-Boyce J, Lindson N, Butler AR, McRobbie H, Bullen C, Begh R, et al. Electronic cigarettes for smoking cessation. *Cochrane Database Syst Rev.* 17 nov 2022;11(11):CD010216.
20. Sæbø G, Scheffels J. Assessing notions of denormalization and renormalization of smoking in light of e-cigarette regulation. *Int J Drug Policy.* nov 2017;49:58-64.
21. Yoong SL, Hall A, Leonard A, McCrabb S, Wiggers J, Tursan d'Espaignet E, et al. Prevalence of electronic nicotine delivery systems and electronic non-nicotine delivery systems in children and adolescents: a systematic review and meta-analysis. *Lancet Public Health.* sept 2021;6(9):e661-73.
22. Merez-Sadowska A, Sitarek P, Zielinska-Blizniewska H, Malinowska K, Zajdel K, Zakonnik L, et al. A Summary of In Vitro and In Vivo Studies Evaluating the Impact of E-Cigarette Exposure on Living Organisms and the Environment. *Int J Mol Sci.* 19 janv 2020;21(2):E652.
23. Hiemstra PS, Bals R. Basic science of electronic cigarettes: assessment in cell culture and in vivo models. *Respir Res.* 07 2016;17(1):127.

24. Cheng T. Chemical evaluation of electronic cigarettes. *Tob Control*. mai 2014;23 Suppl 2:ii11-17.
25. Scheffler S, Dieken H, Krischenowski O, Förster C, Branscheid D, Aufderheide M. Evaluation of E-cigarette liquid vapor and mainstream cigarette smoke after direct exposure of primary human bronchial epithelial cells. *Int J Environ Res Public Health*. 8 avr 2015;12(4):3915-25.
26. Scott A, Lugg ST, Aldridge K, Lewis KE, Bowden A, Mahida RY, et al. Pro-inflammatory effects of e-cigarette vapour condensate on human alveolar macrophages. *Thorax*. 2018;73(12):1161-9.
27. Lerner CA, Sundar IK, Yao H, Gerloff J, Ossip DJ, McIntosh S, et al. Vapors produced by electronic cigarettes and e-juices with flavorings induce toxicity, oxidative stress, and inflammatory response in lung epithelial cells and in mouse lung. *PloS One*. 2015;10(2):e0116732.
28. Garcia-Arcos I, Geraghty P, Baumlin N, Campos M, Dabo AJ, Jundi B, et al. Chronic electronic cigarette exposure in mice induces features of COPD in a nicotine-dependent manner. *Thorax*. déc 2016;71(12):1119-29.
29. Shivalingappa PC, Hole R, Westphal CV, Vij N. Airway Exposure to E-Cigarette Vapors Impairs Autophagy and Induces Aggresome Formation. *Antioxid Redox Signal*. 01 2016;24(4):186-204.
30. Wu Q, Jiang D, Minor M, Chu HW. Electronic cigarette liquid increases inflammation and virus infection in primary human airway epithelial cells. *PloS One*. 2014;9(9):e108342.
31. Hwang JH, Lyes M, Sladewski K, Enany S, McEachern E, Mathew DP, et al. Electronic cigarette inhalation alters innate immunity and airway cytokines while increasing the virulence of colonizing bacteria. *J Mol Med Berl Ger*. 2016;94(6):667-79.
32. Muthumalage T, Prinz M, Ansah KO, Gerloff J, Sundar IK, Rahman I. Inflammatory and Oxidative Responses Induced by Exposure to Commonly Used e-Cigarette Flavoring Chemicals and Flavored e-Liquids without Nicotine. *Front Physiol*. 2017;8:1130.
33. Lee WH, Ong SG, Zhou Y, Tian L, Bae HR, Baker N, et al. Modeling Cardiovascular Risks of E-Cigarettes With Human-Induced Pluripotent Stem Cell-Derived Endothelial Cells. *J Am Coll Cardiol*. 4 juin 2019;73(21):2722-37.
34. Crotty Alexander LE, Drummond CA, Hepokoski M, Mathew D, Moshensky A, Willeford A, et al. Chronic inhalation of e-cigarette vapor containing nicotine disrupts airway barrier function and induces systemic inflammation and multiorgan fibrosis in mice. *Am J Physiol-Regul Integr Comp Physiol*. 31 janv 2018;314(6):R834-47.
35. Basu AK. DNA Damage, Mutagenesis and Cancer. *Int J Mol Sci*. 23 mars 2018;19(4):970.

36. Tommasi S, Bates SE, Behar RZ, Talbot P, Besaratinia A. Limited mutagenicity of electronic cigarettes in mouse or human cells in vitro. *Lung Cancer Amst Neth.* oct 2017;112:41-6.
37. Misra M, Leverette RD, Cooper BT, Bennett MB, Brown SE. Comparative in vitro toxicity profile of electronic and tobacco cigarettes, smokeless tobacco and nicotine replacement therapy products: e-liquids, extracts and collected aerosols. *Int J Environ Res Public Health.* 30 oct 2014;11(11):11325-47.
38. Neilson L, Mankus C, Thorne D, Jackson G, DeBay J, Meredith C. Development of an in vitro cytotoxicity model for aerosol exposure using 3D reconstructed human airway tissue; application for assessment of e-cigarette aerosol. *Toxicol Vitro Int J Publ Assoc BIBRA.* oct 2015;29(7):1952-62.
39. Anthérieu S, Garat A, Beauval N, Soyez M, Allorge D, Garçon G, et al. Comparison of cellular and transcriptomic effects between electronic cigarette vapor and cigarette smoke in human bronchial epithelial cells. *Toxicol In Vitro.* 1 déc 2017;45:417-25.
40. McGrath-Morrow SA, Hayashi M, Aherrera A, Lopez A, Malinina A, Collaco JM, et al. The effects of electronic cigarette emissions on systemic cotinine levels, weight and postnatal lung growth in neonatal mice. *PLoS One.* 2015;10(2):e0118344.
41. Shi H, Fan X, Horton A, Haller ST, Kennedy DJ, Schiefer IT, et al. The Effect of Electronic-Cigarette Vaping on Cardiac Function and Angiogenesis in Mice. *Sci Rep.* 11 mars 2019;9(1):4085.
42. Platel A, Dusautoir R, Kervoaze G, Dourdin G, Gateau E, Talahari S, et al. Comparison of the in vivo genotoxicity of electronic and conventional cigarettes aerosols after subacute, subchronic and chronic exposures. *J Hazard Mater.* 5 févr 2022;423(Pt B):127246.
43. Vardavas CI, Anagnostopoulos N, Kougias M, Evangelopoulou V, Connolly GN, Behrakis PK. Short-term pulmonary effects of using an electronic cigarette: impact on respiratory flow resistance, impedance, and exhaled nitric oxide. *Chest.* juin 2012;141(6):1400-6.
44. Meo SA, Ansary MA, Barayan FR, Almusallam AS, Almehaid AM, Alarifi NS, et al. Electronic Cigarettes: Impact on Lung Function and Fractional Exhaled Nitric Oxide Among Healthy Adults. *Am J Mens Health.* 2019;13(1):1557988318806073.
45. Moheimani RS, Bhetraratana M, Yin F, Peters KM, Gornbein J, Araujo JA, et al. Increased Cardiac Sympathetic Activity and Oxidative Stress in Habitual Electronic Cigarette Users: Implications for Cardiovascular Risk. *JAMA Cardiol.* 1 mars 2017;2(3):278-84.
46. Chatterjee S, Tao JQ, Johncola A, Guo W, Caporale A, Langham MC, et al. Acute exposure to e-cigarettes causes inflammation and pulmonary endothelial oxidative stress in nonsmoking, healthy young subjects. *Am J Physiol Lung Cell Mol Physiol.* 1 août 2019;317(2):L155-66.
47. Winnicka L, Shenoy MA. EVALI and the Pulmonary Toxicity of Electronic

- Cigarettes: A Review. *J Gen Intern Med.* juill 2020;35(7):2130-5.
48. Barna S, Rózsa D, Varga J, Fodor A, Szilasi M, Galuska L, et al. First comparative results about the direct effect of traditional cigarette and e-cigarette smoking on lung alveolocapillary membrane using dynamic ventilation scintigraphy. *Nucl Med Commun.* févr 2019;40(2):153-8.
 49. Polosa R, Morjaria JB, Prosperini U, Russo C, Pennisi A, Puleo R, et al. Health effects in COPD smokers who switch to electronic cigarettes: a retrospective-prospective 3-year follow-up. *Int J Chron Obstruct Pulmon Dis.* 2018;13:2533-42.
 50. Polosa R, Morjaria JB, Caponnetto P, Caruso M, Campagna D, Amaradio MD, et al. Persisting long term benefits of smoking abstinence and reduction in asthmatic smokers who have switched to electronic cigarettes. *Discov Med.* févr 2016;21(114):99-108.
 51. World Health Organization. Regional Office for Europe. Heated tobacco products: a brief [Internet]. Copenhagen: World Health Organization. Regional Office for Europe; 2020. Disponible sur: <https://apps.who.int/iris/handle/10665/350470>
 52. Lavery AA, Vardavas CI, Filippidis FT. Prevalence and reasons for use of Heated Tobacco Products (HTP) in Europe: an analysis of Eurobarometer data in 28 countries. *Lancet Reg Health Eur.* sept 2021;8:100159.
 53. Liu X, Lugo A, Spizzichino L, Tabuchi T, Pacifici R, Gallus S. Heat-not-burn tobacco products: concerns from the Italian experience. *Tob Control.* janv 2019;28(1):113-4.
 54. Andler R, Guignard R, Richard JB, Pasquereau A, Quatremère G, Nguyen-Thanh V. Types de tabac fumés, évolutions et facteurs associés : résultats des Baromètre de Santé publique France 2017 et 2018. 2020;8.
 55. Gallus S, Lugo A, Liu X, Borroni E, Clancy L, Gorini G, et al. Use and Awareness of Heated Tobacco Products in Europe. *J Epidemiol.* 5 mars 2022;32(3):139-44.
 56. Simonavicius E, McNeill A, Shahab L, Brose LS. Heat-not-burn tobacco products: a systematic literature review. *Tob Control.* sept 2019;28(5):582-94.
 57. Szparaga M, Świercz R, Stępnik M. Review of data on chemical content in an aerosol resulting from heating a tobacco or a solution used in e-cigarettes and in the smoke generated from the reference cigarettes. *Toxicol Mech Methods.* juin 2021;31(5):323-33.
 58. Uguna CN, Snape CE. Should IQOS Emissions Be Considered as Smoke and Harmful to Health? A Review of the Chemical Evidence. *ACS Omega.* 5 juill 2022;7(26):22111-24.
 59. St Claire S, Gouda H, Schotte K, Fayokun R, Fu D, Varghese C, et al. Lung health, tobacco, and related products: gaps, challenges, new threats, and suggested research. *Am J Physiol Lung Cell Mol Physiol.* 1 mai 2020;318(5):L1004-7.
 60. Organisation mondiale de la santé. WHO report on the global tobacco epidemic

2021: addressing new and emerging products | The Union [Internet]. [cité 1 mars 2023]. Disponible sur: <https://theunion.org/implementation-hub/resource-library/who-report-on-the-global-tobacco-epidemic-2021-addressing-new-and-emerging-products>

61. ERS. ERS Position Paper on Heated Tobacco Products [Internet]. ERS - European Respiratory Society. 2020 [cité 7 mai 2023]. Disponible sur: <https://www.ersnet.org/news-and-features/news/ers-position-paper-on-heated-tobacco-products/>
62. Znyk M, Jurewicz J, Kaleta D. Exposure to Heated Tobacco Products and Adverse Health Effects, a Systematic Review. *Int J Environ Res Public Health*. 21 juin 2021;18(12):6651.
63. Leigh NJ, Tran PL, O'Connor RJ, Goniewicz ML. Cytotoxic effects of heated tobacco products (HTP) on human bronchial epithelial cells. *Tob Control*. 2018;27(Suppl 1):s26-9.
64. Walczak J, Malińska D, Drabik K, Michalska B, Prill M, John S, et al. Mitochondrial Network and Biogenesis in Response to Short and Long-Term Exposure of Human BEAS-2B Cells to Aerosol Extracts from the Tobacco Heating System 2.2. *Cell Physiol Biochem Int J Exp Cell Physiol Biochem Pharmacol*. 11 mars 2020;54(2):230-51.
65. Malinska D, Szymański J, Patalas-Krawczyk P, Michalska B, Wojtala A, Prill M, et al. Assessment of mitochondrial function following short- and long-term exposure of human bronchial epithelial cells to total particulate matter from a candidate modified-risk tobacco product and reference cigarettes. *Food Chem Toxicol Int J Publ Br Ind Biol Res Assoc*. mai 2018;115:1-12.
66. Sohal SS, Eapen MS, Naidu VGM, Sharma P. IQOS exposure impairs human airway cell homeostasis: direct comparison with traditional cigarette and e-cigarette. *ERJ Open Res*. févr 2019;5(1):00159-2018.
67. Nabavizadeh P, Liu J, Havel CM, Ibrahim S, Derakhshandeh R, Jacob lii P, et al. Vascular endothelial function is impaired by aerosol from a single IQOS HeatStick to the same extent as by cigarette smoke. *Tob Control*. 2018;27(Suppl 1):s13-9.
68. Koike S, Sato K, Sawa M, Inaba Y, Hattori K, Nakadate K, et al. Exposure to Heated Tobacco Products Aerosol Causes Acute Stress Responses in the Lung of Mouse. *Antioxid Basel Switz*. 25 nov 2022;11(12):2329.
69. Bhat TA, Kalathil SG, Leigh N, Muthumalage T, Rahman I, Goniewicz ML, et al. Acute Effects of Heated Tobacco Product (IQOS) Aerosol Inhalation on Lung Tissue Damage and Inflammatory Changes in the Lungs. *Nicotine Tob Res*. 21 déc 2020;23(7):1160-7.
70. Nitta NA, Sato T, Komura M, Yoshikawa H, Suzuki Y, Mitsui A, et al. Exposure to the heated tobacco product IQOS generates apoptosis-mediated pulmonary emphysema in murine lungs. *Am J Physiol Lung Cell Mol Physiol*. 1 mai 2022;322(5):L699-711.

71. Aokage T, Tsukahara K, Fukuda Y, Tokioka F, Taniguchi A, Naito H, et al. Heat-not-burn cigarettes induce fulminant acute eosinophilic pneumonia requiring extracorporeal membrane oxygenation. *Respir Med Case Rep.* 2019;26:87-90.
72. Kamada T, Yamashita Y, Tomioka H. Acute eosinophilic pneumonia following heat-not-burn cigarette smoking. *Respirol Case Rep.* nov 2016;4(6):e00190.
73. Lee A, Lee SY, Lee KS. The Use of Heated Tobacco Products is Associated with Asthma, Allergic Rhinitis, and Atopic Dermatitis in Korean Adolescents. *Sci Rep.* 27 nov 2019;9(1):17699.
74. Tattan-Birch H, Hartmann-Boyce J, Kock L, Simonavicius E, Brose L, Jackson S, et al. Heated tobacco products for smoking cessation and reducing smoking prevalence. *Cochrane Database Syst Rev.* 6 janv 2022;1(1):CD013790.
75. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol.* juill 2016;17(7):451-9.
76. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. *PLoS One.* 16 févr 2011;6(2):e16957.
77. Nicholson JK, Lindon JC, Holmes E. « Metabonomics »: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica Fate Foreign Compd Biol Syst.* nov 1999;29(11):1181-9.
78. Dormoy V, Massfelder T. [Medical perspectives of metabolomics: the example of renal carcinoma]. *Med Sci MS.* mai 2013;29(5):463-8.
79. Kim SJ, Kim SH, Kim JH, Hwang S, Yoo HJ. Understanding Metabolomics in Biomedical Research. *Endocrinol Metab Seoul Korea.* mars 2016;31(1):7-16.
80. Matsuda F. Technical Challenges in Mass Spectrometry-Based Metabolomics. *Mass Spectrom Tokyo Jpn.* 2016;5(2):S0052.
81. Cajka T, Fiehn O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal Chem.* 5 janv 2016;88(1):524-45.
82. Burla B, Arita M, Arita M, Bendt AK, Cazenave-Gassiot A, Dennis EA, et al. MS-based lipidomics of human blood plasma: a community-initiated position paper to develop accepted guidelines. *J Lipid Res.* oct 2018;59(10):2001-17.
83. Emwas AH, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, et al. NMR Spectroscopy for Metabolomics Research. *Metabolites.* 27 juin 2019;9(7).
84. Ohta D, Kanaya S, Suzuki H. Application of Fourier-transform ion cyclotron resonance mass spectrometry to metabolic profiling and metabolite identification. *Curr Opin Biotechnol.* févr 2010;21(1):35-44.
85. Bruno C, Patin F, Bocca C, Nadal-Desbarats L, Bonnier F, Reynier P, et al. The combination of four analytical methods to explore skeletal muscle metabolomics:

- Better coverage of metabolic pathways or a marketing argument? *J Pharm Biomed Anal.* 30 janv 2018;148:273-9.
86. Banerjee S, Mazumdar S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem.* 2012;2012:282574.
 87. Kramer G, Moerland PD, Jeeninga RE, Vlietstra WJ, Ringrose JH, Byrman C, et al. Proteomic analysis of HIV-T cell interaction: an update. *Front Microbiol.* 2012;3:240.
 88. Picache JA, Rose BS, Balinski A, Leapfrog KL, Sherrod SD, May JC, et al. Collision cross section compendium to annotate and predict multi-omic compound identities. *Chem Sci.* 23 janv 2019;10(4):983-93.
 89. Ewing MA, Glover MS, Clemmer DE. Hybrid ion mobility and mass spectrometry as a separation tool. *J Chromatogr A.* 25 mars 2016;1439:3-25.
 90. Ma X. Recent Advances in Mass Spectrometry-Based Structural Elucidation Techniques. *Molecules.* janv 2022;27(19):6466.
 91. Hernández-Mesa M, D'Atri V, Barknowitz G, Fanuel M, Pezzatti J, Dreolin N, et al. Interlaboratory and Interplatform Study of Steroids Collision Cross Section by Traveling Wave Ion Mobility Spectrometry. *Anal Chem.* 07 2020;92(7):5013-22.
 92. Hernández-Mesa M, Ropartz D, García-Campaña AM, Rogniaux H, Dervilly-Pinel G, Le Bizec B. Ion Mobility Spectrometry in Food Analysis: Principles, Current Applications and Future Trends. *Mol Basel Switz.* 25 juill 2019;24(15).
 93. Plachká K, Pezzatti J, Musenga A, Nicoli R, Kuuranne T, Rudaz S, et al. Ion mobility-high resolution mass spectrometry in doping control analysis. Part II: Comparison of acquisition modes with and without ion mobility. *Anal Chim Acta.* 29 août 2021;1175:338739.
 94. Celma A, Ahrens L, Gago-Ferrero P, Hernández F, López F, Lundqvist J, et al. The relevant role of ion mobility separation in LC-HRMS based screening strategies for contaminants of emerging concern in the aquatic environment. *Chemosphere.* oct 2021;280:130799.
 95. Gray CJ, Thomas B, Upton R, Migas LG, Evers CE, Barran PE, et al. Applications of ion mobility mass spectrometry for high throughput, high resolution glycan analysis. *Biochim Biophys Acta BBA - Gen Subj.* 1 août 2016;1860(8):1688-709.
 96. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022; Disponible sur: <https://www.R-project.org/>
 97. Di Guida R, Engel J, Allwood JW, Weber RJM, Jones MR, Sommer U, et al. Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics Off J Metabolomic Soc.* 2016;12:93.

98. Broadhurst D, Goodacre R, Reinke SN, Kuligowski J, Wilson ID, Lewis MR, et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics Off J Metabolomic Soc.* 2018;14(6):72.
99. Sangster T, Major H, Plumb R, Wilson AJ, Wilson ID. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabolomic analysis. *The Analyst.* oct 2006;131(10):1075-8.
100. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc.* 30 juin 2011;6(7):1060-83.
101. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, et al. Global metabolic profiling procedures for urine using UPLC-MS. *Nat Protoc.* juin 2010;5(6):1005-18.
102. Lewis MR, Pearce JTM, Spagou K, Green M, Dona AC, Yuen AHY, et al. Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping. *Anal Chem.* 20 sept 2016;88(18):9004-13.
103. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics.* 8 juin 2006;7:142.
104. Peluso A, Glen R, Ebbels TMD. Multiple-testing correction in metabolome-wide association studies. *BMC Bioinformatics.* 12 févr 2021;22(1):67.
105. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* 4 janv 2018;46(D1):D608-17.
106. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, Palermo A, Warth B, Hermann G, et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal Chem.* 6 mars 2018;90(5):3156-64.
107. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, et al. MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom JMS.* juill 2010;45(7):703-14.
108. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* 6 janv 2023;51(D1):D587-92.
109. Pezzatti J, Boccard J, Codesido S, Gagnebin Y, Joshi A, Picard D, et al. Implementation of liquid chromatography-high resolution mass spectrometry methods for untargeted metabolomic analyses of biological samples: A tutorial. *Anal Chim Acta.* 8 avr 2020;1105:28-44.
110. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al.

Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics Off J Metabolomic Soc.* sept 2007;3(3):211-21.

111. Jewison T, Su Y, Disfany FM, Liang Y, Knox C, Maciejewski A, et al. SMPDB 2.0: big improvements to the Small Molecule Pathway Database. *Nucleic Acids Res.* janv 2014;42(Database issue):D478-484.
112. Chagoyen M, Pazos F. Tools for the functional interpretation of metabolomic experiments. *Brief Bioinform.* nov 2013;14(6):737-44.
113. Wieder C, Frainay C, Poupin N, Rodríguez-Mier P, Vinson F, Cooke J, et al. Pathway analysis in metabolomics: Recommendations for the use of over-representation analysis. *PLoS Comput Biol.* sept 2021;17(9):e1009105.
114. Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res.* juill 2010;38(Web Server issue):W71-77.
115. Barupal DK, Fan S, Fiehn O. Integrating bioinformatics approaches for a comprehensive interpretation of metabolomics datasets. *Curr Opin Biotechnol.* déc 2018;54:1-9.
116. Fu J, Zhang LL, Li W, Zhang Y, Zhang Y, Liu F, et al. Application of metabolomics for revealing the interventional effects of functional foods on metabolic diseases. *Food Chem.* 15 janv 2022;367:130697.
117. Alseekh S, Fernie AR. Metabolomics 20 years on: what have we learned and what hurdles remain? *Plant J.* 2018;94(6):933-42.
118. Alarcon-Barrera JC, Kostidis S, Ondo-Mendez A, Giera M. Recent advances in metabolomics analysis for early drug development. *Drug Discov Today.* 1 juin 2022;27(6):1763-73.
119. Le Gouellec A, Plazy C, Toussaint B. What clinical metabolomics will bring to the medicine of tomorrow. *Front Anal Sci [Internet].* 2023 [cité 3 avr 2023];3. Disponible sur: <https://www.frontiersin.org/articles/10.3389/frans.2023.1142606>
120. Lenski M, Bruno C, Darrouzain F, Allorge D. Métabolomique : principes et applications en toxicologie biologique et médicolégale. *Toxicol Anal Clin [Internet].* 8 févr 2023; Disponible sur: <https://www.sciencedirect.com/science/article/pii/S2352007823000021>
121. Wei S, Wei Y, Gong Y, Chen Y, Cui J, Li L, et al. Metabolomics as a valid analytical technique in environmental exposure research: application and progress. *Metabolomics Off J Metabolomic Soc.* 31 mai 2022;18(6):35.
122. Xu T, Holzappel C, Dong X, Bader E, Yu Z, Prehn C, et al. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med.* 4 mars 2013;11:60.
123. Goettel M, Niessner R, Pluym N, Scherer G, Scherer M. A fully validated GC-

- TOF-MS method for the quantification of fatty acids revealed alterations in the metabolic profile of fatty acids after smoking cessation. *J Chromatogr B Analyt Technol Biomed Life Sci.* 15 janv 2017;1041-1042:141-50.
124. Goettel M, Niessner R, Mueller D, Scherer M, Scherer G, Pluym N. Metabolomic Fingerprinting in Various Body Fluids of a Diet-Controlled Clinical Smoking Cessation Study Using a Validated GC-TOF-MS Metabolomics Platform. *J Proteome Res.* 06 2017;16(10):3491-503.
 125. Ishikawa S, Matsumura K, Kitamura N, Takanami Y, Ito S. Multi-omics analysis: Repeated exposure of a 3D bronchial tissue culture to whole-cigarette smoke. *Toxicol Vitro Int J Publ Assoc BIBRA.* févr 2019;54:251-62.
 126. Vulimiri SV, Misra M, Hamm JT, Mitchell M, Berger A. Effects of mainstream cigarette smoke on the global metabolome of human lung epithelial cells. *Chem Res Toxicol.* 16 mars 2009;22(3):492-503.
 127. Aug A, Altraja A, Altraja S, Laaniste L, Mahlapuu R, Soomets U, et al. Alterations of bronchial epithelial metabolome by cigarette smoke are reversible by an antioxidant, O-methyl-L-tyrosinyl- γ -L-glutamyl-L-cysteinylglycine. *Am J Respir Cell Mol Biol.* oct 2014;51(4):586-94.
 128. Hsu PC, Zhou B, Zhao Y, Resson HW, Cheema AK, Pickworth W, et al. Feasibility of identifying the tobacco-related global metabolome in blood by UPLC-QTOF-MS. *J Proteome Res.* 1 févr 2013;12(2):679-91.
 129. Hsu PC, Lan RS, Brasky TM, Marian C, Cheema AK, Resson HW, et al. Metabolomic profiles of current cigarette smokers. *Mol Carcinog.* 2017;56(2):594-606.
 130. Gu F, Derkach A, Freedman ND, Landi MT, Albanes D, Weinstein SJ, et al. Cigarette smoking behaviour and blood metabolomics. *Int J Epidemiol.* 2016;45(5):1421-32.
 131. Goettel M, Niessner R, Scherer M, Scherer G, Pluym N. Analysis of Urinary Eicosanoids by LC-MS/MS Reveals Alterations in the Metabolic Profile after Smoking Cessation. *Chem Res Toxicol.* 19 2018;31(3):176-82.
 132. Peralbo-Molina A, Calderón-Santiago M, Jurado-Gómez B, Luque de Castro MD, Priego-Capote F. Exhaled breath condensate to discriminate individuals with different smoking habits by GC-TOF/MS. *Sci Rep.* 03 2017;7(1):1421.
 133. Rushing BR, Tilley S, Molina S, Schroder M, Sumner S. Commonalities in Metabolic Reprogramming between Tobacco Use and Oral Cancer. *Int J Environ Res Public Health.* 18 août 2022;19(16):10261.
 134. Aug A, Altraja S, Kilk K, Porosk R, Soomets U, Altraja A. E-Cigarette Affects the Metabolome of Primary Normal Human Bronchial Epithelial Cells. *PloS One.* 2015;10(11):e0142053.
 135. Ren X, Lin L, Sun Q, Li T, Sun M, Sun Z, et al. Metabolomics-based safety evaluation of acute exposure to electronic cigarettes in mice. *Sci Total Environ.* 15

sept 2022;839:156392.

136. Madison MC, Landers CT, Gu BH, Chang CY, Tung HY, You R, et al. Electronic cigarettes disrupt lung lipid homeostasis and innate immunity independent of nicotine. *J Clin Invest.* 1 oct 2019;129(10):4290-304.
137. Chaumont M, Tagliatti V, Channan EM, Colet JM, Bernard A, Morra S, et al. Short halt in vaping modifies cardiorespiratory parameters and urine metabolome: a randomized trial. *Am J Physiol Lung Cell Mol Physiol.* 01 2020;318(2):L331-44.
138. Hsiao YC, Matulewicz RS, Sherman SE, Jaspers I, Weitzman ML, Gordon T, et al. Untargeted Metabolomics to Characterize the Urinary Chemical Landscape of E-Cigarette Users. *Chem Res Toxicol.* 13 mars 2023;
139. Wang Q, Ji X, Rahman I. Dysregulated Metabolites Serve as Novel Biomarkers for Metabolic Diseases Caused by E-Cigarette Vaping and Cigarette Smoking. *Metabolites.* 29 mai 2021;11(6):345.
140. Alqahtani S, Cooper B, Spears CA, Wright C, Shannahan J. Electronic nicotine delivery system-induced alterations in oral health via saliva assessment. *Exp Biol Med Maywood NJ.* sept 2020;245(15):1319-25.
141. Moshensky A, Du M, Shin J, Advani I, Gunge D, Mathew D, et al. Vaping-induced metabolomic signatures in the circulation of mice are driven by device type, e-liquid, exposure duration and sex. *ERJ Open Res.* juill 2021;7(3):00229-2021.
142. Smith MR, Jarrell ZR, Orr M, Liu KH, Go YM, Jones DP. Metabolome-wide association study of flavorant vanillin exposure in bronchial epithelial cells reveals disease-related perturbations in metabolism. *Environ Int.* févr 2021;147:106323.
143. Jarrell ZR, Smith MR, He X, Orr M, Jones DP, Go YM. Firsthand and Secondhand Exposure Levels of Maltol-Flavored Electronic Nicotine Delivery System Vapors Disrupt Amino Acid Metabolism. *Toxicol Sci Off J Soc Toxicol.* 16 juill 2021;182(1):70-81.
144. Liu G, Lin CJ, Yates CR, Prasad GL. Metabolomics analysis identified reduced levels of xenobiotics, oxidative stress, and improved vitamin metabolism in smokers switched to Vuse Electronic Nicotine Delivery System. *Nicotine Tob Res Off J Soc Res Nicotine Tob.* 9 nov 2020;
145. Zanetti F, Titz B, Sewer A, Lo Sasso G, Scotti E, Schlage WK, et al. Comparative systems toxicology analysis of cigarette smoke and aerosol from a candidate modified risk tobacco product in organotypic human gingival epithelial cultures: A 3-day repeated exposure study. *Food Chem Toxicol.* 1 mars 2017;101:15-35.
146. Titz B, Boué S, Phillips B, Talikka M, Vihervaara T, Schneider T, et al. Effects of Cigarette Smoke, Cessation, and Switching to Two Heat-Not-Burn Tobacco Products on Lung Lipid Metabolism in C57BL/6 and Apoe^{-/-} Mice-An Integrative Systems Toxicology Analysis. *Toxicol Sci Off J Soc Toxicol.* févr 2016;149(2):441-57.

147. Titz B, Szostak J, Sewer A, Phillips B, Nury C, Schneider T, et al. Multi-omics systems toxicology study of mouse lung assessing the effects of aerosols from two heat-not-burn tobacco products and cigarette smoke. *Comput Struct Biotechnol J*. 2020;18:1056-73.
148. Beauval N, Antherieu S, Soyez M, Gengler N, Grova N, Howsam M, et al. Chemical Evaluation of Electronic Cigarettes: Multicomponent Analysis of Liquid Refills and their Corresponding Aerosols. *J Anal Toxicol*. 1 oct 2017;41(8):670-8.
149. Pezzatti J, González-Ruiz V, Codesido S, Gagnebin Y, Joshi A, Guillarme D, et al. A scoring approach for multi-platform acquisition in metabolomics. *J Chromatogr A*. 10 mai 2019;1592:47-54.
150. Ross DH, Cho JH, Xu L. Breaking Down Structural Diversity for Comprehensive Prediction of Ion-Neutral Collision Cross Sections. *Anal Chem*. 17 mars 2020;92(6):4548-57.
151. Wen Y, Amos RIJ, Talebi M, Szucs R, Dolan JW, Pohl CA, et al. Retention Index Prediction Using Quantitative Structure-Retention Relationships for Improving Structure Identification in Nontargeted Metabolomics. *Anal Chem*. 7 août 2018;90(15):9434-40.
152. Zhou Z, Shen X, Tu J, Zhu ZJ. Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry. *Anal Chem*. 15 nov 2016;88(22):11084-91.
153. Zhou Z, Tu J, Zhu ZJ. Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. *Curr Opin Chem Biol*. 1 févr 2018;42:34-41.
154. Zhou Z, Luo M, Chen X, Yin Y, Xiong X, Wang R, et al. Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nat Commun*. 28 août 2020;11(1):4334.
155. Rainey MA, Watson CA, Asef CK, Foster MR, Baker ES, Fernández FM. CCS Predictor 2.0: An Open-Source Jupyter Notebook Tool for Filtering Out False Positives in Metabolomics. *Anal Chem*. 20 déc 2022;94(50):17456-66.
156. Plante PL, Francovic-Fontaine É, May JC, McLean JA, Baker ES, Laviolette F, et al. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. *Anal Chem*. 16 avr 2019;91(8):5191-9.
157. Zhou Z, Xiong X, Zhu ZJ. MetCCS predictor: a web server for predicting collision cross-section values of metabolites in ion mobility-mass spectrometry based metabolomics. *Bioinforma Oxf Engl*. 15 juill 2017;33(14):2235-7.
158. Zhou Z, Tu J, Xiong X, Shen X, Zhu ZJ. LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision To Support Ion Mobility-Mass Spectrometry-Based Lipidomics. *Anal Chem*. 5 sept 2017;89(17):9559-66.
159. Blaženović I, Shen T, Mehta SS, Kind T, Ji J, Piparo M, et al. Increasing Compound Identification Rates in Untargeted Lipidomics Research with Liquid

Chromatography Drift Time-Ion Mobility Mass Spectrometry. *Anal Chem.* 18 sept 2018;90(18):10758-64.

160. Grisoni F, Ballabio D, Todeschini R, Consonni V. Molecular Descriptors for Structure-Activity Applications: A Hands-On Approach. *Methods Mol Biol Clifton NJ.* 2018;1800:3-53.
161. Stanstrup J, Neumann S, Vrhovšek U. PredRet: prediction of retention time by direct mapping between multiple chromatographic systems. *Anal Chem.* 15 sept 2015;87(18):9421-8.
162. Falchi F, Bertozzi SM, Ottonello G, Ruda GF, Colombano G, Fiorelli C, et al. Kernel-Based, Partial Least Squares Quantitative Structure-Retention Relationship Model for UPLC Retention Time Prediction: A Useful Tool for Metabolite Identification. *Anal Chem.* 4 oct 2016;88(19):9510-7.
163. Creek DJ, Jankevics A, Breitling R, Watson DG, Barrett MP, Burgess KEV. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Anal Chem.* 15 nov 2011;83(22):8703-10.
164. Bonini P, Kind T, Tsugawa H, Barupal DK, Fiehn O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. *Anal Chem.* 2 juin 2020;92(11):7515-22.
165. Liapikos T, Zisi C, Kodra D, Kademoglou K, Diamantidou D, Begou O, et al. Quantitative structure retention relationship (QSRR) modelling for Analytes' retention prediction in LC-HRMS by applying different Machine Learning algorithms and evaluating their performance. *J Chromatogr B.* 15 févr 2022;1191:123132.
166. Gritti F. Perspective on the Future Approaches to Predict Retention in Liquid Chromatography. *Anal Chem.* 13 avr 2021;93(14):5653-64.
167. Naylor BC, Catrow JL, Maschek JA, Cox JE. QSRR Automator: A Tool for Automating Retention Time Prediction in Lipidomics and Metabolomics. *Metabolites.* 9 juin 2020;10(6):E237.
168. Dunn WB, Erban A, Weber RJM, Creek DJ, Brown M, Breitling R, et al. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics.* 1 mars 2013;9(1):44-66.
169. Bittremieux W, Wang M, Dorrestein PC. The critical role that spectral libraries play in capturing the metabolomics community knowledge. *Metabolomics Off J Metabolomic Soc.* 19 nov 2022;18(12):94.
170. Park SH, De Pra M, Haddad PR, Grosse S, Pohl CA, Steiner F. Localised quantitative structure-retention relationship modelling for rapid method development in reversed-phase high performance liquid chromatography. *J Chromatogr A.* 4 janv 2020;1609:460508.
171. Tsugawa H, Ikeda K, Takahashi M, Satoh A, Mori Y, Uchino H, et al. A lipidome atlas in MS-DIAL 4. *Nat Biotechnol.* oct 2020;38(10):1159-63.

172. Mollerup CB, Mardal M, Dalsgaard PW, Linnet K, Barron LP. Prediction of collision cross section and retention time for broad scope screening in gradient reversed-phase liquid chromatography-ion mobility-high resolution accurate mass spectrometry. *J Chromatogr A*. 23 mars 2018;1542:82-8.
173. Celma A, Bade R, Sancho JV, Hernandez F, Humphries M, Bijlsma L. Prediction of Retention Time and Collision Cross Section (CCSH+, CCSH-, and CCSNa+) of Emerging Contaminants Using Multiple Adaptive Regression Splines. *J Chem Inf Model*. 28 nov 2022;62(22):5425-34.
174. Ross DH, Cho JH, Zhang R, Hines KM, Xu L. LiPydomics: A Python Package for Comprehensive Prediction of Lipid Collision Cross Sections and Retention Times and Analysis of Ion Mobility-Mass Spectrometry-Based Lipidomics Data. *Anal Chem*. 17 nov 2020;92(22):14967-75.
175. Lenski M, Maallem S, Zarcone G, Garçon G, Lo-Guidice JM, Anthérieu S, et al. Prediction of a Large-Scale Database of Collision Cross-Section and Retention Time Using Machine Learning to Reduce False Positive Annotations in Untargeted Metabolomics. *Metabolites*. 15 févr 2023;13(2):282.
176. Zarcone G, Lenski M, Martinez T, Garçon G, Allorge D, Nesslany F, et al. Comparison of toxicity and transcriptomic effects of emissions from heated tobacco products, electronic cigarettes and tobacco cigarettes in human bronchial epithelial cells. *Toxicol Lett*. 1 sept 2021;350:S124-5.
177. World Health Organization, WHO Tobacco Free Initiative. Standard operating procedure for intense smoking of cigarettes [Internet]. WHO TobLabNet official method SOP 01. Geneva: World Health Organization; 2012 [cité 11 avr 2023]. 7 p. Disponible sur: <https://apps.who.int/iris/handle/10665/75261>
178. Ballman KV, Grill DE, Oberg AL, Therneau TM. Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinforma Oxf Engl*. 1 nov 2004;20(16):2778-86.
179. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. juin 2016;15(2):155-63.
180. consortium M. MassBank/MassBank-data: Release version 2022.12 [Internet]. Zenodo; 2022 [cité 23 avr 2023]. Disponible sur: <https://zenodo.org/record/7436494>
181. Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminformatics*. 2016;8:61.
182. Wu Y, Li L. Sample normalization methods in quantitative metabolomics. *J Chromatogr A*. 22 janv 2016;1430:80-95.
183. Stabbert R, Schäfer KH, Biefel C, Rustemeier K. Analysis of aromatic amines in cigarette smoke. *Rapid Commun Mass Spectrom RCM*. 2003;17(18):2125-32.
184. Bie Z, Lu W, Zhu Y, Chen Y, Ren H, Ji L. Rapid determination of six carcinogenic primary aromatic amines in mainstream cigarette smoke by two-dimensional online

solid phase extraction combined with liquid chromatography tandem mass spectrometry. *J Chromatogr A*. 27 janv 2017;1482:39-47.

185. Yu J, Wang S, Zhao G, Wang B, Ding L, Zhang X, et al. Determination of urinary aromatic amines in smokers and nonsmokers using a MIPs-SPE coupled with LC-MS/MS method. *J Chromatogr B Analyt Technol Biomed Life Sci*. 1 mai 2014;958:130-5.
186. Vivarelli F, Canistro D, Cirillo S, Elias RJ, Granata S, Mussoni M, et al. Unburned Tobacco Cigarette Smoke Alters Rat Ultrastructural Lung Airways and DNA. *Nicotine Tob Res Off J Soc Res Nicotine Tob*. 5 nov 2021;23(12):2127-34.
187. Nichols WK, Mehta R, Skordos K, Macé K, Pfeifer AMA, Carr BA, et al. 3-methylindole-induced toxicity to human bronchial epithelial cell lines. *Toxicol Sci Off J Soc Toxicol*. févr 2003;71(2):229-36.
188. Weems JM, Lamb JG, D'Agostino J, Ding X, Yost GS. Potent mutagenicity of 3-methylindole requires pulmonary cytochrome P450-mediated bioactivation: a comparison to the prototype cigarette smoke mutagens B(a)P and NNK. *Chem Res Toxicol*. 15 nov 2010;23(11):1682-90.
189. Weems JM, Cutler NS, Moore C, Nichols WK, Martin D, Makin E, et al. 3-Methylindole is mutagenic and a possible pulmonary carcinogen. *Toxicol Sci Off J Soc Toxicol*. nov 2009;112(1):59-67.
190. Pfau W, Skog K. Exposure to beta-carbolines norharman and harman. *J Chromatogr B Analyt Technol Biomed Life Sci*. 25 mars 2004;802(1):115-26.
191. Harris AC, Muelken P, LeSage MG. β -Carbolines found in cigarette smoke elevate intracranial self-stimulation thresholds in rats. *Pharmacol Biochem Behav*. nov 2020;198:173041.
192. Celikyurt IK, Utkan T, Gocmez SS, Hudson A, Aricioglu F. Effect of harmane, an endogenous β -carboline, on learning and memory in rats. *Pharmacol Biochem Behav*. janv 2013;103(3):666-71.
193. Abelaira HM, Réus GZ, Scaini G, Streck EL, Crippa JA, Quevedo J. β -Carboline harmine reverses the effects induced by stress on behaviour and citrate synthase activity in the rat prefrontal cortex. *Acta Neuropsychiatr*. déc 2013;25(6):328-33.
194. Aricioglu F, Altunbas H. Harmane induces anxiolysis and antidepressant-like effects in rats. *Ann N Y Acad Sci*. déc 2003;1009:196-201.
195. Li F, Liu Q, Cai W, Shao X. Analysis of Scopoletin and Caffeic Acid in Tobacco by GC-MS After a Rapid Derivatization Procedure. *Chromatographia*. 1 avr 2009;69(7):743-8.
196. Abraham NG, Kappas A. Pharmacological and clinical aspects of heme oxygenase. *Pharmacol Rev*. mars 2008;60(1):79-127.
197. Even B, Fayad-Kobeissi S, Gagliolo JM, Motterlini R, Boczkowski J, Foresti R, et al. Heme oxygenase-1 induction attenuates senescence in chronic obstructive

- pulmonary disease lung fibroblasts by protecting against mitochondria dysfunction. *Aging Cell*. déc 2018;17(6):e12837.
198. Song Y, Li R, Zhang Y, Wei J, Chen W, Chung CKA, et al. Mass spectrometry-based metabolomics reveals the mechanism of ambient fine particulate matter and its components on energy metabolic reprogramming in BEAS-2B cells. *Sci Total Environ*. 15 févr 2019;651:3139-50.
 199. Ma J, Zhong M, Xiong Y, Gao Z, Wu Z, Liu Y, et al. Emerging roles of nucleotide metabolism in cancer development: progress and prospect. *Aging*. 5 mai 2021;13(9):13349-58.
 200. De Vitto H, Arachchige DB, Richardson BC, French JB. The Intersection of Purine and Mitochondrial Metabolism in Cancer. *Cells*. 30 sept 2021;10(10):2603.
 201. Drazic A, Winter J. The physiological role of reversible methionine oxidation. *Biochim Biophys Acta*. août 2014;1844(8):1367-82.
 202. Solanki HS, Babu N, Jain AP, Bhat MY, Puttamallesh VN, Advani J, et al. Cigarette smoke induces mitochondrial metabolic reprogramming in lung cells. *Mitochondrion*. mai 2018;40:58-70.
 203. McElroy JP, Carmella SG, Heskin AK, Tang MK, Murphy SE, Reisinger SA, et al. Effects of cessation of cigarette smoking on eicosanoid biomarkers of inflammation and oxidative damage. *PloS One*. 2019;14(6):e0218386.
 204. Lundström SL, Balgoma D, Wheelock ÅM, Haeggström JZ, Dahlén SE, Wheelock CE. Lipid mediator profiling in pulmonary disease. *Curr Pharm Biotechnol*. juill 2011;12(7):1026-52.
 205. Huang RY, Chen GG. Cigarette smoking, cyclooxygenase-2 pathway and cancer. *Biochim Biophys Acta*. avr 2011;1815(2):158-69.
 206. Engel KM, Schiller J, Galuska CE, Fuchs B. Phospholipases and Reactive Oxygen Species Derived Lipid Biomarkers in Healthy and Diseased Humans and Animals - A Focus on Lysophosphatidylcholine. *Front Physiol*. 2021;12:732319.
 207. Jain V. Role of Polyamines in Asthma Pathophysiology. *Med Sci Basel Switz*. 6 janv 2018;6(1):4.
 208. Barnes S, Benton HP, Casazza K, Cooper SJ, Cui X, Du X, et al. Training in metabolomics research. II. Processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. *J Mass Spectrom JMS*. août 2016;51(8):535-48.
 209. Rathahao-Paris E, Alves S, Junot C, Tabet JC. High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics*. 14 nov 2015;12(1):10.
 210. Ivanisevic J, Want EJ. From Samples to Insights into Metabolism: Uncovering Biologically Relevant Information in LC-HRMS Metabolomics Data. *Metabolites*. 17 déc 2019;9(12).

211. Perez De Souza L, Alseekh S, Brotman Y, Fernie AR. Network-based strategies in metabolomics data analysis and interpretation: from molecular networking to biological interpretation. *Expert Rev Proteomics*. avr 2020;17(4):243-55.
212. Gika H, Virgiliou C, Theodoridis G, Plumb RS, Wilson ID. Untargeted LC/MS-based metabolic phenotyping (metabonomics/metabolomics): The state of the art. *J Chromatogr B Analyt Technol Biomed Life Sci*. 1 juin 2019;1117:136-47.

ANNEXE

Annexe : Résultats des analyses statistiques et de l'identification des composés discriminants pour les 4 conditions testées

ESI	Pic	ID	Nom du composé	Identifiant HMDB	Score VIP (PLS-DA)	HTP		3R4F		MB-18W		MB-30W	
						ANOVA p-value	Log2(FC) TO vs T1						
NEG	0.78_472.9154m/z	4	Non identifié	-	1,2	0,0148	-1,0	-1,3	NS	NS	NS	NS	NS
NEG	0.79_114.9875m/z	4	Non identifié	-	1,0	0,0447	-0,6	-0,6	NS	NS	NS	NS	NS
NEG	0.79_500.9167m/z	4	Non identifié	-	1,1	0,0140	-1,0	-1,3	NS	NS	NS	NS	NS
NEG	0.80_272.9572m/z	4	Non identifié	-	1,3	0,0391	-0,4	-0,6	NS	NS	NS	NS	NS
NEG	0.91_405.1395m/z	4	Non identifié	-	1,4	0,0326	-0,7	-0,5	NS	NS	NS	NS	NS
NEG	0.93_565.0441m/z	2	Uridine diphosphate glucose	HMDB0000286	1,3	0,0005	-1,4	-1,8	NS	NS	NS	NS	NS
NEG	0.94_578.2861m/z	4	Non identifié	-	1,4	0,0192	-1,1	-1,2	NS	NS	NS	NS	NS
NEG	0.97_607.0776n	2	Uridine diphosphate-N-acetylglucosamine	HMDB0000290	1,1	0,0016	-1,1	-1,2	NS	NS	NS	NS	NS
NEG	0.97_628.0517m/z	2	Uridine diphosphate-N-acetylglucosamine	HMDB0000290	1,0	0,0068	-2,3	-2,0	NS	NS	NS	NS	NS
NEG	0.99_606.2042m/z	4	Non identifié	-	1,2	0,0230	-1,6	-2,1	NS	NS	NS	NS	NS
NEG	0.99_606.2934m/z	4	Non identifié	-	1,1	0,0130	-1,7	-1,9	NS	NS	NS	NS	NS
NEG	1.00_506.9926n	2	dGTP	HMDB0001440	1,4	0,0058	-0,6	-1,0	NS	NS	NS	NS	NS
NEG	1.00_572.1473n	4	Non identifié	-	1,2	0,0166	-1,4	-1,7	NS	NS	NS	NS	NS
NEG	1.00_606.1479m/z	4	Non identifié	-	1,3	0,0124	-1,5	-1,8	NS	NS	NS	NS	NS
NEG	1.00_482.9586m/z	1	Uridine triphosphate	HMDB00285	1,4	0,0094	-0,8	-1,2	NS	NS	NS	NS	NS
NEG	1.04_191.0547336m/z	1	Quinic acid	HMDB03072	1,7	0,0005	2,8	3,1	NS	NS	NS	NS	NS
NEG	1.08_427.0267n	1	Adenosine 3',5'-diphosphate	HMDB00061	1,5	0,0097	1,2	1,5	NS	NS	NS	NS	NS
NEG	1.08_604.0656m/z	1	Guanosine diphosphate mannose	HMDB01163	1,3	0,0042	-1,3	-1,5	NS	NS	NS	NS	NS
NEG	1.09_429.0553m/z	2	CMP-2-aminoethylphosphonate	HMDB0060067	0,9	0,0112	-0,7	-1,3	NS	NS	NS	NS	NS
NEG	1.10_275.0200m/z	4	Non identifié	-	1,3	0,0367	NS	-2,7	NS	NS	NS	NS	NS
NEG	1.10_333.0247m/z	4	Non identifié	-	1,3	0,0392	NS	-3,1	NS	NS	NS	NS	NS
NEG	1.10_405.0264m/z	4	Non identifié	-	1,1	0,0008	-1,1	-2,1	NS	NS	NS	NS	NS
NEG	1.59_742.0631m/z	1	NADP	HMDB00217	1,2	0,0242	-0,4	-0,7	NS	NS	NS	NS	NS
NEG	1.60_347.0374m/z	1	Inosine 5'-monophosphate	HMDB00175	1,6	0,0379	3,7	3,7	NS	NS	NS	NS	NS
NEG	1.60_384.0518n	4	Non identifié	-	0,6	0,0012	-0,9	-1,3	NS	NS	NS	NS	NS
NEG	1.61_620.0151m/z	4	Non identifié	-	1,0	0,0437	-0,5	-0,9	NS	NS	NS	NS	NS
NEG	1.63_319.0746m/z	4	Non identifié	-	1,0	0,0147	-2,3	-1,9	NS	NS	NS	NS	NS
NEG	1.64_148.0420m/z	1	Methionine	HMDB00696	1,3	0,0028	-1,2	-1,3	NS	NS	NS	NS	NS
NEG	1.64_216.0295m/z	4	Non identifié	-	1,1	0,0175	-1,2	-1,0	NS	NS	NS	NS	NS
NEG	2.34_240.0499m/z	4	Non identifié	-	1,4	0,0026	-1,2	-1,4	NS	NS	NS	NS	NS
NEG	2.34_379.9976m/z	4	Non identifié	-	1,3	0,0126	-0,9	-1,1	NS	NS	NS	NS	NS
NEG	2.35_321.0676n	2	Beta-Citryl-L-glutamic acid	HMDB0013220	1,4	0,0031	-1,0	-1,3	NS	NS	NS	NS	NS
NEG	2.96_612.1481n	2	Oxidized glutathione	HMDB0003337	1,7	NS	NS	NS	0,0447	NS	1,2	NS	NS

ESI	Pic	ID	Nom du composé	Identifiant HMDB	Score VIP (PLS-DA)	HTP		3R4F			MB-18W			MB-30W		
						ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1
NEG	7.39_197.0399m/z	4	Non identifié	-	1,2	0,0011	-1,6	-1,6	NS	NS	NS	NS	NS	NS	NS	NS
NEG	7.40_129.0540m/z	4	Non identifié	-	1,3	0,0005	-1,5	-1,5	NS	NS	NS	NS	NS	NS	NS	NS
NEG	8.28_68.6490m/z	4	Non identifié	-	1,1	0,0326	-0,5	-1,2	NS	NS	NS	NS	NS	NS	NS	NS
NEG	8.29_65.9652m/z	4	Non identifié	-	0,7	0,0217	-0,4	-0,6	NS	NS	NS	NS	NS	NS	NS	NS
NEG	8.29_66.0318m/z	4	Non identifié	-	0,9	0,0108	-0,4	-0,7	NS	NS	NS	NS	NS	NS	NS	NS
NEG	9.38_176.0102m/z	4	Non identifié	-	1,7	NS	NS	NS	0,0277	6,0	8,6	NS	NS	NS	NS	NS
NEG	9.38_353.0075m/z	4	Non identifié	-	1,9	0,0016	8,7	9,1	0,0385	5,7	7,7	NS	NS	NS	NS	NS
NEG	13.88_512.2963m/z	4	Non identifié	-	1,5	0,0095	1,5	1,8	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.10_545.2712n	4	Non identifié	-	1,7	0,0389	3,2	3,8	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.11_498.2602m/z	2	LPE 20:5	HMDB0011519	1,8	0,0200	2,4	2,7	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.20_351.1825m/z	4	Non identifié	-	1,2	0,0453	-1,1	-2,3	0,0276	NS	-1,6	NS	NS	NS	NS	NS
NEG	14.27_538.3115m/z	4	Non identifié	-	1,6	0,0033	1,3	1,8	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.57_439.2346m/z	4	Non identifié	-	1,1	0,0186	-1,3	-1,8	0,0258	-0,7	-1,6	NS	NS	NS	NS	NS
NEG	14.72_564.3259m/z	4	Non identifié	-	1,6	0,0042	2,7	3,6	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.75_588.3279m/z	4	Non identifié	-	1,8	0,0007	3,8	4,9	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.98_526.2911m/z	2	LPE 22:5	HMDB0011525	1,8	0,0146	2,5	2,9	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.98_573.3028n	4	Non identifié	-	1,7	0,0326	2,7	3,2	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.99_526.3259m/z	4	Non identifié	-	1,6	0,0242	1,5	1,9	NS	NS	NS	NS	NS	NS	NS	NS
NEG	14.99_564.3277m/z	4	Non identifié	-	1,5	0,0453	1,4	2,0	NS	NS	NS	NS	NS	NS	NS	NS
NEG	15.34_540.3697m/z	4	Non identifié	-	1,4	0,0249	0,5	0,8	NS	NS	NS	NS	NS	NS	NS	NS
NEG	15.46_566.3853m/z	4	Non identifié	-	1,6	0,0160	1,0	1,3	NS	NS	NS	NS	NS	NS	NS	NS
NEG	15.46_590.3428m/z	4	Non identifié	-	1,8	0,0007	2,0	2,9	NS	NS	NS	NS	NS	NS	NS	NS
NEG	15.47_506.3213m/z	2	LPE 20:1	HMDB0011512	1,6	0,0450	0,9	1,4	NS	NS	NS	NS	NS	NS	NS	NS
NEG	15.48_429.2980m/z	4	Non identifié	-	1,8	0,0220	-7,9	-8,4	0,0402	-6,3	-8,3	NS	NS	NS	NS	NS
NEG	15.69_506.3213m/z	2	LPE 20:1	HMDB0011512	1,5	0,0308	1,0	1,9	NS	NS	NS	NS	NS	NS	NS	NS
NEG	16.24_355.2701n	4	Non identifié	-	1,7	0,0012	16,0	17,0	0,0011	18,8	20,2	NS	NS	NS	NS	NS
NEG	16.40_618.3766m/z	4	Non identifié	-	1,8	0,0018	4,4	5,4	NS	NS	NS	NS	NS	NS	NS	NS
NEG	16.55_368.2779m/z	4	Non identifié	-	1,8	0,0093	10,4	11,4	0,0258	6,9	8,8	NS	NS	NS	NS	NS
NEG	16.63_325.2599n	4	Non identifié	-	1,7	0,0000	10,7	11,7	0,0258	10,1	11,8	NS	NS	NS	NS	NS
NEG	16.86_357.2858n	4	Non identifié	-	1,9	0,0000	9,3	10,3	0,0022	7,9	9,2	NS	NS	NS	NS	NS
NEG	17.25_271.2265m/z	4	Non identifié	-	1,5	0,0260	1,2	2,1	NS	NS	NS	NS	NS	NS	NS	NS
NEG	17.26_327.2753n	4	Non identifié	-	1,9	0,0002	10,6	11,3	0,0013	5,6	7,8	NS	NS	NS	NS	NS

ESI	Pic	ID	Nom du composé	Identifiant HMDB	Score VIP (PLS-DA)	HTP		3R4F		MB-18W		MB-30W	
						ANOVA p-value	Log2(FC) TO vs T1						
POS	0.88_168.0507n	4	Non identifié	-	1,1	0,0108	-2,3	-2,0	NS	NS	NS	NS	NS
POS	0.89_451.1205m/z	4	Non identifié	-	0,9	0,0345	-1,1	-0,9	NS	NS	NS	NS	NS
POS	0.96_629.0623n	4	Non identifié	-	1,0	0,0009	-1,0	-1,1	NS	NS	NS	NS	NS
POS	0.96_404.0019n	2	Uridine 5'-diphosphate	HMDB0000295	0,8	0,0392	-1,4	-1,3	NS	NS	NS	NS	NS
POS	0.98_489.1138m/z	1	Cytidine 5' diphosphocholine	HMDB0001413	1,7	0,0057	5,6	7,0	NS	NS	NS	NS	NS
POS	0.99_506.9954n	2	dGTP	HMDB0001440	1,5	0,0123	-0,7	-1,1	NS	NS	NS	NS	NS
POS	1.03_623.0651m/z	4	Non identifié	-	1,0	0,0326	-1,7	-2,1	NS	NS	NS	NS	NS
POS	1.04_551.9662m/z	4	Non identifié	-	1,1	0,0392	-0,5	-0,8	NS	NS	NS	NS	NS
POS	1.07_321.0689n	2	Beta-Citryl-L-glutamic acid	HMDB0013220	1,3	0,0009	-1,3	-1,5	NS	NS	NS	NS	NS
POS	1.07_192.0265n	1	Isocitric Acid	HMDB0000193	0,9	0,0042	-1,0	-1,4	NS	NS	NS	NS	NS
POS	1.07_194.0440m/z	4	Non identifié	-	1,2	0,0286	-1,6	-2,1	NS	NS	NS	NS	NS
POS	1.07_612.1507n	2	Oxidized glutathione	HMDB0003337	0,7	NS	NS	NS	0,0410	NS	1,4	NS	NS
POS	1.11_250.0931m/z	2	Deoxyguanosine	HMDB0000085	1,2	0,0115	-1,4	-1,6	NS	NS	NS	NS	NS
POS	1.11_298.0961m/z	4	Non identifié	-	1,1	0,0070	-1,7	-2,0	NS	NS	NS	NS	NS
POS	1.11_400.1463m/z	4	Non identifié	-	1,5	0,0005	-1,3	-1,6	NS	NS	NS	NS	NS
POS	1.65_179.1539m/z	4	Non identifié	-	1,4	0,0043	20,9	19,2	NS	NS	NS	NS	NS
POS	2.01_227.0902n	1	Deoxycytidine	HMDB000014	1,4	0,0034	-0,5	-1,0	NS	NS	NS	NS	NS
POS	2.29_321.0690n	2	Beta-Citryl-L-glutamic acid	HMDB0013220	1,3	0,0018	-1,3	-1,6	NS	NS	NS	NS	NS
POS	2.33_138.0291n	4	Non identifié	-	0,8	0,0091	-2,7	-2,0	NS	NS	NS	NS	NS
POS	2.50_701.0641n	4	Non identifié	-	1,1	0,0274	-1,2	-2,1	NS	NS	NS	NS	NS
POS	2.98_612.1497n	2	Oxidized glutathione	HMDB0003337	0,9	0,0427	NS	0,9	0,0204	NS	1,7	NS	NS
POS	3.63_132.0806m/z	2	3-Methylindole	HMDB0000466	2,0	0,0018	4,8	6,6	NS	NS	NS	NS	NS
POS	3.66_163.1228m/z	4	Non identifié	-	2,2	0,0005	3,8	4,4	NS	NS	NS	NS	NS
POS	3.82_229.1781n	2	N1,N8-Diacetylspermidine	HMDB00041947	1,3	0,0007	4,6	6,9	0,0345	2,8	6,5	NS	NS
POS	3.83_230.1859m/z	2	N1,N8-Diacetylspermidine	HMDB00041947	1,4	0,0012	3,2	4,9	NS	NS	NS	NS	NS
POS	4.22_221.1280m/z	2	3-Indolebutyric acid	HMDB0002096	1,9	0,0001	8,7	9,7	0,0011	6,3	8,8	NS	NS
POS	4.53_143.0734n	2	1-Naphthylamine	HMDB0243964	1,6	0,0000	15,4	15,6	0,0076	10,0	12,0	NS	NS
POS	5.39_203.0344n	4	Non identifié	-	1,1	0,0203	-1,0	-0,8	NS	NS	NS	NS	NS
POS	5.62_163.1228m/z	1	Nicotine	HMDB0001934	1,8	0,0000	16,6	17,3	0,0001	12,0	14,3	NS	NS
POS	6.54_191.1176m/z	2	N-Methylserotonin	HMDB0004369	1,7	0,0005	10,6	11,3	0,0258	6,6	9,1	NS	NS
POS	6.59_190.0840n	2	3-Hydroxyisuberic acid	HMDB0000325	1,8	0,0015	6,7	6,8	NS	NS	NS	NS	NS
POS	6.74_472.1638n	4	Non identifié	-	1,5	0,0161	-1,7	-1,6	NS	NS	NS	NS	NS

ESI	Pic	ID	Nom du composé	Identifiant HMDB	Score VIP (PLS-DA)	HTP		3R4F		MB-18W		MB-30W	
						ANOVA p-value	Log2(FC) TO vs T1						
POS	6.93_241.1334m/z	4	Non identifié	-	1,7	0,0018	9,1	9,0	0,0405	5,6	7,3	NS	NS
POS	6.95_371.1701n	4	Non identifié	-	1,6	0,0190	-2,1	-2,0	0,0258	-1,8	-1,4	NS	NS
POS	6.98_326.1454m/z	4	Non identifié	-	1,5	0,0024	-2,5	-3,0	NS	NS	NS	NS	NS
POS	7.04_189.1019m/z	4	Non identifié	-	1,7	0,0002	11,2	11,6	NS	NS	NS	NS	NS
POS	7.20_395.0171m/z	4	Non identifié	-	1,3	0,0468	-1,2	-1,6	NS	NS	NS	NS	NS
POS	7.22_372.1776m/z	4	Non identifié	-	1,0	0,0036	-2,8	-4,2	0,0385	-1,7	-2,3	NS	NS
POS	7.26_187.0629n	2	Indoleacrylic acid	HMDB0000734	1,0	0,0095	-0,9	-0,9	NS	NS	NS	NS	NS
POS	7.26_146.0601m/z	4	Non identifié	-	1,0	0,0453	-1,3	-1,0	NS	NS	NS	NS	NS
POS	7.26_484.0903n	4	Non identifié	-	1,0	0,0347	-1,9	-1,6	NS	NS	NS	NS	NS
POS	7.26_486.1796n	4	Non identifié	-	1,7	0,0274	-2,7	-2,5	NS	NS	NS	NS	NS
POS	7.28_240.1491m/z	4	Non identifié	-	1,8	0,0005	3,7	4,9	NS	NS	NS	NS	NS
POS	7.37_527.0454m/z	4	Non identifié	-	1,1	0,0261	-6,0	-6,0	NS	NS	NS	NS	NS
POS	7.38_168.0185n	4	Non identifié	-	0,9	0,0111	-1,5	-1,7	NS	NS	NS	NS	NS
POS	7.38_404.0248n	4	Non identifié	-	1,0	0,0085	-2,0	-2,4	NS	NS	NS	NS	NS
POS	7.41_203.1176m/z	4	Non identifié	-	1,6	0,0105	8,4	8,4	NS	NS	NS	NS	NS
POS	7.49_195.1145m/z	4	Non identifié	-	1,6	0,0242	1,6	2,3	NS	NS	NS	NS	NS
POS	7.54_321.1305m/z	4	Non identifié	-	1,2	0,0034	-1,5	-1,9	0,0258	-1,1	-1,5	NS	NS
POS	7.54_439.0432m/z	4	Non identifié	-	1,2	0,0065	-1,7	-2,0	NS	NS	NS	NS	NS
POS	7.56_385.1850n	4	Non identifié	-	1,6	NS	NS	NS	0,0355	-4,5	-4,8	NS	NS
POS	7.63_301.0793m/z	4	Non identifié	-	1,1	0,0229	2,4	NS	NS	NS	NS	NS	NS
POS	7.81_483.0689m/z	4	Non identifié	-	1,3	0,0340	-2,0	-2,1	NS	NS	NS	NS	NS
POS	7.83_611.2745m/z	4	Non identifié	-	1,3	0,0007	-3,7	-3,8	NS	NS	NS	NS	NS
POS	7.84_1160.5992n	4	Non identifié	-	1,3	0,0005	-2,5	-2,8	0,0258	-1,0	-1,4	NS	NS
POS	8.13_574.1854m/z	4	Non identifié	-	1,0	0,0140	-2,0	-2,3	NS	NS	NS	NS	NS
POS	8.13_582.1723m/z	4	Non identifié	-	1,0	0,0143	-1,9	-2,6	0,0301	-1,3	-1,4	NS	NS
POS	8.14_514.2478m/z	4	Non identifié	-	1,1	0,0090	-1,5	-1,7	0,0310	-1,0	-1,0	NS	NS
POS	8.14_616.1653m/z	4	Non identifié	-	1,1	0,0168	-2,0	-3,2	NS	NS	NS	NS	NS
POS	8.16_533.2251m/z	4	Non identifié	-	1,1	0,0116	-2,2	-2,3	NS	NS	NS	NS	NS
POS	8.18_246.1697m/z	2	Isovaleryl carnitine	HMDB0000688	1,8	0,0186	-1,0	-1,4	NS	NS	NS	NS	NS
POS	8.24_988.3096m/z	4	Non identifié	-	1,1	0,0347	-0,6	-0,8	NS	NS	NS	NS	NS
POS	8.25_705.9359m/z	4	Non identifié	-	1,1	0,0492	-0,5	-0,9	NS	NS	NS	NS	NS
POS	8.25_823.4225m/z	4	Non identifié	-	1,1	0,0347	-0,5	-0,8	NS	NS	NS	NS	NS

ESI	Pic	ID	Nom du composé	Identifiant HMDB	Score VIP (PLS-DA)	HTP			3R4F			MB-18W			MB-30W		
						ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2
POS	8.25_954.8758n	4	Non identifié	-	1,2	0,0175	-0,7	-0,9	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	8.25_988.1096m/z	4	Non identifié	-	1,2	0,0392	-0,4	-0,9	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	8.25_988.5098m/z	4	Non identifié	-	1,1	0,0470	-0,6	-0,8	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	8.53_169.0760m/z	2	Beta-carboline	HMDB0012897	1,9	0,0010	9,5	10,3	0,0001	7,6	8,8	NS	NS	NS	NS	NS	NS
POS	8.73_215.1184m/z	4	Non identifié	-	1,6	0,0015	2,1	3,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	8.82_183.0914m/z	2	Harman	HMDB0035196	1,9	0,0001	6,7	7,0	#####	6,4	7,5	NS	NS	NS	NS	NS	NS
POS	8.99_290.1990n	4	Non identifié	-	1,8	0,0045	13,2	13,0	0,0258	10,5	11,7	NS	NS	NS	NS	NS	NS
POS	9.00_215.1177m/z	2	Melatonin	HMDB0001389	1,6	0,0000	11,1	12,4	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	9.09_290.1992n	4	Non identifié	-	1,8	0,0000	9,5	10,7	0,0020	5,8	8,1	NS	NS	NS	NS	NS	NS
POS	9.22_260.1851m/z	2	Hexanoylcarnitine	HMDB0000705	1,5	0,0274	-1,1	-1,0	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	9.24_197.1068m/z	4	Non identifié	-	1,7	0,0137	11,7	11,0	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	9.26_239.1174m/z	4	Non identifié	-	1,8	0,0274	8,4	7,9	0,0258	6,3	7,7	NS	NS	NS	NS	NS	NS
POS	9.34_193.0494m/z	2	Scopoleptin	HMDB0034344	1,8	0,0005	13,1	13,5	0,0020	7,0	9,5	NS	NS	NS	NS	NS	NS
POS	9.62_362.2321m/z	4	Non identifié	-	1,8	0,0188	-2,2	-3,5	0,0077	-2,6	-4,7	NS	NS	NS	NS	NS	NS
POS	10.23_246.1729n	4	Non identifié	-	1,8	0,0003	12,6	13,2	0,0011	7,9	10,2	NS	NS	NS	NS	NS	NS
POS	11.43_236.2004m/z	2	Theaspirane	HMDB0036823	1,5	0,0036	-2,4	-3,3	0,0258	-2,0	-2,6	NS	NS	NS	NS	NS	NS
POS	11.44_344.3517m/z	4	Non identifié	-	1,9	NS	NS	NS	NS	NS	NS	0,0315	6,9	8,3	NS	NS	NS
POS	11.64_381.2028m/z	4	Non identifié	-	1,7	0,0007	12,5	12,6	0,0305	5,5	7,5	NS	NS	NS	NS	NS	NS
POS	11.96_239.1174m/z	4	Non identifié	-	1,7	0,0029	10,4	9,5	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	12.20_274.2043n	4	Non identifié	-	1,8	NS	NS	NS	0,0258	17,2	18,8	NS	NS	NS	NS	NS	NS
POS	12.52_496.3309m/z	4	Non identifié	-	1,5	NS	NS	NS	0,0431	-1,9	-2,5	NS	NS	NS	NS	NS	NS
POS	12.54_262.1781n	2	Glycerol 5-hydroxydecanoate	HMDB0032297	1,0	0,0392	NS	1,7	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	12.75_299.0191m/z	4	Non identifié	-	1,5	0,0242	-3,0	-8,3	0,0258	NS	-11,6	NS	NS	NS	NS	NS	NS
POS	12.78_391.1874m/z	3	Eicosanoid	-	1,8	0,0000	11,3	11,8	0,0020	6,7	8,9	NS	NS	NS	NS	NS	NS
POS	12.85_253.1335m/z	2	Rollipyrrole	HMDB0037554	1,7	0,0248	11,8	10,3	0,0082	10,9	11,5	NS	NS	NS	NS	NS	NS
POS	12.93_377.2081m/z	3	Eicosanoid	-	2,2	0,0191	11,1	11,0	0,0000	10,1	10,8	NS	NS	NS	NS	NS	NS
POS	13.03_391.1889m/z	3	Eicosanoid	-	1,8	0,0001	5,4	6,4	0,0081	2,1	4,2	NS	NS	NS	NS	NS	NS
POS	13.15_391.1875m/z	3	Eicosanoid	-	1,6	0,0005	9,2	9,7	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	13.17_377.2074m/z	3	Eicosanoid	-	1,6	0,0482	12,2	11,0	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	13.27_566.3961m/z	4	Non identifié	-	1,5	0,0427	6,1	6,3	0,0258	3,6	3,8	NS	NS	NS	NS	NS	NS
POS	13.34_296.2582m/z	4	Non identifié	-	1,3	NS	NS	NS	0,0258	-3,4	-6,7	NS	NS	NS	NS	NS	NS
POS	13.36_336.2287n	3	Eicosanoid	-	1,8	0,0002	9,6	10,5	0,0011	6,4	8,3	NS	NS	NS	NS	NS	NS

ESI	Pic	ID	Nom du composé	Identifiant HMDB	Score VIP (PLS-DA)	HTP			3R4F			MB-18W			MB-30W		
						ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2	ANOVA p-value	Log2(FC) TO vs T1	Log2(FC) TO vs T2
POS	13.37_465.2850n	2	LysoPC(14:1)	HMDB0010380	1,6	0,0220	2,2	2,8	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	13.52_377.2078m/z	3	Eicosanoïd	-	1,9	0,0018	11,4	11,5	0,0455	6,6	9,3	NS	NS	NS	NS	NS	NS
POS	13.61_583.2540m/z	2	Biliverdin	HMDB01008	1,3	NS	NS	NS	0,0274	1,6	2,1	NS	NS	NS	NS	NS	NS
POS	13.79_338.2413n	4	Non identifié	-	1,8	0,0133	8,3	7,9	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	13.86_467.3004n	2	LysoPC(14:0/0:0)	HMDB0010379	1,5	0,0142	1,3	1,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	13.97_359.1982m/z	3	Eicosanoïd	-	1,8	0,0018	9,4	9,2	0,0020	3,2	5,1	NS	NS	NS	NS	NS	NS
POS	14.07_499.2692n	2	LysoPE(0:0/20:5)	HMDB0011489	1,7	0,0192	2,4	3,1	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.08_303.2418m/z	4	Non identifié	-	1,7	0,0090	8,4	7,7	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.12_541.3164n	2	LysoPC(20:5)	HMDB0010397	1,8	0,0007	5,6	7,2	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.18_373.1770m/z	3	Eicosanoïd	-	1,3	0,0035	14,9	14,0	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.19_253.2637m/z	4	Non identifié	-	1,1	0,0326	4,2	4,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.24_493.3160n	2	LysoPC(16:1/0:0)	HMDB0010383	1,6	0,0013	1,2	1,7	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.27_518.3212m/z	2	LysoPC(16:0)	HMDB0010382	1,7	0,0043	2,5	3,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.42_359.1980m/z	3	Eicosanoïd	-	1,8	0,0003	17,1	16,1	0,0014	10,5	11,5	NS	NS	NS	NS	NS	NS
POS	14.66_322.2497n	3	Eicosanoïd	-	1,3	0,0112	9,4	8,8	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.68_519.3316n	2	LysoPC(18:2)	HMDB0010386	1,7	0,0005	2,1	2,9	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.68_522.2142n	4	Non identifié	-	1,0	0,0274	NS	-3,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.72_543.3316n	2	LysoPC(20:4)	HMDB0010395	1,8	0,0033	3,0	4,3	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.74_281.2949m/z	4	Non identifié	-	1,1	0,0274	10,8	11,3	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.76_466.3730m/z	4	Non identifié	-	1,2	0,0439	-4,2	-7,8	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.83_508.3388m/z	2	LPC 17:1	0	1,4	0,0340	NS	2,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.85_510.3983m/z	2	MG(10:0/0:0/0:0)	HMDB0072866	1,1	0,0352	-3,3	-7,4	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.90_554.4247m/z	4	Non identifié	-	1,2	0,0018	-6,2	-9,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.95_521.3129n	4	Non identifié	-	1,1	0,0326	1,6	1,9	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	14.96_519.3317n	2	LysoPC(18:2)	HMDB0010386	1,6	0,0093	1,0	1,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	15.00_569.3475n	2	LysoPC(22:5)	HMDB0010403	1,8	0,0021	4,7	6,6	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	15.04_295.3104m/z	4	Non identifié	-	1,3	0,0427	7,4	8,4	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	15.07_495.3320n	2	LysoPC(16:0)	HMDB0010382	1,5	0,0212	0,8	1,2	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	15.19_259.1691m/z	4	Non identifié	-	1,6	0,0267	11,8	9,7	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	15.21_584.3104m/z	2	LysoPC(20:3)	HMDB0010393	1,6	0,0203	2,7	3,5	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	15.22_547.3573m/z	2	1-Stearoyl glycerophosphocholine	HMDB0094688	1,7	0,0036	3,1	4,4	NS	NS	NS	NS	NS	NS	NS	NS	NS
POS	15.24_453.2849n	2	LysoPE(16:0/0:0)	HMDB0011503	1,8	0,0115	-1,8	-2,0	NS	NS	NS	NS	NS	NS	NS	NS	NS

ESI	Pic	ID	Nom du composé	Identifiant HMDB	Score VIP (PLS-DA)	HTP		3R4F		MB-18W		MB-30W	
						ANOVA p-value	Log2(FC) TO vs T1						
POS	15.27_318.2190n	3	Eicosanoid	-	1,7	0,0114	11,2	0,0095	4,6	7,3	NS	NS	NS
POS	15.30_497.3424m/z	4	Non identifié	-	1,1	0,0487	NS	NS	NS	NS	NS	NS	NS
POS	15.43_521.3476n	2	LysoPC(18:1)	HMDB0002815	1,5	0,0166	0,9	1,3	NS	NS	NS	NS	NS
POS	15.43_545.3462n	2	LysoPC(20:3)	HMDB0010393	1,8	0,0008	2,0	2,7	NS	NS	NS	NS	NS
POS	15.43_524.3606m/z	4	Non identifié	-	1,5	0,0154	1,1	1,6	NS	NS	NS	NS	NS
POS	15.58_344.3306m/z	4	Non identifié	-	1,8	0,0016	9,7	10,8	NS	NS	NS	NS	NS
POS	15.65_502.3256m/z	2	LysoPC(P-16:0)	HMDB0010407	1,6	0,0094	4,0	4,5	NS	NS	NS	NS	NS
POS	15.66_479.3361n	2	LysoPC(P-16:0)	HMDB0010407	1,6	0,0036	3,1	3,7	NS	NS	NS	NS	NS
POS	15.67_525.3412m/z	4	Non identifié	-	1,4	0,0035	3,2	3,3	NS	NS	NS	NS	NS
POS	15.79_548.3696m/z	2	LysoPC(20:2)	HMDB0010392	1,5	0,0104	2,5	3,2	NS	NS	NS	NS	NS
POS	15.94_573.3417n	4	Non identifié	-	1,4	0,0095	4,8	5,1	NS	NS	NS	NS	NS
POS	15.96_547.3629n	2	LysoPC(20:2)	HMDB0010392	1,5	0,0191	2,6	3,5	NS	NS	NS	NS	NS
POS	15.99_548.3700m/z	2	LysoPC(20:2)	HMDB0010392	1,6	0,0143	1,8	2,7	NS	NS	NS	NS	NS
POS	15.99_549.3421n	4	Non identifié	-	1,3	0,0192	3,1	3,5	NS	NS	NS	NS	NS
POS	16.15_312.3620m/z	4	Non identifié	-	1,4	0,0022	5,9	7,5	NS	NS	NS	NS	NS
POS	16.18_327.3495n	4	Non identifié	-	1,0	0,0498	3,6	4,4	NS	NS	NS	NS	NS
POS	16.30_326.5412m/z	4	Non identifié	-	1,2	0,0447	1,7	1,7	NS	NS	NS	NS	NS
POS	16.36_573.3789n	4	Non identifié	-	1,8	0,0005	3,4	4,7	NS	NS	NS	NS	NS
POS	16.80_549.3785n	2	LysoPC(20:1)	HMDB0010391	1,3	0,0183	1,5	2,2	NS	NS	NS	NS	NS
POS	16.90_508.3750m/z	2	LysoPC(P-18:0)	HMDB0013122	1,5	0,0030	5,4	6,0	NS	NS	NS	NS	NS

PUBLICATION

Article

Prediction of a Large-Scale Database of Collision Cross-Section and Retention Time Using Machine Learning to Reduce False Positive Annotations in Untargeted Metabolomics

Marie Lenski ^{1,2,*}, Saïd Maallem ¹, Gianni Zarcone ¹, Guillaume Garçon ¹, Jean-Marc Lo-Guidice ¹, Sébastien Anthérieu ¹ and Delphine Allorge ^{1,2}

¹ ULR 4483, IMPECS—IMPact de l'Environnement Chimique sur la Santé humaine, CHU Lille, Institut Pasteur de Lille, Université de Lille, F-59000 Lille, France

² CHU Lille, Unité Fonctionnelle de Toxicologie, F-59037 Lille, France

* Correspondence: marie.lenski@univ-lille.fr



Citation: Lenski, M.; Maallem, S.; Zarcone, G.; Garçon, G.; Lo-Guidice, J.-M.; Anthérieu, S.; Allorge, D. Prediction of a Large-Scale Database of Collision Cross-Section and Retention Time Using Machine Learning to Reduce False Positive Annotations in Untargeted Metabolomics. *Metabolites* **2023**, *13*, 282. <https://doi.org/10.3390/metabo13020282>

Academic Editors: Sarah Wille and Andrea E. Steuer

Received: 9 January 2023

Revised: 7 February 2023

Accepted: 12 February 2023

Published: 15 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Metabolite identification in untargeted metabolomics is complex, with the risk of false positive annotations. This work aims to use machine learning to successively predict the retention time (Rt) and the collision cross-section (CCS) of an open-access database to accelerate the interpretation of metabolomic results. Standards of metabolites were tested using liquid chromatography coupled with high-resolution mass spectrometry. In CCSBase and QSRR predictor machine learning models, experimental results were used to generate predicted CCS and Rt of the Human Metabolome Database. From 542 standards, 266 and 301 compounds were detected in positive and negative electrospray ionization mode, respectively, corresponding to 380 different metabolites. CCS and Rt were then predicted using machine learning tools for almost 114,000 metabolites. R² score of the linear regression between predicted and measured data achieved 0.938 and 0.898 for CCS and Rt, respectively, demonstrating the models' reliability. A CCS and Rt index filter of mean error ± 2 standard deviations could remove most misidentifications. Its application to data generated from a toxicology study on tobacco cigarettes reduced hits by 76%. Regarding the volume of data produced by metabolomics, the practical workflow provided allows for the implementation of valuable large-scale databases to improve the biological interpretation of metabolomics data.

Keywords: mass spectrometry-based metabolomics; ion mobility-mass spectrometry; metabolomics data analysis; machine learning; collision cross-section; retention time

1. Introduction

The metabolome of a biological system is influenced by physiological, pathological, or environmental conditions [1]. As it gathers the final products of the cellular process, the exhaustive measurement of metabolite changes by metabolomics provides dynamic and sensitive information closely linked to its phenotype. Investigating the metabolome allows the identification of metabolic fingerprints that can then be used as biomarkers and/or provides new mechanistic perspectives leading to a particular phenotype [2,3]. Recent developments in mass spectrometry (MS) technology, informatics, and analytical chemistry have made it possible to comprehensively analyze the metabolome [4,5] with a high level of sensitivity [6] compared to nuclear magnetic resonance-based metabolomics. Additionally, high throughput analyses like high-resolution MS in full scan mode allow the rapid achievement of large-scale studies of hundreds of compounds (untargeted metabolomics), which is an evident benefit over analyses that target a restricted number of metabolites or pathways (targeted metabolomics).

Untargeted metabolomics is a multistep process involving first sample collection, preparation, and analysis that generates data, then data processing and chemometrics that generate a candidate list of features, and finally, metabolite identification [7]. This final step

gives biological meaning to MS data [8]. A consensus by the Chemical Analysis Working Group of the Metabolomics Standards Initiative (MSI) reported different levels of confidence in the annotation depending on the method of identification used [9]. A confident and definitive identification (level 1) is hit when two or more orthogonal properties fit with data from authentic standard compounds in identical analytical settings. When the latter are unavailable, a comparison of experimental data with public libraries could lead to a putative compound annotation (level 2) or class annotation (level 3). Finally, unknown features discriminated with spectral data are classified at the lowest confidence level (level 4). Therefore, feature annotation is achieved by comparing experimental measurements to existing in-house or external databases of known metabolites to generate potential candidates [10]. Several commercial or open-source databases containing spectral data in libraries (Human Metabolome Database (HMDB) [11], Metlin [12] . . .) were developed and are continuously updated by the scientific community. However, confident and unequivocal structure identification could quickly be an issue when a candidate is not found (limited number of spectra) or when several candidates are proposed (false positives), increasing the probability of misidentification [13,14]. Therefore, it becomes important to use other readily obtained physicochemical properties for better metabolite identification.

Ion mobility-mass spectrometry (IMS-MS) is a fast two-dimensional separation of ions based on their mobility in a buffer gas. Importantly, this mobility is structure-dependent and is not affected by equipment or experimental factors (matrix effects, variations in mobile phase composition, and chromatography settings, ionization mode, acquisition settings . . .), unlike retention time (R_t) and mass spectrum [15], resulting in a high degree of repeatability and therefore facilitating database queries [16]. The physical property measured in IMS-MS is the collision cross-section (CCS). Thus, it provides the orthogonal separation to improve signal-to-noise, resolution, and isomeric metabolite separation [17], participating in the reduction of misidentification. However, the favorable contribution of CCS is currently limited by the poor availability of CCS reference values [18–22]. New experimental and computational approaches to predict those parameters for a large number of compounds is highly valuable. Several studies have developed, or applied machine learning-based prediction approaches [23]. Softwares like AllCCS [22], CCS Predictor [24], DeepCCS [25], MetCCS Predictor [26], or LipidCCS predictor [27] can efficiently generate a model when molecular descriptors are provided [15]. Molecular descriptors are numeric information generated by mathematical treatment of compound structures that characterize the physico-chemical properties of metabolites (ex: polarity, LogP . . .) [28]. In contrast, CSSBase is a web interface (<https://CCSbase.net>) (accessed on 6 May 2022) that provides access to a ready-to-use predictive model, allowing rapid prediction of CCS values directly from SMILES structures (Simplified Molecular Input Line Entry System representation), using a cluster-based prediction model [29]. This platform allows a broad coverage of chemical structure diversity and can thus be easily used in existing metabolomics workflows.

The R_t of a compound is defined by its chemical interactions with the chosen mobile phase and stationary phase. Metabolite retention can be improved by optimizing solvent gradient elution, nature, and dimensions of the chromatographic column or chromatographic settings [30]. R_t is often decisive in feature annotation but usually relies on the availability of authentic chemical standards that are applied to experimental conditions. In untargeted metabolomics, the transferability of R_t database between laboratories is not achievable because of the absence of standardized assays across different laboratories. Multiple machine learning models for the prediction of R_t have already been described, including quantitative structure–retention relationship (QSRR) models [31–35]. QSRR strategies have been used to accelerate the method development process by comparing predicted separation with different columns [36] or to enhance the confidence of identifications [37]. Software packages, such as the QSRR Automator [38], exist to automate R_t prediction model creation. Structure and chromatographic data from known metabolites, obtained from their SMILES and from chemical standards analyzed using a particular LC method, are used to generate a model. It identifies relations between chromatographic reten-

tion and the molecular descriptors, theoretically allowing to predict R_t for any metabolite whose molecular descriptors can be calculated [39].

In the present work, we aim to describe the workflow permitting the generation of a large-scale in-house database of R_t and CCS predicted with published machine learning models. Integration of these data with other sources of information, such as accurate mass, MS_e fragmentation, and isotope pattern for facilitating the identification of compounds, is illustrated in an application to toxicology data.

2. Materials and Methods

2.1. Chemicals and Standards

Solutions used were: acetonitrile (UPLC-MS grade, Waters, Milford, MA, USA), methanol (UPLC-MS grade, Waters), Milli-Q purified water (Millipore, Burlington, MA, USA), formic acid (UPLC-MS grade, Honeywell, Charlotte, NC, USA), ammonium formate (Reagent-grade, Sigma-Aldrich, St. Louis, MO, USA), and chloroform (VWR Chemicals, Radnor, PA, USA). Chemical standards (MSMLS) were purchased from Sigma-Aldrich. This library was chosen for the broad chemical and functional diversity of metabolites included. It contains 634 standard metabolites sampled into seven 96-well plates at 5 µg per well, including 37 duplicates. An associated spreadsheet with information, such as metabolite identification, molecular formula, and SMILES was used to build our targeted database. The compounds were dissolved using two different solutions (5% methanol for plates 1–5 and chloroform:methanol:water 1:1:0.3 for plates 6–7) following the manufacturer's instructions to obtain a 20 µg mL⁻¹ concentration. Stock solutions were pooled with a maximum of 12 compounds to obtain 56 solutions at 1.6 µg mL⁻¹ to perform simple multiplex injections for LC-MS analysis.

2.2. LC-MS Conditions

Analyses were conducted on a liquid chromatograph system coupled to high-resolution mass spectrometry (LC-HRMS). Chromatographic separation was obtained with the following characteristics: Instrument: Acquity UPLC I-Class system (Waters); column: Acquity UPLC HSS T3 (1.8 µm, 150 × 2.1 mm; Waters); column temperature: 50 °C; flow rate: 0.4 mL min⁻¹; autosampler temperature: 10 °C; volume of injection: 15 µL. Separation was performed in a gradient elution mode. Mobile phases for the multistep gradient in the positive mode were solution A: aqueous solution of ammonium formate (3 mM) with 0.1% formic acid and solution B: acetonitrile with 0.1% formic acid (*v/v*). The elution gradient was: 100% A for 1 min, 0–1% B for 1 min, 1–3% B for 2 min, 3–99% B for 13 min, 99% B for 3 min, 99–0% B for 0.5 min, and 100% A for 2.5 min. Detection was performed on a Vion IMS-QToF mass spectrometer (Waters) with the following settings: ionization source: electrospray operating in positive (ESI+) and negative (ESI-) modes; source temperature: 120 °C; desolvation temperature: 600 °C; cone gas flow: 50 L h⁻¹; desolvation gas flow: 1000 L h⁻¹; capillary voltage: 0.5 kV in ESI+ or 2 kV in ESI-; *m/z* range: 50–1000; scan time: 0.25 s; lock mass reference: leucine enkephalin (*m/z* 556.2766) solution at 200 ng mL⁻¹; infusion intervals: 5 min; acquisition mode: high-definition MS_e; low collision energy: 6 V; high collision energy ramp: 14–56 V; IMS drift gas and collision gas: nitrogen; ion mobility and mass calibrations solution: Major Mix IMS/ToF Calibration Kit (Waters). These parameters allow for achieving a mass resolving power of >20,000 FWHM.

Data analysis of the mixes was semi-automatically performed through the Unifi software (version 1.9.4.053 Waters MS Technologies, Manchester, UK) to obtain the R_t , response, and CCS of the standards after manual verification of the peak integration. Adducts considered were [M+H]⁺, [M+K]⁺, [M+Na]⁺, [M+Cl]⁻, [M+HCOO]⁻, and [M+CH₃OO]⁻. Accurate mass, R_t , CCS, and fragmentation patterns were used to build a targeted database of LC and MS properties.

2.3. CCS Prediction

CCSBase is an electronic interface (<https://CCSbase.net>) (accessed on 6 May 2022) for accessing the CCS predictive model [29]. It calculates the predictive CCS values of adducts using SMILES. Original performances of CCSBase were described by Ross et al. [29] with an R^2 score, a mean absolute error, and a root mean squared error at 0.991, 3.83 Å², and 5.48 Å², respectively. In the study, all adducts considered by CCSBase, namely [M+H]⁺, [M+K]⁺, [M+Na]⁺, [M+Na-2H]⁻, [M+NH₄]⁺, [M]⁺, [M-H]⁻, and [M]⁻, were taken in account. A batch prediction was performed for a dataset of metabolite structures freely available in HMDB v4.0 [11], which gathers up to 114,000 human metabolites, covering the majority of untargeted metabolomic data sets. A linear regression was performed comparing predicted and measured CCS of adducts from standard compounds. Measured CCS were included in this comparison if the standard were listed in HMDB, and if they presented common adducts with predictions. The coefficient of determination R^2 between the predicted and the experimental CCS data and mean absolute error permitted to evaluate the model. The best fit of linear regression was calculated, with an interval of ± 2 standard deviations (SD). All statistical analyses and figure production of this manuscript were conducted under R language and environment [40].

2.4. Rt Prediction of Small Molecules

QSRR Automator [38] builds regression retention models. Based on their SMILES, chemical structures were converted into their numerical representation by expressing them through structural descriptors produced by informatic algorithms of QSRR Automator. First, using a defined training data set, the machine learning algorithm learns the “rule” between molecular descriptors and their experimental Rt values to establish prediction models and select the best model. QSRR algorithm identifies descriptors that positively impact model performance. Selection and optimization of regression algorithms were carried out by automated procedures and evaluated thanks to the R^2 score and mean absolute error. Then, the external validation data set is used to validate and evaluate the prediction error. Cross-validation ($n = 10$) provides an estimate of the accuracy of the Rt prediction for compounds that were not used in its development or optimization, evaluated thanks to the R^2 score, mean absolute error, and SD. Once a valid model was selected, Rt predictions were performed for metabolites from the HMDB v4.0. The best fit of linear regression was calculated with an interval of ± 2 SD.

2.5. Reduction of the Occurrence of False Positive Annotations in Untargeted Metabolomics: Application to Toxicology Data

We analyzed LC-HRMS data from an ongoing study assessing the potential toxicity of tobacco cigarette fumes on human bronchial epithelial BEAS-2B cells to demonstrate the relevance of the predicted large-scale database of collision cross-section and retention time to metabolomics. The exposure protocol was adapted from Dusautoir et al. [41]. Briefly, BEAS-2B cells cultured at air-liquid interface was exposed to four puffs of tobacco cigarette emissions or to sterile air (negative control) in four replicates per exposure. Twenty-four hours after exposure, cell metabolism was quenched by the addition of ice-cold methanol:water (80:20, *v/v*) mixture. Cells were harvested using a cell scraper. Deproteinization was performed by adding the same methanolic mixture, vortexing, and centrifuging at 14,000 × *g* at +4 °C for 15 min. Supernatants were concentrated to dryness with speedvac and reconstituted before injection in a water:methanol (90:10, *v/v*) mixture. After metabolomic analyses, LC-HRMS data were analyzed with Progenesis Q1 (Nonlinear Dynamics, UK) for feature extraction. Data normalization and statistical analyses were conducted under the R environment [40] on the features detected in ESI⁺ and ESI⁻. When searching against HMDB, two identification strategies were evaluated: (1) with an *m/z* (tolerance set at 5 mDa), isotope and fragmentation match only, and (2) with an *m/z*, isotope, fragmentation, CCS, and Rt match of the created predicted large-scale database.

3. Results

3.1. Analysis of Standard Compounds and Generation of an In-House Database

A total of 542 standards were originally used in the candidate database. Each set of data was manually examined for errors. Figure 1 describes the workflow used for the targeted database construction. The candidate sorting step has allowed the detection of 266 and 301 compounds in ESI+ or ESI−, respectively, corresponding to a total of 380 different metabolites. Accurate mass, Rt, CCS, and fragmentation were used to build a targeted database of LC and MS properties.

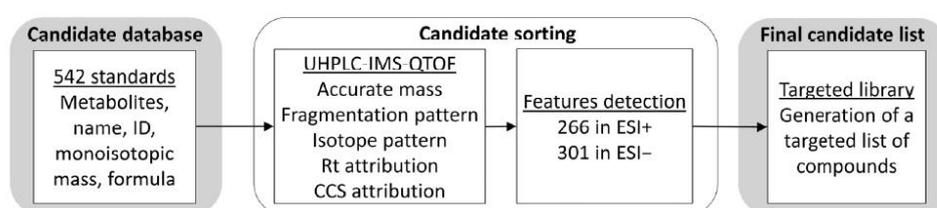


Figure 1. Workflow and results of the targeted-based metabolomic approach developed.

Metabolites belong to different chemical classes and are mainly carbohydrates, carboxylic acids, lipids, nucleotides, and organoheterocyclic compounds (Figure 2) involved in different pathways that reflect the metabolic status in biological matrices of interest.

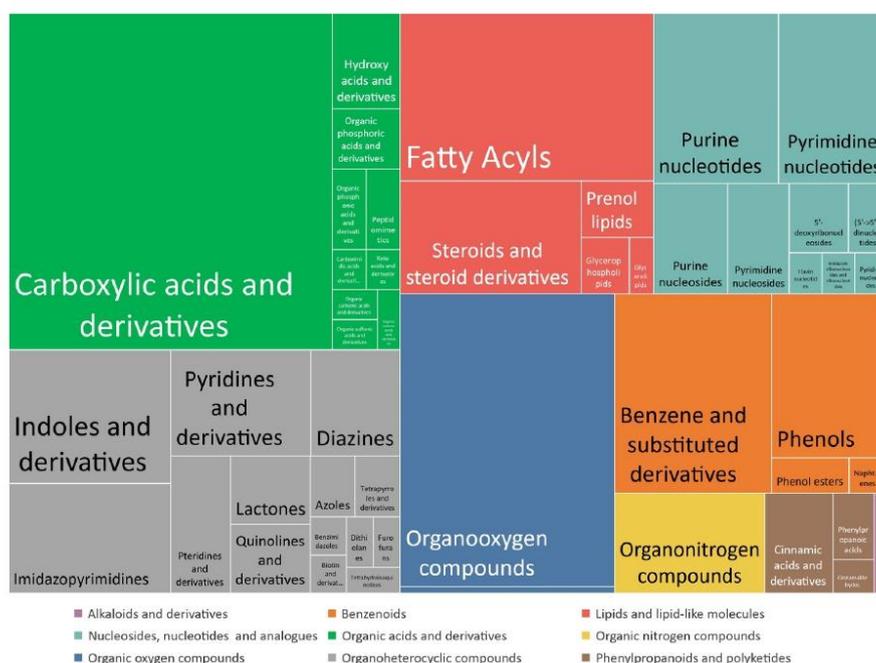


Figure 2. Chemical taxonomy of metabolites included in the targeted library and proportion of each superclass and class.

3.2. CCS Prediction and Creation of a CCS Database

The CCS database was generated according to the workflow described in Figure 3. Predictions were performed for almost 114,000 metabolites from the HMDB v4.0 database, generating 916,104 CCS adduct values. Results were validated with a validation set composed of 501 measured CCS adduct values from 297 standard compounds in both ionization modes.

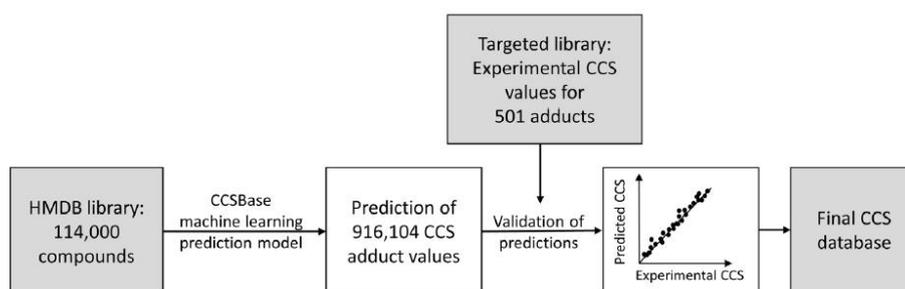


Figure 3. Workflow of CCS prediction.

Figure 4 illustrates the match between predicted and experimentally determined CCS. Outliers were kept as part of the data set in the absence of any evidence that they were the result of an error. The R^2 score of the linear regression achieves 0.938 and the mean absolute error was calculated at 3.94 \AA^2 , while the SD reaches 6.11 \AA^2 or 3.36%. The predicted CCS = $0.95 \times \text{measured CCS} + 7.92$. The resulting output table of CCS allowed us to build our large-scale in-house reference database.

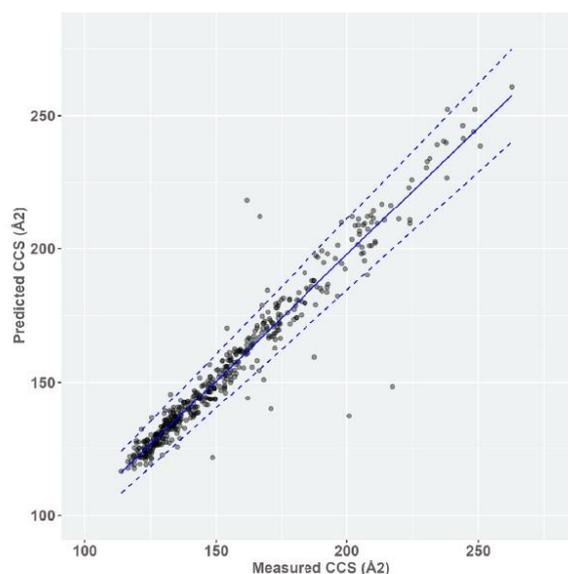


Figure 4. Predicted CCS (\AA^2) using CCSBase algorithm [29] by measured CCS (\AA^2). Validation was performed on the CCS of 501 adducts from 297 standard compounds measured in ESI⁻ and ESI⁺. Linear regression: adjusted $R^2 = 0.938$. Blue line: best fit of linear regression. Dashed blue lines best fit $\pm 6.72\%$ (2 SD). Predicted CCS = $0.95 \times \text{measured CCS} + 7.92$.

3.3. Rt Prediction and Creation of an Rt Database

The Rt prediction workflow is described in Figure 5. In total, 204 compounds from the developed method were selected for the QSRR model; 114 were detected in both ESI+ and ESI−, while 90 were detected only in one ionization mode (45 for each ionization mode). Seven compounds were excluded due to incomplete data in molecular descriptors. Support vector regression (SVR) algorithm based on 113 molecular descriptors presented the best performances, with the R^2 score at 0.999 and the mean absolute error at 0.10 min for the training set. The validation set tested by cross-validation ($n = 10$) validated the model with the following performances: mean of cross-validation R^2 score 0.898, mean absolute error 0.81 min, and standard deviation of the mean absolute error 0.15 min. Detailed results are presented in Table S1 and Figure S1. Rt predictions were performed for almost 114,000 metabolites from HMDB v4.0. The resulting output table of Rt allowed us to build our large-scale in-house reference database.

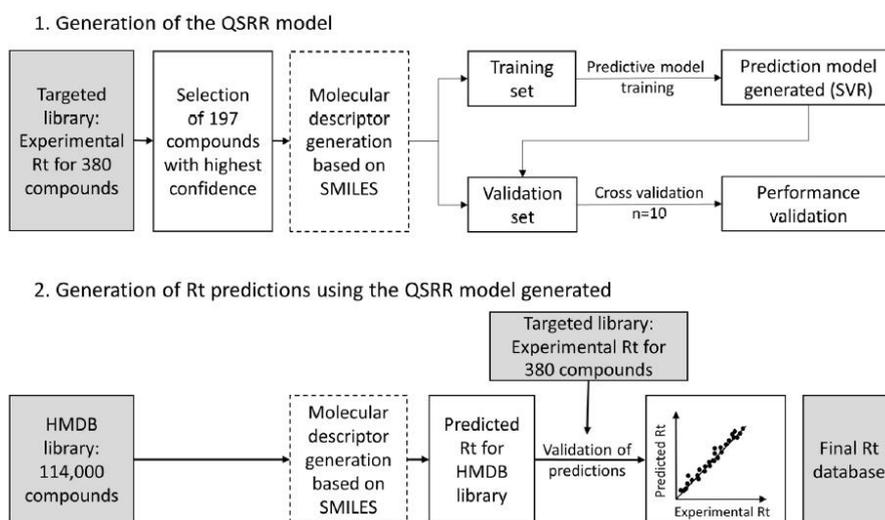


Figure 5. Workflow of Rt prediction.

3.4. Reduction of the Occurrence of False Positive Annotations in Untargeted Metabolomics: Application to Toxicology Data

We assessed the potential toxicity associated to tobacco cigarette fumes on the human bronchial epithelial BEAS-2B cells using metabolomics. Among the 3591 features detected in ESI+ and ESI−, 51 features were significantly deregulated by cigarette smoke compared to controls and needed to be identified. As illustrated in Figure 6a, 46 out of 51 features had one hit or more (90%). The number of hits exceeded 10 hits for the major part of the features. For the method combining m/z , CCS, and Rt match search, CCS and Rt match tolerances were set at 16 \AA^2 and 1.1 minutes, respectively, according to the determined CCS and Rt index filter expressed as mean error ± 2 SD. Only 37 out of 51 features (72%) had one or more metabolite hits. Seventy-six percent of hits were filtered using the predicted large-scale database (Figure 6b). The percentage of features with only one hit significantly increased with the additional CCS and Rt match (+53%), while the percentage of features with more than 10 hits decreased in the same conditions (−39%) (Figure 6c). For further identification, possible candidates for each compound are ranked by Progenesis QI on an overall score based on the m/z match, isotope similarity, fragmentation score, CCS, and Rt error (data not shown).

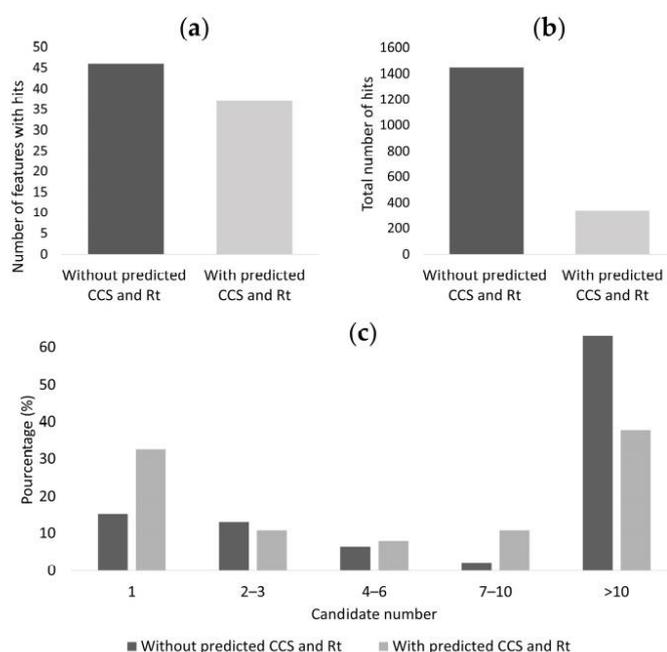


Figure 6. Contribution of the predicted CCS and Rt database to metabolite identification in untargeted metabolomics applied to toxicology. From the two match methods: (a) number of features with hits, (b) total number of hits (c), percentage distribution of features with different metabolite hits.

4. Discussion

LC-HRMS is an uncontested powerful analytical approach employed in both targeted and untargeted metabolomics. We used an UPLC-IMS-QTOF to create an accurate in-house database using a commercial library of metabolite standards. Based on experimental results, we used two existing prediction models (CCSBASE and QSRR Automator) to predict CCS and Rt values of a large-scale database to increase confidence in metabolite annotation. Associating Rt and CCS is relevant as it can provide complementary information coming from chromatographic and ion mobility separation or even replace other orthogonal properties (isotope similarity and fragmentation score) for putative compound annotation. Moreover, for all metabolites, predictions were performed for protonated and deprotonated ions as well as adducts, each having the same Rt but a different CCS [42]. The co-occurrence of adducts is common when analyzing heterogeneous biological samples [43]. Gathering predictions for multiple metabolic features of the same metabolite is valuable information, allowing for cross-validation of identification. The predicted database is presented in Table S2.

The comparison of mass spectra, Rt, CCS, and accurate mass of a feature with experimental data acquired from standard compounds measured under the same analytical settings permits achieving the highest level of identification confidence. An in-house database of 380 different metabolites was generated, allowing a confident and definitive identification (level 1 according to the MSI [9]). Metabolites were separated using an Acquity UPLC HSS T3 column that possesses superior polar-compound retention and aqueous mobile phase compatibility compared to more classical stationary phases. It could be used for the retention of mid-polar to apolar analytes. The Sigma library used contains several polar metabolites that cannot be retained in those analytical conditions, explaining

the lack of detection for some metabolites [44]. Moreover, with up to 114,000 chemicals deposited in HMDB, only a few percent of these compounds could be covered with authentic standards. Therefore, structure identification in untargeted metabolomics analyses remains a significant challenge. By predicting chromatographic R_t and CCS from experimentally acquired data, this targeted library represents a starting point to potentially give access to a detailed sample composition for future untargeted metabolomic studies. Using this methodology, we were able to drastically expand the number of metabolites at level 2 or level 3 annotations [9].

Predictions using machine learning are data-driven approaches providing predictions for metabolites with corresponding properties [45]. After the training of the model, predictions could be generated immediately for other compounds.

For CCS predictions, CCSBase is a machine learning-based prediction model built from a combined database, enabling to cover an important variety of structural compounds, participating in the transferability of this model. Indeed, large-scale CCS predictions were validated with our experimental data with a low bias and a high R^2 at 0.938. A CCS index filter defined as mean error ± 2 SD, i.e., maximum 16.16 \AA^2 , could be used as the threshold for excluding false positives. This match tolerance, reflecting the deviation of analytes or family of analytes or type of adducts, is relatively large compared to other work demonstrating that median relative errors as low as 3 to 5% are reachable using other models [22,24–27]. However, excluding false positive identifications with a CCS match higher than the defined threshold remains of great importance when considering the number of possible matches when using m/z match, isotope similarity, and fragmentation score only. Moreover, this additional separation process participates in better detection of compounds presenting contaminant mass spectra due to the co-elution or a poor abundance.

For R_t predictions, the workflow was different as we trained and validated an accurate machine-learning model based on compounds with various physicochemical properties. The training set allowed the model to be trained, while the test set made of unknown data for the trained model allowed the model to be validated. With this strategy, the model was estimated with small error differences in favour of minimum overfitting. An R_t index filter defined as mean error ± 2 SD, i.e., maximum 1.11 min, could be used as a threshold for eliminating the majority of misidentified compounds. Outliers could be due to software bias, random noise in the data used or errors in the attribution of standards. Naylor et al. described the QSRR Automator's original performances on various chromatographic columns. They showed errors in predictions within 1 min for the majority of predictions, and within 2 min for almost all predictions [38]. QSRR here performs comparably to previously published methods [34,36]. In relatively short run time methods, as in our method, many metabolites have very close R_t , including isomers with R_t that fall within 10 s of each other. Our database does not permit the distinction between those metabolites but is adequate to differentiate between clearly separated compounds of the same mass and reduce false positive identifications, leading to an advanced biological interpretation of results. Even if the generated model was based on compounds with various physicochemical properties separated and identified with an optimized method, particular attention should be dedicated to avoiding inaccurate results, including (i) compounds not retained in the column (ii) compounds retained after the observed R_t of the training data (iii) compounds with physicochemical properties that differ from the training set. For example, the in-house database presented here was generated from a large variety of chemical standards dedicated to metabolomics analyses but did not include complex high molecular weight compounds. Biased predictions for those metabolites should be excluded. The R_t database that we created is strictly related to our chromatographic conditions, so it can be directly useful only for those who decide to strictly adapt our choice of column, mobile phases, and flow rates. In reversed-phase chromatography, authors suggested that R_t from a defined method can be projected in other chromatographic settings as soon as the elution order of metabolites is preserved [46]. Most of the time, laboratories employ a distinct chromatographic setting depending on the separation required. We here presented

a practical workflow, with the objective of generating QSRR models and predictions for every set of LC conditions.

Some limitations of the present study must be mentioned. The number of compounds is relatively small, resulting in a limited number of experimental data that could influence the performance of models. However, this limitation is counterbalanced by the quality of data since we included data from authentic standards with the highest confidence possible. The resulting performances could have been further validated by performing a side-by-side comparison with other existing machine-learning tools. Such a comparison has already been described elsewhere [22,24–27]. Instead of that, the chosen strategy consisted of emphasizing the usefulness of our workflow with concrete application on biological data. Finally, by associating *Rt* predictions with CCS predictions, the generated large-scale database is strictly related to the instrumental configuration, but the workflow could be largely generated to other experimental conditions.

Most prediction models or workflows previously reported discuss one or the other predicted property (CCS or *Rt*), while only a few associate multi-dimensional information for metabolite annotation [47–50]. Interestingly, all of them are dedicated to lipids or exogenous compounds, while our workflow predicted a database including small molecules found in the human body, including water-soluble or lipid-soluble endogenous metabolites and exogenous compounds. Regarding the tremendous interest of the scientific community in metabolomics, providing a practical workflow is of large importance for analytical chemists or biologists who cannot develop machine learning models but who want to improve the biological interpretation of their metabolomics data. The usefulness of our combined large-scale predicted database was demonstrated with an application of biological data generated from a toxicology study on tobacco cigarettes. The results demonstrated that the introduction of CCS and *Rt* values for metabolite identification could significantly reduce false positive identifications, with the benefit of narrowing the search scope and improving the identification accuracy.

5. Conclusions

In this study, a workflow was introduced to remove false positive annotations in non-targeted metabolomics studies. The procedure includes the implementation of a combined CCS and *Rt*-restricted database starting from a commercial library of metabolite standards. This experimental database has been used to predict CCS and *Rt* of a large-scale dataset using existing machine learning tools. As illustrated by an application on a metabolomic study on tobacco cigarette toxicity, the presented workflow reduces the occurrence of false positive annotations in untargeted metabolomics and adds confidence to the identification of metabolites. This database has been integrated into the protocol used in our laboratory for untargeted metabolomics analyses and is freely downloadable. When making the assumption that the created database could be a representative subset of compounds present in the human metabolome, biological interpretation of metabolomics data is notably improved, giving new insights into biomarker research or mechanisms that generate a specific phenotype. We suggest using our data as a methodological starting point for the development of a large-scale in-house reference database based on artificial intelligence tools, providing a practical and effective workflow to improve the predictive confidence of metabolomic studies at a large-scale level.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/metabo13020282/s1>, Figure S1: Performances of the SVR model using QSRR Automator algorithm; Table S1: Linear regression of Predicted *Rt* (min) using QSRR Automator algorithm by measured *Rt* (min) of compounds used from the developed method. This figure was generated by QSRR Automator from all the data available for model construction (including training data); Table S2: Large-scale in-house reference database.

Author Contributions: Conceptualization, M.L. and S.M.; methodology, S.M. and M.L.; investigation, M.L., S.M., G.Z., S.A. and D.A.; resources, D.A., G.G. and J.-M.L.-G.; writing—original draft preparation, M.L.; writing—review and editing, S.A., D.A., G.G., G.Z. and J.-M.L.-G.; visualization, S.M. and M.L.; supervision, D.A. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: Graphical abstract was modified from SMART (Servier Medical Art). <http://smart.servier.com> (accessed on 6 May 2022).

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

Abbreviations

CCS	collision cross-section
ESI	electrospray ionization
HMDB	human metabolome database
HRMS	high-resolution mass spectrometry
IMS	ion mobility spectrometry
LC	liquid chromatography
<i>m/z</i>	mass to charge ratio
MS	mass spectrometry
MSI	metabolomics standards initiative
MSMLS	mass spectrometry metabolite library of standards (Sigma-Aldrich)
QSRR	quantitative structure retention relationships
QTOF	quadrupole time-of-flight
R _t	retention time
SD	standard deviation
SMILES	simplified molecular input line entry system
SVR	support vector regression
UHPLC	ultra high-performance liquid chromatography

References

- Roessner, U.; Bowne, J. What Is Metabolomics All About? *BioTechniques* **2009**, *46*, 363–365. [CrossRef] [PubMed]
- Beger, R.D.; Dunn, W.; Schmidt, M.A.; Gross, S.S.; Kirwan, J.A.; Cascante, M.; Brennan, L.; Wishart, D.S.; Oresic, M.; Hankemeier, T.; et al. Metabolomics Enables Precision Medicine: “A White Paper, Community Perspective”. *Metabolomics Off. J. Metabolomic Soc.* **2016**, *12*, 149. [CrossRef] [PubMed]
- Trifonova, O.P.; Maslov, D.L.; Balashova, E.E.; Lokhov, P.G. Current State and Future Perspectives on Personalized Metabolomics. *Metabolites* **2023**, *13*, 67. [CrossRef]
- Ma, X. Recent Advances in Mass Spectrometry-Based Structural Elucidation Techniques. *Molecules* **2022**, *27*, 6466. [CrossRef] [PubMed]
- Zarrouk, E.; Lenski, M.; Bruno, C.; Thibert, V.; Contreras, P.; Privat, K.; Ameline, A.; Fabresse, N. High-Resolution Mass Spectrometry: Theoretical and Technological Aspects. *Toxicol. Anal. Clin.* **2022**, *34*, 3–18. [CrossRef]
- Patti, G.J.; Yanes, O.; Siuzdak, G. Innovation: Metabolomics: The Apogee of the Omics Trilogy. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263–269. [CrossRef] [PubMed]
- Barnes, S.; Benton, H.P.; Casazza, K.; Cooper, S.J.; Cui, X.; Du, X.; Engler, J.; Kabarowski, J.H.; Li, S.; Pathmasiri, W.; et al. Training in Metabolomics Research. II. Processing and Statistical Analysis of Metabolomics Data, Metabolite Identification, Pathway Analysis, Applications of Metabolomics and Its Future. *J. Mass Spectrom. JMS* **2016**, *51*, 535–548. [CrossRef]
- Nash, W.J.; Dunn, W.B. From Mass to Metabolite in Human Untargeted Metabolomics: Recent Advances in Annotation of Metabolites Applying Liquid Chromatography-Mass Spectrometry Data. *TrAC Trends Anal. Chem.* **2019**, *120*, 115324. [CrossRef]
- Sumner, L.W.; Amberg, A.; Barrett, D.; Beale, M.H.; Beger, R.; Daykin, C.A.; Fan, T.W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J.L.; et al. Proposed Minimum Reporting Standards for Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics Off. J. Metabolomic Soc.* **2007**, *3*, 211–221. [CrossRef]
- Yi, Z.; Zhu, Z.-J. Overview of Tandem Mass Spectral and Metabolite Databases for Metabolite Identification in Metabolomics. *Methods Mol. Biol. Clifton NJ* **2020**, *2104*, 139–148. [CrossRef]

11. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [[CrossRef](#)] [[PubMed](#)]
12. Guijas, C.; Montenegro-Burke, J.R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A.E.; et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* **2018**, *90*, 3156–3164. [[CrossRef](#)] [[PubMed](#)]
13. Wen, Y.; Amos, R.I.J.; Talebi, M.; Szucs, R.; Dolan, J.W.; Pohl, C.A.; Haddad, P.R. Retention Index Prediction Using Quantitative Structure-Retention Relationships for Improving Structure Identification in Nontargeted Metabolomics. *Anal. Chem.* **2018**, *90*, 9434–9440. [[CrossRef](#)] [[PubMed](#)]
14. Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z.-J. Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 11084–11091. [[CrossRef](#)]
15. Blaženović, I.; Shen, T.; Mehta, S.S.; Kind, T.; Ji, J.; Piparo, M.; Cacciola, F.; Mondello, L.; Fiehn, O. Increasing Compound Identification Rates in Untargeted Lipidomics Research with Liquid Chromatography Drift Time-Ion Mobility Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 10758–10764. [[CrossRef](#)]
16. Hinnekamp, V.; Klein, J.; Meckelmann, S.W.; Balsaa, P.; Schmidt, T.C.; Schmitz, O.J. Comparison of CCS Values Determined by Traveling Wave Ion Mobility Mass Spectrometry and Drift Tube Ion Mobility Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 12042–12050. [[CrossRef](#)]
17. Zhang, X.; Kew, K.; Reisdorph, R.; Sartain, M.; Powell, R.; Armstrong, M.; Quinn, K.; Cruickshank-Quinn, C.; Walmsley, S.; Bokatzian, S.; et al. Performance of a High-Pressure Liquid Chromatography-Ion Mobility-Mass Spectrometry System for Metabolic Profiling. *Anal. Chem.* **2017**, *89*, 6384–6391. [[CrossRef](#)]
18. Zheng, X.; Aly, N.A.; Zhou, Y.; Dupuis, K.T.; Bilbao, A.; Paurus, V.L.; Orton, D.J.; Wilson, R.; Payne, S.H.; Smith, R.D.; et al. A Structural Examination and Collision Cross Section Database for over 500 Metabolites and Xenobiotics Using Drift Tube Ion Mobility Spectrometry. *Chem. Sci.* **2017**, *8*, 7724–7736. [[CrossRef](#)]
19. Righetti, L.; Bergmann, A.; Galaverna, G.; Rolfsson, O.; Paglia, G.; Dall'Asta, C. Ion Mobility-Derived Collision Cross Section Database: Application to Mycotoxin Analysis. *Anal. Chim. Acta* **2018**, *1014*, 50–57. [[CrossRef](#)]
20. Picache, J.A.; Rose, B.S.; Balinski, A.; Leaptrot, K.L.; Sherrod, S.D.; May, J.C.; McLean, J.A. Collision Cross Section Compendium to Annotate and Predict Multi-Omic Compound Identities. *Chem. Sci.* **2019**, *10*, 983–993. [[CrossRef](#)]
21. Hernández-Mesa, M.; Le Bizec, B.; Monteau, F.; García-Campaña, A.M.; Dervilly-Pinel, G. Collision Cross Section (CCS) Database: An Additional Measure to Characterize Steroids. *Anal. Chem.* **2018**, *90*, 4616–4625. [[CrossRef](#)] [[PubMed](#)]
22. Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z.-J. Ion Mobility Collision Cross-Section Atlas for Known and Unknown Metabolite Annotation in Untargeted Metabolomics. *Nat. Commun.* **2020**, *11*, 4334. [[CrossRef](#)] [[PubMed](#)]
23. Zhou, Z.; Tu, J.; Zhu, Z.-J. Advancing the Large-Scale CCS Database for Metabolomics and Lipidomics at the Machine-Learning Era. *Curr. Opin. Chem. Biol.* **2018**, *42*, 34–41. [[CrossRef](#)]
24. Rainey, M.A.; Watson, C.A.; Asef, C.K.; Foster, M.R.; Baker, E.S.; Fernández, F.M. CCS Predictor 2.0: An Open-Source Jupyter Notebook Tool for Filtering Out False Positives in Metabolomics. *Anal. Chem.* **2022**, *94*, 17456–17466. [[CrossRef](#)] [[PubMed](#)]
25. Plante, P.-L.; Francovic-Fontaine, É.; May, J.C.; McLean, J.A.; Baker, E.S.; Laviolette, F.; Marchand, M.; Corbeil, J. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. *Anal. Chem.* **2019**, *91*, 5191–5199. [[CrossRef](#)]
26. Zhou, Z.; Xiong, X.; Zhu, Z.-J. MetCCS Predictor: A Web Server for Predicting Collision Cross-Section Values of Metabolites in Ion Mobility-Mass Spectrometry Based Metabolomics. *Bioinform. Oxf. Engl.* **2017**, *33*, 2235–2237. [[CrossRef](#)]
27. Zhou, Z.; Tu, J.; Xiong, X.; Shen, X.; Zhu, Z.-J. LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision To Support Ion Mobility-Mass Spectrometry-Based Lipidomics. *Anal. Chem.* **2017**, *89*, 9559–9566. [[CrossRef](#)]
28. Grisoni, F.; Ballabio, D.; Todeschini, R.; Consonni, V. Molecular Descriptors for Structure-Activity Applications: A Hands-On Approach. *Methods Mol. Biol. Clifton NJ* **2018**, *1800*, 3–53. [[CrossRef](#)]
29. Ross, D.H.; Cho, J.H.; Xu, L. Breaking Down Structural Diversity for Comprehensive Prediction of Ion-Neutral Collision Cross Sections. *Anal. Chem.* **2020**, *92*, 4548–4557. [[CrossRef](#)]
30. Rainville, P.D.; Wilson, I.D.; Nicholson, J.K.; Isaac, G.; Mullin, L.; Langridge, J.I.; Plumb, R.S. Ion Mobility Spectrometry Combined with Ultra Performance Liquid Chromatography/Mass Spectrometry for Metabolic Phenotyping of Urine: Effects of Column Length, Gradient Duration and Ion Mobility Spectrometry on Metabolite Detection. *Anal. Chim. Acta* **2017**, *982*, 1–8. [[CrossRef](#)]
31. Stanstrup, J.; Neumann, S.; Vrhovšek, U. PredRet: Prediction of Retention Time by Direct Mapping between Multiple Chromatographic Systems. *Anal. Chem.* **2015**, *87*, 9421–9428. [[CrossRef](#)] [[PubMed](#)]
32. Falchi, F.; Bertozzi, S.M.; Ottonello, G.; Ruda, G.F.; Colombano, G.; Fiorelli, C.; Martucci, C.; Bertorelli, R.; Scarpelli, R.; Cavalli, A.; et al. Kernel-Based, Partial Least Squares Quantitative Structure-Retention Relationship Model for UPLC Retention Time Prediction: A Useful Tool for Metabolite Identification. *Anal. Chem.* **2016**, *88*, 9510–9517. [[CrossRef](#)] [[PubMed](#)]
33. Creek, D.J.; Jankevics, A.; Breitling, R.; Watson, D.G.; Barrett, M.P.; Burgess, K.E.V. Toward Global Metabolomics Analysis with Hydrophilic Interaction Liquid Chromatography-Mass Spectrometry: Improved Metabolite Identification by Retention Time Prediction. *Anal. Chem.* **2011**, *83*, 8703–8710. [[CrossRef](#)] [[PubMed](#)]
34. Bonini, P.; Kind, T.; Tsugawa, H.; Barupal, D.K.; Fiehn, O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. *Anal. Chem.* **2020**, *92*, 7515–7522. [[CrossRef](#)]

35. Liapikos, T.; Zisi, C.; Kodra, D.; Kademoglou, K.; Diamantidou, D.; Begou, O.; Pappa-Louisi, A.; Theodoridis, G. Quantitative Structure Retention Relationship (QSRR) Modelling for Analytes' Retention Prediction in LC-HRMS by Applying Different Machine Learning Algorithms and Evaluating Their Performance. *J. Chromatogr. B* **2022**, *1191*, 123132. [[CrossRef](#)]
36. Park, S.H.; De Pra, M.; Haddad, P.R.; Grosse, S.; Pohl, C.A.; Steiner, F. Localised Quantitative Structure-Retention Relationship Modelling for Rapid Method Development in Reversed-Phase High Performance Liquid Chromatography. *J. Chromatogr. A* **2020**, *1609*, 460508. [[CrossRef](#)]
37. Goryński, K.; Bojko, B.; Nowaczyk, A.; Buciniński, A.; Pawliszyn, J.; Kaliszan, R. Quantitative Structure-Retention Relationships Models for Prediction of High Performance Liquid Chromatography Retention Time of Small Molecules: Endogenous Metabolites and Banned Compounds. *Anal. Chim. Acta* **2013**, *797*, 13–19. [[CrossRef](#)]
38. Naylor, B.C.; Catrow, J.L.; Maschek, J.A.; Cox, J.E. QSRR Automator: A Tool for Automating Retention Time Prediction in Lipidomics and Metabolomics. *Metabolites* **2020**, *10*, 237. [[CrossRef](#)]
39. Gritti, F. Perspective on the Future Approaches to Predict Retention in Liquid Chromatography. *Anal. Chem.* **2021**, *93*, 5653–5664. [[CrossRef](#)]
40. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
41. Dusautoir, R.; Zarcone, G.; Verrielle, M.; Garçon, G.; Fronval, I.; Beauval, N.; Allorge, D.; Riffault, V.; Locoge, N.; Lo-Guidice, J.-M.; et al. Comparison of the Chemical Composition of Aerosols from Heated Tobacco Products, Electronic Cigarettes and Tobacco Cigarettes and Their Toxic Impacts on the Human Bronchial Epithelial BEAS-2B Cells. *J. Hazard. Mater.* **2021**, *401*, 123417. [[CrossRef](#)]
42. Dunn, W.B.; Erban, A.; Weber, R.J.M.; Creek, D.J.; Brown, M.; Breitling, R.; Hankemeier, T.; Goodacre, R.; Neumann, S.; Kopka, J.; et al. Mass Appeal: Metabolite Identification in Mass Spectrometry-Focused Untargeted Metabolomics. *Metabolomics* **2013**, *9*, 44–66. [[CrossRef](#)]
43. Bittremieux, W.; Wang, M.; Dorrestein, P.C. The Critical Role That Spectral Libraries Play in Capturing the Metabolomics Community Knowledge. *Metabolomics Off. J. Metabolomic Soc.* **2022**, *18*, 94. [[CrossRef](#)] [[PubMed](#)]
44. Pezzatti, J.; González-Ruiz, V.; Codesido, S.; Gagnebin, Y.; Joshi, A.; Guillaume, D.; Schappler, J.; Picard, D.; Boccad, J.; Rudaz, S. A Scoring Approach for Multi-Platform Acquisition in Metabolomics. *J. Chromatogr. A* **2019**, *1592*, 47–54. [[CrossRef](#)] [[PubMed](#)]
45. Liebal, U.W.; Phan, A.N.T.; Sudhakar, M.; Raman, K.; Blank, L.M. Machine Learning Applications for Mass Spectrometry-Based Metabolomics. *Metabolites* **2020**, *10*, 243. [[CrossRef](#)] [[PubMed](#)]
46. Domingo-Almenara, X.; Guijas, C.; Billings, E.; Montenegro-Burke, J.R.; Uritboonthai, W.; Aisporna, A.E.; Chen, E.; Benton, H.P.; Siuzdak, G. The METLIN Small Molecule Dataset for Machine Learning-Based Retention Time Prediction. *Nat. Commun.* **2019**, *10*, 5811. [[CrossRef](#)] [[PubMed](#)]
47. Tsugawa, H.; Ikeda, K.; Takahashi, M.; Satoh, A.; Mori, Y.; Uchino, H.; Okahashi, N.; Yamada, Y.; Tada, I.; Bonini, P.; et al. A Lipidome Atlas in MS-DIAL 4. *Nat. Biotechnol.* **2020**, *38*, 1159–1163. [[CrossRef](#)] [[PubMed](#)]
48. Møllerup, C.B.; Mardal, M.; Dalsgaard, P.W.; Linnert, K.; Barron, L.P. Prediction of Collision Cross Section and Retention Time for Broad Scope Screening in Gradient Reversed-Phase Liquid Chromatography-Ion Mobility-High Resolution Accurate Mass Spectrometry. *J. Chromatogr. A* **2018**, *1542*, 82–88. [[CrossRef](#)]
49. Celma, A.; Bade, R.; Sancho, J.V.; Hernandez, F.; Humphries, M.; Bijlsma, L. Prediction of Retention Time and Collision Cross Section (CCSH+, CCSH-, and CCSNa+) of Emerging Contaminants Using Multiple Adaptive Regression Splines. *J. Chem. Inf. Model.* **2022**, *62*, 5425–5434. [[CrossRef](#)]
50. Ross, D.H.; Cho, J.H.; Zhang, R.; Hines, K.M.; Xu, L. LiPydomics: A Python Package for Comprehensive Prediction of Lipid Collision Cross Sections and Retention Times and Analysis of Ion Mobility-Mass Spectrometry-Based Lipidomics Data. *Anal. Chem.* **2020**, *92*, 14967–14975. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Lastname : Lenski **Surname :** Marie

Title of the thesis: Development and application of a metabolomic approach for identification of early markers of toxicity of tobacco, electronic-cigarette and heated tobacco product

Key words: Cigarette, electronic cigarette, heated tobacco products, *in vitro* experimental toxicology, metabolomics, analytical development, machine learning.

Abstract: Combustible cigarette smoking is an established risk factor for various pathologies including lung cancer. Electronic cigarettes and heated tobacco are new products that have appeared on the market in recent years and are sometimes used as a smoking cessation aid. However, harmlessness or toxicity of electronic cigarettes or heated tobacco products has not yet been proved. This study aims to explore and compare the metabolome of human lung cells exposed to emissions of electronic cigarette, heated tobacco products, or tobacco cigarette, in order to highlight their specific metabolic fingerprints and potential early markers of toxicity.

The first objective of this thesis was to develop an untargeted metabolomic method. Liquid chromatography and high-resolution mass spectrometry methods were developed and optimised using chemical standards. Machine learning tools were used to build a reference database of almost 114 000 metabolites, facilitating interpretation of untargeted metabolomics data.

The second objective of this doctoral project was to analyze the metabolome of immortalized human lung epithelial cells (BEAS-2B) cultured at air-liquid interface and exposed to air (control) or emissions from reference 3R4F cigarette (1 min or 2 min), heated tobacco products (30 min or 60 min), or electronic cigarette set up at low (18 W) or high (30 W) power (30 min or 60 min) using the Vitrocell Smoking machine. Duration of exposures were chosen based on comparable sub-toxic doses (> 80% cell viability) previously reported, to evaluate a potential time-dependant effect. Chemometric analysis of metabolomics data highlighted that both 3R4F and heated tobacco product emissions significantly affected metabolome, whereas no difference was observed after electronic cigarette exposures regardless of its power and duration of exposure, compared to controls. The metabolomic fingerprint highlighted following exposure to tobacco products consisted of exogenous compounds, some of which are cancerogenic, as well as endogenous metabolites, which are markers of effects. Their dysregulation indicates alterations in various metabolic pathways, including oxidative stress, energy metabolism and lipid metabolism.

This metabolomic strategy gives new insights to discover potential markers of disrupted metabolic pathways after cigarette or heated tobacco product exposure, both tobacco products. These *in vitro* results strongly suggest a lower toxicity of electronic cigarette compared to those of 3R4F cigarette and heated tobacco product emissions in the BEAS-2B cell line. The dysregulated metabolites are involved in metabolic pathways that are also altered in respiratory diseases, confirming that the toxicity of heated tobacco products should not be underestimated. The present data agree with transcriptomic results obtained on the same model and constitute a baseline for further cross-disciplinary research by integrative biology. Studies on longer-term effects are needed.

Nom : Lenski **Prénom :** Marie

Titre de la thèse : Développement et application d'une approche de métabolomique pour l'identification de biomarqueurs précoces de toxicité du tabac, de la cigarette électronique et du tabac chauffé.

Mots-clés : Cigarette, cigarette électronique, tabac chauffé, toxicologie expérimentale *in vitro*, métabolomique, développement analytique, *machine learning*.

Résumé : Le tabagisme est un facteur de risque établi pour diverses pathologies, dont le cancer du poumon. La cigarette électronique et le tabac chauffé sont de nouveaux produits arrivés sur le marché depuis quelques années et utilisés parfois comme aide au sevrage tabagique. Pourtant l'innocuité de ces produits n'est pas établie actuellement, du fait d'un manque d'études toxicologiques approfondies. Nos travaux visent à explorer et à comparer le métabolome de cellules pulmonaires humaines exposées à des émissions de cigarette électronique, de tabac chauffé ou de cigarettes de tabac, afin de mettre en évidence leurs empreintes métaboliques spécifiques et de potentiels biomarqueurs précoces de toxicité.

Le premier objectif de ce travail était de développer une méthode de métabolomique non-ciblée. Les méthodes de chromatographie liquide et de spectrométrie de masse haute résolution ont été développées et optimisées à partir de standards de métabolites. L'utilisation d'outils de *machine learning* a permis de mettre en place une base de données de référence d'environ 114 000 métabolites, facilitant l'analyse des données de métabolomique.

Le deuxième objectif de ce projet doctoral consistait à analyser le métabolome de cellules épithéliales pulmonaires humaines immortalisées (BEAS-2B) cultivées à l'interface air-liquide et exposées à de l'air stérile (contrôle) ou aux émissions d'une cigarette de référence 3R4F (1 min ou 2 min), de tabac chauffé (30 min ou 60 min), ou de cigarette électronique réglée à faible (18 W) ou moyenne (30 W) puissance (30 min ou 60 min) en utilisant la machine à fumer Vitrocell. Les durées d'exposition ont été choisies sur la base de doses sub-toxiques comparables (> 80 % de viabilité cellulaire) précédemment rapportées, afin d'évaluer un effet potentiel dépendant du temps. L'analyse chimométrique des données de métabolomique a permis de souligner que les émissions de 3R4F et de tabac chauffé ont affecté de manière significative le métabolome par rapport aux contrôles, alors qu'aucune différence n'a été observée après les expositions à la cigarette électronique, quelles que soient sa puissance d'utilisation et la durée d'exposition. La signature métabolomique mise en évidence à la suite de l'exposition aux produits du tabac était constituée de composés exogènes, dont certains cancérigènes, ainsi que de métabolites endogènes, marqueurs d'effets. Leur dérégulation signe des altérations de divers voies métaboliques et notamment le stress oxydant, le métabolisme énergétique et le métabolisme des lipides.

Cette stratégie métabolomique offre de nouvelles perspectives pour une meilleure compréhension des variations du métabolisme cellulaire après une exposition à la cigarette ou au tabac chauffé, deux produits du tabac. Globalement, ces analyses *in vitro* suggèrent une toxicité moindre des aérosols de cigarette électronique par rapport à celle des émissions de la cigarette 3R4F et du tabac chauffé dans la lignée cellulaire BEAS-2B. Les métabolites dérégulés sont impliqués dans des voies métaboliques également altérées lors de pathologies respiratoires, confirmant qu'il ne faut pas sous-estimer la toxicité des produits du tabac chauffé. Ces données sont en accord avec les profils transcriptomiques obtenus sur le même modèle et constituent une base pour de futures recherches transversales par biologie intégrative. Des études sur les effets à plus long terme sont nécessaires.