UNIVERSITÉ DE TOURS

ÉCOLE DOCTORALE : SSBCV

INSERM UMR 1253

iBrain – Imagerie, Biomarqueurs et Thérapie

Equipe 3

THÈSE présentée par :

Cécile MARTIAS

soutenue le : **16 décembre 2021**
pour obtenir le grade de : **Docteur de l'université de Tours**

Discipline/ Spécialité : Chimie

Phénotypage métabolique multi-matrices biologiques et multi-plateformes analytiques

THÈSE dirigée par :
M. EMOND Patrick PU-PH, Université de Tours

Co-encadrée par :
Mme NADAL-DESBARATS Lydie MCU, Université de Tours

RAPPORTEURS :
Mme LE MOYEC Laurence Pr, Université d'Evry Val d'Essone
M. MAUNIT Benoit Pr, Université de Clermont Auvergne

JURY :

Mme BERTRAND-MICHEL Justine	Ingénieur de recherche	I2MC Toulouse
M. CHOURPA Igor	Professeur des universités	Université de Tours
M. EMOND Patrick	PU-PH	Université de Tours
Mme LE MOYEC Laurence	Professeur des universités	Université d'Evry Val d'Essone
M. MAUNIT Benoit	Professeur des universités	Université de Clermont Auvergne
Mme NADAL-DESBARATS Lydie	Maitre de conférences	Université de Tours
Mme PLE Karen	Chargé de recherche	Université d'Orléans

Remerciements

Ce projet de thèse a été financé par la région Centre-Val de Loire et ce travail a été réalisé à l'université de Tours, au sein du laboratoire INSERM UMR 1253 dirigé par Pr Catherine Belzung que je remercie.

Je tiens d'abord à remercier les membres du jury qui ont accepté d'évaluer ce travail.

Je voudrais remercier mon directeur de thèse, Pr Emond Patrick, pour sa confiance pendant ce projet et sa façon à toujours pousser la réflexion plus loin.

Je remercie mon encadrante Dr Lydie Nadal-Desbarats pour sa disponibilité à toutes épreuves, sa patience, ses conseils et pour toutes ses heures de réflexion à en avoir des cheveux blancs. Merci du soutien pendant toutes ces 3 années surtout pendant le confinement.

Nadine Baroukh qui m'a beaucoup aidée à structurer mes idées mais également pour toutes ses connaissances en biologie. Je suis enfin arrivée au bout du tourbillon en étant alignée ! Je tiens également à remercier grandement Karen Plé pour sa disponibilité que ce soit par mail ou par visio et pour tous ses échanges qui m'ont énormément aidé en chimie organique. Merci Julie G. pour tous ses échanges sur le monde de l'élevage, j'en ai beaucoup appris !

Je remercie grandement toute l'équipe Métabo, Sylvie Mavel, Antoine Lefevre, Frédéric Montigny pour toutes les fois où vous m'avez redonné confiance en la métabolomique. Merci Antoine L. de m'avoir laissé la garde de ton « bébé » pendant tes vacances. J'en ai pris soin !

Merci à Julie B. je me souviendrais de ton rire communicatif. Merci bobo pour tous ces chouettes moments passés. Merci Zuhal sans toi mes journées de manip' n'auraient pas été les mêmes tu es ma SAUVEUSE de manips !

Je remercie chaleureusement mes compagnons de galère/bureau Alex et Camille pour tous ses moments de bavardage scientifique (ou pas), de grignotage de chocobon, de rigolade, de réflexion sur la société actuelle x), à nos séances professionnelles de patinage ! Vous avez été là pour moi à me conseiller face à mes problèmes de chimie. Camille ma copine de souris ! … On en a fait des choses ! Je dirais juste « Casse la démarche ! ». Je remercie également Antoine P. pour tous ses conseils en informatique je reviens de loin !

Je remercie toute l'équipe de l'étage 4 (Géraldine, Clovis, Sophie, Sylvie Chalon, Laurent, Claire) mais aussi ceux de l'étage 3. C'est bon je ne vais plus déranger la tranquillité de l'étage avec mes rires. Merci Linda pour toutes ses heures à rechercher des produits chimiques à travers le monde.

Je remercie également mes amis qui m'ont bcp soutenue pendant ses 3 années Claudia, Miléna les Loutres , les Perclures mais surtout Anne-Laure pour ton soutien inestimable. Ca y est ! C'est fini les audios/messages dépressifs tous les jours. ALL THE WAY UP !

Merci à ma famille, mes parents, mes frères, mes belles-sœurs, neveux et nièces qui ont été d'un soutien inestimable. Merci de m'avoir fait confiance pendant toutes ses années d'études et de croire en moi. Je ne serais pas arrivée au bout sans vous !

Mr Manuel Taibauhlt Gère qu'est-ce que je serais devenue sans ton appui moral et physique ? Ces 3 années n'ont vraiment pas été faciles mais c'est bon ! Si tu lis ça c'est que c'est enfin le JDC ! I'm BACK !

Résumé

Un des objectifs de la métabolomique est d'identifier et quantifier l'ensemble des molécules (métabolome) d'échantillons prélevés d'un individu. Cette image métabolique instantanée permet de caractériser l'état de santé de cet individu puis de suivre son évolution. Par ailleurs, la comparaison de cet état métabolique à des situations pathologiques permet l'identification de biomarqueurs de diagnostic, de pronostic et le suivi de la réponse à un traitement. Ces objectifs nécessitent de disposer d'outils d'exploration du métabolome qui puissent être standardisés, les plus larges possibles et quantitatifs. L'objectif de ce projet est la mise au point de procédures standardisées et quantitatives d'exploration du métabolome chez l'humain et le caprin.

Dans une première partie, nous nous sommes proposés de développer une méthodologie robuste capable de décrire le métabolome le plus complet possible d'un individu. Pour se faire, différents types d'échantillons ont été utilisés : sang, urine, fèces et salive chez l'Homme ; sang, lait et fèces chez la chèvre. L'ensemble de ces prélèvements a été analysé à l'aide de plateformes analytiques multiples (Résonance Magnétique Nucléaire et Spectrométrie de Masse). Afin de répondre à des contraintes d'analyses à haut débit et transférables en routine de biologie médicale, nous avons recherché un compromis entre la réduction de l'étape pré-analytique et la description métabolique la plus exhaustive possible. Ainsi, une préparation unique de chaque échantillon pour les plateformes LC (RP-LC et HILIC) – Q Orbitrap et ^{1}H-RMN a été privilégiée. Les critères de choix de la meilleure préparation sont le nombre de métabolites détectés et la variabilité analytique obtenue. L'ensemble des expériences et résultats obtenus pour les prélèvements humains est décrit dans la publication (Martias et al. Molecules 2021, 26(14), 4111). Ce travail propose également la représentation des données sous la forme d'une carte métabolomique individuelle. Celle-ci pourrait être utilisée ultérieurement lors d'étude clinique. La même approche méthodologique a été effectuée pour l'analyse

du métabolome caprin. Avec comme objectif final de pouvoir détecter la présence de maladies telles que les mammites. Ce travail est réalisé en partenariat avec la société Allice (Nouzilly, France) et est présenté dans la publication (Martias et al. Metabolites 2021, 11(10), 681). Cette étude propose l'accès de façon fiable à un métabolome sous forme d'atlas et peut permettre la recherche de biomarqueurs de diagnostic chez le caprin, pour le suivi de l'état de santé des troupeaux pour une détection très précoce de la mise en place de pathologie dans un élevage.

Dans une seconde partie, nous avons débuté un travail afin de répondre à la problématique de quantification du métabolome par spectrométrie de masse. Les méthodologies proposées utilisent des dérivations de groupes chimiques tels que les acides carboxyliques au moyen d'étiquettes chimiques (Tag). Quatre Tags utilisant le couplage peptidique ont été testés avec 3 agents de couplage. Ces protocoles ont été testés dans un premier temps avec des standards afin de sélectionner les meilleurs étiquettes et agent de couplage. Cependant, cette approche de marquage en deux étapes induit des variabilités importantes de rendement d'incorporation. Un développement méthodologique mono-étape sans agent de couplage est à privilégier.

L'ensemble de ce travail a permis la mise en évidence des avantages de l'exploration du métabolome par une approche multi-matrices et multiplateformes chez l'humain et le caprin. Ces méthodes permettent d'obtenir une couverture métabolique importante de manière robuste. Cette approche pourrait être transférée en santé animale ou humaine. Afin d'avoir une information longitudinale, le développement de méthodes de quantification est une étape clé. Dans ce but, ce développement quantitatif sera poursuivi avec des protocoles mono-étape.

Abstract

One of the aim of metabolomics is to identify and quantify the entire scope of molecules (metabolome) from samples withdrawn from an individual. This instantaneous metabolic image allows to characterize the health status of this individual and to follow his evolution. In addition, the comparison of this metabolic state with pathological situations allows the identification of diagnostic biomarkers, prognosis but also the following of a treatment response. These objectives require the availability of tools for the exploration of metabolome that can be standardized, as exhaustive as possible and quantitative. The objective of this project is development of standardized operating procedure and quantitative procedures for the exploration of metabolome for humans and for goats.

In the first part, we focus on the development of a robust methodology able to describe the most complete metabolome as possible of an individual. To achieve our goal, different types of samples were used: blood, urine, faeces and saliva from humans and blood, milk and faeces have been used for goats. All samples were analyzed using multiple analytical platforms (Nuclear Magnetic Resonance and Mass Spectrometry). The objective was to meet the constraints of high throughput analyzes that can be transferred to routine medical biology. We looked for a compromise between reducing the pre-analytical step while preserving the most exhaustive metabolic description possible. Thus, a unique preparation of each sample was privileged for the LC (RP-LC and HILIC) - Q Orbitrap and for 1H-NMR platforms. The process to select the suitable preparation took in consideration the number of detected metabolites and the analytical variability. All the experiments and results obtained for human samples are described in the publication (Martias et al. Molecules 2021, 26 (14), 4111). This work also proposes the representation of data in the form of an individual metabolomic map. This map could be used later in some clinical study. The same methodological approach was carried out for the analysis of the goat metabolome. With the final objective of being able to detect the

presence of diseases such as mastitis. This activity has been carried out in partnership with the company Allice (Nouzilly, France) and is presented in the publication (Martias et al. Metabolites 2021, 11 (10), 681). The study provides reliable access to a metabolome in the form of an atlas. It may allow the search for diagnostic biomarkers in goats, for monitoring the health status of herds. It could also permit very early detection of the development of pathology in a goat farm.

In a second part, we initiated the process to find an approach for the quantification problem of the metabolome by mass spectrometry. The proposed methodologies use derivations of chemical groups such as carboxylic acids by means of chemical labels (Tag). Four Tags using peptide coupling were tested with 3 coupling agents. These protocols were first tested with acid standards in order to select the suitable Tag and the suitable coupling agent. However, this two-steps labeling approach induces significant variability in incorporation yield. A single-step methodological development without coupling agent is to be privileged.

All of this work has demonstrated the advantages of exploring the metabolome using a multi-matrix and multiplatform approach for humans and for goats. These methods allow to obtain a significant metabolic coverage. Thus it describes a large number of metabolic pathways in a robust manner. This approach could be transferred to animal or to human health. In order to have longitudinal information on animal and human health, the development of quantification methods is a key step. For this purpose, the development of quantitative methods will be continued with the single-step protocols.

Table des matières

Remerciements 3

Résumé 7

Abstract 9

Table des matières 13

Abréviations 17

Liste des tableaux 20

Liste des figures 23

Chapitre 1 : Introduction générale – Etat de l'art 29

I. Historique 29

II. Généralités 30

1. Métabolites et Métabolome 30
2. Applications de la métabolomique 32

III. Déroulement général d'une analyse métabolomique 33

IV. Collection et préparation des échantillons 34

1. Collection des échantillons 34
   - 1.1. Echantillon sanguin 34
   - 1.2. Echantillon urinaire 35
   - 1.3. Echantillon salivaire 35
   - 1.4. Echantillon fécal 36
2. Acheminement au laboratoire 36
3. Stockage des échantillons 36
   - 3.1. Echantillon sanguin 37
   - 3.2. Echantillon urinaire 38
   - 3.3. Echantillon fécal 38
4. Préparation des échantillons 39
   - 4.1. Lyophilisation 39
   - 4.2. Centrifugation 40
   - 4.3. Dilution 41
   - 4.4. Précipitation des protéines 41
   - 4.5. Extraction liquide-liquide 42
   - 4.6. Extraction en phase solide (SPE) 44

V. La spectroscopie par résonnance magnétique nucléaire (RMN) 45

1. Le spectromètre RMN 45
2. Acquisition des données 48
   - 2.1. RMN unidimensionnelle (1D) 48
   - 2.2. RMN multidimensionnelle (2D) 49
3. Traitement des données 50
   - 3.1. Pré-traitement 50
   - 3.2. Normalisation 51

4. Logiciels de traitement 53
5. Avantages et inconvénients de la RMN 55

VI. La spectrométrie de masse (MS) 56
1. Le spectromètre de masse 56
1.1. Les sources d'ionisation 56
1.2. Les analyseurs 60
1.3. Couplage d'analyseurs / approche en tandem 65
1.4. Couplage avec la chromatographie gazeuse 67
1.5. Couplage à la chromatographie liquide 67
1.5.1. Chromatographie en phase inverse (RP) 68
1.5.2. Chromatographie à interactions hydrophiles (HILIC) 68
2. Acquisition des données 69
2.1. Acquisition des données non ciblées 69
2.2. Acquisition des données ciblées 71
3. Logiciels de traitement et prétraitement des données 72
4. Traitement des données 76
4.1. Identification 76
4.2. Normalisation des données 78
5. Avantages et inconvénients de la spectrométrie de masse 80

VII. Approches multiples 80
1. Approche multi-matrices 80
2. Approche multiplateformes 82

VIII. Quantification 86
1. Introduction 86
2. La quantification relative 88
3. La quantification absolue 89
3.1. Courbe de calibration 89
3.2. Etalon interne 90
3.2.1. Par synthèse chimique 91
3.2.2. Echange isotopique 92
3.2.3. Par microorganisme 92
3.2.4. Incorporation des isotopes par kit commercial 93
3.3. Dérivation chimique 94
3.3.1. Conception des étiquettes chimiques 94

IX. Objectifs de la thèse 95

Chapitre 2 : Développement de méthodes standardisées et exploration multiplateformes et multi matrices du métabolome 99

I. Contexte général 99
A. Chez l'humain 100
1. Contexte et objectif 100
2. Principaux résultats et conclusion 101
A.1. Publication: Optimization of sample preparation for metabolomics exploration of urine, feces, blood and saliva in humans using NMR and UHPLC-HRMS platforms 102
B. Chez le modèle animal caprin 119
1. Contexte et objectif 119

2. Principaux résultats et conclusion 120
B.1. Publication: Analytical methodology for a metabolome atlas in goat's plasma, milk and feces using ^{1}H-NMR and UHPLC-MS 120

II. Conclusion générale et perspectives 134

Chapitre 3 : Quantification de métabolites par dérivation chimique 137

I. Introduction 137
1. Intérêt de la dérivation chimique en spectrométrie de masse 137
2. A quoi sert la dérivation chimique ? 139
2.1. Séparation LC 140
2.2. Détection 141
2.3. Stabilité 142
2.4. Quantification 142

II. Dérivation des fonctions chimiques 143
1. Fonction acide carboxylique 143
1.1. Dérivés de pipérazine et amine : couplage en deux étapes 144
1.1.1. Première étape : Activation du groupement carboxyle 144
1.1.2. Problème d'épimérisation/racémisation 151
1.1.3. Deuxième étape : ajout de l'étiquette chimique 153
1.2. Dérivé bromé : couplage en une étape 155
2. Fonctions amine et phénol 157
2.1. Dérivés du Dansyle 157
2.2. Dérivé isothiocyanate 160
3. Fonctions alcool et thiol 161
4. Fonctions carbonyles 163

III. Plan expérimental 164

IV. Matériels et méthodes 165
1. Produits chimiques et réactifs 165
1.1. Standards de métabolites 165
1.2. Etiquettes chimiques, agents de couplage et bases 167
2. Protocole de marquage des fonctions acides 168
2.1. Marquage en deux étapes avec les dérivés pipérazine et amine 168
2.1.1. Protocole CMPI 168
2.1.2. Protocole HOAt/EDC 169
2.1.3. Protocole HATU 170
2.2. Marquage en une étape avec le DmPABr 172
3. Protocole de marquage des fonctions amine et phénol par les étiquettes chlorure de Dansyle et PITC 173
3.1. Protocole chlorure de Dansyle 173
3.2. Protocole PITC 174

V. Acquisition des données 175
1. Paramètres LC 175
2. Paramètres MS 175

VI. Résultats et discussion 176
1. Marquage en deux étapes par les dérivés pipérazine et amine de la fonction acide carboxylique 176
1.1. Cas des acides gras et triacide 177

1.2. Cas des acides aminés 180
1.3. Cas des α-hydroxy acides 186
1.4. Bilan du marquage en deux étapes 191
2. Marquage en une étape par le DmPABr de la fonction acide carboxylique 193
2.1. Cas des acides gras et triacides 193
2.2. Cas des acides aminés et α-hydroxy acides 195
2.3. Acquisition MS^2 198
2.3.1. Cas des acides gras triacides 198
2.3.2. Cas des acides aminés et α-hydroxy acides 201
2.3.3. Bilan 205
2.4. Bilan du marquage en une étape 206
3. Comparaison du DmPABr avec des étiquettes chimiques connues : chlorure de Dansyl et PITC 207
3.1. L'étiquette DmPABr marque-t-elle les fonctions amine et alcool ? 208
3.2. L'étiquette DmPABr est-elle plus efficace que le chlorure de Dansyl et le PITC ? 210
3.2.1. Le pseudo-rendement de marquage 211
3.3. Bilan 213

VII. Conclusion 215

Conclusion générale 218

ANNEXE 222

REFERENCES 226

Abréviations

2D: bidimensionnelle
AAE: Agent Anti-épimérisation
ACN : Acétonitrile
AF: Acide Formique
AGC: Automatic Gain Control / Contrôle Automatique du Gain
APC : halogénure de 4-(2-(triméthylammonio)éthoxy)benzénamine
APCI : Atmospheric pressure chemical ionization / ionisation chimique à pression atmosphérique
BOP: Benzotriazol-1-yloxy-tris(dimethylamino)- phosphonium hexafluorophosphate
$CHCl_3$: Chloroforme
CI: Chemical ionization / ionisation chimique
CMPI: 2-Chloro-1-methylpyridinium iodide
COSY: Correlation Spectroscopy / Spectroscopie bidimensionnelle correlation homonucléaire
CPMG: Carr-Purcell-Meiboom-Gil
CV : coefficient de variation
DAPPZ : 1-((4-dimethylaminophenyl)carbonyl) piperazine
DCC : Dicyclohexylcarbodiimide
DDA : Data Dependant acquisition / Acquisition en donnée dépendante
DENSYL: chlorure de 5-naphtalène-1-sulfonYL
DIA : Data Independent acquisition / Acquisition en donnée indépendante
DMED : 2-Dimethylaminoethylamine
DmPABr : 4-(Dimethylamino)phenacyl bromide
DMPP : 2,4-Dimethoxy-6-(piperazin-1-yl)pyrimidine
DMPT : 2,4-Dimethoxy-6-(piperazin-1-yl)s-triazine
DMT-MM: Sel de 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)- 4-methylmorpholinium
EDC: 1-ethyl-3-(3-diméthylaminopropyl)carbodiimide
EI : electron ionization / impact électronique
ESI : Electrospray Ionization / ionisation par électrospray
FID : Free Induction Decay / Décroissance libre de l'induction
FT-ICR-MS : Fourier Transform Ion Cyclotron Resonance Mass Spectrometry / Spectrométrie de masse à resonance cyclonique ionique
GC: Gas Chromatography / Chromatographie gazeuse
H/D: Hydrogène/Deutérium
HATU: 2-(7-Aza-1H-benzotriazol-1-yl)-N,N,N',N'- tetramethylaminium tetrafluoroborate/ hexafluorophosphate
HBTU: 2-(1H-Benzotriazol-1-yl)N,N,N,N-tetramethylaminium hexafluorophosphate
HCD: Higher-energy Collision Dissociation / Dissociation induite par collision après activation non résonnante
HILIC: Hydrophilic interaction chromatography / Chromatographie d'interaction hydrophile

HIQB: 2-(2-Hydrazinyl-2-oxoethyl) isoquinolin-2-ium bromide
HMBC : Heteronuclear Multiple Bond Correlation / Spectroscopie bidimensionnelle avec correlation hétéronucléaire
HMDB : Human Metabolome Database / Base de données du metabolome humain
HMPA: hexamethylphosphoramide
HOAt: 1-hydroxy-7-azabenzotriazole
HOBt: 1-Hydroxybenzotriazole
HODhat: Hydroxy-3,4-dihydro-4-oxo-1,2,3- aza-benzotriazine
HSQC : Heteronuclear Single Quantum Correlation / Spectroscopie bidimensionnelle avec correlation hétéronucléaire
LC: Liquid Chromatography / Chromatographie liquide
LLE: Liquid-liquid extraction / Extraction liquide/liquide
LOESS: Locally Estimated Scatterplot Smoothing / regression locale
m/z: masse sur charge
MALDI: matrix-assisted laser desorption/ionization
MASCl: 10 méthyl-acridone-2-sulfonyl chloride
MeOH: Méthanol
MRM: Multiple Reaction Monitoring
MS : Mass spectrometry / Spectrométrie de masse
MS/MS : Spectrométrie en tandem
MTBE : methyl-tert-butyl ether
NMR/RMN: Nuclear Magnetic Resonance / Résonance magnétique nucléaire
NOESY–PRESAT : Nuclear Overhauser Effect Spectroscopy
pH : potentiel hydrogène
PITC : Isothiocyanate de phényle
PPT: précipitation de protéines
PyBOP : Benzotriazol-1-yloxy-tripyrrolidino-phosphonium hexafluorophosphate
PyBrOP: Bromo-tripyrrolidino-phosphonium hexafluorophosphate
QC: Quality Control/ Contrôle qualité
Q-Orbitrap: quadripôle – Orbitrap
QqQ: Triple quadripôle
Q-TOF: Quadrupole – Time of Flight/ quadripôle- Temps de Vol
RP: Reverse phase / phase inverse
RP-LC: Reverse phase Liquide Chromatography / chromatographie liquide en phase inverse
RT : Retention Time / Temps de rétention
SIMS : Secondary-ion mass spectrometry / Spectrométrie de masse des ions secondaires
SPE : Solide-phase Extraction / Extraction sur phase solide
TBTU: 2-(1H-Benzotriazol-1-yl)N,N,N,N-tetramethylaminium tetrafluoroborate
TOCSY:Total Correlation Spectroscopy / Spectroscopie par corrélation totale
TOF: Time Of Flight / Temps de vol
TSP: triméthylsilylpropanoique

W4M: Workflow4Metabolomics
XCMS: Various forms (X) of Chromatography Mass Spectrometry
YMDB: Yeast Metabolome Database / Base de données du métabolome des levures

Liste des tableaux

Tableau 1 :Liste non exhaustive des logiciels de traitement utilisées pour le traitement des spectres RMN 54

Tableau 2 : Comparaison des caractéristiques des principales sources utilisées en spectrométrie de masse (www.thermofisher.com , www.agilent.com) 57

Tableau 3 : Comparaison des caractéristiques des principaux analyseurs utilisés en spectrométrie de masse (données extraites des constructeurs : www.waters.com , www.bruker.com et www.thermofisher.com) 61

Tableau 4 :Liste non exhaustive des logiciels de traitement de données en spectrométrie de masse 73

Tableau 5 : Bases de données utilisées pour l'identification de métabolites (https://metlin.scripps.edu , https://hmdb.ca, https://lmdb.ca, https://www.genome.jp/kegg, https://pubchem.ncbi.nlm.nih.gov) 77

Tableau 6 : Résumé des principaux agents de couplage utilisés lors du marquage en deux étapes des fonctions acides carboxyliques 146

Tableau 7 : Agents anti-épimérisation (AAE) les plus utilisés lors d'études métabolomiques et de synthèse peptidique 152

Tableau 8 : Etiquettes de dérivation chimique pour le marquage des fonctions acides carboxylique en deux étapes 153

Tableau 9 : Structure chimique, conditions réactionnelles et utilisation de l'étiquette DmPABr pour le marquage en une étape des fonctions acides carboxyliques 156

Tableau 10 : Structure chimique, conditions réactionnelles et utilisations des étiquettes chimiques utilisées pour une réaction de dansylation 158

Tableau 11 : Structure chimique, conditions réactionnelles et utilisation de l'étiquette PITC dans le kit de dosage Biocrates 160

Tableau 12 : Structure chimique, conditions réactionnelles et applications des étiquettes chimiques utilisées pour le marquage des thiols et alcools 162

Tableau 13 : Structure chimique, conditions réactionnelles et utilisations des étiquettes chimiques utilisées pour le marquage des carbonyles 163

Tableau 14 :Structures chimiques, formules brutes et m/z des étiquettes chimiques DMED, DMPP, DMPT, DAPPZ, chlorure de Dansyle et PITC utilisées dans notre étude.....167

Tableau 15 : Structures chimiques, formules brutes et m/z des agents de couplages, AAE et bases utilisés dans notre étude 168

Tableau 16 : Nombre de moles des standards,base TEA, agent de couplage CMPI et étiquettes chimiques (DMED, DAPPZ, DMPT et DMPP) utilisé lors du protocole CMPI 169

Tableau 17 :Nombre de moles des standards, agent de couplage EDC, AAE HOAt et étiquettes chimiques DMED, DAPPZ, DMPT et DMPP utilisé lors du protocole EDC/HOAt.....170

Tableau 18 : Nombre de moles des standards, base TEA, agent de couplage HATU et étiquettes chimiques DMED, DAPPZ, DMPT et DMPP utilisé lors du protocole HATU.....171

Tableau 19 : Nombre de moles des standards,base TEOA et étiquette chimique DmPABr utilisé lors du protocole de marquage des fonctions acides par le DmPABr 172

Tableau 20 : Nombre de moles des standards et étiquette chimique chlorure de Dansyle utilisé lors du protocole de marquage des fonctions amine/phénol par le chlorure de Dansyle 173

Tableau 21 : Nombre de moles des standards, base pyridine et étiquette chimique PITC utilisé lors du protocole de marquage des fonctions amine/phénol par le PITC 174

Tableau 22 : Paramètres utilisés pour la détection positive en full-scan (a) et paramètres de la source ESI (b).....176

Tableau 23 : m/z attendus pour le marquage par les dérivés pipérazines et amine de la fonction acide des acides gras et triacides sous la forme [M+H]+ 177

Tableau 24 : m/z attendus pour le marquage du DMED, DAPPZ, DMPT et DMPP de la fonction acide des acides aminés sous la forme $[M+H]^+$.....181

Tableau 25 : m/z attendus des réactions secondaires de la fonction acide des acides aminés avec les agents de couplage (CMPI, HATU et EDC) et l'AAE (HOAt) 184

Tableau 26 : m/z attendus pour le marquage par le DMED, DAPPZ, DMPT et DMPP de la fonction acide des α-hydroxy acides sous la forme $[M+H]^+$ 186

Tableau 27 : m/z attendus des réactions secondaires de la fonction acide des α-hydroxy acides avec les agents de couplage (CMPI, HATU et EDC) et l'AAE (HOAt) 188

Tableau 28 :m/z des standards marquées 1,2,3 ou 4 fois par l'étiquette DmPABr (TAG) en détection positif sous forme $[M+H]^+$.....193

Tableau 29 : m/z des standards marquées 1,2,3 ou4 fois par l'étiquette DmPABr (TAG) en détection positif sous forme $[M+H]^+$.....195

Tableau 30 : Tableau du fragment MS^2 majoritaire détecté après fragmentation en mode MRM des acides gras et triacides couplés au DmPABr (TAG) 1 à 3 fois. RT = temps de rétention.....198

Tableau 31 : Tableau du fragment MS^2 majoritaire détecté après fragmentation en mode MRM des acides aminés et α-hydroxy acides couplés au DmPABr (TAG) 1 à 4 fois. RT = temps de rétention.....201

Tableau 32 : principaux fragments retrouvés lors de la MS^2 du DmPABr couplé aux composés.....205

Tableau 33 : m/z des standards de dérivés indoliques marqués 1,2,3 ou 4 fois par l'étiquette DmPABr (TAG) en détection positive sous forme $[M+H]^+$.....208

Tableau 34 : Tableau du fragment MS^2 majoritaire détecté après fragmentation en mode MRM des dérivés indoliques couplés au DmPABr (TAG) 1 à 2 fois 209

Liste des figures

Figure 1 : Sciences « omiques » 29

Figure 2 : Domaines d'application de la métabolomique (adapté de Damiani et al. [10]) 32

Figure 3 : Schéma général du déroulement d'une analyse métabolomique (adapté de Segers et al. [14]) 33

Figure 4 :Conditions de stockage des échantillons biologiques [17] 37

Figure 5 : Sélectivité des méthodes de purification d'échantillons biologiques 39

Figure 6 : Exemple d'un spectre RMN des deux doublets de la valine avec son déplacement chimique (δ) , sa constante de couplage J, sa multiplicité et son intensité 48

Figure 7 : Représentation schématique des parties d'un spectromètre de masse. 56

Figure 8 : Représentation schématique d'une source électrospray [83] 58

Figure 9 :Représentation schématique d'un analyseur quadripolaire (https://www.shimadzu.com) 61

Figure 10 : Représentation d'un analyseur Orbitrap [84] 63

Figure 11 : Schéma d'un quadripôle couplé à un Orbitrap (Q-Orbitrap) [84] 66

Figure 12 : Représentation schématique d'une acquisition en full-scan ,Data dependent acquisition (DDA) et Data independent acquisition (DIA) 69

Figure 13 : Mode d'acquisition des données en approche ciblée en utilisant un spectromètre de masse schématisé pour un analyseur triple quadrupôle 71

Figure 14 : Organigramme montrant la stratégie générale de prétraitement et l'analyse des données LC-MS par XCMS [100] 74

Figure 15 : Ordre suggéré pour l'ajout de QC et de mélange de standard (test mix) lors d'une étude utilisant LC-MS [104] 78

Figure 16 : Pourcentage de publications recensé entre 2010 et 2020 via Science Direct combinant au moins 2 matrices en métabolomique 81

Figure 17 : Nombre de publications par année retrouvé après recherche sur pubmed.ncbi.nlm.nih.gov [81]..... 83

Figure 18 : Nombre d'articles scientifiques répertorié sur science direct entre 2010 et 2020 pour chaque plateforme analytique et combinant au moins 2 plateformes « metabolomics and plateforme » 83

Figure 19 : Schématique configuration des différents modes de travail LC-RMN couplée en parallèle à un spectromètre de masse en (a) mode en ligne/flux continu (b) mode stop-flow (c) mode de stockage/ boucle [118]..... 84

Figure 20 : Représentation des approches quantitatives relative (A) et absolue incluant la calibration (B), l'approche par étalon interne (C) et la dérivation chimique (D) utilisées de nos jours en métabolomique MS..... 88

Figure 21 : Parties devant composer une étiquette chimique..... 94

Figure 22 : Workflow général (A) utilisé lors d'étude métabolomique par une analyse LC-MS et par dérivation chimique en LC-MS (B) [124] 138

Figure 23 : Workflow général de la dérivation chimique utilisée en LC-MS lors d'approche métabolomique [172] 143

Figure 24 : Réactions entre la fonction acide carboxylique et l'étiquette chimique lors d'une réaction acido-basique (A) et lors de la formation d'une liaison amide (B). 145

Figure 25 : Schéma réactionnel d'activation d'une fonction acide carboxylique par un sel de phosphonium (BOP) [184]..... 147

Figure 26 : Schéma réactionnel d'activation d'une fonction acide carboxylique contenu dans un acide aminé par un sel d'uronium (HATU ou HBTU) [184] 148

Figure 27 : Réaction d'activation d'une fonction acide carboxylique par le DMT-MM [184]..... 149

Figure 28 : Schéma réactionnel d'activation d'une fonction acide carboxylique par un carbodiimide lors d'une épimérisation (A), sans AAE (B) avec un AAE (C). Schéma inspiré par [184] 150

Figure 29 : Schéma d'épimérisation dans le cas d'un acide carboxylique ayant une fonction en position α 151

Figure 30 : Schéma réactionnel du DMPP avec une fonction acide [43] 154

Figure 31 : Schéma réactionnel du DAPPZ avec une fonction acide appartenant à des acides gras [44] 154

Figure 32 : Schéma réactionnel du DMED avec une fonction acide [177]..... 155

Figure 33 : Schéma réactionnel du DmPABr et d'une fonction acide carboxylique [197]..... 156

Figure 34 : Schéma réactionnel du marquage des fonctions amine et phénol par le chlorure de Dansyle. X= ^{12}C ou ^{13}C [203]..... 159

Figure 35 : Liste des standards utilisés dans notre étude selon leurs classes chimiques..... 166

Figure 36 : Schéma réactionnel de la formation d'une liaison peptidique à l'aide de l'agent de couplage CMPI..... 168

Figure 37 : Schéma réactionnel lors de la formation d'une liaison peptidique à l'aide de l'agent de couplage HOAt/EDC inspiré de Leng et al. [43] 169

Figure 38 : Schéma réactionnel d'une N-acylation utilisant l'agent de couplage HATU..... 171

Figure 39 : Schéma réactionnel de la formation d'un ester à partir du DmPABr ... 172

Figure 40 : Schéma réactionnel d'une dansylation réalisée pour une amine primaire (A) et pour un phénol (B) à l'aide du chlorure de Dansyle 173

Figure 41 : Schéma réactionnel de dérivation du PITC avec une fonction amine primaire 174

Figure 42: Résultats de marquage des acides gras et triacides par les dérivés pipérazine et amine. L'ordonnée représente l'aire brute des pics et l'abscisse l'étiquette chimique considérée.AG= agent de couplage, CMPI, HATU, EDC ou HOAt et TAG=DMED, DAPPZ, DMPT ou DMPP 178

Figure 43 : spectres MS^2 de l'acide subérique marqué une fois par le DMED (a), l'acide laurique marqué avec le DMED (b), l'acide acétique marqué avec le DAPPZ (c) et l'acide acétique marqué avec le DMPT (d) 179

Figure 44 : Résultats de marquage des acides aminés par les dérivés pipérazine et amine. L'ordonnée représente l'aire brute des pics et l'abscisse l'étiquette chimique considérée. AG= agent de couplage, CMPI, HATU, EDC ou HOAt et TAG=DMED, DAPPZ, DMPT ou DMPP..... 182

Figure 45 : Spectre MS² de l'acide hippurique couplé au DMPT sous les conditions HATU.....183

Figure 46 : Spectre MS² de l'alanine (a) et de la phénylalanine (b) couplées au CMPI 185

Figure 47 : Schéma de la liaison hydrogène dans la structure de l'acide aspartique 186

Figure 48: Résultats de marquage des α-hydroxy acides par les dérivés pipérazine et amine. L'ordonnée représente l'aire brute des pics et l'abscisse l'étiquette chimique considérée AG= agent de couplage, CMPI, HATU, EDC ou HOAt et TAG=DMED, DAPPZ, DMPT ou DMPP 187

Figure 49 : Spectre MS² de l'acide lactique couplé au DMPT sous les conditions HATU.....188

Figure 50 : Spectre MS² du couplage avec le CMPI dans le cas de l'acide lactique (a) et l'acide hydroxyphenyllactique (b).....189

Figure 51 : Structures chimiques de l'acide lactique (a) et l'acide hydroxyphenyllactique (b) couplés au CMPI sur leurs fonctions hydroxyle et carboxylique 189

Figure 52 : Chromatogrammes de l'acide lactique pour un m/z de 182.081 (a) et de l'acide hydroxyphenyllactique pour un m/z de 274.107 (b) couplés au CMPI 190

Figure 53 : Histogrammes des résultats de marquage par l'étiquette DmPABr dans le cas des acides gras et triacides. L'axe des ordonnées représente l'aire brute sous la courbe de chaque molécule marquée. Le code couleur 1,2 et3 représente le nombre de fois où la molécule est marquée par le DmPABr. 195

Figure 54 : Histogrammes des résultats de l'étiquette DmPABr dans le cas des α-hydroxy acides et acides aminés. L'axe des ordonnées représente l'aire brute sous la courbe de chaque molécule marquée. Le code couleur 1,2,3 et 4 représente le nombre de fois la molécule est marquée par le DmPABr. 197

Figure 55 : spectres MS² de l'acide citrique marqué par l'étiquette chimique DmPABr, une fois (a) , deux fois (b) et (c) et trois fois (d).....199

Figure 56 :Pics chromatographiques correspondant à l'acide hippurique marqué deux fois par le DmPABr (m/z de 502.2) 202

Figure 57 : Spectre MS² de l'acide hippurique marqué deux fois par le DmPABr au temps de rétention 9.25min (a), 9.96min (b) et 10.64min (c) 203

Figure 58 : Spectre MS² de l'acide lactique couplé au DmPABr une fois (a) et deux fois (b).....204

Figure 59 : Spectre MS² de la cystéine couplé au DmPABr une fois (a) et (b), trois fois (c) et quatre fois (d) 205

Figure 60 : Histogramme des résultats du marquage par l'étiquette DmPABr dans le cas des dérivés indoliques. L'axe des ordonnées représente l'aire brute sous la courbe de chaque molécule marquée. Le code couleur 1et 2 représente le nombre de fois où la molécule est marquée par le DmPABr. 209

Figure 61 : figure représentant le pseudo-rendement de marquage de chaque molécule en fonction du DmPABr, chlorure de Dansyle et PITC.....212

Chapitre 1 : Introduction générale – Etat de l'art

I. Historique

La caractérisation et la compréhension du fonctionnement d'un organisme peuvent être réalisées à l'aide de sciences « omiques ». La métabolomique est la science « omique » la plus récente qui complète la génomique, transcriptomique et protéomique (figure 1). La métabolomique permet de caractériser à un instant donné l'état métabolique d'un organisme biologique au moyen de l'exploration de son métabolome. Alors que la génomique, la transcriptomique et la protéomique permettent de faire état d'un potentiel biochimique et dans une certaine mesure de prédire ce qui peut survenir au niveau biochimique [1]; la métabolomique génère un cliché métabolique qui résulte de la cascade omique tout en considérant son interaction avec l'environnement.

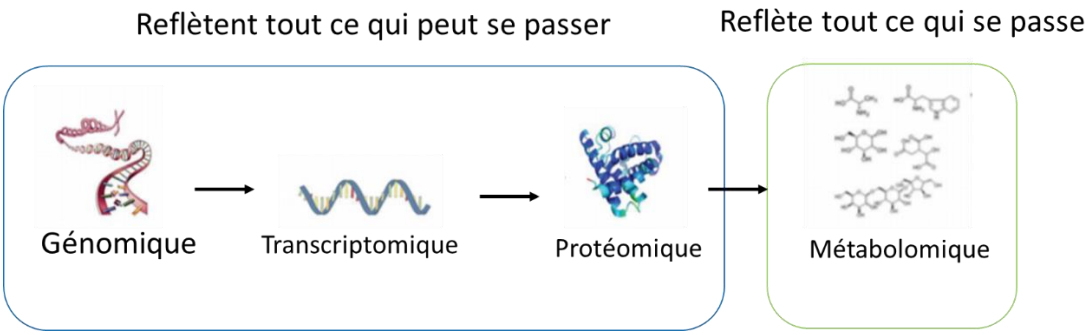


Figure 1 : Sciences « omiques »

En 1957, RJ Williams est l'un des premiers à proposer le concept de profil métabolique [2]. A l'aide de la chromatographie sur papier, il obtient des profils métaboliques (profils de migration) de patients atteints de schizophrénie. Ces profils métaboliques sont réalisés à partir de prélèvements urinaire et salivaire. Grâce aux avancées technologiques successives, des profils métaboliques qualitatifs et quantitatifs ont pu être obtenus par analyse en chromatographie

gazeuse (GC) par Pauling [3] et Horning [4] en 1971. Ces travaux portent sur les profils métaboliques de stéroïdes, acides organiques, drogues et leurs métabolites, présents dans les constituants urinaires par la chromatographie gazeuse couplée à la spectrométrie de masse (GC MS). Conjointement au développement de la GC, la technologie par (Résonance Magnétique Nucléaire) RMN se développe et est utilisée par Nicholson et son équipe. En 1984, les premiers profils métaboliques sont obtenus par RMN ^{1}H sur prélèvements sanguin et urinaire de patients diabétiques [5]. Quatre années plus tard, en 1988, Oliver introduit pour la première fois le terme « métabolome » et « métabolomique » [6]. Au début des années 2000, la métabolomique connait une évolution grâce à son application en biologie plus précisément sur les plantes qui sont considérées comme ayant un métabolome large [7]. Depuis, des bases de données comme METLIN :metabolite mass spectral database (https://metlin.scripps.edu), HMDB: Human Metabolome Database (https://hmdb.ca), YMDB: the Yeast Metabolome Database, Livestock metablome data base (https://lmdb.ca), fecal métabolome data base (https://fecalmetabolome.ca), KEGG : Kyoto Encyclopedia of Genes and Genomes (https://www.genome.jp/kegg), PubChem (https://pubchem.ncbi.nlm.nih.gov)... ont été créées afin de répertorier tous les métabolites identifiés dans les organismes biologiques. Depuis 2010, la métabolomique est de plus en plus utilisée avec plus de 50 000 articles contenant le mot métabolomique publiés (source : PubMed).

II. Généralités

1. Métabolites et Métabolome

Le métabolome d'un organisme correspond à l'ensemble des composés issus du métabolisme de celui-ci. Le métabolisme (primaire ou secondaire) d'un organisme correspond à l'ensemble des réactions chimiques (anabolisme et catabolisme) qui se déroule en son sein pour lui permettre de se maintenir en vie, de se reproduire, de se développer et d'interagir avec son environnement. Ces réactions chimiques ont lieu à l'aide de composé organique dit métabolite de masse inférieure à 1500

Da. Les métabolites issus du métabolisme primaire sont des molécules indispensables au fonctionnement de l'organisme. Ils sont directement impliqués dans les principaux processus de développement normal et de reproduction de l'organisme. Ce sont par exemple les acides aminés, les acides carboxyliques, les nucléotides ou encore les vitamines et certains lipides. Un métabolite secondaire n'est pas directement impliqué dans les processus vitaux de la cellule mais assurent des fonctions écologiques importantes. Les métabolites secondaires sont principalement retrouvés chez les plantes. La métabolomique correspond à une approche technologique et méthodologique qui permet d'analyser le plus grand nombre de métabolites possible, donnant accès à des informations concernant les métabolismes propres à l'organisme mais également ceux qui résultent de l'interaction de l'organisme avec son environnement. Les métabolites exogènes constituant le xénométabolome [8] sont des métabolites issus de l'environnement d'un organisme et viennent compléter la vue globale du métabolome d'un organisme.

Ainsi, le métabolome est la collection de tous les métabolites présents dans une unité biologique. Sa dimension peut varier en nombre et en quantité car il s'agit d'un système dynamique qui change au cours du temps, mais aussi selon les espèces, le stade physiologique ou pathologique.

Chez l'homme il y a plus de 220 945 métabolites, selon HMDB [9] répertoriés dans différentes catégories (carbohydrates, acides aminés, lipides, nucléotides, glucides…). En raison de ce nombre, de la très grande variabilité moléculaire, mais également de leur localisation (urines, tissus, sang, selles…) aucune méthode actuelle ne permet de détecter et quantifier tous les métabolites d'un organisme. A partir d'un type de prélèvement, il faut adapter la préparation d'échantillon (pré analytique) en fonction des objectifs de l'étude puis investiguer en utilisant des technologies d'analyses physico-chimiques complémentaires afin de couvrir de manière la plus large et sensible possibles le métabolome.

2. Applications de la métabolomique

La métabolomique est une approche scientifique incluant des disciplines fondamentales telles que la chimie, la biologie et la bio-informatique. Les potentiels d'applications de la métabolomique sont nombreux (figure 2).

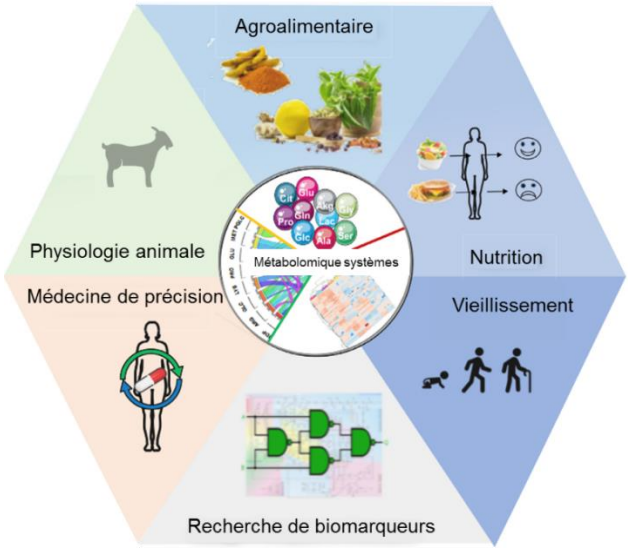


Figure 2 : Domaines d'application de la métabolomique (adapté de Damiani et al. [10])

La métabolomique est devenue une approche pour la **découverte de biomarqueurs de métabolites** en comparant les profils métaboliques de différents phénotypes afin d'identifier des changements métaboliques spécifiques. Ces profils métaboliques peuvent provenir de systèmes biologiques tels que les plantes, les animaux et les Hommes [11, 12]. La comparaison de profils métaboliques conduit à la compréhension de la **physiologie**, de la toxicologie et de la progression de la maladie. La métabolomique est également appliquée dans le domaine de la pharmacologie afin de développer et valider de nouveaux médicaments notamment pour la **médecine de précision**. Dans le domaine de l'**agroalimentaire**, la métabolomique est appliquée en **nutrition**, suivi nutritionnel personnalisé, sécurité alimentaire, fraudes, traçabilité, OGM… De nos jours, de nouveaux domaines d'applications se développent tels que celui de l'environnement avec par exemple l'étude des microorganismes des nuages par

l'institut de chimie de Clermont-Ferrand [13]. Cependant dans cette thèse nous développerons l'application de la métabolomique dans le domaine de la santé (humaine et animale).

III. Déroulement général d'une analyse métabolomique

Les analyses métabolomiques consistent à une suite d'étapes comprenant la collecte des échantillons, leur préparation, leur analyse, leur détection et annotation puis de l'interprétation biologique (figure 3).



Figure 3 : Schéma général du déroulement d'une analyse métabolomique (adapté de Segers et al. [14])

La première étape consiste à la collecte des échantillons qui consiste à rassembler les fluides biologiques mais également les données associées (heure de prélèvement, récipient, état de santé). La préparation des échantillons est la deuxième étape qui a pour objectif l'extraction des métabolites. Le choix de la technique d'extraction dépend du type d'échantillon et du but de l'étude. L'étape suivante est l'analyse des échantillons réalisée par spectrométrie de masse (MS) et/ou par résonnance magnétique nucléaire (RMN). Selon les objectifs de l'étude et l'information métabolomique souhaitée différents types d'approches peuvent être envisagées (quantitative, qualitative, ciblée pour un ensemble de métabolite, non-ciblée …). La quatrième étape consiste à la détection et à l'annotation des

métabolites. La dernière étape est l'interprétation biologique qui peut être réalisée par des études statistiques, des analyses de voies métaboliques.

Ce projet de thèse se focalise sur le développement de méthode analytique comprenant les étapes de préparation et d'analyse des échantillons. Ces étapes seront développées en détail dans la suite du manuscrit.

IV. Collection et préparation des échantillons

1. Collection des échantillons

La collection des échantillons est une étape cruciale car elle peut avoir des répercussions sur la pertinence des résultats obtenus à la fin du processus analytique comme le manque de reproductibilité, l'altération de la quantification et de l'identification des métabolites. Les objectifs d'une étude peuvent donc être conditionnés par cette étape. C'est pourquoi le moment dans la journée du prélèvement est fixé en fonction de l'objectif de recherche. L'échantillon peut être collecté soit de manière aléatoire, soit de manière séquentielle comme par exemple lors d'étude nutritionnelle un prélèvement à jeun puis toutes les 30min est réalisé, soit à heure fixe afin d'étudier le statut d'un individu et de limiter l'impact des fluctuations nycthémérales. Le choix du tube de collection est à déterminer avant le prélèvement et dépend de l'échantillon.

1.1. Echantillon sanguin

A partir d'un prélèvement sanguin il est possible d'obtenir du plasma ou du sérum. Le plasma est obtenu par prélèvement sur un tube contenant un anticoagulant qui inhibe les facteurs de coagulation ou limite leur action par chélation des ions Ca^{2+}.La présence d'anticoagulant (EDTA acide éthylène diamine tétra-acétique, citrate, héparine) peut interférer avec les signaux des métabolites détectés en MS et en RMN. Par exemple, l'EDTA présent en quantité plus importante que les métabolites

peut couvrir leurs signaux [15]. Le sérum est lui obtenu par prélèvement sur tube sec ou tube à gel. Dans le cas des tubes secs, la coagulation se fait naturellement pendant 30 à 60 minutes à température ambiante puis une centrifugation est réalisée pour récupérer le sérum. Les tubes de prélèvement sanguin avec gel et activateur de coagulation permettent d'obtenir du sérum. Le gel, utilisé dans les tubes, a une densité définie de manière à se situer entre la densité des cellules sanguines et la densité du sérum. Lors de la centrifugation, le gel durcit et forme une barrière entre les globules rouges et cellules sanguines qui sont précipités au fond du tube et le sérum est situé au-dessus du gel en haut du tube. Le temps de coagulation est compris entre 10 et 30 min.

1.2. Echantillon urinaire

Le prélèvement d'échantillon urinaire s'effectue généralement dans un flacon en polypropylène pouvant contenir des conservateurs tels que l'acide borique ou l'azide de sodium. En effet, un développement bactérien peut avoir lieu dans l'urine ce qui viendrait modifier le métabolome urinaire. Cependant ces ajouts peuvent modifier l'échantillon avec par exemple la formation d'adduits visibles en spectrométrie de masse ce qui peut engendrer une source de variabilité supplémentaire.

1.3. Echantillon salivaire

Le prélèvement de salive est réalisé à l'aide d'une Salivette® (Sarstedt, Nümbrecht, Allemagne). Celle-ci se compose d'un tube stérile contenant un tampon en coton à mâcher ou à frotter sur les joues. Ce tampon doit être conservé en bouche jusqu'à imprégnation complète. Il est ensuite placé dans le tube. Une autre méthode de collection est mise en place par SalivaBio n'utilisant pas le tampon de prélèvement. Cette méthode consiste à placer le flacon de prélèvement avec un embout dans la bouche afin de laisser s'accumuler la salive dans la bouche et par l'inclinaison de la

tête vers l'avant permettant de guider délicatement la salive jusqu'au fond du tube de prélèvement jusqu'à l'obtention du volume requis (www.salivabio.com).

1.4. Echantillon fécal

Le prélèvement des fèces est fait dans des flacons pour coprologie en polypropylène transparent à usage unique contenant généralement une cuillère pour permettre un meilleur prélèvement des matières fécales. Il est également possible d'utiliser des sacs en plastique stérilisés.

2. Acheminement au laboratoire

En général, les échantillons d'urine, de plasma/sérum et de matières fécales doivent être transférés dans des cryotubes immédiatement après le prélèvement et conservés à -80 °C dans un congélateur ou dans des réservoirs d'azote liquide jusqu'à l'analyse. Généralement, les prélèvements sanguins sont réalisés en laboratoire et sont donc immédiatement mis au frais. En cas d'échantillons d'urine ou de selles prélevés sur 24 heures au domicile du volontaire, les échantillons doivent être conservés au réfrigérateur du volontaire à 4 °C afin de réduire les activités enzymatiques et bactériennes. Une fois la collecte des échantillons d'urine ou de selles sur 24 heures terminée, les échantillons doivent être transportés au laboratoire, de préférence sur de la glace, ou dans un sac réfrigérant [16]. D'après le constructeur SalivaBio®, le prélèvement doit être immédiatement conservé à au moins -20°C. Si cette congélation n'est pas possible, l'échantillon salivaire peut être conservé idéalement moins de 2h à 4°C.

3. Stockage des échantillons

Au laboratoire, les échantillons doivent être préparés en fonction de leur destination et conservés à − 80 °C (figure 4). Pour une stabilité à long terme des fluides biologiques, les cycles de congélation/décongélation doivent être évités. Ainsi, une fois le prélèvement effectué, l'échantillon est aliquoté puis congelé à basse

température. L'échantillon peut être placé dans un bain d'azote liquide (-196°C) ou bien dans un congélateur à température très basse (-80°C). Ces deux dispositifs permettent de stocker les échantillons sur des périodes relativement longues compatibles avec le temps nécessaire pour l'inclusion et la collecte de tous les échantillons d'une étude. Il est alors important de noter la durée de stockage des échantillons, le nombre de cycles de congélation/décongélation (s'il n'y a pas eu de préparation d'aliquot avant la congélation). En effet, la stabilité des protéines et la dégradation des molécules peut être affectée durant ces conditions de stockage.

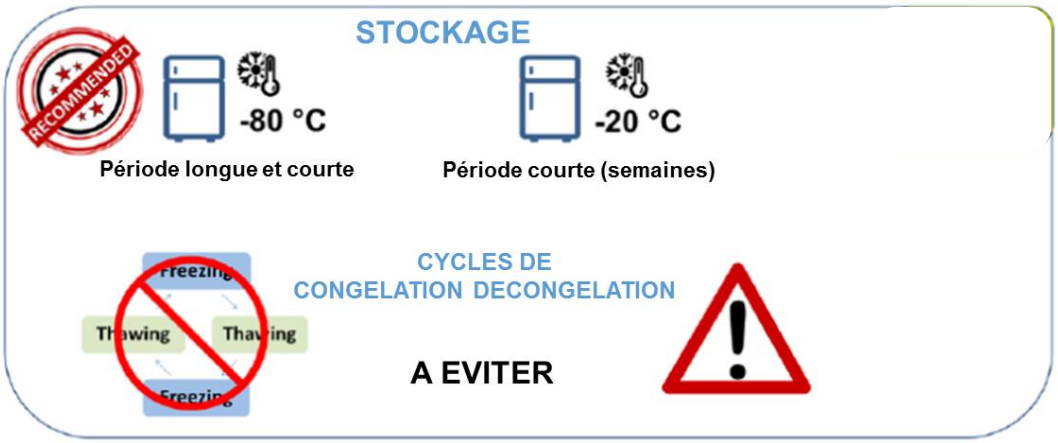


Figure 4 :Conditions de stockage des échantillons biologiques [17]

Dans de nombreux cas, la disponibilité des congélateurs à -80 °C est limitée, donc une congélation temporaire à –20 °C est envisageable avant l'analyse des échantillons. Pour cette raison, de nombreuses études ont été menées ces dernières années dans le but d'évaluer la stabilité de l'ensemble du métabolome dans différentes conditions de stockage.

3.1. Echantillon sanguin

Dans le cas du sang, il est conseillé de congeler le plus rapidement possible l'échantillon de plasma ou de sérum obtenu. Un stockage à -80° C doit être préféré à un simple stockage à - 20 °C [18]. En effet, il a été montré que certains métabolites

tels que le glucose ou la proline présentent des variations significatives de concentration lors d'un stockage à -20°C [19]. Dans le cas du plasma, le temps et les conditions de stockage de l'échantillon entre son prélèvement et l'étape de centrifugation apparaît comme étant le premier facteur de variabilité dans la préparation des échantillons en métabolomique [20]. Cette étude a montré que si l'échantillon est laissé à température ambiante pendant plus d'une heure, le métabolome est altéré du fait de l'activité des érythrocytes.

3.2. Echantillon urinaire

Pour les prélèvements urinaires, il a été montré que, lors d'une analyse RMN et LC-MS, la conservation des échantillons urinaires à 4°C est possible pendant 24h à 72h sans dégradation significative [21]. Il a également été montré que les échantillons urinaires sont stables 1 mois après stockage à -20°C et -80°C [22] . Toutefois, une étude en GC-MS a montré que l'aire des métabolites volatils dans les urines diminuent de moitié au cours du temps s'ils ne sont pas congelés [23].

3.3. Echantillon fécal

Le métabolome fécal avant extraction est également altéré si celui-ci est stocké à température ambiante. L'étude de Phua et al. montre que seulement 10% des métabolites sur 268 sont stables à température ambiante (29°C) pendant 1 jour tandis que 24% sont stables à -80°C pendant 1 semaine [24].

Tous ces résultats suggèrent que les différents types de prélèvement peuvent être utilisés mais qu'il est essentiel de conserver un même et unique type d'échantillons et un même protocole de prélèvement pour tous les échantillons d'une étude. L'ensemble des paramètres concernant le prélèvement, l'aliquotage et la conservation doivent être scrupuleusement consignés car ils influencent les résultats obtenus.

4. Préparation des échantillons

Les échantillons biologiques contiennent de nombreuses molécules telles que les lipides, les protéines ou les métabolites. Selon les objectifs du projet, des propriétés physiques et chimiques des échantillons (polarité, solubilité), il est nécessaire d'adapter l'étape d'extraction des métabolites. Il est même parfois nécessaire d'effectuer une combinaison de protocoles d'extraction. Toutefois, une préparation d'échantillons biologiques doit être simple, reproductible et rapide avec un minimum d'étapes car à chaque étape est associé un rendement d'extraction. La figure 5 présente les principales préparations utilisées en métabolomique de la moins sélective à la plus sélective.

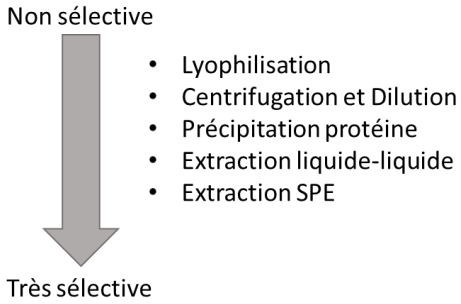


Figure 5 : Sélectivité des méthodes de purification d'échantillons biologiques

4.1. Lyophilisation

La lyophilisation est une opération de déshydratation qui consiste à éliminer sous forme gazeuse l'eau contenue dans un échantillon solide préalablement congelé. Ce changement d'état s'appelle la sublimation. Cette technique s'effectue à température et à pression basse ce qui permet de dénaturer le moins possible les échantillons biologiques. Cette technique est souvent utilisée pour les échantillons biologiques « solides » contenant de l'eau tel que le cerveau [25] ou les fèces [26]. Cette étape est importante pour une analyse reproductible, car selon la provenance

biologique de l'échantillon la quantité d'eau contenue peut différer. La lyophilisation des échantillons biologiques permet de prélever la même quantité de matière sèche, donc de métabolites par échantillon en s'affranchissant du taux d'hydratation du prélèvement tout en les concentrant. Une étude comparative en ^{1}H-RMN des fèces de rat congelées ou lyophilisées a montré que la lyophilisation permet d'augmenter le signal des métabolites notamment la classe des acides aminés. [27].

4.2.Centrifugation

La centrifugation est un procédé de séparation des composés d'un mélange en fonction de leur différence de densité en les soumettant à une force centrifuge. Le mélange à séparer est constitué soit de deux phases liquides, soit de particules solides en suspension dans un fluide qui seront donc précipitées. Avant tout traitement, les fluides biologiques sont centrifugés afin d'éliminer ces particules solides en suspension. La centrifugation est une étape réalisée après la collecte de l'échantillon mais également lors de l'étape d'extraction des métabolites. Elle permet par exemple de précipiter dans le culot du tube de centrifugation les sels qui se sont éventuellement formés lors de l'étape d'extraction et/ou les particules solides/fibres contenues dans les fèces.

Les paramètres de centrifugation sont à définir et à conserver pour tous les échantillons d'une seule et même cohorte. En effet, des études ont montré que celle-ci influence les résultats obtenus lors des analyses. Par exemple pour une analyse RMN, la ligne de base peut être déformée et altérer la quantification des métabolites. Alexander Gardner et al. [28] montre que l'acide lactique contenue dans un échantillon salivaire est surestimé lorsque l'échantillon n'est pas centrifugé. En effet, le déplacement chimique de l'acide lactique est proche de la zone spectrale où les pics des macromolécules sortent. Elles déforment donc le signal

de l'acide lactique. La centrifugation permet d'éliminer ces macromolécules du milieu réactionnel et du spectre RMN.

4.3. Dilution

La dilution est un procédé consistant à obtenir une solution de concentration inférieure à celle de départ par ajout de solvant. Ce procédé est adapté aux échantillons dont on souhaite diminuer la concentration en molécule. Cela peut être par exemple les échantillons fortement concentrés en métabolites tels que l'urine et ceux contenant des macromolécules tels que la salive. Ces échantillons sont dilués afin de limiter l'effet matrice, la suppression d'ionisation avant l'injection et la saturation du détecteur. Le facteur de dilution doit être adapté à la sensibilité de l'instrument et à l'échantillon considéré. Par exemple, pour une analyse RMN la salive est faiblement diluée (9 :1) à l'aide d'une solution au pH contrôlé [29]. Le facteur de dilution utilisé lors d'analyse en spectrométrie de masse est plus important que lors d'analyse RMN car sa sensibilité est plus élevée. Ainsi, l'urine est dilué à 1/10 par de l'eau ultrapure [30].

4.4. Précipitation des protéines

Les protéines sont des molécules sensibles à leur environnement et peuvent être éliminées d'un échantillon par le changement de leur solubilité dans une solution aqueuse. Ce changement de solubilité permet de les précipiter, soit par des conditions ioniques différentes soit en changeant le pH de la solution [31]. Pour l'extraction des métabolites d'un fluide biologique, l'ajout de solvants organiques miscibles à l'eau tels que l'éthanol, le méthanol ou l'acétonitrile est utilisé. En effet, la couche de solvatation (composé d'eau) autour de la protéine diminue à mesure que la teneur en solvant organique augmente. Ces couches d'hydratation sont réduites de même que la constante diélectrique de l'eau provoquant un rapprochement des protéines donc un agrégat par l'action de forces

électrostatiques actives. Des études ont été menées et révèlent que selon l'échantillon le MeOH ou l'ACN est à privilégier. Par exemple lors d'une étude de Nagana-Gowda en 2014 [32], une comparaison entre le méthanol et l'acétonitrile sur la précipitation des protéines contenues dans le sérum a montré par une analyse RMN quantitative que l'efficacité des précipitations avec ces 2 solvants sont différentes. Dans les conditions de cette étude l'acétonitrile altère quantitativement le signal des métabolites en RMN et les auteurs concluent qu'il est préférable d'utiliser le méthanol. Toutefois, l'utilisation de l'acétonitrile est privilégiée par Schultz pour éliminer les protéines contenues dans la salive lors d'une analyse RMN [33].Le mélange MeOH : H_2O (4 :1) est efficace pour une extraction des métabolites contenues dans les fèces pour une approche non ciblée en LC-MS [30]. De nos jours, la précipitation des protéines pour l'urine est peu utilisée.

4.5. Extraction liquide-liquide

L'extraction liquide-liquide (LLE) est une méthode de purification basée sur la différence de solubilité d'un soluté dans deux phases liquides non miscibles. Elle permet la séparation de deux ou plusieurs constituants d'un mélange en mettant à profit leurs différences d'affinité pour deux phases liquides non miscibles. L'extraction liquide-liquide permet une approche multi-omique. En effet, la phase aqueuse permet d'extraire préférentiellement les métabolites et la phase organique les lipides. En métabolomique et lipidomique, trois types d'extractions biphasiques sont connus : Folch [34], Bligh-Dyer [35] et Matyash [36]. Les extractions Folch et Bligh-Dyer utilisent toutes les deux un mélange $CHCl_3$:MeOH :H_2O en proportion différente, respectivement (8 :4 :3) et (2 :2 :1,8) [37]. Lors de ces extractions, la phase aqueuse se trouve au-dessus (du fait de la densité plus importante du chloroforme comparée à l'eau) et une couche de protéines est formée entre les deux phases. Cette configuration rend le prélèvement de la phase organique difficile et celle-ci peut être contaminée par la couche de protéines lors du prélèvement. L'extraction de Matyash a été développé pour faciliter le prélèvement

de la phase organique, elle est préférentiellement utilisée lors d'analyses lipidiques. Le chloroforme (d=1,49) est remplacé par le methyl-tert-butyl ether (d=0,74) (MTBE) ce qui inverse les phases aqueuse et organique par rapport aux extractions de Folch et Bligh-Dyer [38] . Ce phénomène est dû à la densité du MTBE qui est plus faible que celle de l'eau et du méthanol. La couche de protéine se retrouve précipitée au fond du mélange ce qui permet de limiter les contaminations lors du prélèvement des phases.

Ce type d'extraction de métabolites est adapté aux échantillons contenant des macromolécules et des lipides tels que le sang et les fèces. Le sang (plasma/serum) est généralement préparé par un mélange MeOH/$CHCl_3$ [39] dans les proportions (1 :1). Les métabolites fécaux sont également extraits à l'aide d'un mélange biphasique mais en y ajoutant de l'eau afin de solubiliser toute la matière sèche que compose les fèces. Une étude menée par Lamichhane *et al*. en 2015 par RMN sur des selles humaines [40] utilisant le mélange methanol/Chloroforme/H_2O (1 :1 :1) montre que les deux phases permettent d'extraire différents types de métabolites. En effet, la phase organique constituée par le chloroforme contient des molécules non polaires qui ont été assignées à des lipides ou des groupements lipidiques, tandis que la phase aqueuse contient des acides gras à chaines courtes mais également des acides organiques et de nombreux métabolites.

Toutefois, la LLE présente des inconvénients tels que la formation d'émulsion, la solubilité mutuelle des 2 phases, la possibilité de contamination d'une phase par l'autre lors du prélèvement. Ces désagréments peuvent être limités par l'ajout de sel à la phase aqueuse, le refroidissement ou chauffage du récipient d'extraction, par filtration ou/et par centrifugation. Un autre inconvénient est le temps nécessaire pour réaliser cette étude et la répétabilité d'extraction qui est faible, entrainant de forte variation de concentration de métabolites d'une extraction à une autre. Cette extraction de métabolites est donc peu adaptée aux études à haut débit.

4.6. Extraction en phase solide (SPE)

L'extraction en phase solide (SPE) est une méthode de préparation d'échantillons qui permet de concentrer et de pré-purifier des analytes à partir d'échantillon liquide. Les composés sont extraits par adsorption sur un support solide (phase stationnaire), puis élués avec un solvant approprié. Les interactions moléculaires mises en jeu lors de cette extraction sont les mêmes que celles de la chromatographie liquide. Le choix de la phase stationnaire dépend d'abord de la matrice de l'échantillon. Dans le cas des échantillons biologiques, des supports hydrophobes conviennent parfaitement car ces échantillons sont majoritairement aqueux. Dans un deuxième temps, le choix de la phase mobile va dépendre des caractéristiques physico-chimiques des solutés recherchés et notamment de leur polarité. La polarité d'un soluté est difficile à définir, mais elle peut être toutefois reliée à la notion d'hydrophobie définie à partir de la constante de partage octanol-eau (P_{oct}). Les composés polaires ont un P_{oct} inférieur à 1 et les apolaires ont un P_{oct} supérieure à 3. C'est en fonction de ce classement que peut s'effectuer la sélection des supports [41].

Pour la métabolomique, la SPE peut être utilisée seule ou couplée à d'autres types d'extraction. L'étude non ciblée de Rico *et al.* compare différents types d'extraction de métabolites plasmiques et montre que la SPE permet d'extraire le moins de métabolites 890 « features » contre 1905 pour une précipitation de protéine MeOH (3 :1) [42]. Cette méthode d'extraction des métabolites est donc sélective et permet dans la plupart des études de cibler une famille de métabolites grâce à la diversité des cartouches disponibles. Par exemple les études de Leng *et al.* et Ogawa *et al.* montrent qu'une SPE permet d'extraire spécifiquement les acides gras contenus, respectivement, dans l'urine [43] et dans la salive [44]. D'autres familles de molécules peuvent être spécifiquement extraites telles que les composés organiques volatils extraits des fèces pour une étude en GC-MS [45].

Cette technique souvent comparée à la précipitation de protéines peut être utilisée comme alternative et est facilement automatisée. Elle permet de purifier l'échantillon en éliminant les macromolécules, ce qui allonge la durée de vie des techniques analytiques telle que la colonne en chromatographie liquide. Cependant, l'un des défis majeurs de la métabolomique est l'analyse d'un grand nombre d'échantillons. La SPE nécessite un temps supplémentaire lors de la phase de pré-analytique et un coût élevé, ce qui augmente le prix unitaire d'une analyse métabolomique et allonge le temps du pré-analytique. Ainsi cette méthode d'extraction est à privilégier lors d'étude composée de peu d'échantillons. Un gain de temps peut être envisagé en réalisant la SPE online avant une analyse LC.

V. La spectroscopie par résonnance magnétique nucléaire (RMN)

1. Le spectromètre RMN

La résonnance magnétique nucléaire (RMN) est une technique spectroscopique permettant de déterminer des informations structurales par l'interprétation du comportement de l'aimantation nucléaire (moment magnétique du noyau) en présence d'un champ magnétique statique intense B_0 (quelques Teslas) Elle est basée sur les propriétés magnétiques de certains noyaux ayant un spin nucléaire intrinsèque non nul (I). Les spins nucléaires de ces atomes placés dans un champ magnétique externe (B_0) s'orienteront selon B_0. Chaque noyau est défini par son nombre quantique magnétique de spin m_s. Un noyau de spin I non nul peut prendre $2I+1$ orientations par rapport à la direction de ce champ. Par exemple, le proton (^1H) qui possède un spin nucléaire $I = ½$ prend 2 orientations, celles-ci correspondent au nombre quantique magnétique de spin m_s : $m_s = +½$ lorsque le spin nucléaire est dans la même direction que le champ appliqué, il est alors dans un état énergétique favorable α et ms = -½ lorsque le spin nucléaire a une orientation antiparallèle au champ appliqué, il est alors dans une position énergétique défavorable β. Un

noyau de spin *I*=0 possède un seul état et n'est donc pas détectable par RMN. La transition entre les deux états énergétiques demande un apport d'énergie qui est fourni par l'envoi d'une radiofréquence spécifique $\vec{B}_{RF}$. Elle entrainera l'excitation des spins nucléaires par absorption d'une certaine quantité d'énergie correspondant au passage de l'état α à l'état β. Le noyau peut être représenté par un vecteur moment magnétique μ proportionnel au moment cinétique (P), $\vec{\mu} = \gamma\vec{P}$ avec γ le rapport gyromagnétique du noyau considéré. Le moment cinétique est proportionnel au spin du noyau par $\vec{P} = \frac{h}{2\pi}\vec{I}$, avec h la constante de Planck . Ainsi le moment magnétique d'un noyau est proportionnel au spin du noyau par $\vec{\mu} = \gamma \frac{h}{2\pi}\vec{I}$. Parmi les noyaux détectables en RMN, le proton a le rapport gyromagnétique le plus fort (42,576 MHz/T). Ce moment magnétique aura dans B_0 un mouvement de précession suivant la direction de B_0, à la vitesse dictée par la relation de Larmor : $\omega_0 = \gamma B_0$. Pour avoir résonance, il faut que ω_{RF} soit égale à ω_0. D'un point de vu macroscopique, l'aimantation basculera dans le plan transverse, l'aimantation pourra ainsi être décomposée en une composante longitudinale (le long de B_0 soit M_z) et d'une composante transverse (M_{xy}). L'arrêt de B_{RF} entraine le retour spontané à l'équilibre du système. L'enregistrement du retour à l''équilibre est une onde sinusoïdale amortie (FID) qui est propre à chaque proton différent (si le noyau utilisé est le ^1H) présent l'échantillon analysé. Les noyaux tels que ^1H, ^{13}C, ^{15}N, ^{19}F et ^{31}P ont un spin nucléaire non nul (*I*=1/2), ils ne possèdent donc que deux états de spin qui en font des spectres simples souvent utilisés en RMN. Cependant, le ^1H et le ^{13}C sont les noyaux principalement utilisés en métabolomique de par leur sensibilité et leur abondance naturelle. Le passage du domaine temporel (FID) au domaine fréquentiel (spectre) se fera par une transformation de Fourier (TF). Celle-ci permet d'extraire toutes les fréquences différentes contenues dans la FID et de les positionner sur un axe de fréquence.

Les résultats de RMN sont reproductibles et offrent la possibilité de fournir des informations structurelles sur les groupements chimiques des molécules [46] [47].

Le spectre RMN, (exemple figure 6) comporte plusieurs paramètres permettant l'identification des composés qui sont :

- le **déplacement chimique** exprimé en ppm : chaque groupement a sa propre fréquence de résonance dans B_0 en fonction de son environnement chimique et électronique. Pour cette raison, tous les protons différents d'un échantillon résonneront avec leur propre fréquence et auront leur propre FID. Après TF, chaque proton aura sa propre position sur l'axe de fréquence appelé aussi déplacement chimique (δ).

- la **constante de couplage (J)** exprimée en Hz : mesure la force de l'interaction (couplage scalaire) entre deux spins nucléaires séparés par des électrons de liaisons. Cette information renseigne sur la structure et la conformation de la molécule.

- la **multiplicité** : d'un groupement rend compte du nombre de voisins qui couplent avec ce groupement. Le couplage se mesure par la constante de couplage (J).

- l'**intensité/amplitude du signal** représente la quantité d'aimantation (soit le nombre de noyaux) qui résonnent. Elle est proportionnelle au nombre de noyaux du groupement et à la concentration de la molécule. L'intensité du signal permet de quantifier les métabolites à l'aide du rapport de proportionnalité entre le rapport signal sur bruit et la quantité de métabolites.

Le spectre suivant (figure 6) représente les déplacements chimiques des deux groupements CH_3 de la molécule de valine à pH de 7.4. Les pics RMN de ces deux groupements sont sous forme de doublet car ils ont tous les deux un H comme voisin (CH_B). La multiplicité du signal est de N+1 avec n le nombre de voisin du groupement.

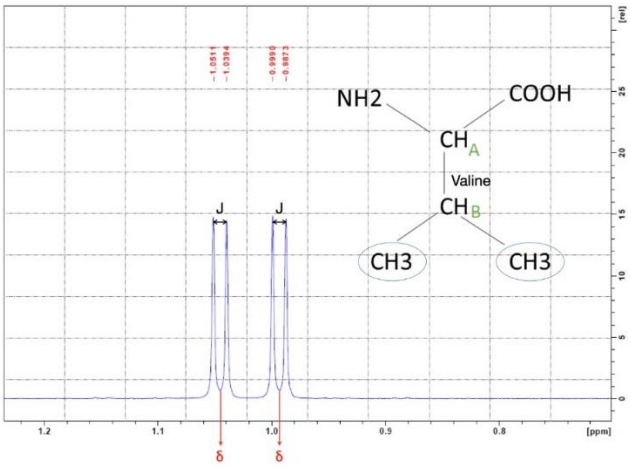


Figure 6 : Exemple d'un spectre RMN des deux doublets de la valine avec son déplacement chimique (δ), sa constante de couplage J, sa multiplicité et son intensité

Cette technique analytique est non destructive pour les molécules et permet des analyses rapides d'une large gamme de métabolites [46-49] avec des limites de détection de l'ordre du µM ou du nM [46, 50].

2. Acquisition des données

2.1. RMN unidimensionnelle (1D)

L'acquisition en RMN 1D est utilisée le plus couramment. L'interprétation spectrale peut être affectée par la présence de solvant et/ou de macromolécules restantes dans l'échantillon biologique. Par conséquence, un grand nombre de séquences d'impulsion permettent d'analyser les milieux biologiques telles que la séquence 1D NOESY–PRESAT (Nuclear Overhauser Effect Spectroscopy), ou encore la séquence CPMG (Carr-Purcell-Meiboom-Gil).

La 1D NOESY–PRESAT[51] permet de supprimer efficacement le signal de l'eau afin de visualiser les métabolites en concentration faible (de l'ordre du µM) contenus dans les échantillons biologiques utilisés en métabolomique.

La séquence CPMG [52] présente l'avantage de pouvoir diminuer la résonance de grosses molécules telles que les protéines et les lipides. En effet ces macromolécules induisent un élargissement des signaux du fait de leur temps de relaxation T_2 très inférieur à celui des métabolites dont la masse moléculaire est <1kDa. Cette séquence permet d'éliminer le signal des molécules possédant un temps de relaxation T_2 court. Les premiers points d'acquisition de la FID sont donc éliminés volontairement, ce qui engendre une quantification partielle avec une perte de la proportionnalité entre l'amplitude et la concentration des autres composés. Couplée à une pré saturation de l'eau, la séquence CPMG permet donc de réduire très fortement les signaux liés aux protéines et est donc particulièrement adaptée à l'étude métabolomique d'échantillons biologiques liquides tels que les urines ou le plasma qui peuvent être riches en protéines et/ou en lipides.

2.2. RMN multidimensionnelle (2D)

Les échantillons biologiques (sérum/plasma, fèces, urine, salive) sont des milieux riches contenant de nombreux métabolites. Le recouvrement des signaux en RMN 1D peut rendre parfois l'identification des composés très difficile. L'utilisation de la RMN 2D est une alternative pour permettre de séparer les métabolites grâce à la 2ème dimension. Les signaux sont dispersés et leur identification est ainsi facilitée. Pour la métabolomique, deux types de séquences 2D peuvent être utilisés en regardant les interactions ^{1}H-^{1}H et les interactions ^{1}H-^{13}C.

Les interactions ^{1}H-^{1}H sont réalisées à l'aide de séquences COSY (COrrelation spectroscopY) (corrélation entre ^{1}H proches vicinaux ou géminés) [53] et TOCSY (Total Correlation Spectroscopy, corrélations entre tous les protons d'un système donné et pas seulement entre les protons vicinaux ou géminés [54]. Ces deux types de séquence sont régulièrement utilisés lors d'étude métabolomique.

Les séquences utilisant des interactions hétéronucléaires comme ^{1}H-^{13}C HSQC (Heteronuclear Single Quantum Correlation) ou ^{1}H-^{13}C HMBC (Heteronuclear Multiple Bond Correlation) sont de plus en plus utilisées en métabolomique car la fenêtre spectrale du carbone (environ 220 ppm) permet d'étaler et de résoudre les pics sur la 2ème dimension [55-57]

D'autres types de séquence bidimensionnelle peuvent être utilisés tels que la DOSY (Diffusion Ordered Spectroscopy [56], ou encore 2-D J résolue (met en évidence les constantes de couplage) [53], qui sont un compromis entre une meilleure sensibilité et une dispersion spectrale limitée du proton (environ 12 ppm).

Un gain de résolution avec une identification améliorée des composés dans un mélange peut être obtenu avec RMN 2D [58]. Cependant, la sensibilité et le débit sont affectés négativement [59] . Pour pallier à cela, une séquence NUS (non uniform sampling) a été développé en 2015 [60]. Le NUS permet de diminuer les temps d'analyse en diminuant le nombre de point d'acquisition en faisant un échantillonnage non uniforme du signal temporel permettant ainsi une identification des molécules plus rapide qu'une RMN 2D classique..

3. Traitement des données

3.1. Pré-traitement

La première étape d'une analyse RMN est la transformation de la FID (free induction decay dans le domaine temporel) en un spectre de fréquence (dans le domaine fréquentiel) grâce à une transformée de Fourier qui permet d'extraire toutes les fréquences de la FID et de les positionner sur un axe de fréquence. Cette étape est généralement réalisée automatiquement par les logiciels de traitement (voir IV – 1).

La déformation de la ligne de base peut rendre très difficile l'identification des pics et peut conduire à des erreurs significatives dans toute mesure quantitative. La

correction de la ligne de base [61, 62] est une étape essentielle pour obtenir un spectre RMN de bonne qualité. Cette étape peut être réalisée manuellement ou automatiquement tout dépend du logiciel de traitement utilisé.

Beaucoup de paramètres tels que le pH, la température et la concentration en sel peuvent affecter le déplacement chimique des groupements moléculaires et avoir des conséquences sur l'interprétation et la quantification des molécules. Pour réduire cet effet le binning ou bucketing method [63] et l'alignement des pics sont utilisés [64].

L'alignement des pics est réalisé avant la méthode binning afin de réduire l'éventuelle dérive par rapport au déplacement chimique attendu des groupements d'un échantillon à un autre. Dans la méthode de binning, les spectres sont subdivisés en plusieurs régions de tailles égales ou variables appelées buckets. L'intégration de l'aire sous la courbe du spectre d'origine correspond à la superficie totale de chaque bucket. Ainsi, l'effet mineur du décalage de pic peut être atténué à travers les échantillons. Les applications de traitement incluent la méthode de binning équidistante par laquelle les spectres sont d'abord divisés en régions de taille égale puis intégrés.

3.2. Normalisation

L'étape de normalisation analytique est essentielle pour pouvoir comparer tous les spectres d'une même étude. Elle est introduite pour contrôler d'éventuelles variations des signaux dues à des raisons techniques. En effet, les spectres peuvent être enregistrés en utilisant un nombre différent de scans ce qui fait varier les intensités absolues des spectres. La méthode standard de normalisation des spectres RMN est la normalisation à l'aire totale [27, 65], en particulier lorsque différents échantillons ont un nombre similaire de métabolites. W. Gronwald et son équipe ont fait une étude comparative de 11 méthodes de normalisation

applicables dans le cas d'un échantillon urinaire analysé en RMN [66]. Parmi ces méthodes, la normalisation par quotient probabiliste a été spécialement développée pour pallier aux effets de dilution dans les fluides biologiques comme l'urine. Le principe de cette méthode est dans un premier temps de normaliser chaque spectre par son intensité totale puis de prendre le spectre médian comme spectre de référence. Ensuite, l'aire de chaque pic du spectre est divisée par l'aire du pic correspondant du spectre de référence. Le quotient médian de tous les quotients est déterminé et enfin tous les pics du spectre sont normalisés au quotient médian [67] .

Une autre méthode est la normalisation à une référence interne. Cette référence peut être le pic d'une molécule comprise dans l'échantillon. Par exemple, pour un échantillon urinaire le spectre est souvent normalisé par le pic de la créatinine [68]. Une référence interne telle qu'un composé ajouté en quantité connue dans la solution peut être utilisée. L'échantillon et la référence doivent remplir les conditions suivantes : aucune réaction chimique entre elles et/ou avec le solvant, une solubilité complète et aucun signal de chevauchement. Les dérivés silylés, le TSP (triméthylsilylpropanoique) [69, 70], l'hexamethyldisilane [71] ou le DSS (sodium 3-(trimethylsilyl)-1-propane sulfonate) [72, 73], sont très utilisés en ^{1}H-RMN de par leur déplacement chimique proche de 0ppm. Une alternative à l'ajout d'une référence interne est l'utilisation d'un signal électronique ERETIC™ (Electronic Reference To access In vivo Concentration) [74]. Cette méthode a été initiée pour éviter d'utiliser des composés de références en RMN in vivo. Le concept de base vient de la méthode « calibrator » proposée par Mahon [75]. La référence est non pas un signal RMN réel mais une pseudo-FID synthétisée par un dispositif électronique et transmis à l'intérieur de la sonde RMN lors de la réception du signal RMN de l'échantillon. Si ce signal électronique a la même forme et la même fréquence qu'une FID il donne lieu après transformée de Fourier à un pic supplémentaire dans le spectre. Cette méthode est appliquée en imagerie dans le domaine biomédical [74, 76].

Récemment l'étude de Zoelch *et al*. montre la possibilité de quantifier des métabolites cérébraux in vivo sur 21 volontaires sains [77].

4. Logiciels de traitement

Les étapes de traitement des données RMN peuvent être réalisées manuellement ou automatiquement. De nos jours, il existe de nombreux logiciels qui sont spécialisés dans une ou plusieurs de ces étapes (tableau 1). Ces logiciels RMN sont disponibles en open-source ou commerciaux et présentent des avantages comme des inconvénients.

Tableau 1 :Liste non exhaustive des logiciels de traitement utilisées pour le traitement des spectres RMN

Logiciel	Source	Traitement des données	Normalisation	Identification	Quantification
Topspin (Bruker-Biospin)	Brucker TopSpin	Correction de la ligne de base, phasage	oui	non	Oui
NMRprocFlow (D. Jacob *et al.*) [78]	Open source	Correction de la ligne de base, suppression du solvant et contaminants, alignement des pics	oui	non	Oui
ASICS (Tardivel *et al.*) [79]	Package gratuit du logiciel R	Correction de la ligne de base, phasage, suppression du solvant et contaminants, alignement des pics	oui	oui	Non
Chenomx NMR (NMR suite 7.7)	Chenomx	Correction de la ligne de base, phasage	non	Oui	Non
Amix tool-Kit (Bruker)	Brucker Topspin	non	Aire totale, standard externe (TSP), pic choisi	non	oui

Le logiciel Topspin 3.6 software créé par Bruker-Biospin (Karlsruhe, Germany) est un logiciel qui permet l'acquisition, la collecte des signaux et leur traitement. Des extensions sont téléchargeables tels que Amix tool-kit (Analysis of MIXture, version 3.2, Bruker Karlsruhe, Germany) pour permettre la normalisation et le découpage des spectres. Le logiciel open-source NMRprocflow [78] est un outil qui permet le traitement de données automatique tel que la correction de la ligne de base, l'alignement des pics ou le découpage des spectres. Le package R ASICS [79] permet un traitement automatisé des spectres RMN allant de l'ajustement de la phase jusqu'à l'identification des pics. L'identification des pics peut être réalisée également par ChenomX NMR (NMR Suite 7.7, ChenomX version 8.6). Ce logiciel d'analyse fonctionne avec des bibliothèques de référence spectrales complètes pour identifier et mesurer les concentrations de composés visibles en RMN.

L'identification des molécules s'effectue selon 3 critères. Le déplacement chimique des pics, il faut que tous les groupements moléculaires d'un composé soient présents dans le spectre d'acquisition. Il faut que les intensités relatives des groupements à l'intérieur d'une molécule soient respectées. De plus il faut que la multiplicité et les constantes de couplage du spectre de référence soient les mêmes que dans le spectre à identifier.

5. Avantages et inconvénients de la RMN

La RMN est une technique d'analyse non destructive, quantitative et robuste. Elle a une sensibilité de détection jusqu'au micromolaire [80] et permet d'avoir des informations structurales des métabolites. De plus, elle présente l'avantage de fournir des données robustes dans le temps c'est-à-dire que les résultats sont comparables entre différentes plateformes et à différents temps d'acquisition. Toutefois, les spectres obtenus peuvent être compliqués à interpréter du au chevauchement des signaux. En effet, la complexité et l'abondance des métabolites

présents dans un échantillon biologique rendent l'identification des signaux difficiles.

VI. La spectrométrie de masse (MS)

Depuis le début des années 2000, les études de métabolomique utilisant la spectrométrie de masse sont de plus en plus nombreuses [81]. Cette technique analytique montre ses avantages pour l'obtention d'empreinte métabolique complète notamment pour l'analyse de fluides biologiques. Ce projet de thèse a été réalisé à l'aide d'une source ESI couplée à un analyseur hybride Q-Orbitrap. Ainsi dans cette partie nous détaillerons en priorité la description de ce type de spectromètre de masse.

1. Le spectromètre de masse

La spectrométrie de masse est une technique physico-chimique d'analyse basée sur l'ionisation de molécules suivie d'un tri en fonction du rapport masse/charge. Un spectromètre de masse est composé de trois modules (figure 7).



Figure 7 : Représentation schématique des parties d'un spectromètre de masse

1.1. Les sources d'ionisation

La source d'ionisation est la partie du spectromètre de masse où sont produits des ions gazeux à partir des molécules introduites. La nature de la source utilisée dépend de plusieurs critères décrits dans le tableau 2.

Tableau 2 : Comparaison des caractéristiques des principales sources utilisées en spectrométrie de masse (www.thermofisher.com , www.agilent.com)

Source	Ionisation	Type d'échantillon	Processus ionisation	Molécules
EI	Dure	Gaz Liquide	Arrachement d'un électron par collision	Petites
CI	Douce	Gaz Liquide	Bombardement électronique et transfert de charge	Petites
ESI	Douce	Liquide	Désolvatation des ions	Petites, Macro, Protéines
APCI	Douce	Gaz Liquide	Dissociation induite par collision	Petites, Thermolabiles, Macro
SIMS	Dure	Solide	Bombardement d'ions primaires	Petites, Macro
MALDI	Douce	Solide	Transfert de charge d'une matrice	Macro, Peptides, Protéines
DESI	Douce	Solide	Désolvatation des ions	Petites, Macro, Protéines

L'**impact électronique (EI)** consiste à "bombarder" les molécules par un faisceau d'électrons de haute énergie. L'« impact » d'un électron sur la molécule M a pour effet d'arracher un électron à cette dernière, conduisant ainsi à la formation d'un ion radicalaire $M^{+\cdot}$. L'énergie des électrons incidents est telle que l'ion $M^{+\cdot}$ acquiert une grande quantité d'énergie interne qui l'amène généralement à se fragmenter spontanément en ions plus petits baptisés ions fils ou ions fragments. Les électrons sont produits par le chauffage d'un filament métallique (tungstène ou rhénium, le plus souvent) et accélérés par une différence de potentiel de 70 V, ce qui leur confère une énergie cinétique de 70 eV. Le fait de disposer d'un standard mondial pour l'énergie des électrons ionisants permet de comparer des spectres de masse réalisés sur différents appareils et d'avoir recours à des bases de données partagées (comprenant plusieurs dizaines de milliers de spectres).

L'**ionisation chimique (CI)** utilise un gaz réactif (ammoniac, méthane, isobutane) qui est ionisé par un faisceau d'électrons. Ce gaz est à une pression telle que les ions $R^{+\cdot}$ formés réagissent immédiatement avec les molécules R non ionisées pour former des ions RH^+ à nombre pair d'électrons. Ce sont ces derniers qui, dans un deuxième temps, vont réagir avec l'analyte M en lui transférant un proton pour former des ions MH^+. Il s'agit d'une méthode plus "douce" que l'ionisation par EI et qui permet l'observation plus aisée des ions moléculaires.

L'**ionisation par électrospray (ESI)** est une méthode d'ionisation atmosphérique la plus utilisée aujourd'hui en métabolomique [82] (figure 8). Le principe d'ionisation de cette source dite « douce » est de faire passer les composés d'un état liquide à un état gazeux. A pression atmosphérique, certains composés ne sont pas volatiles il est donc nécessaire d'utiliser un nébuliseur et un différentiel de voltage entre la sortie de l'aiguille du nébuliseur et l'entrée du capillaire pour les ioniser. La génération d'ions peut être décrite en trois étapes.

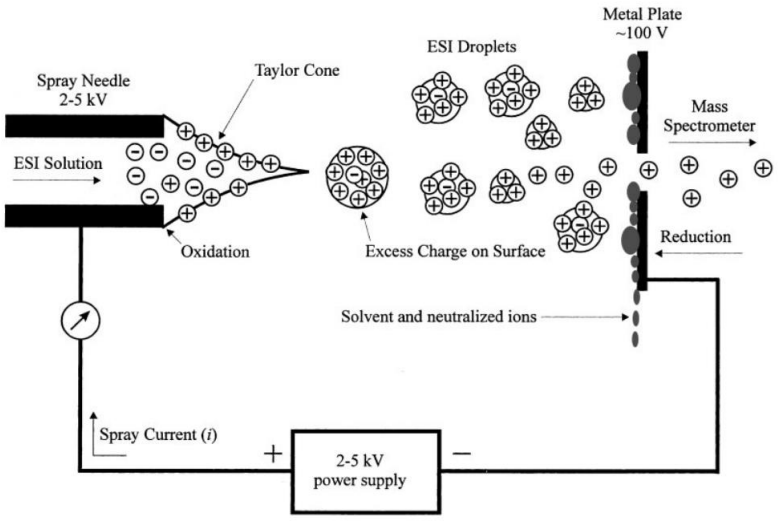


Figure 8 : Représentation schématique d'une source électrospray [83]

La première est l'étape de nébulisation. Le composé dissous dans la phase mobile est introduit dans le nébuliseur. Sous l'action du gaz du nébuliseur, l'échantillon est

dispersé sous forme de gouttelettes qui sont ensuite dirigées dans le capillaire. La différence de potentiel entre le nébuliseur et le capillaire entraîne la fixation de charges sur les gouttelettes (cône de Taylor). La deuxième étape est celle de la désorption. Le solvant des gouttelettes ainsi chargées s'évapore progressivement provoquant une augmentation des charges à la surface jusqu'à la limite de Rayleigh. Cette limite est atteinte lorsque les répulsions coulombiennes entre les charges égalent les forces cohésives de la tension de surface du liquide. Les gouttelettes vont ainsi exploser et produire de fines gouttelettes. La troisième et dernière étape est l'évaporation complète du solvant qui entraine l'ionisation et le passage sous forme gazeuse des molécules contenues dans les gouttelettes.

La **source APCI** (ionisation chimique à pression atmosphérique) utilise un nébuliseur pneumatique à azote pour former un aérosol à l'intérieur d'une chambre (le vaporiseur) chauffée à des températures élevées (entre 350 °C et 500 °C). Dans le vaporiseur, l'effluent de la colonne de chromatographie (solvants + analytes) est évaporé et se mélange avec le gaz nébuliseur pour amener les analytes vers l'électrode à décharge couronne constituée d'une aiguille et d'une chambre de nébulisation qui sert de contre-électrode. Une différence de potentiel élevée (entre ±3 et ±6 kV) est appliquée sur cette électrode, provoquant une décharge couronne de ~2-3 μ A. Cette décharge est une source constante d'électrons pour le processus d'ionisation en APCI : elle ionise l'air ambiant et crée un plasma (gaz ionisé), autour de la pointe de l'aiguille. Les ions radicalaires du plasma participent ensuite à des réactions chimiques qui donnent lieu à l'ionisation des molécules d'analyte.

L'ionisation peut également être réalisée en phase solide à l'aide de sources d'ions primaires comme en **SIMS** (spectrométrie de masse à ionisation secondaire). La source SIMS permet d'obtenir des informations élémentaires, isotopiques ou moléculaires de la surface d'un échantillon. Elle consiste à former des ions secondaires en bombardant la surface de l'échantillon à analyser avec un faisceau d'ions primaires. Cette technique d'ionisation est relativement violente et le signal

est dominé par des ions fragments. L'analyse est généralement effectuée sur l'échantillon sans préparation.

La source **MALDI** (désorption/ionisation laser assistée par matrice) est dite « douce », la désorption et l'ionisation des molécules sont réalisées à l'aide d'un faisceau laser aidées par la présence d'un composé organique (dit matrice) déposé sur l'échantillon. Un échantillon solide, dispersé dans une grande quantité de matrice (cristalline) est irradié par des photons émis par un laser dont la longueur d'onde est située dans la bande d'absorption de la matrice. L'irradiation de la matrice et de l'échantillon provoque la création d'une grande quantité d'énergie dans la phase condensée, par excitation électronique des molécules de la matière. Des ions sont formés par transfert de protons ou d'électrons entre la matière photoexcitée et l'échantillon à analyser qui se désorbent.

La source **DESI** se caractérise par sa capacité à ioniser de petites à grandes molécules sans fragmentation significative comme la source ESI. Un mélange électriquement chargé d'eau et de solvant organique est pulvérisé vers la surface analysée. Ce processus de propagation du liquide entraîne la formation de gouttelettes qui sont ensuite projetées par l'arrivée de nouvelles. Le contact solvant-surface entraine la dissolution des molécules d'analyte de la surface et les ionise de la même manière que dans l'ESI. Les ions produits sont transportés vers l'analyseur de masse par des forces hydrodynamiques.

1.2. Les analyseurs

Le choix d'un analyseur dépend du type d'information souhaité mais également du type de source. Les analyseurs peuvent se différencier selon plusieurs paramètres : la fréquence d'échantillonnage, la résolution (capacité à séparer deux ions de m/z proches) et la précision en masse (exprimée en ppm, c'est le rapport entre la différence de la masse théorique et celle mesurée sur la masse théorique) (tableau 3).

Tableau 3 : Comparaison des caractéristiques des principaux analyseurs utilisés en spectrométrie de masse (données extraites des constructeurs : www.waters.com , www.bruker.com et www.thermofisher.com)

Exemples	Analyseur	Avantages	Inconvénients	Précision en masse	Fréquence d' échantillonnage	Résolution
Xevo TQ-XS (Waters)	Quadripôle	Rapidité Quantitatif	Précision en masse	<1ppm	40 Hz (à m/z 200)	240 000 (à m/z 200)
LTQ XL linear ion trap (Thermo Fisher Scientific)	Piège à ions	Rapidité Quantitatif	Couplage à un autre analyseur	1,5ppm	80Hz (à m/z 609)	300 000 (à m/z 609)
tims TOF (Bruker)	TOF	Rapidité	Précision en masse	<1ppm	100 Hz (à m/z 1222)	240 000 (à m/z 200)
Quadripôle-Orbitrap™ (Thermo Fisher Scientific)	Orbitrap	Précision en masse	Lenteur Quantitatif	<1ppm	12 Hz (à m/z 200)	280 000 (à m/z 200)
SolariX MRMS (Bruker)	FT-ICR MS	Précision en masse	Infusion directe	<0.6ppm	1 Hz (m/z 200)	60 000 000 (à m/z 702)

Un **quadripôle** est composé de quatre électrodes cylindriques ou hyperboliques reliées électriquement deux à deux (figure 9) et soumises à un potentiel sinusoïdal ϕ_0. Ce potentiel est la combinaison d'une tension continue (V_{DC}) et une tension alternative (V_{AC}) de fréquence angulaire ω. $\phi_0 = V_{DC} + V_{AC}\cos(\omega t)$

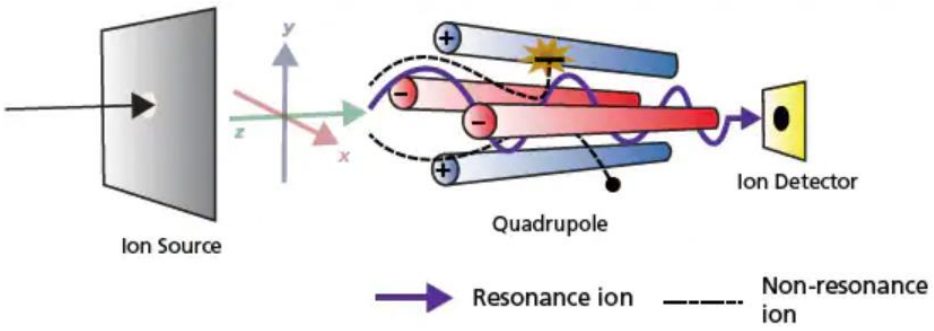


Figure 9 :Représentation schématique d'un analyseur quadripolaire (https://www.shimadzu.com)

Ces électrodes sont soumises à deux potentiels. Les 2 électrodes opposées sont soumises aux mêmes tensions créant ainsi un champ quadripolaire. Deux types de trajectoire ionique sont possibles. Une trajectoire dite « stable » où seuls les ions oscillant autour de l'axe de l'analyseur sont conduits jusqu'à l'extrémité de

l'analyseur. Une trajectoire dite « instable » où les ions s'écartent de l'axe et sont arrêtés lorsqu'ils entrent en collision avec les électrodes. Cette particularité permet de cibler des gammes de m/z selon l'analyse souhaitée. La stabilité des ions au sein du quadripôle dépend de leur rapport m/z et des tensions appliquées aux électrodes.

Les **pièges à ions linéaires** sont composés de quatre électrodes à section hyperbolique. Les ions vont d'abord être piégés à l'intérieur des quatre électrodes puis par application de radiofréquences plus ou moins élevées, les ions vont avoir des oscillations plus ou moins stables en fonction de leur rapport m/z. Cela implique que les ions avec les oscillations instables vont être éjectés du système pour être détectés.

La spectrométrie à **temps de vol** (Time of flight ou TOF) est une méthode de spectrométrie de masse dans laquelle les ions sont accélérés par un champ électrique de valeur connue. Il résulte de cette accélération que les ions de même charge électrique acquièrent la même quantité de mouvement. La vitesse des ions, par contre, dépend du rapport masse sur charge. On mesure le temps mis par une particule chargée pour atteindre un détecteur situé à une distance connue. Ce temps dépendra du rapport m/z de la particule considérée. Ce sont les particules les plus lourdes qui seront accélérées aux vitesses les plus basses. La détermination du rapport m/z découle de ce temps de vol et de la connaissance des autres paramètres expérimentaux comme la position du détecteur et la tension d'accélération. Cet analyseur est rapide, à résolution variable selon les appareils. Il est l'analyseur en masse le plus souvent utilisé pour les analyses à haute cadence (notamment pour l'imagerie par spectrométrie de masse).

Le développement du premier analyseur de masse **Orbitrap** a été motivé par le désir de construire un nouveau spectromètre qui éviterait les défauts des instruments antérieurs, tels que la complexité et la taille des analyseurs à résonance

cyclotron ionique à transformée de Fourier (FT-ICR) ; la faible sensibilité, plage dynamique et (à l'époque) résolution des analyseurs de temps de vol orthogonal (TOF) ; et la précision de masse limitée des analyseurs à piège à ions [84].

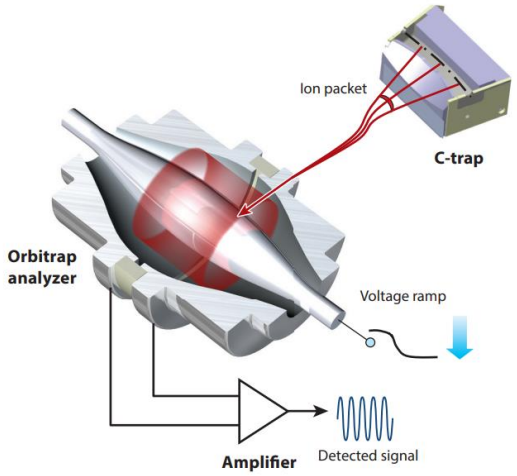


Figure 10 : Représentation d'un analyseur Orbitrap [84]

L'Orbitrap est composé d'une électrode centrale et d'une électrode externe en forme de cylindre. Toutes les deux sont alignées le long de l'axe horizontal (figure 10). Entre les deux électrodes se trouve une chambre de mesure composé d'un vide 10^{-8} torr généré par un système de pompage. Les deux électrodes sont connectées à des bornes de voltages différents, ce qui permet de faire varier leur voltage de façon indépendante et donc de générer un champ électrique. Les paquets d'ions entrent dans l'analyseur Orbitrap de façon tangentielle et tournent autour de l'électrode centrale. L'attraction électrostatique dirigée vers l'électrode centrale est compensée par la force centrifuge des ions. Le mouvement des ions dans l'Orbitrap est modélisé par des équations mathématiques décrites par Alexander Makarov [85].La vitesse angulaire de déplacement d'un ion le long de l'axe z est fonction de sa charge q et de sa masse m et de la courbure du champ k : $\omega = \sqrt{\left(\frac{q}{m}\right)k}$ en rad/sec [85]. La fréquence d'oscillation d'un ion sur l'axe z crée un courant induit qui est

mesuré par l'électrode externe. Dans le cas de l'analyse d'un mélange d'ions, le signal mesuré est alors la combinaison des fréquences des différents ions piégés. Une transformation de Fourrier permet de décomposer le signal de synthèse en ses différentes composantes ; des traitements additionnels du signal permettent de reconstituer le spectre de masse. Un dispositif de stockage d'ion a été créé pour permettre l'accumulation d'ion avant injection dans l'analyseur Orbitrap. Ce stockage externe nommé C-Trap (figure 11) a été développé pour mettre le couplage de l'orbitrap qui a un fonctionnement discontinu avec des sources d'ionisations douces telles que l'électrospray [86-88]. Une fois couplés avec succès à l'ionisation par électrospray, des mélanges complexes ont pu être analysés avec le spectromètre Orbitrap. La C-trap est un piège électrostatique qui permet de focaliser les ions au moyen d'une radiofréquence et d'une pression de gaz inerte. Elle permet d'éjecter les ions orthogonalement vers l'Orbitrap selon deux critères définis par l'AGC (automatic gain control) soit un nombre d'ion compris dans la trappe soit un intervalle de temps [85].

Le **FT-ICR-MS** est un analyseur qui consiste à soumettre les ions à un fort champ magnétique dans une cellule ICR (Ion Cyclotron Resonance). Ce champ entraine les ions à adopter un mouvement circulaire dans la cellule dont la fréquence dépend du rapport m/z de l'ion. L'application d'un champ électrique de même fréquence cyclonique d'un ion présent conduit à l'absorption d'énergie par l'ion, augmentant ainsi son énergie cinétique et de ce fait son rayon de mouvement d'oscillation. Il s'agit du phénomène de résonance de l'ion. Plus l'impulsion sera importante plus l'orbite suivie sera grande et donc plus les ions se rapprocheront des plaques de détection. Le signal est traité par transformée de Fourier pour obtenir un spectre de masse. Cet analyseur a un temps d'acquisition long, il n'est pas utilisé en routine clinique et est peu compatible au couplage avec les techniques séparatives. Il est souvent couplé à un système d'introduction directe [89]. Le FT-ICR se distingue particulièrement par ses très hautes résolution et précision de mesure de masse très utiles pour identifier sans ambiguïté des métabolites inconnus.

Pour conclure, il convient de noter que chaque instrument présente des avantages et des inconvénients qui lui sont propres (Tableau 3). Les différents instruments disponibles peuvent donc apparaitre comme complémentaires les uns des autres.

1.3. Couplage d'analyseurs / approche en tandem

Afin d'améliorer leurs performances quantitatives et d'acquérir des spectres MS/MS nécessaires à l'identification de métabolites, les spectromètres de masse comportent généralement plusieurs analyseurs. Un des analyseurs sert à filtrer les ions et l'autre permet de les analyser. L'un des spectromètres de masse en tandem le plus connu et le plus utilisé pour des approches de quantification ciblée est le triple quadripôle (QqQ). Comme son nom l'indique, il est composé de trois quadripôles en série ayant chacun leur rôle (filtration, fragmentation, analyseur). Par exemple, les ions sont analysés par le premier analyseur, puis passent dans le deuxième analyseur dit chambre de collision où ils sont fragmentés. Les fragments passent ensuite dans un troisième analyseur de masse où ils sont analysés. Il existe différentes méthodes d'acquisition qui sont détaillées dans la partie Acquisition des données – Ciblée. Ce type d'instrument permet d'augmenter la sensibilité et la spécificité de détection par rapport à un quadripôle simple. Ces instruments sont très utilisés pour les analyses ciblées quantitatives en raison de leur grande gamme dynamique [90].

Des spectromètres de masse « hybrides » ont été développés afin d'avoir un pouvoir résolutif plus important qu'un triple quadripôle. La résolution et la précision de mesure du Q-TOF sont fonction du TOF. L'association du quadripôle au TOF permet de combiner la robustesse du quadripole pour la sélection d'ions parents avec la précision de mesure du temps de vol) [91]. Le **Q-Orbitrap** est un analyseur hybride plus précis que le Q-TOF. En effet, l'Orbitrap permet une meilleure précision en masse que le TOF mais avec une fréquence d'échantillonnage plus faible. L'instrument Q Exactive a été développé par Thermo Fisher ® [92] (figure 11).

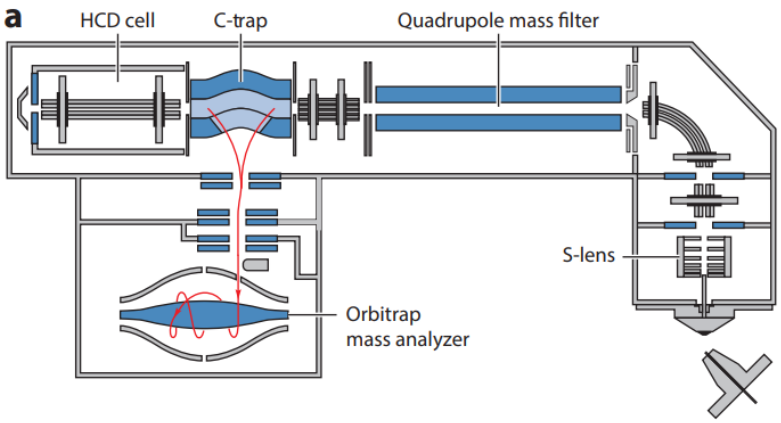


Figure 11 : Schéma d'un quadripôle couplé à un Orbitrap (Q-Orbitrap) [84]

Cet analyseur hybride est composé d'un quadripôle utilisé comme filtre de masse et permet l'isolement des ions précurseurs. L'Orbitrap est utilisé pour la détection en balayage complet ou en mode ciblé lors d'acquisition de spectres MS^2 utilisant ainsi une cellule HCD pour la fragmentation. La cellule HCD est une cellule de collision quadripolaire munie de miroirs électrostatiques à ses extrémités pour permettre de retenir puis expulser en sens inverse les ions qu'elle contient. Cet analyseur a montré ses avantages notamment lors d'étude métabolomique non ciblée. De Paepe et son équipe l'utilise lors d'étude métabolomique non ciblée d'échantillons complexes tels que l'urine, le plasma et les fèces [30]. Le Q-Orbitrap combiné à une technique séparative en phase liquide a permis la détection de 9647 composés dans l'urine, 9672 dans les fèces et 6122 dans le plasma. Ce type d'étude permet de se rapprocher de l'objectif premier de la métabolomique qui est la détection la plus large possible du métabolome d'un échantillon pouvant aboutir à la découverte et l'identification de biomarqueurs potentiels. Ainsi, l'utilisation d'analyseur hybride avec des méthodes séparatives liquides ou gazeuses confère des propriétés avantageuses pour l'étude des matrices biologiques ou de systèmes complexes.

1.4. Couplage avec la chromatographie gazeuse

Les fluides biologiques sont des milieux très complexes ainsi, pour une analyse en métabolomique la spectrométrie de masse nécessite un couplage avec une méthode séparative. Généralement, cette méthode séparative peut être la chromatographie gazeuse (GC). La GC couplée à la spectrométrie de masse est une technique qui permet de séparer les composés volatils ayant une masse inférieure à 1000 Da s'ils sont thermostables. C'est une technique sensible généralement associée à une source à ionisation par impact électronique ou par ionisation chimique. L'énergie transmise aux molécules lors de l'impact électronique étant importante, les spectres de fragmentation obtenus sont riches en informations ce qui a permis d'établir des bibliothèques de spectres applicables à n'importe quel appareil. La principale limite de cette technique est qu'elle ne peut analyser que des métabolites volatils et thermostables. Elle nécessite généralement une étape de préparation des échantillons (dérivation chimique) afin de rendre les métabolites les plus volatils possibles. Ce traitement des échantillons est assez long et peut conduire à une perte de métabolites. De plus la GC-MS ne permet pas d'observer des composés très polaires (Lee et al. 2013) tels que les nucléotides.

1.5. Couplage à la chromatographie liquide

La chromatographie liquide (LC) est de nos jours la technique la plus largement utilisée en métabolomique grâce aux nombreux types de colonnes que l'on peut utiliser. Les différents types de colonnes permettent d'analyser une large gamme de métabolites. L'efficacité de séparation de la colonne dépend de la longueur et du diamètre de la colonne, de la nature et de la taille des particules de la phase stationnaire ainsi que de la phase mobile. À l'heure actuelle deux principales colonnes sont utilisées en métabolomique.

1.5.1. Chromatographie en phase inverse (RP)

Les colonnes sont composées d'une phase stationnaire contenant de la silice sur laquelle des chaines alkyles sont greffées. Les composés sont alors retenus via des interactions hydrophobes. Selon la nature et la taille des chaines alkyles greffées, la rétention des composés varie. La phase stationnaire greffée avec des chaines alkyles avec 18 carbones dite C_{18} est très utilisée en métabolomique pour l'analyse de composés à polarités moyennes et/ou apolaires [93]. La phase mobile est majoritairement composée de solvants aqueux tel que l'eau qui est le plus souvent utilisé. Lors d'une analyse, le gradient d'élution débute avec une forte proportion en eau puis les composés sont élués par ajout de solvant organique (MeOH ou ACN).

1.5.2. Chromatographie à interactions hydrophiles (HILIC)

La chromatographie à interactions hydrophiles (HILIC) est également très utilisée pour les composés polaires [48]. Les colonnes HILIC sont composées d'une phase stationnaire de silice greffée avec par exemple un groupement diol, amino ou sulfobétaine afin de rendre la phase stationnaire polaire et qu'une couche d'eau puisse être absorbée à la surface de cette phase. HILIC combine les phases stationnaire et mobile polaires [93]. La teneur élevée en solvant organique de la phase mobile HILIC se traduit par une bonne compatibilité avec la détection MS [48]. Un des inconvénients souvent évoqué pour cette technique est le temps assez long de rééquilibrage de la colonne entre deux échantillons afin d'obtenir les analyses les plus robustes possibles.

Ces deux colonnes sont donc complémentaires et la combinaison des deux permet de couvrir une grande partie du métabolome. Cependant cette approche nécessite un temps d'analyse plus long car un même échantillon doit être injecté plusieurs fois.

2. Acquisition des données

L'acquisition des données par spectrométrie de masse peut être réalisée de façon non ciblée ou ciblée. Cette partie décrit ces deux types d'acquisitions dans le cas d'un analyseur hydride Q-Orbitrap.

2.1. Acquisition des données non ciblées

En métabolomique, chaque type d'échantillon a son empreinte qui peut être définie de trois manières différentes par la MS (figure 12).

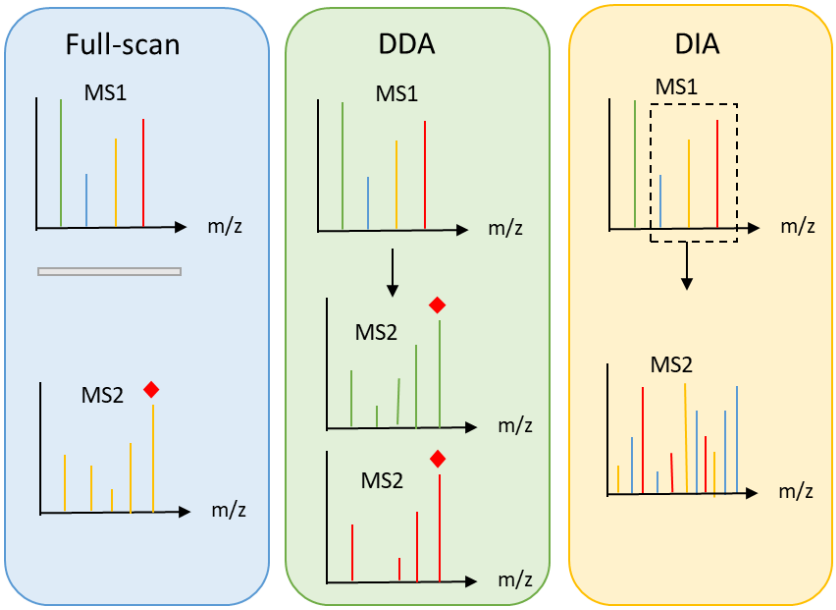


Figure 12 : Représentation schématique d'une acquisition en full-scan ,Data dependent acquisition (DDA) et Data independent acquisition (DIA)

L'acquisition **full scan** est une méthode de screening qui consiste à mesurer le rapport m/z (masse sur charge) de tous les ions détectés dans un échantillon. Cette acquisition peut être associée à une prise d'empreinte physico-chimique de l'échantillon. Cette approche est relativement simple et permet de comparer des profils de spectres métabolomiques entre eux. L'avantage de ce mode d'acquisition est qu'il permet la détection de la majorité des m/z des ions. Toutefois, l'absence de

MS^2 ne permet pas une identification formelle des ions. En effet, pour un même m/z plusieurs molécules peuvent correspondre seule une idenfication putative peut être faite. Cette acquisition est souvent suivie par une analyse ciblant l'ion dont on veut connaître la MS^2 pour confirmer son identification.

L'acquisition **DDA** (Data Dependent Acquisition) est une approche complémentaire à une full-scan. En effet, le spectromètre de masse permute automatiquement entre le mode full-scan et la MS^2. Cette permutation se fait selon les critères définis par l'utilisateur, le seuil d'intensité d'un ion, tels que la reconnaissance d'un profil isotopique particulier, le défaut de masse ou la perte d'un neutre. Le nombre de MS^2 réalisé dans chaque cycle est défini par l'utilisateur mais également par le système chromatographique utilisé. En effet, le temps d'élution de chaque composé est à considérer afin d'avoir assez d'information pour l'obtention d'une MS^2 correcte et permettant une quantification acceptable. Ce type d'application commence à se développer dans les études de métabolomique. Elle est surtout utilisée sur des instruments à haute vitesse d'acquisition tels que des instruments à temps de vol [94] mais certaines études utilisent l'Orbitrap avec une résolution non maximale [95]. L'avantage de cette technique est qu'elle permet d'obtenir des spectres MS^2 relativement purs car les ions parents sont sélectionnés individuellement (sauf en présence de composés isobares). L'interprétation des spectres MS^2 obtenus est donc facilitée. Il existe cependant un risque que certains ions d'intérêt ne soient pas fragmentés [96]. En effet, malgré l'application de différents critères permettant d'éliminer un certain nombre ions, les ions d'intérêt de faible intensité ont peu de chance d'être sélectionnés lors d'une analyse globale.

L'approche **DIA** (Data Independent Acquisition) consiste à fragmenter tous les ions sans critères de sélection en plusieurs fenêtres de masse de petite taille [97]. La taille des fenêtres de masse et donc le nombre de spectres MS^2 nécessaires pour couvrir toute la gamme de masse sont adaptés à la vitesse d'acquisition de l'instrument. Elle peut être ajustée de façon à utiliser des fenêtres de plus petite largeur sur les

gammes de masse où de nombreux métabolites sont attendus et des fenêtres de plus grande largeur là où peu de métabolites sont attendus [98]. Ce type d'acquisition entraine la création d'une taille importante d'informations MS et MS² nécessitant l'utilisant de logiciel spécialisé pour décorreler les spectres MS².

2.2. Acquisition des données ciblées

L'approche ciblée est actuellement la plus utilisée lors d'études métabolomiques car elle permet d'identifier formellement des métabolites. Dans ce mode, le m/z et le temps de rétention sont renseignés et en fonction de ces informations, les ions ciblés sont spécifiquement isolés et fragmentés. Généralement, ce type d'acquisition est précédé par une analyse en full-MS afin de préalablement détecter les ions présents dans un échantillon.

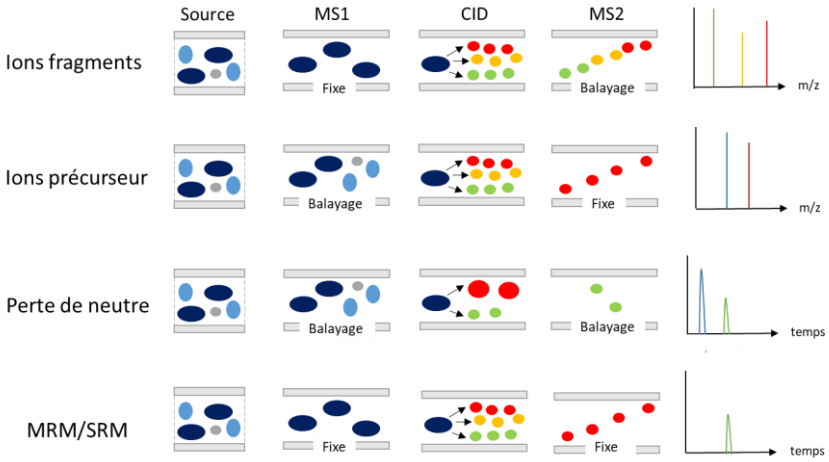


Figure 13 : Mode d'acquisition des données en approche ciblée en utilisant un spectromètre de masse schématisé pour un analyseur triple quadrupôle

L'analyse MS/MS permet d'avoir plus d'informations sur la structure du métabolite analysé et d'aider à son identification. Elle peut être menée selon quatre modes différents selon l'information recherchée (figure 13). La figure 13 représente schématiquement les 4 modes d'acquisitions en mode ciblé en utilisant un triple Quad. Ce type d'acquisition peut être réalisée par d'autres analyseurs tels que Q-

Orbitrap. Dans ce cas l'analyseur Orbitrap est représenté par le module MS2 (figure 13). Le mode ion fragment est le plus utilisé pour obtenir des informations structurales. L'ion à étudier est sélectionné en focalisant le premier analyseur sur son rapport m/z. Les fragments formés dans la cellule de collision sont séparés et analysés. Les modes ion précurseur et perte de neutre sont d'un usage plus restreint et permettent de mettre en évidence des ions ayant des particularités communes. Le mode Multiple Reaction Monitoring (MRM) est souvent utilisé lors d'étude quantitative car il présente une double sélectivité. L'ion parent à étudier est sélectionné et fragmenté, le second analyseur est focalisé sur l'ion produit. Les deux analyseurs étant fixés à des tensions constantes, la sensibilité de détection est améliorée par rapport aux autres modes de balayage, faisant de la MRM un mode de choix pour la quantification.

3. Logiciels de traitement et prétraitement des données

Une fois acquis, les signaux bruts (chromatogrammes, spectres) sont prétraités par des logiciels pour faciliter l'identification et la quantification (absolue, relative) des composés (Tableau 4). Le but commun de ces outils est de détecter l'ensemble des pics de l'analyse chromatographique en les associant aux données de spectrométrie de masse afin de créer une matrice à deux dimensions. Les données tridimensionnelles sont donc réduites à deux dimensions en associant un rapport m/z à un temps de rétention. La matrice brute obtenue à partir de ce traitement des données recense l'ensemble des variables détectées (en colonnes) et leurs intensités dans chacun des échantillons dans lesquels ces variables ont été mesurées (en lignes).

Tableau 4 :Liste non exhaustive des logiciels de traitement de données en spectrométrie de masse

	Accès	Traitement des données	Normalisation	Identification	Identification à partir de MS^2
Xcalibur (ThermoFisher Scientific)	Logiciel constructeur payant (ThermoFisher Scientific)	Détection des pics directement sur les données brutes	non	Oui à partir d'une base de donnée locale	Non
XCMS (Smith et al.) [99]	Package gratuit du logiciel R	Correction de la ligne de base, détection	non	Connexion à metlin	non
MSDial (Tsugawa et al.) [102]	Open source	Détection des pics, alignements des pics, Déconvolution des pics pour le traitement MS^2	non	Oui à partir d'une base de donnée locale	oui
W4M (Giacomoni et al.) [101]	Open source	Correction de la ligne de base, détection	oui	Connexion à HMDB, lipidmaps	oui

Ce prétraitement implique la réduction du bruit, la correction du temps de rétention, la détection et l'intégration des pics et l'alignement du chromatogramme. Xcalibur est un logiciel commercial complet de Thermo Scientific qui permet d'effectuer la configuration de la méthode, l'acquisition, le traitement et l'intégration de données. L'utilisation de ce logiciel est compréhensible de par la visualisation des pics chromatographiques et des spectres lors du traitement des données. Toutefois cet avantage peut entrainer un biais car l'intégration des données est dépendante de l'appréciation de l'opérateur. Ainsi un pic chromatographique peut être non intégré alors qu'il est bien présent. Le traitement des données peut être ainsi automatisé. Le package R XCMS [99] est un outil performant et automatisé mais nécessitant une compréhension du codage sous R. L'extraction des données brutes se déroule en plusieurs étapes (figure 14). La conception de cette méthode est modulaire et permet de supprimer ou ajouter des étapes de traitement supplémentaires selon les jeux de données.

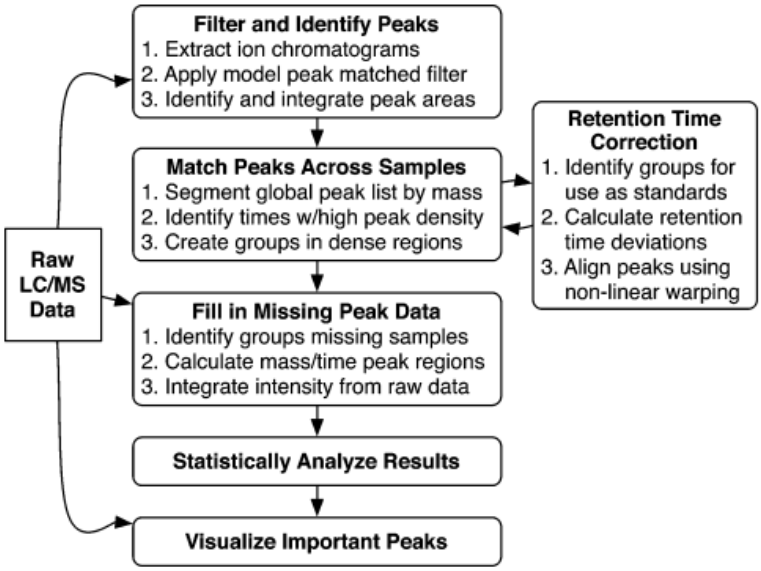


Figure 14 : Organigramme montrant la stratégie générale de prétraitement et l'analyse des données LC-MS par XCMS [100]

Le jeu de données LC-MS est introduit en format raw ou .mzXML. La première étape consiste à la détection des pics « Filter and Identifiy Peaks ». Il s'agit à la fois

d'informations spectrales avec la détection des différents signaux de masse présents à un instant t mais aussi d'informations chromatographiques avec la détection des pics correspondant aux différents ions détectés au cours du temps. L'algorithme de détection de pic consiste à la découpe des données LC/MS en tranches d'une fraction d'unité de masse (0,1 m/z) de large modulable. A l'intérieur de chaque tranche, le signal est déterminé en prenant l'intensité maximale du m/z à chaque point de temps dans la tranche. Des couples m/z et RT (temps de rétention) sont ainsi créés. L'étape d'après « match peaks accross samples » consiste à rechercher dans l'ensemble des échantillons les différents couples. Lorsqu'un couple est retrouvé dans plusieurs échantillons il est considéré comme un « group » permettant ainsi d'éliminer les couples redondants d'un échantillon à l'autre. L'étape « retention time correction » est optionnelle elle permet de corriger la dérive du temps de rétention à l'intérieur d'un même groupe. En effet, lors d'une analyse des petites variations des temps de rétention peut être observées. Une étape de correction de ces temps est donc réalisée pour permettre de grouper de façon optimale les signaux d'un même ion présent sur les différents échantillons. La dernière étape dans la détection de pics est « Fill in missing peak data » qui consiste à la recherche et l'ajout de pics manquants. En effet, les intensités des pics d'un composé pouvant varier d'un échantillon à l'autre, il est possible que certains pics ne soient pas détectés sur tous les échantillons lors de la première détection sans a priori. Ils sont alors recherchés de manière ciblée pour permettre de vérifier que la première détection n'a pas oublié un signal. Après ces étapes de prétraitement, un tableau de données est obtenu avec des intensités spectrales correspondant aux rapports m/z et RT associé en lignes et les observations en colonnes. Des analyses statistiques peuvent être effectuées lors de l'étape « statistically analyze Results ». Cette étape est utile lorsque le jeu de données possède plus d'un groupe et permet d'étudier simultanément plusieurs variables (tous les couples détectés) et visent à sélectionner les variables d'intérêt d'un groupe d'échantillons. Cette analyse statistique permet donc d'aider à la visualisation des pics importants « Visualize important peaks ». Ce package est utilisé par l'open source

Workflow4Metabolomics W4M [101]. Elle a été développé en 2014 par une collaboration française entre les domaines de la métabolomique et la bio-informatique. W4M permet d'avoir une interface plus accessible pour les non-initiés au codage tout en gardant les avantages de XCMS. Cependant, le traitement des analyses MS² reste encore incomplète par ces deux logiciels notamment pour les analyses DIA. L'open source MSDial [102] est développée en 2015 afin de faciliter l'identification et la quantification des acquisitions MS² notamment celles réalisées en mode DIA. Elle permet de déconvoluer les pics des chromatogrammes et d'associer l'analyse MS² à l'ion précurseur qui lui correspond.

4. Traitement des données

4.1. Identification

L'identification des pics détectés peut être réalisée selon différents niveaux de confiance [103]. Le niveau 1 correspond aux composés identifiés de façon certaine. Pour qu'un composé puisse être considéré comme identifié, il faut qu'il soit validé par au moins deux critères. Il peut par exemple s'agir du temps de rétention et de la masse précise ou de la masse précise et du spectre de fragmentation. La comparaison de ces critères doit être faite par rapport à un composé de référence analysé dans les mêmes conditions expérimentales. Le niveau 2 correspond aux composés putativement annotés. Il s'agit généralement de composés pour lesquels une annotation est très probable, mais pour laquelle le standard n'est pas disponible commercialement ou de composés qui sont annotés à partir de base de données externes au laboratoire menant l'étude. Le niveau 3 correpond aux composés dont l'annotation n'est pas complète, et/ou pour lesquels plusieurs annotations sont proposées, mais pour lesquels les informations disponibles permettent de les classer au sein d'une famille chimique. Le niveau 4 correspond aux composés inconnus. Il s'agit des variables détectées pour lesquels les

informations disponibles ne permettent pas de tirer de conclusions claires quant à la nature du composé.

L'identification des données est réalisée à l'aide de base de données internes au laboratoire ou des bases de données en ligne. Le tableau 5 résume une partie des bases de données en ligne permettant l'identification des métabolites.

Tableau 5 : Bases de données utilisées pour l'identification de métabolites (https://metlin.scripps.edu , https://hmdb.ca, https://lmdb.ca, https://www.genome.jp/kegg, https://pubchem.ncbi.nlm.nih.gov)

Base de données	Nombre de composés	Applications	Commentaires
METLIN	Environ un million de composés	Lipides, steroides, bacteries, metabolites, petits peptides, carbohydrates, drogues metabolites	Identification avec ion précurseur, perte de neutre, fragments, masse précise
HMDB	220 945 métabolites	Métabolites, drogues, polluants environnementaux	Structure chimique, spectres MS et RMN
YMDB	16042 petites molécules	Petites molécules, enzymes et leurs transporteurs	Propriétés chimiques, spectres RMN MS et MS/MS
KEGG	Environ 150 000 composés	Genes, voies,enzymes, métabolites	Voies métaboliques, génomes, représentation des systèmes biologiques
PUBCHEM	Plusieurs millions de composés	Petites molécules, nucléotides, lipides, carbohydrates, peptides, macromolécules modifiées chimiquement	Structure chimique, propriétés physico-chimiques, biochimique, pharmacologie, production, toxicologie

Ces 5 bases de données en ligne renseignent des informations différentes en fonction des molécules. Metlin est une base de données axée sur l'identification des molécules analysées en MS nottamment grâce aux fragments des spectres MS^2. Les bases de données HMDB et YMDB permettent la visualisation des molécules par leur structure chimique, spectres MS et RMN mais également leur rôle dans les systèmes biologiques. Ce rôle dans les systèmes biologiques est décrit également par la base KEGG qui permet également une visualisation des voies biologiques dans lesquelles les molécules sont identifiées.

Cette diversité de base de données ne permet pas, de nos jours, une description exhaustive des métabolites potentiellement retrouvés dans un échantillon, réduisant ainsi l'exploration du métabolome.De plus, un problème d'homogénité dans les conditions d'obtention des spectres RMN ou MS (sources et analyseurs différents) entraine un biais quant à l'identification des métabolites à partir de ces bases de données.

4.2. Normalisation des données

Trois caractéristiques du système analytique doivent être stables : le temps de rétention, l'intensité du signal et la précision de la masse pour que les analyses métaboliques soient répétables. Au cours d'une séquence d'acquisition LC-MS, l'encrassement de la source peut altérer le processus d'ionisation notamment par des diminutions de l'intensité du signal. Ainsi la variabilité biologique à étudier peut être modifiée aléatoirement à cause de la variabilité analytique. Ainsi avant de commencer tout type d'analyse statistique, la normalisation analytique des données est effectuée pour réduire les biais et les variations techniques. La correction de cette dérive analytique est souvent réalisée par l'utilisation d'échantillon de contrôle qualité (Quality control, QC). Les échantillons biologiques donnent souvent des résultats non représentatifs, principalement en raison de petites modifications du temps de rétention chromatographique et de l'intensité du signal [104-106], il est conseillé d'effectuer cinq à dix échantillons QC au début du lot d'échantillons.

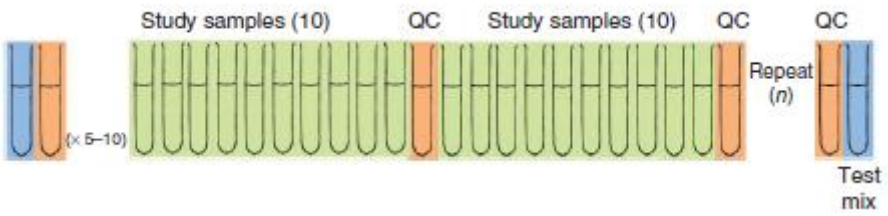


Figure 15 : Ordre suggéré pour l'ajout de QC et de mélange de standard (test mix) lors d'une étude utilisant LC-MS [104]

En règle générale, un échantillon QC est préparé en mélangeant des aliquots (par exemple, 5 à 10 µL) de chaque échantillon d'une même étude. Ce mélange fournit ainsi un échantillon avec une représentation fidèle de l'étendue des métabolites présents dans l'ensemble des échantillons [104, 106, 107] Ces échantillons de QC sont ensuite injectés à intervalles réguliers (par exemple, tous les cinq à dix échantillons [104]) tout au long de l'analyse afin de fournir des données à partir desquelles la répétabilité peut être évaluée (figure 15). Le QC, les échantillons et les mélanges de standard permettent d'obtenir des informations concernant la stabilité du système en mettant en évidence les problèmes résultant d'une contamination du système ou d'une défaillance de l'instrument (baisse de sensibilité, décalages du temps de rétention ou des changements dans la précision de la masse [104]).

Cet ajout de QC permet de mettre en évidence la dérive du signal et d'adapter la méthode de normalisation des données. L'article de G. Hughes *et al. [108]* répertorie 5 méthodes de normalisation envisageables lors d'une étude en spectrométrie de masse. Celles-ci peuvent être la normalisation à l'aire totale [109], normalisation médiane de l'intensité du métabolite [110] ou bien des algorithmes de correction du signal [111] tels que la méthode LOESS (localy estimated scatterplot smoothing). Il existe également, la normalisation par l'ajout d'un ou plusieurs standards internes (avant l'extraction) ou externes (à la fin de l'extraction). Cependant, cette méthode est souvent utilisée lors d'études métabolomiques mais est sujette au phénomène de « cross contribution » [112]. En effet, des métabolites élués en même temps que les standards peuvent influer sur leur ionisation à cause des phénomènes de suppression d'ions.

Ainsi, à ce jour, il n'y a pas de méthode de normalisation permettant d'éliminer la variabilité analytique tout en gardant la variabilité biologique. Il est recommandé d'utiliser une méthode de normalisation en fonction de la dérive connue de l'analyse.

5. Avantages et inconvénients de la spectrométrie de masse

La spectrométrie de masse est une technique sensible, permettant la mesure de concentrations en dessous de µM, spécifique, versatile (applicable à des matrices biologiques très diverses), universelle (prise en charge d'analytes aux propriétés physico-chimiques très variées) et offre une précision/justesse. La spectrométrie de masse permet la détection des métabolites ayant des concentrations inférieures au nanomolaire [113]. Cette sensibilité supérieure permet de travailler à partir d'une quantité réduite d'échantillon, de l'ordre de la dizaine de microlitres contre quelques centaines de microlitres en RMN. Le second avantage est l'utilisation de technique séparative. De ce fait la distinction des composés est meilleure par rapport à une infusion directe où le signal des molécules moins ionisables ou moins concentrées est diminué/supprimé. Cependant ce couplage entraine des temps d'analyses par échantillons plus important ce qui rend l'analyse MS plus longue qu'une analyse RMN. Une limite de la MS est la reproductibilité des analyses dû à la nécessité d'ioniser les composés. Cette étape d'ionisation est difficile à contrôler (effet matrice, sensible à la température, solvant d'élution) ce qui rend difficile la comparaison entre plateforme des résultats obtenus.

VII. Approches multiples

1. Approche multi-matrices

La grande majorité des études donnant des résultats prometteurs en terme de recherche de biomarqueurs et d'élucidation de voies métaboliques ont tendance à se concentrer sur une seule matrice biologique. Ce type d'approche peut entrainer le manque ou la non détection de molécules marqueurs cruciales de voies physiologiques [114] limitant l'information métabolique associée à un individu. En effet par des effets compensatoires, la dérégulation des voies métaboliques lors d'une pathologie peut être plus facilement détectable dans un compartiment biologique plutôt qu'un autre, ainsi l'analyse de l'un ou de l'autre peut induire un

diagnostic erroné. Lenz et al. en 2003 mettent en évidence sur les mêmes sujets sains une très faible variabilité inter-sujets et inter-jours sur des échantillons de plasmas alors que cette variabilité est plus importante sur les échantillons d'urines. Ces données montrent que chaque compartiment biologique est impacté différemment par les facteurs environnementaux. La combinaison des données de diverses matrices permet de révéler des corrélations intéressantes entre les matrices, ce qui peut aider à comprendre les mécanismes métaboliques et à qualifier les biomarqueurs révélés dans un contexte clinique. L'étude de la littérature réalisée via le site Science Direct montre qu'entre 2010 et 2020 il y a 12 503 articles en métabolomique dont 3226 multi matrices (figure 16). Parmi ces études multi matrices, 79% des articles combinent au moins 2 matrices utilisant l'urine et le sang, car ces deux matrices sont historiquement celles les plus utilisées en routine clinique. D'autres matrices comme les fèces sont étudiées en combinaison avec le sang (18%) et peu d'études (1%) utilisent la salive combinée à l'urine ou les fèces ou le sang.

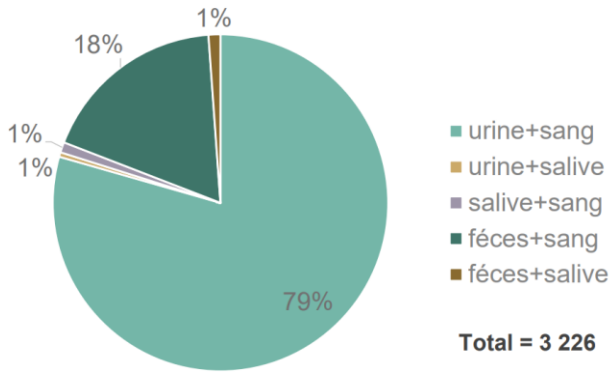


Figure 16 : Pourcentage de publications recensé entre 2010 et 2020 via Science Direct combinant au moins 2 matrices en métabolomique

L'utilisation de 3 matrices a été rapportée par Zhao *et al.* [115], réalisant ainsi un profilage métabolique des fèces, urine et sérum dans une étude métabolique ciblée sur le microbiome. L'utilité d'une approche multi matrices est discutée dans le travail de De Paepe [30] en 2018, qui montre la complémentarité des matrices urine, sang et fèces lors d'une approche non ciblée en LC-MS. Ce type d'approche est

intéressant car permet d'avoir un grand nombre de données mais reste limitant pour une application en clinique par la non identification des « features ».

2. Approche multiplateformes

L'utilisation combinée de la RMN et de la MS est exploitée depuis longtemps pour la caractérisation structurale de métabolites d'intérêts notamment dans le domaine des produits naturels [116]. Toutefois encore actuellement, la majorité des études publiées s'appuie sur une seule des deux techniques. L'une des raisons la plus courante est que les chercheurs utilisent la technologie disponible dans leurs laboratoires ou instituts [81]. La figure 17 représente le nombre de publications par année retrouvé depuis 2001. Trois groupes de mots ont permis d'établir cette histogramme récapitulatif. Le groupe de mots 1 (« metabolomics » « metabonomics » « metabolic profiling ») avec le groupe de mots 2 (« nuclear magnetic resonance » ou « nmr ») ont permis de représenter les barres oranges. Le groupe de mots 1 avec le groupe mots 3 (« mass spectrometry » ou « ms ») a permis de représenter les barres bleues et les 3 groupes de mots ensemble pour les barres grises. Celle-ci montre que la proportion des études menées avec la MS versus la RMN au cours du temps augmente. Cette augmentation est principalement due aux différences de coût et à la sensibilité des techniques. Par ailleurs, nous pouvons remarquer une augmentation du nombre des études utilisant les deux technologies, toutefois ce nombre d'études multiplateformes reste très faible par rapport à celles utilisant la RMN ou la MS.

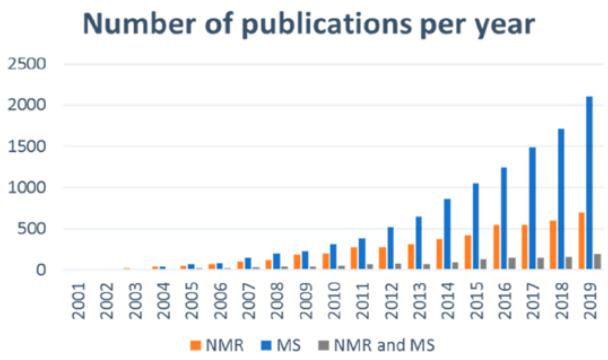


Figure 17 : Nombre de publications par année retrouvé après recherche sur pubmed.ncbi.nlm.nih.gov [81]

La combinaison des deux plateformes analytiques (MS et RMN) permet d'améliorer l'identification des molécules, la précision de l'analyse, la quantification et la couverture métabolique. La spectrométrie de masse est communément couplée avec la chromatographie liquide (LC) ou gazeuse (GC). La figure 18 représente le pourcentage d'articles sur 12737 répertoriés entre 2010 et 2020 en fonction des techniques analytiques pour la recherche des mots clés « metabolomics and plateforme ».

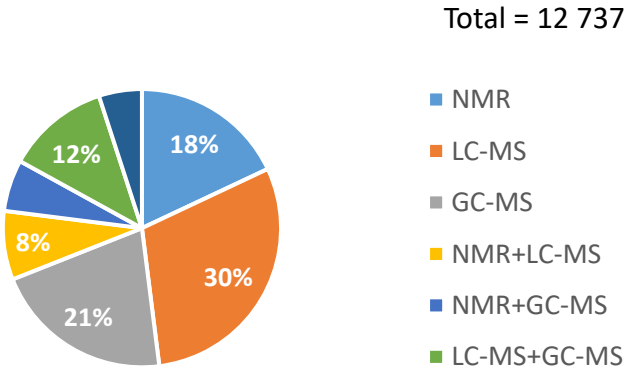


Figure 18 : Nombre d'articles scientifiques répertorié sur science direct entre 2010 et 2020 pour chaque plateforme analytique et combinant au moins 2 plateformes « metabolomics and plateforme »

Le couplage de la spectrométrie de masse avec la chromatographie liquide ou gazeuse est largement utilisé dans les études métabolomiques par rapport aux

analyses avec la RMN seule. Vient ensuite le couplage LC-MS et GC-MS puis NMR et LC-MS. On retrouve en faible proportion le couplage NMR avec la LC-MS ou la GC-MS. Cette association de plateforme est souvent limitée par l'accessibilité des techniques analytiques par l'opérateur. Tous les laboratoires d'analyses ne sont pas équipés des deux plateformes. Une autre limite à ce couplage peut être due à la préparation d'échantillon qui est différente entre la RMN et la MS. En effet, ces deux plateformes ont des préparations significativement différentes. Toutefois, des études ont prouvé l'efficacité d'une approche multi-plateformes. Par exemple Diémé et al. a montré que des analyses RMN, C18-MS et HILIC-MS d'un même échantillon urinaire ont permis de caractériser une signature métabolique du spectre autistique. [117].

Logistiquement une approche multiplateforme peut être réalisée de deux manières. L'échantillon peut être injecté séparément dans 2 appareillages ou injecté avec les appareillages en série. La combinaison de la MS et la RMN en série peut être effectuée de 3 manières différentes (voir figure 19).

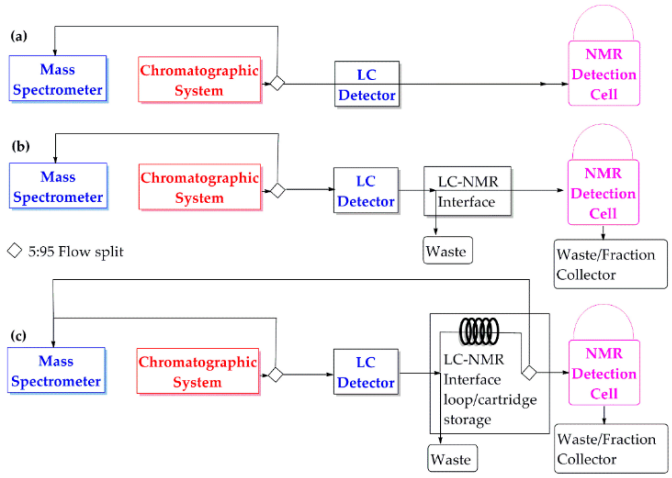


Figure 19 : Schématique configuration des différents modes de travail LC-RMN couplée en parallèle à un spectromètre de masse en (a) mode en ligne/flux continu (b) mode stop-flow (c) mode de stockage/ boucle [118]

Dans le mode en ligne (figure 19a) le détecteur LC est connecté directement à la sonde RMN. Au fur et à mesure que les composés sont élués, les spectres RMN sont

acquis en continu. Le système chromatographique est utilisé pour séparer et déplacer les composants à travers la sonde RMN pour la détection. Le flux continu est la configuration la plus simple car il n'implique aucune synchronisation entre la séparation et le système de détection. Cependant il a une faible sensibilité en raison du fait que les pics élués sont associés à un court temps d'exposition dans la cellule de détection. De plus, les déplacements chimiques des protons d'une molécule étant dépendants du solvant ; une évolution de la composition du solvant d'élution (gradient LC) peut entrainer des difficultés d'interprétation des spectres RMN.

La configuration du mode stop-flow est similaire au mode en ligne (figure 19b) avec l'ajout d'une vanne à l'interface entre le LC détecteur et la sonde RMN. Ce mode permet donc de détecter uniquement les pics sélectionnés grâce à la vanne et a une meilleure sensibilité par rapport au mode en ligne. En effet, seul les pics d'intérêt sont analysés en RMN ce qui permet d'augmenter le temps d'analyse donc le nombre de scans des spectres RMN.

Dans ce mode, la sortie du détecteur LC doit être connectée directement à un système de stockage de l'échantillon sous la forme d'une boucle ou d'une cartouche (extraction en phase solide SPE) (figure 19c). Une fois le pic détecté par le détecteur LC il est dirigé et collecté dans les boucles d'échantillonnage ou la cartouche SPE. Les pics pré-collectés ou purifiés sont ensuite dirigés vers la sonde RMN.

Ces 3 types de configuration demandent une mise en place instrumentale spécifique afin que les 2 plateformes soient connectées entre elles. Ce qui semble difficilement réalisable dû à la place requise pour les deux plateformes. De plus, cette configuration entraine une perte de sensibilité qui peut être compenser en utilisant ces deux plateformes séparément.

L'injection de l'échantillon en dérivation permet d'éviter des inconvénients logistiques et de sensibilité mais rallonge les temps d'analyses car il doit être injecté sur chaque plateforme et le pré-analytique doit être adapté à chaque plateforme.

VIII. Quantification

1. Introduction

Les informations quantitatives des métabolites constituent un atout inestimable dans la compréhension des réactions et processus biologiques se produisant dans un organisme. La métabolomique quantitative permet à la fois d'avoir des identités et des concentrations des composés, elle permet une identification explicite des biomarqueurs et une interprétation immédiate des voies ou processus métaboliques [119, 120]. Elle permet également une comparaison fiable des résultats inter-jour et entre laboratoires. Pour ces raisons, des efforts sont faits pour collecter des données quantitatives et les intégrer dans des bases de données telles que HMDB, YMDB. Les données quantitatives aident également à obtenir des modèles statistiques plus robustes, discriminants et prédictifs [121]. Les méthodes de quantification basées sur la RMN et la MS sont bien établies [121, 122].

La quantification par RMN est relativement simple et repose sur l'utilisation d'un seul standard interne comme par exemple le TSP (sodium d4-3- (triméthylsilyl) - propionate), à condition que les données soient acquises et traitées dans des conditions appropriées appelées qNMR [80]. Dans le cas des mélanges complexes, où les chevauchements de pics empêchent une quantification précise, des procédures quantitatives basées sur la RMN 2D ont été développées [123].

La quantification MS est plus chronophage car elle nécessite un développement supplémentaire pour cibler les métabolites d'intérêt et sans phénomène de suppression d'ions. Les analyses LC-ESI-MS sont souvent impactées par les effets matrice en raison de la coexistence et la co-élution de nombreux composés

présents dans les matrices biologiques complexes [124]. Il est pourtant primordial en métabolomique quantitative de pouvoir comparer les échantillons analysés de façon fiable et précise. Les approches ciblées les plus robustes reposent sur l'utilisation d'étalons internes marqués par isotopes stables et à l'utilisation de courbes d'étalonnage. Cependant, les étalons internes ne sont pas toujours disponibles et sont coûteux.

L'idée de combiner les techniques RMN et MS pour obtenir des dosages quantitatifs plus robustes et surmonter les difficultés rencontrées par les dosages basés sur MS est relativement récente [123]. Une première approche a été décrite comme une « quantification MS guidée par RMN » [125]. La quantification absolue par RMN des métabolites d'intérêt a été réalisée pour un échantillon de sérum qui a été sélectionné au hasard pour être l'échantillon de référence. Les concentrations obtenues sont ensuite fixées comme références pour les MS^2 [125]. Grâce à cette approche sans étalon interne marqué, 30 métabolites contenus dans le sérum ont été quantifiés avec succès.

Ainsi, contrairement aux approches MS, les analyses RMN permettent d'avoir une approche quantitative absolue de par la présence d'un pic de référence. Pour garantir la robustesse des analyses MS, deux types d'approches quantitatives peuvent être utilisées (figure 20). L'approche relative correspond à la comparaison des niveaux d'abondance entre deux échantillons, métabolite par métabolite. L'approche absolue permet la détermination des concentrations de métabolites dans les extraits.

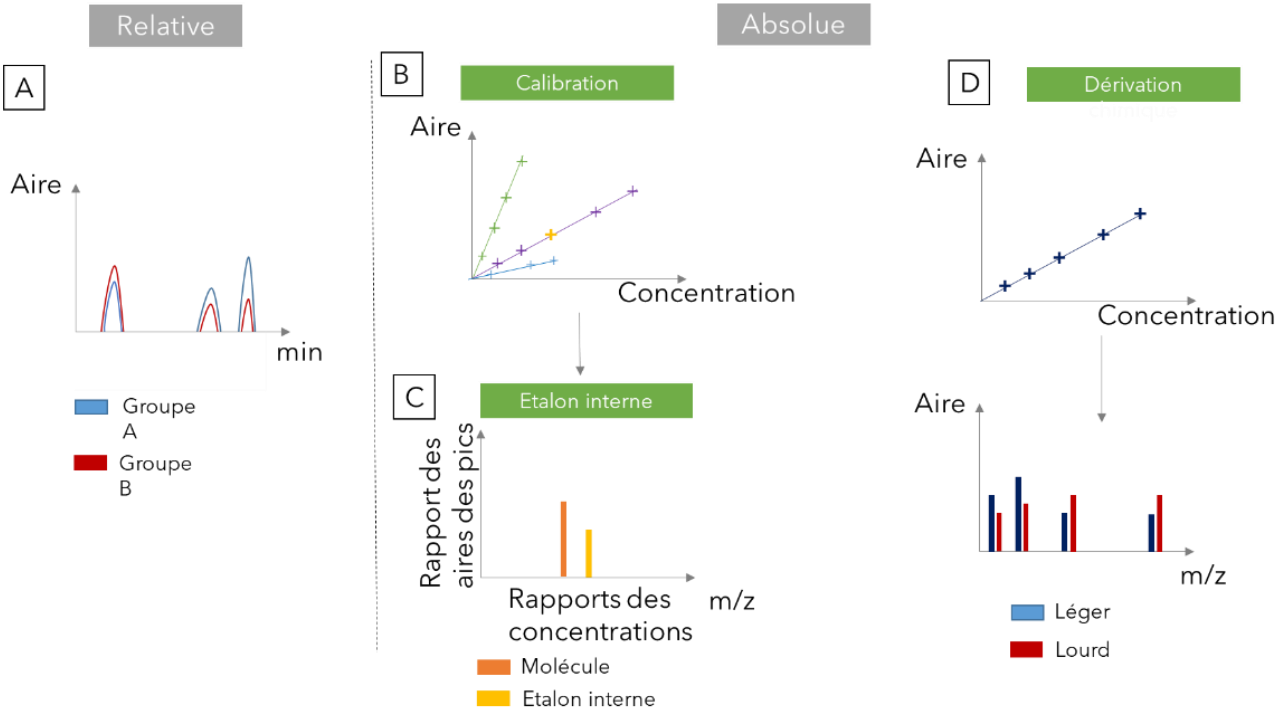


Figure 20 : Représentation des approches quantitatives relative (A) et absolue incluant la calibration (B), l'approche par étalon interne (C) et la dérivation chimique (D) utilisées de nos jours en métabolomique MS

2. La quantification relative

La quantification relative est de loin la plus largement appliquée dans les études de profilage métabolomique non ciblé. La MS et la RMN peuvent être utilisées pour ce type d'approche. Elle consiste à comparer les niveaux d'abondance d'un même métabolite d'un échantillon à un autre (voir figure 20 A). Cette comparaison est généralement faite entre un ou plusieurs groupes ayant des profils métaboliques différents par exemple un groupe contrôle et un groupe malade. L'étude non ciblée de Xi *et al. [126]* en LC-MS permet de différencier 3 métabolomes celui de vaches laitières saines, celles atteintes d'inflammation des mamelles (mammite) avec des signes cliniques et celles n'ayant pas de signes cliniques. Ce même type d'étude est réalisé pour des maladies métaboliques telles que le TDAH (Trouble Déficit de

l'Attention/Hyperactivité) [127]. La poursuite de ce type d'études permettra d'identifier de potentiels biomarqueurs permettant l'établissement de diagnostic précoce. La quantification relative est le seul moyen d'exprimer les niveaux et les changements de niveau d'analytes non annotés.

3. La quantification absolue

La quantification absolue consiste à convertir l'abondance d'un ion en concentration massique (par exemple µg/mL) ou molaire (par exemple µmol/mL) [128]. La quantification absolue permet de comparer les échantillons injectés lors d'une même séquence, mais également lors d'analyses réalisées à des temps différents ou par des laboratoires différents.

Trois types d'approches peuvent être envisagés : par calibration (figure 20 B) par utilisation d'étalons internes (figure 20 C) ou par dérivation chimique (figure 20 D).

3.1. Courbe de calibration

L'approche par courbe de calibration permet d'estimer la réponse du détecteur en fonction de la concentration du standard analysé. La courbe d'étalonnage est une représentation graphique de la relation entre l'aire du pic (axe Y) et la concentration de l'analyte (axe X) (figure 20B). Dans un premier temps, des solutions de standards doivent être préparées (5 au minimum) contenant des analytes de concentrations connues. La courbe de calibration obtenue est modélisée par une régression linéaire ou une équation polynomiale. L'analyte est ensuite injecté et son aire mesurée est placée sur la droite d'étalonnage afin de déterminer sa concentration dans l'échantillon. L'aire de l'analyte dont la concentration est à déterminer doit se trouver au milieu de la gamme définie au préalable. La gamme doit être ajustée si celle-ci se trouve en dehors de la courbe d'étalonnage.

La quantification par courbe de calibration des métabolites fait face à de nombreuses contraintes. De nombreux signaux de métabolites dans de tels mélanges sont non linéaires pour diverses raisons (l'interaction ionique, suppression d'ions …) ce qui complique considérablement la quantification. Dans une matrice biologique les concentrations de métabolites sont très larges [129]. Les courbes doivent couvrir une grande gamme dynamique de concentrations (variabilité des gammes de concentrations) entrainant parfois la saturation du détecteur MS. De plus, le nombre de métabolites détectés puis annotés dans une matrice biologique reste important, il est très complexe de quantifier un grand nombre de ces composés simultanément [128]. De plus, la quantification est particulièrement problématique dans le cas de l'étalonnage, où la quantification des standards est effectuée dans un mélange beaucoup plus simple que celui de l'extrait biologique. Couramment, une dizaine de métabolites sont quantifiés simultanément lors d'une analyse quantitative ciblée. De nombreuses méthodes analytiques ont été développées afin de répondre à ces besoins notamment avec l'utilisation d'étalons internes marqués.

3.2. Etalon interne

La quantification par étalon interne marqué est aujourd'hui la méthode de référence pour la quantification absolue des métabolites [130]. Un étalon interne marqué est un composé incorporant différents isotopes stables « lourds » et possédant les mêmes propriétés physico-chimiques que son homologue non marqué. Ces critères lui permettent d'avoir la même réponse analytique que ce soit au niveau de la chromatographie ou de l'ionisation en spectrométrie de masse. La faible abondance naturelle de certains isotopes stables tels que le carbone-13, l'azote-15 ou le deutérium, permet d'envisager leur utilisation pour le marquage de composés organiques d'intérêt. Le nombre d'isotopes ajoutés à la structure de l'étalon interne est à déterminer. Dans le cas du deutérium, il est préconisé d'incorporer à la structure de la molécule au minimum deux atomes de deutérium. Ceci permet

d'éviter toute superposition avec le massif isotopique du composé non marqué, en particulier lié à la contribution naturelle de l'isotope C13 du composé non marqué. Un ajout excessif d'isotopes dans la structure peut entrainer la modification des propriétés physico-chimiques de la molécule marquée et donc engendrer un décalage du temps de rétention [131].

Afin d'obtenir une grande précision, l'étalon marqué préalablement quantifié par une courbe d'étalonnage doit être ajouté aux échantillons d'intérêt à une concentration similaire à celle retrouvée dans l'échantillon. La concentration du métabolite d'intérêt est déterminée par le rapport des réponses analytiques de l'étalon interne ajouté et du métabolite d'intérêt multiplié par la concentration connue de l'étalon interne.

L'obtention de molécules marquées aux isotopes stables peut être réalisée par synthèse chimique, échange isotopique, par utilisation de microorganismes et par l'utilisation de kits commerciaux.

3.2.1. Par synthèse chimique

L'ensemble des approches développées pour introduire un ou des isotopes stables dans les structures chimiques doivent répondre à deux critères principaux : la rapidité et l'efficacité d'incorporation des isotopes. Différentes stratégies d'incorporation d'isotopes stables par voie chimique dans les structures de molécules à quantifier ont été développées. Une méthode rapide de marquage se résume le plus souvent à l'application de protocole mono-étape soit par dérivation chimique (à l'aide de couplage marqué) ou par échange isotopique (le plus souvent hydrogène/ deutérium).

Compte tenu du grand nombre de composés constituant le métabolome et la grande variété de leurs structures chimiques, la production de composés marqués aux isotopes stables par synthèse totale est inenvisageable car trop longue et

fastidieuse. Il est possible de retrouver dans le commerce, que SigmaAldrich, VWR ou encore agilent, une large gamme de standards internes marqués.

3.2.2. Echange isotopique

L'échange isotopique consiste à remplacer un atome par son isotope stable « lourd » dans la structure d'une molécule. Les hydrogènes sont des atomes faciles à échanger notamment lorsqu'ils sont liés à un hétéroatome tel qu'un oxygène ou azote [132] où ils sont plus labiles. L'échange isotopique peut donc être réalisé à l'aide de solvant deutéré lors de la préparation d'échantillon ou dans la phase mobile de l'analyse LC-MS. Toutefois, ce type d'approche est réversible c'est-à-dire qu'un hydrogène peut s'échanger avec un deutérium et la réaction inverse peut également se produire. Ainsi, pour être sûr de l'étalon interne injecté dans le système chromatographique, il est donc préférable d'effectuer les échanges H/D sur des carbones que sur des hétéroatomes. Ce qui entraine un échange plus stable car les hydrogènes sont moins labiles sur un carbone. Pour se faire, ce type de réaction est catalysé par un métal tel que le paladium, platine et/ou ruthérium [133, 134]. Ils permettent d'effectuer ses échanges H/D de manière plus ou moins sélectifs selon la position de l'hydrogène (aromatique ou aliphatique). Les échanges isotopiques $^{12}C/^{13}C$ ne sont pas décrits par la littérature. Ce type de réaction doit être plus compliqué à mettre en œuvre.

Ces réactions d'échange isotopique H/D sont utilisées pour la création de standards internes marqués utilisés notamment en LC-MS et peuvent être une alternative à une synthèse chimique. Toutefois, les échanges isotopiques sont parfois longs (24h) ajoutant un temps considérable lors de la phase pré analytique.

3.2.3. Par microorganisme

L'obtention de métabolites marqués, le plus souvent au ^{13}C ou ^{15}N, est réalisable par action métabolique (ou enzymatique) d'un organisme (petits mammifère,

champignons, plantes …) alimenté en nutriments marqués. Cette méthode consiste à faire croitre un organisme en présence de nutriments marqués afin que celui-ci produise l'ensemble des molécules constitutives de son métabolome dans une version marquée. Originellement, cette méthode a été développée pour l'analyse fluxomique. Elle s'intéresse à la mesure de la dynamique du métabolisme et représente toutes les interactions jusqu'à un résultat final de toutes les interactions métaboliques [135]. Principalement, ce marquage a été développé avec des plantes [136] ou des microorganismes [137]. Dans le cas des végétaux, les versions marquées de métabolites de plantes sont générées par leur mise en croissance sous atmosphère contrôlée de $^{13}CO_2$ [138]. Une application en protéomique montre qu'il est possible de nourrir des rongeurs avec de la spiruline ^{13}C et/ou^{15}N afin d'obtenir des dérivés lourds [139]. Toutefois, le marquage complet d'un organisme est limité par le coût important de ce type d'approche. Il présente l'avantage de donner accès à un nombre important de composés marqués simultanément et offre la possibilité de mettre en évidence les voies métaboliques impliquées dans la biotransformation de chaque métabolite [140]. Toutefois, les métabolomes marqués produits ne sont pas similaires à ceux retrouvés chez l'Homme. Ce biais limite en grande partie l'utilisation de ce type d'approche dans le cas d'une étude clinique.

3.2.4. Incorporation des isotopes par kit commercial

Des kits commerciaux ont été développés afin d'augmenter le nombre de métabolites quantifiables en une analyse. Le kit biocrates (AbsoluteIDQ® p180 kit) permet d'effectuer des approches ciblées avec la quantification de 600 molécules issues de 26 classes chimiques telles que les acides aminés, les phosphatidylcholines.... Ce kit a par exemple permis de quantifier et de générer une signature métaboliques caractéristique de patients infectés par le SRAS-COV2 par rapport aux patients non infectés hospitalisés dans le même département [141]. Le Metabolomics QReSS (Quantification, Retention, and System Suitability) Kits a été développé pour les approches métabolomiques ciblées et non ciblées. Cependant, l'utilisation de ces kits nécessite un équipement provenant de fournisseur précis par

exemple Sciex pour le QReSS kit. Le laboratoire doit donc s'équiper et engendrer des frais supplémentaires pour l'utilisation d'un kit.

3.3. Dérivation chimique

La dérivation chimique est utilisée depuis plusieurs décennies en métabolomique quantitative [142]. Cette technique repose sur la conversion d'une fonction chimique contenue par le métabolite d'intérêt en un groupement fonctionnel le plus souvent marqué d'isotopes stables (« lourds ») (voir figure 20 D). Cette conversion est réalisée à l'aide d'une molécule nommée étiquette chimique. Le processus de dérivation chimique est souvent appliqué avec deux versions de la même étiquette chimique, une version « légère » ou non marquée et une version « lourde » c'est-à-dire comportant des isotopes stables « lourds ». Le principe d'utilisation des étiquettes est représenté schématiquement par la figure 20 D. Les échantillons sont divisés en deux groupes, un groupe dérivé avec l'étiquette « légère » et un autre avec l'étiquette « lourde » (groupe contrôle ou mélange de standards en quantité connue). Les échantillons dérivés avec l'étiquette « lourde » sont ensuite mélangés aux échantillons « légers » et le mélange est injecté pour une analyse LC-MS. Les composés « lourds » vont servir d'étalon interne pour permettre la quantification de ceux dérivés dans l'échantillon « léger ». La partie introductive du chapitre 3 de ce projet de thèse est entièrement consacrée aux avantages et inconvénients de la dérivation chimique.

3.3.1. Conception des étiquettes chimiques

La création d'une étiquette chimique se compose de 4 parties (figure 21).



Figure 21 : Parties devant composer une étiquette chimique

Le groupe réactif est responsable de la réaction entre le réactif et le groupe fonctionnel du métabolite ciblé. La fraction codée par l'isotope est la partie qui pourra être modifiée pour contenir les atomes isotopiques stables. Par exemple un atome d'hydrogène H est remplacé par un deutérium ^{2}D ou ^{12}C en ^{13}C. L'amélioration de la détection en MS est induite par l'ajout d'un groupement facilement ionisable. Dans le cas du mode positif l'introduction d'un groupement chimique ayant une forte affinité proton (groupements amine, alkyle…) ou un groupement chargé positivement améliore l'ionisation. Pour une meilleure rétention sur les colonnes en analyse LC, l'étiquette chimique doit également contenir une partie qui aidera à sa rétention lors de son passage dans la colonne LC. Par exemple, l'ajout d'un groupement hydrophobe tel qu'un aromatique aide à la rétention d'une molécule en analyse phase inverse.

IX. Objectifs de la thèse

La métabolomique se développe en raison de son potentiel d'identification des biomarqueurs et à suivre l'évolution d'un individu, pour détecter une pathologie le plus tôt possible et pour prédire et suivre la réponse au traitement. Bien que la littérature montre clairement le potentiel de la métabolomique pour phénotyper l'état de santé d'une personne, il subsiste de nombreuses limites techniques quant à son transfert à l'usage clinique et en élevage. L'objectif de ce projet est d'établir un métabotype complet d'un individu en créant son empreinte métabolique afin de suivre son état de santé au cours de sa vie.

La description la plus exhaustive du métabolome d'un individu vient des échantillons utilisés. Ainsi, l'objectif premier est d'utiliser l'ensemble des prélèvements couramment disponible en clinique (sang, urine, salive et fèces) et en élevage caprin (sang, lait et fèces). La plupart des études en métabolomique ont

tendance à exploiter une seule matrice biologique pour mettre en évidence des biomarqueurs et proposer des voies métaboliques dérégulées. Ce type d'approche est réductrice quant à l'information métabolique associée à un individu. En effet, la combinaison des données métabolomiques provenant de plusieurs matrices biologiques peut aider à comprendre les mécanismes physiopathologiques à l'échelle d'un individu en apportant une information métabolique plus complète provenant de plusieurs compartiments biologiques. L'utilisation des deux plateformes analytiques en métabolomique permettront de répondre à la diversité moléculaire ainsi qu'à la grande amplitude de concentrations biologiques. Les analyses ciblées et non ciblées en RMN et en MS permettront d'augmenter la couverture métabolique de chaque type d'échantillon. La combinaison de ces informations pourra être visualisée à l'aide de cartes métaboliques individuelles.

L'approche non ciblée en métabolomique par spectrométrie de masse à haute résolution souffre d'un certain manque de robustesse analytique et validation méthodologique. Cela conduit souvent à une couverture du métabolome faible et peu fiable en raison de l'absence de protocoles standardisés. L'identification et la quantification de métabolites sont absolument nécessaires pour obtenir à la fois un biomarqueur fiable et des informations biologiques quantitatives facilement comparables entre les laboratoires.

Pour cela, le second objectif est de développer une méthode quantitative large en spectrométrie de masse pour quantifier des molécules de classes chimiques différentes. L'objectif de cette tâche vise à utiliser la dérivation de groupes chimiques afin de produire des mélanges complexes marqués par des isotopes stables (^{2}H, ^{13}C) utilisables comme étalons internes lors d'études métabolomiques.

Ces développements technologiques participeront au développement d'une méthodologie visant à obtenir pour un individu une carte métabolique la plus exhaustive possible pour la recherche et le suivi de biomarqueurs dans un contexte

de production animale ou de routine clinique dans le cadre de médecine de précision. Ce développement répond aux problématiques de préparations d'échantillon, de conditions de stockage, de stabilité, d'extraction, de développement de séquences d'acquisition et d'évaluation de la robustesse.

Chapitre 2 : Développement de méthodes standardisées et exploration multiplateformes et multi matrices du métabolome

I. Contexte général

Le métabolome est considéré comme le point final ultime de la cascade (génomique, transcriptomique, protéomique). Impacté par les interactions complexes entre l'hôte, son microbiote et d'autres facteurs tels que l'alimentation, le stress, le sexe et l'âge et mode de vie [143, 144], le métabolome reflète le phénotype biologique d'un individu et est très souvent ciblé dans la recherche en sciences de la vie pour capturer les changements induits par le métabolisme. La métabolomique est devenue une approche populaire qui a permis de faire des progrès considérables en santé humain [143], en sciences pharmaceutiques [145], en agriculture ou santé et production animale [146]. C'est une approche pour la découverte de biomarqueurs en comparant les profils métaboliques de différents phénotypes afin d'identifier les changements métaboliques spécifiques, conduisant à la compréhension de la physiologie, de la toxicologie et de la progression de la maladie. Ces dernières années en recherche clinique et en élevage, la métabolomique est un outil de choix pour la recherche de biomarqueurs de pathologie [117, 147]. Cependant, ce type d'études est souvent limité à l'exploration d'une seule matrice analysée à l'aide d'une plateforme analytique (Voir chapitre 1 analyses multiples). Notre objectif est de pouvoir caractériser le métabolome d'un individu à partir d'un métabotype le plus complet possible et d'obtenir une couverture la plus large du métabolome grâce à l'utilisation simultanée des plateformes analytiques. Pour se faire l'exploration du métabolome est faite à partir de matrices biologiques accessibles et communément utilisées en santé humaine et animale. La diversité et le nombre de matrices permettent de cibler de façon plus large les métabolites issus d'un individu. Pour mettre en évidence diverses classes chimiques et avoir accès à des amplitudes de concentrations très larges, il est

nécessaire d'utiliser en parallèle la spectrométrie de masse couplée à la LC (RP et HILIC) et la ^{1}H-RMN.

A. Chez l'humain

1. Contexte et objectif

Au cours de la dernière décennie, beaucoup de données ont été générées à l'aide de la métabolomique, ce qui a permis de mieux comprendre de nombreux systèmes biologiques [11, 12] et des maladies telles que le diabète [148], la maladie de Crohn, les maladies inflammatoires intestinales [149] et les troubles psychiatriques [150]. La métabolomique implique un workflow spécifique [151] à partir du travail pré-analytique (collecte d'échantillons, prétraitement des échantillons et stockage des échantillons) suivi d'un travail analytique (traitement des échantillons, acquisition, normalisation des données RMN, GC-MS, LC-MS) et se terminant par le traitement des données (prétraitement des données, analyse statistique, identifications des métabolites, quantification des métabolites) permettant d'identifier des biomarqueurs et/ou comprendre les dérèglements biologiques [152]. Chaque étape de ce travail est cruciale pour identifier des biomarqueurs fiables pouvant être transférables en clinique. De plus, la métabolomique combinée aux méthodes bio informatiques ouvrent la voie à l'analyse des données à grande échelle et donnent accès aux analyses des voies métaboliques associées.

Actuellement, la plupart des études de métabolomique clinique sont réalisées sur une seule matrice (sang ou urine, ou tout autre fluide ou tissu biologique) ou en utilisant une plateforme analytique, soit par résonance magnétique nucléaire (RMN) soit par chromatographie couplée à la spectrométrie de masse (MS). A partir de cette situation, on peut se demander comment décrire au plus près un métabolome humain avec seulement un ou deux bio fluides si l'on considère la complexité biologique et la régulation systématique entre différents compartiments

biologiques (sang et voies digestives ou urinaires). En raison de la complexité, de la taille du métabolome humain, de la diversité physico-chimiques des métabolites, il semble qu'aucune plateforme analytique unique ne puisse être utilisée pour cribler tous les métabolites dans un échantillon biologique [153]. Une autre préoccupation est la préparation des échantillons, dans la littérature, différentes procédures ont été proposées en fonction des bio fluides [154] et des techniques analytiques utilisées [26, 154]. Actuellement, il n'y a pas de consensus, pas de procédures opératoires standards (SOP) pour préparer des échantillons biologiques. Ce manque de SOP génère des écueils lors de la comparaison des études inter-laboratoires. De plus, le manque de standardisation des procédures de préparations des échantillons biologiques a entrainé un faible transfert des études métabolomiques en routine clinique. Ce faible transfert entraine un manque de validation dans la découverte de biomarqueurs sur des cohortes inter-centres.

L'objectif dans un premier temps est de standardiser l'extraction métabolique de 4 matrices (urine, sang, salive et fèces) humaines pour une exploration en ^{1}H-RMN et LC-MS (RP et HILIC).

2. Principaux résultats et conclusion

Une fois la fusion des données issues des 3 plateformes faites, le protocole d'extraction approprié pour chaque matrice est déterminé en fonction du nombre de métabolites extraits, de leur CV et du nombre de voies métaboliques décrites. Ces critères d'évaluation ont pu déterminer que le mélange Urine : MeOH (1 :8) est le plus adapté pour l'exploration la plus exhaustive possible du métabolome urinaire avec 201 métabolites détectés, 143 métabolites sériques détectés avec l'extraction Serum :MeOH (1 :8), 136 métabolites salivaires détectés avec l'extraction Salive : ACN (1 :2) et 264 métabolites fécaux détectés avec l'extraction MeOH :H_2O (1 :1). Cette standardisation de la phase pré analytique permet d'établir un phénotypage métabolique le plus large possible avec une identification niveau

2 [103] permettant la construction d'une carte métabolique à un temps t. Ainsi cette carte facilitera la visualisation d'un métabotype individuel et la recherche de biomarqueurs pour un diagnostic et un suivi longitudinal.

A.1. Publication: Optimization of sample preparation for metabolomics exploration of urine, feces, blood and saliva in humans using NMR and UHPLC-HRMS platforms

molecules

MDPI

Article

Optimization of Sample Preparation for Metabolomics Exploration of Urine, Feces, Blood and Saliva in Humans Using Combined NMR and UHPLC-HRMS Platforms

Cécile Martias [1], Nadine Baroukh [1], Sylvie Mavel [1], Hélène Blasco [1,2], Antoine Lefèvre [1], Léa Roch [3], Frédéric Montigny [1], Julie Gatien [3], Laurent Schibler [3], Diane Dufour-Rainfray [1,2], Lydie Nadal-Desbarats [1,*] and Patrick Emond [1,2]

1 UMR 1253, iBrain, University of Tours, Inserm, 37044 Tours, France; cecile.martias@univ-tours.fr (C.M.); nadine.baroukh@univ-tours.fr (N.B.); sylvie.mavel@univ-tours.fr (S.M.); helene.blasco@univ-tours.fr (H.B.); antoine.lefevre@univ-tours.fr (A.L.); frederic.montigny@univ-tours.fr (F.M.); diane.dufour@univ-tours.fr (D.D.-R.); patrick.emond@univ-tours.fr (P.E.)
2 CHRU Tours, Medical Biology Center, 37000 Tours, France
3 ALLICE, Phenotyping Station, 37380 Nouzilly, France; lroch8@gmail.com (L.R.); julie.gatien@allice.fr (J.G.); laurent.schibler@allice.fr (L.S.)
* Correspondence: lydie.nadal@univ-tours.fr; Tel.: +33-(0)-2-4736-6164

Abstract: Currently, most clinical studies in metabolomics only consider a single type of sample such as urine, plasma, or feces and use a single analytical platform, either NMR or MS. Although some studies have already investigated metabolomics data from multiple fluids, the information is limited to a unique analytical platform. On the other hand, clinical studies investigating the human metabolome that combine multi-analytical platforms have focused on a single biofluid. Combining data from multiple sample types for one patient using a multimodal analytical approach (NMR and MS) should extend the metabolome coverage. Pre-analytical and analytical phases are time consuming. These steps need to be improved in order to move into clinical studies that deal with a large number of patient samples. Our study describes a standard operating procedure for biological specimens (urine, blood, saliva, and feces) using multiple platforms (^{1}H-NMR, RP-UHPLC-MS, and HILIC-UHPLC-MS). Each sample type follows a unique sample preparation procedure for analysis on a multi-platform basis. Our method was evaluated for its robustness and was able to generate a representative metabolic map.

Keywords: biological fluids; extraction protocol; UHPLC-HRMS; ^{1}H-NMR; metabolomic card

check for updates

Citation: Martias, C.; Baroukh, N.; Mavel, S.; Blasco, H.; Lefèvre, A.; Roch, L.; Montigny, F.; Gatien, J.; Schibler, L.; Dufour-Rainfray, D.; et al. Optimization of Sample Preparation for Metabolomics Exploration of Urine, Feces, Blood and Saliva in Humans Using Combined NMR and UHPLC-HRMS Platforms. *Molecules* **2021**, *26*, 4111. https://doi.org/10.3390/molecules26144111

Academic Editor: Giorgio Marrubini

Received: 15 June 2021
Accepted: 28 June 2021
Published: 6 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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/).

1. Introduction

The metabolome is considered to represent the ultimate endpoint of the biological cascade (genomic, transcriptomic, proteomic). Impacted by the complex interactions between the host and its microbiota and other factors such as diet, stress, gender, age, and lifestyle [1,2], the metabolome reflects the biological phenotype of an individual and is very often targeted in life science research to capture metabolic-induced changes in physiopathological conditions to highlight dysregulated biological pathways. Metabolomics focuses on the wide variety of low molecular weight metabolites (<1500 Da) present in biological samples (cells, tissues, or biological fluids). Metabolomics has become a popular approach for metabolite biomarker discovery by comparing the metabolic profiles from different phenotypes in order to identify specific metabolic changes, leading to the understanding of physiology, toxicology, and disease progression. In the last decade, a lot of data have been generated using metabolomics, resulting in a better understanding of many biological systems [3,4] and diseases such as diabetes [5–7], Crohn's disease and inflammatory bowel disease [8–10], cancer [11–13], cardiovascular disease [4], atherosclerosis [14], obesity [15–17], and psychiatric disorders [18].

Metabolomics implies a specific workflow [19] starting from the pre-analytical work (sample collection, sample pre-processing as metabolic quenching, centrifugation, filtration, and sample storage) followed by analytical work (sample processing, pre-acquisition normalization, data acquisition (NMR, GC-MS, LC-MS)) and ending with data processing (data pre-processing, statistical analysis, metabolite identification, metabolite quantification) to identify biomarkers and/or to understand biological dysregulations [20]. Each step of this workflow is crucial for identifying reliable biomarkers to be transferable in clinical studies. In addition, metabolomics combined with bioinformatics methods open a path for large-scale data analysis and give access to the analyses of associated metabolic pathways.

Currently, most of the clinical metabolomics studies are completed on a single biological matrix (either blood or urine, or any other biological fluids or tissues) or using one analytical platform, either nuclear magnetic resonance (NMR) or chromatography–mass spectrometry (MS), which is one of the most widely applied technologies. In the literature, blood (serum or plasma) is described as one of the most frequently used biological materials in clinical metabolomics research. Blood (serum and/or plasma) moves many metabolites through the body and may provide a lot of metabolite information about physiological and pathophysiological conditions in a particular biological system [21]. Urine gives another alternative to blood, and it is necessary for metabolic interpretation. Urine contains mainly metabolic breakdown products from foods, drinks, drugs, environmental contaminants, endogenous waste metabolites, and bacterial by-products [21], which makes it ideal for early detection of a wide range of disease [22,23]. Urine may be used as a reflection of the endogenous and exogenous metabolic profiles at a precise moment.

The use of fecal samples in metabolomics is increasing. This growth is driven by the hypothesis that the fecal metabolome may capture the complex interactions between the gut microbiome, the host, and the diet [24]. The role of gut microbiota has been correlated with multiple gastrointestinal diseases, obesity [25], rheumatoid arthritis [26], psychiatric diseases such as autism [25] or depression [25], cardiovascular diseases, and type 2 diabetes [6]. The vast majority of clinical studies tend to focus on a single biological matrix, even if combining data from various matrices should reveal interesting correlations across compartments. From this situation, one could wonder how to describe as closely as possible a human metabolome with only one or two biofluids if we consider the biological complexity and systemic regulation between different biological compartments (blood and digestive or urinary tracts).

Owing to the complexity and size of the human metabolome and the diverse physical and chemical properties of metabolites, it appears that no single analytical platform can be used to screen all metabolites in a biological sample [27].

Another concern is sample preparation, in the literature, different procedures were proposed depending on the biofluids [28–31] and on the analytical techniques used [28,32–34]. Depending on the kind of molecules sought, foreseeable consequences would be based on the laboratory results. Currently, there is no consensus, no standard operating procedure (SOP) to prepare biological samples. This lack of SOP generates pitfalls when comparing inter-lab studies. Actually, metabolomics is very much research laboratory-based and needs to move out of academic laboratories into clinical routine.

Although, some studies have already investigated metabolomics data from multiple fluids [24], the information has stayed limited to a unique analytical platform. In contrast, clinical studies investigating the human metabolome that combine multi-analytical platforms have focused on a single biofluid [35]. Combining data from various matrices captured with NMR and UHPLC-MS platforms would permit for the enhancement of the metabolic compound coverage. This proof of concept would help to reveal potential new biomarkers in clinical contexts with correlations across matrices.

Currently, there are no study combining multiple matrices and multiple platforms. Thus, the present study describes a standard operating procedure for four biological specimens (urine, blood saliva, and feces) using multiple platforms (NMR, UHPLC(RP)-MS, and UHPLC(HILIC)-MS). Here, we aimed to find a unique sample preparation procedure

for each biological sample type to be use on multiple platforms and achieve the widest metabolite coverage to trace a large number of metabolic pathways. Combining multiple analytical methods with several numbers of biofluids should provide a better characterization of a metabolic human global profile and should enable us to illustrate a patient's metabolic card.

2. Results and Discussion

The global analytical workflow applied for each biological sample to select the optimal preparation protocol is shown in Figure 1.

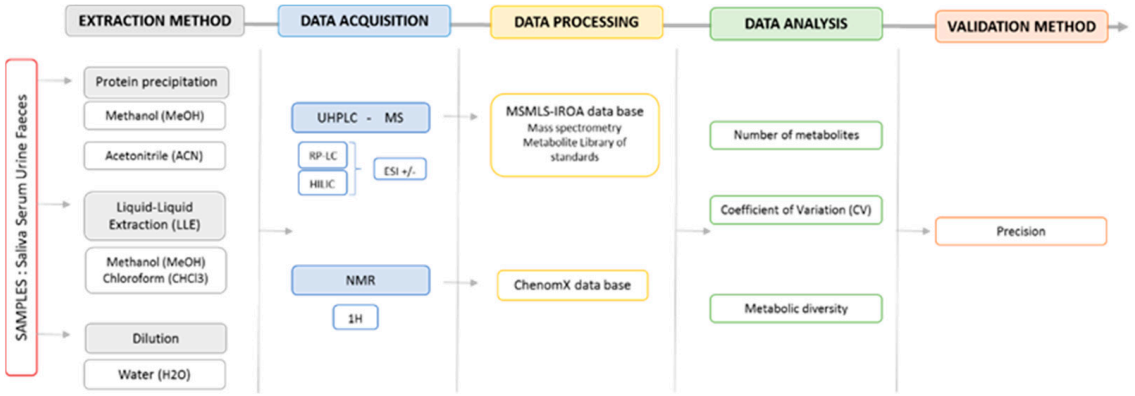


Figure 1. Schematic presentation of the global workflow of preparation procedures applied to urine, saliva, blood and feces pools for the selection of the optimum protocol for the maximum metabolome coverage by LC-MS and ^{1}H-NMR.

2.1. Optimization of Sample Preparations

In this study, we aimed to obtain a single preparation procedure that can be used to explore the metabolome by NMR, ESI+ and ESI- RP-LC-MS, and ESI+ and ESI- HILIC-LC-MS platforms, given the better metabolic coverage in term of number of metabolites detectable and the highest reproducibility (CV%) of polar metabolites.

2.1.1. Urine

We focused on comparing the efficiency of different preparation protocols (see Supplementary Data Figure S1) such as monophasic extraction with two different solvents (MeOH or ACN) or dilution preparation (1:2 for NMR and 1:10 for LC-MS) as has already been described in the literature [28,36]. Each preparation protocol was able to provide a large number of metabolites, as shown in Figure 2a. Untargeted data are provided in the Supplementary Data for guidance. Indeed, we focused on biological information made up of compounds identified at the level 1 of confidence, which is a minimum prerequisite to be used for a biochemical interpretation in a clinical routine. However, these untargeted data did not allow us to assess the redundancy of the metabolic information between analytical platforms (see Supplementary Data Table S1). According to the data-fusion step, the chemical redundancy was removed, and considering all platforms, we detected over 176 compounds with the preparation of urine/ACN (1:8), among which 145 metabolites had a CV < 30%. We found over 215 compounds with the preparation urine/MeOH (1:8), from which 201 metabolites had a CV < 30%. With the dilution protocol of over 205 compounds, 197 metabolites showed a CV < 30% (Table S2). The comparison of the two monophasic preparations of MeOH versus ACN, indicated that MeOH was the more appropriate one to use. For MeOH monophasic preparation and the dilution preparation, the results were similar. According to the CV < 10%, the dilution had a better reproducibility, showing 73% of CV < 10%, while urine/MeOH (1:8) had 62.2% of CV < 10%. Regarding the number of compounds with CV < 30%, 131 were accessible with the three preparations and, interestingly, some of them were found to be unique to a specific preparation. We detected 22 unique compounds associated with the urine/MeOH (1:8) preparation, 19

with the urine/H_2O preparation, while only 5 unique compounds were assessed with the urine/ACN (1:8) preparation (Figure 2a). This result indicated that urine/ACN (1:8) provided minor compound diversity.

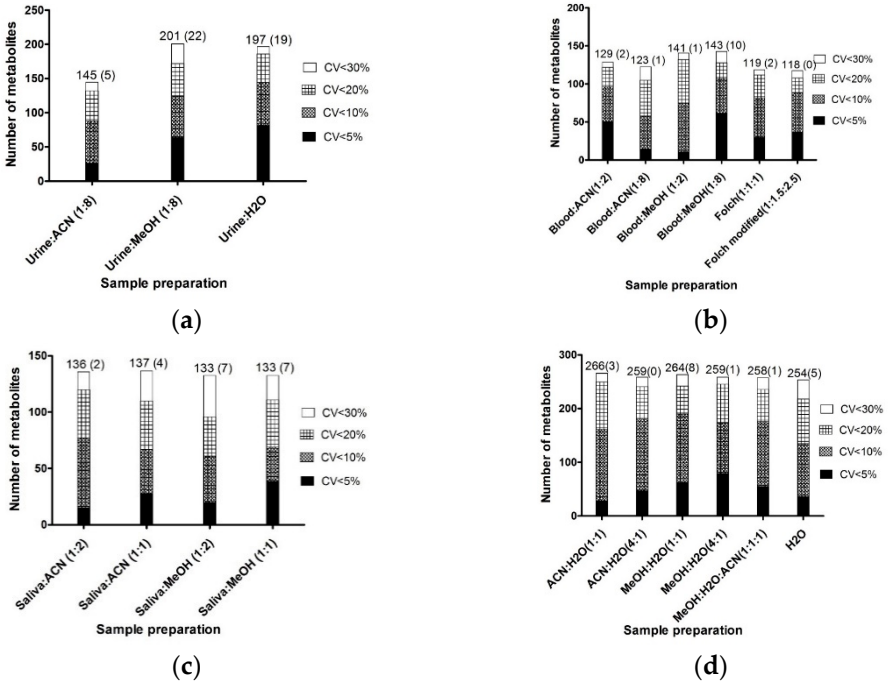


Figure 2. Histogram showing the total number of metabolites extracted for each preparation and their respective reproducibility evaluated by CV (coefficient of variation) in urine (**a**), blood (**b**), saliva (**c**), and feces (**d**). The numbers of metabolites found to be unique from each preparation condition are in brackets.

In order to select the most relevant preparation procedure for urine sample optimization, we looked at the metabolomics data processing level using the Network Explorer module in MetaboAnalyst 4.0 and found additional important biological information. Metabolic pathways were then bioinformatically predicted for each set of metabolites (see Supplementary Data Table S3). Results indicated that the urine/MeOH (1:8) preparation had the most exhaustive metabolic coverage with 12 metabolic pathways described.

In term of the total number of metabolites extracted, reproducibility, and metabolic pathway description, the urine/MeOH (1:8) preparation with 201 chemicals extracted (CV < 30%) including 22 unique compounds had the largest coverage of the metabolite sets.

2.1.2. Blood

A total of six preparation protocols were tested with blood samples (see Supplementary Data Figure S2): four monophasic extractions using two different solvents, ACN or MeOH within 1:2 and 1:8 proportions, and two biphasic extractions, blood/MeOH/$CHCl_3$ at (1:1:1) and (1:1.5:2.5) ratios. All extraction protocols succeeded in providing a wide number of metabolites, as shown in Figure 2b. As mentioned in the urine section, the untargeted data are shown in the Supplementary Data (see Supplementary Data Table S4). Once the chemical redundancy was eliminated, the monophasic extraction protocols detected over 139 compounds on all platforms, with 129 metabolites showing a CV < 30% in the blood/ACN (1:2) preparation. In the blood/ACN (1:8) preparation, over 135 compounds were detected on all platforms, 123 metabolites had a CV < 30%. In the blood/MeOH (1:2) preparation, over 147 compounds were detected on all platforms, 141 metabolites had a CV < 30%. With the blood/MeOH (1:8) preparation, over 157 compounds were found on all platforms, 143 metabolites had a CV < 30%. The biphasic protocols extracted over 127 compounds detectable on all platforms, 119 metabolites had a CV < 30% with

the (blood/MeOH/$CHCl_3$) (1:1:1) platform, while over 126 compounds were detectable on all platforms and 118 metabolites had a CV < 30% with the (blood/MeOH/$CHCl_3$) (1:1.5:2.5) (see Supplementary Data Table S5). Comparing the six preparations, results were similar in term of numbers of metabolites. According to the CV < 30% (Figure 2b), methanol precipitations (1:2) or (1:8) had the best metabolite recoveries with, respectively, 141 and 143 metabolites, while acetonitrile precipitations (1:2) and (1:8) had, respectively, 129 and 123 metabolites. The biphasic extraction (blood/MeOH/$CHCl_3$) (1:1:1) had 119 metabolites with a CV < 30% while (blood/MeOH/$CHCl_3$) (1:1.5:2.5) had 118 metabolites with a CV < 30%. In terms of reproducibility, 53.2% of CV were <10% with blood/MeOH (1:2), while 75.3% of CV were <10% with blood/MeOH (1:8). Regarding the number of unique compounds for each preparation (Figure 2b), only one compound was unique for blood/ACN (1:8) and blood/MeOH (1:2), and two were unique for blood/ACN (1:2). The blood/MeOH (1:8) preparation had 10 unique compounds, meaning that this preparation provided the most information in terms of compound diversity. The biphasic extractions provided none and two unique compounds, respectively, for the (1:1:1) and (1:1.5:2.5) extraction conditions.

Similar to the process for urine, Network Explorer was used to estimate metabolic pathways that might be involved (see Supplementary Data Table S6). Each preparation covered nine metabolite sets (p-value < 0.05) with six or seven pathways with a p-value < 0.01 depending on the preparation. As we can see, few differences was found.

In terms of the number of metabolites extracted, reproducibility, and metabolic pathway description, the blood/MeOH (1:8) preparation with 143 chemicals extracted (CV < 30%) including 10 unique compounds had the best metabolite coverage. The blood/MeOH (1:8) preparation was the most suitable preparation.

2.1.3. Saliva

Saliva preparations were tested with four preparation protocols (see Supplementary Data Figure S3). All preparations were able to provide wide metabolite coverage, as shown in Figure 2c. As mentioned in the urine section, the untargeted data are shown in the supplementary section (see Supplementary Data Table S7). Once the chemical redundancy was eliminated, the preparations saliva/ACN (1:1) and saliva/ACN (1:2) detected a very similar number of metabolites; over 148 compounds or 147 were detectable on all platforms, with 137 metabolites and 136 having a CV < 30% with the preparations saliva/ACN (1:1) and saliva/ACN (1:2), respectively (see Supplementary Data Table S8). With the preparations saliva/MeOH (1:1) and saliva/MeOH (1:2), 145 and 146 compounds were detectable on all platforms, respectively, with 133 metabolites having a CV < 30% with the two methanol preparations. Comparing the four preparation precipitations in term of numbers of metabolites, results were similar. In term of reproducibility, 48.9% of CV were < 10% with saliva/ACN (1:1) while 56.6% of CV were <10% with saliva/ACN (1:2). Regarding the number of unique compounds for each preparation (Figure 2c), four unique compounds were found with the saliva/ACN (1:1) preparation and two unique compounds were found with the saliva/ACN (1:2), saliva/MeOH (1:1), and saliva/MeOH (1:2) preparations.

Once again, Network Explorer was used to predict metabolic pathways (see Supplementary Data Table S9). Each preparation studied allowed us to cover eight metabolite sets (p-value < 0.01) except the saliva/MeOH (1:2) preparation, which showed nine metabolite sets, as shown in Table S9.

In terms of the number of metabolites extracted, reproducibility, and metabolic pathway description, the saliva/ACN (1:2) preparation with 136 chemicals extracted (CV < 30%) including two unique compounds was the most suitable preparation method for saliva samples.

2.1.4. Feces

Six extraction protocols (see Supplementary Data Figure S4) were tested on freeze-dried feces, (ACN/H_2O, MeOH/H_2O) with two ratios (1:1 and 4:1), and (MeOH/ACN/H_2O) (1:1:1) and H_2O (Figure 2d).

The six extraction protocols were able to provide wide metabolite coverage, as shown in Figure 2d. As mentioned in the urine section, the untargeted data are shown in the Supplementary Data (see Supplementary Data Table S10). Once the chemical redundancy was eliminated, the MeOH/H_2O (4:1 and 1:1) preparations had a total of 270 and 278 compounds with 259 and 264 molecules with a CV < 30%, respectively. The ACN/H_2O (4:1 and 1:1) preparations had a total of 272 and 281 compounds with 259 and 266 molecules with a CV < 30% respectively. The MeOH/ACN/H_2O and H_2O preparations had a total of 275 and 269 compounds with 258 and 254 molecules with a CV < 30%, respectively. In terms of the number of compounds (Table S11), MeOH/H_2O and ACN/H_2O ratio (1:1) preparations seemed to be the most effective ones. Comparing these two preparations in term of numbers of metabolites with CV < 30%, results were quite similar. In term of reproducibility ACN/H_2O (1:1) had 60.5% of CV < 10% and MeOH/H_2O (1:1) had 72% of metabolites with CV < 10%. Regarding the number of unique compounds for each preparation (Figure 2d), eight compounds were found only with MeOH/H_2O (1:1), while only three were unique with ACN/H_2O (1:1).

Network Explorer was used again (see Supplementary Data Table S12). Each preparation covered between 10 and 12 metabolite sets (p-value < 0.01), as shown in Table S12.

In term of the number of metabolites extracted, reproducibility, and metabolic pathway description, the MeOH/H_2O (1:1) preparation with 264 chemicals extracted (CV < 30%) including 8 unique compounds was the most suitable method preparation for urine samples.

2.2. Overall Synthesis on Optimization of Sample Preparation

The present study focused on urine, blood, saliva, and feces preparations, which is a critical point in global metabolic profiling, to be analyzed by combined NMR and UHPLC-HRMS platforms in order to describe an individual from their different compartments.

According to the literature, urine is usually diluted in water [36] or with buffered solution [37] depending on the analytical tool used. NMR needs a pH adjustment before acquisition to avoid chemical-shift bias in the automated analysis. Using NMR, it is difficult to quantify metabolites if some macromolecules such as proteins are still present in the sample because of the baseline distortion. UHPLC-MS is more sensitive than is NMR and metabolites are first separated by liquid chromatography before the analysis by mass spectrometry. Comparing solvent precipitation (urine/methanol (1:8)) to the dilution preparation (quick and simple), we obtained as many metabolites as in the diluted preparation with good reproducibility, and the NMR spectra did not need realignment and the baselines were not distorted, allowing for quantification if needed. In our study, combining NMR with UHPLC-MS, urine/methanol (1:8) precipitation gives the most rapid, reproducible metabolomics information and allowed for good metabolic pathway coverage.

Several papers showed that blood is usually prepared with a protein precipitation using acetonitrile or methanol [24,31,36] or a methanol/chloroform precipitation [38] for macromolecule removal. SPE, hybrid-SPE, or micro-extraction [31] can be used, but those methods are time-consuming and costly, when considering large cohort exploration. In our study, using NMR with UHPLC-MS, blood/methanol (1:8) precipitation gives the most rapid, reproducible metabolomics information and allowed for good metabolic pathway coverage.

In the literature, saliva usually is centrifuged and either diluted with a buffered solution for NMR exploration [39] or used with a methanol protein precipitation in LC-MS. Recently, in a targeted metabolomics study using UPLC-MS/MS on the saliva of children, Schultz et al. [40] removed mucins and bacterial debris by centrifugation and proteins by acetonitrile-mediated precipitation. In our study, combining NMR with UHPLC-MS, the preparation saliva/acetonitrile (1:2) was the most suitable sample preparation.

In the literature, feces is described as a complex matrix with very diversified sample preparations [41]. For NMR exploration, the most commonly used extracting solvent is phosphate buffered saline or distilled water [42,43]. For LC-MS analysis, fecal water [34] or lyophilized feces are dissolved in ultrapure water before solvent precipitation [36,44]. Moosmang et al. [45] were the first to establish a protocol for the extraction of human fecal samples for NMR and LC-MS techniques. In their study, they concluded that water was the solvent of choice for fecal sample extraction. In our study with lyophilized feces, methanol/water (1:1) as the extraction solvent yields better results in terms of the number of detected metabolites as well as reproducibility.

2.3. Intraday Precision

According to the optimization of sample preparations, the intraday precision was evaluated. Repeated samples (n = 16) were prepared with the chosen extraction preparation and the variation coefficients of metabolites were calculated (Table 1).

Table 1. Numbers of metabolites recovered in repeated samples compared to metabolites found in the chosen extraction protocol. Variation coefficients (CVmean) and (CVmin–Cvmax) were calculated from the fusion of the three platforms.

	Urine	**Blood**	**Saliva**	**Feces**
Sample preparation chosen	Urine/MeOH (1:8)	Blood/MeOH (1:8)	Saliva/ACN (1:2)	MeOH/H_2O (1:1)
Number of metabolites	151	104	90	208
CV mean % (min–max)	7% (2–53%)	8% (3–53%)	7% (2–26%)	11% (3–40%)

We aimed to validate the selected preparation for each biological compartment by an intraday precision. For the optimization of the sample preparation method, a CV < 30% was considered acceptable regarding the analytical variability. In our validation step, only three metabolites had a CV higher than 30% in blood and urine and five for the feces matrix. All the saliva metabolites had a CV under 30%.

Since all repeated samples used during the validation step should have the same composition, the variability of their signal was only due to the analytical source. Therefore, metabolites should be present in all 16 repeated measures to be kept. The difference was due to the data processing, which was drastic and showed more consequences when the number of repetitions was larger. Indeed, the probability of deleting a metabolite that has a zero-fill among the 16 repeated measures was greater than that in the extraction samples (n = 5). Analytically, metabolites with CV below 30% were the most robust and should be considered to metabolically map an individual.

2.4. Platform Complementarity

The complementarity of the analytically different platforms was assessed through a Venn diagram (see Supplementary Data Figure S5) and using selected preparations for each matrix. Dealing with the four matrices, NMR allowed us to detect between 5% and 13% of the metabolic information, HILIC and RP provided 29–52% and 20%–38%, respectively. Finally, for each biological sample, 75.6% (67%–82%) of the metabolic information was platform specific. UHPLC-MS provided more information than did NMR because this analytical technique is more sensitive than NMR but enables a quantitative aspect. The multiplatform approach is time-consuming but increases the numbers of metabolites, accuracy, and reliably. Metabolites could have a different CV according to the analytical tool used, so the best CV should be kept. Human cohorts are considerable in size, consequently, using a unique sample preparation for NMR and UHPLC-MS analysis would save time.

2.5. Matrix Complementarity, Metabolic Coverage, and Mapping

The complementarity of the four matrices was assessed through an upset diagram (Figure 3), which provides an efficient way to visualize intersections of multiple sets compared to the traditional approach of a Venn diagram. Among the three visualization modes of intersections [46], we decided to represent the upset plot with the degree mode. In Figure 3, the horizontal bar chart shows the entire metabolite size of each sample set (90 in saliva, 104 in blood, 151 in urine, and 208 in feces). The fecal metabolome provides the highest specificity, due to the impact of the microbiota metabolism. Degree one contains only exclusive intersections (metabolites) with exactly one participating set (sample type) without overlapping. The vertical bar chart shows the number of unique metabolite sets (66 metabolites in feces, 25 in urine, 8 in saliva, and 6 in blood). Table S13 in Supplementary Data shows the detailed list of metabolites detected in each intersection of the upset. The unique compounds reflect the complementarity of the metabolic coverage provided by each matrix. Degree two contains metabolites shared exclusively by two sets (sample type). With 31 metabolites shared, urine and feces gave the higher score of intersection. This finding is surprising given the expected correlation between urine and blood compared to urine and feces. Metabolites in blood have a short half-life because of the rapid glomerular filtration. Yet, metabolites accumulated in urine, leading to their increased concentration.

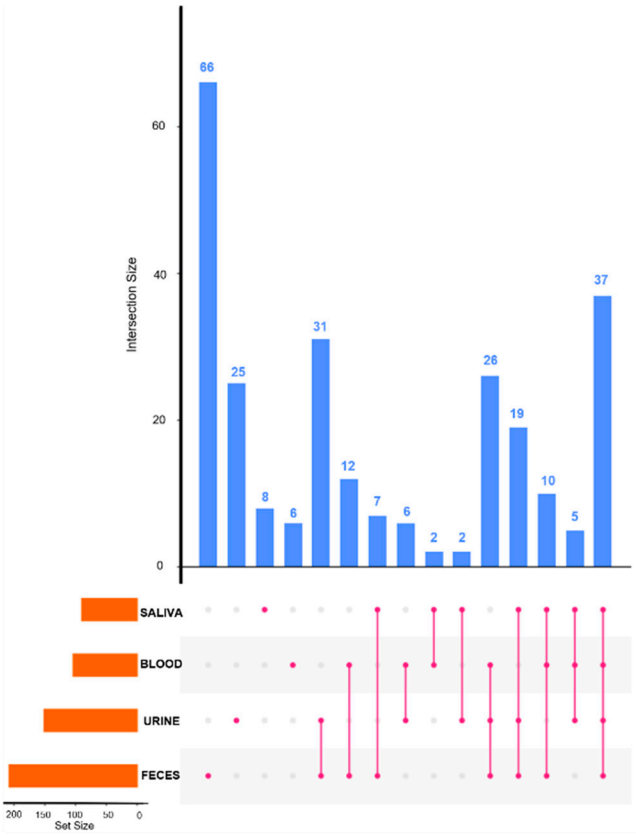


Figure 3. Upset diagram to show the complementarity in term of metabolite coverage across the different biological matrices (Set Size = number of total compounds per matrix, Intersection Size = number of compounds detected in different matrices).

Degree three provides metabolites shared exclusively by three sets (sample type). By considering three matrices, the intersection with urine, feces, and blood gives the maximum metabolites shared (26 metabolites), which seems logical because saliva is secreted by salivary glands, which are not directly link to the other compartments.

Degree four provides metabolites shared exclusively by the four sets (sample type). In all the four matrices, 37 identical metabolites were detected. These 37 common com-

pounds reflect the overlapping matrices and can be very informative for metabolic and physiopathological flux analysis at the organism level and to evaluate the biological compartment exchanges. These observations need to be confirmed in future studies by applying this methodology to a unique individual (we worked here with pooled samples).

While matrices complementarity and metabolic coverage was assessed earlier, metabolic mapping needs to be ascertained. iPath3.0 provides data mapping capabilities. Based on the metabolites recovered from urine, blood, saliva, and feces, we established a global metabolic card to metabolically illustrate an individual (Figure 4). This card was obtained by relocating metabolites from the intraday precision step (see Table 1). The gradient blue dots represent metabolites found in 1, 2, 3, or 4 matrices (blood, urine, saliva, feces). Considering lists of unique metabolites in each matrix described in Figure 3, the pathways described by a single matrix are for energy metabolism and metabolism of other amino acids. The metabolites common to the four matrices mainly identified nucleotide metabolism, metabolism of cofactors and vitamins, and amino acid metabolism.

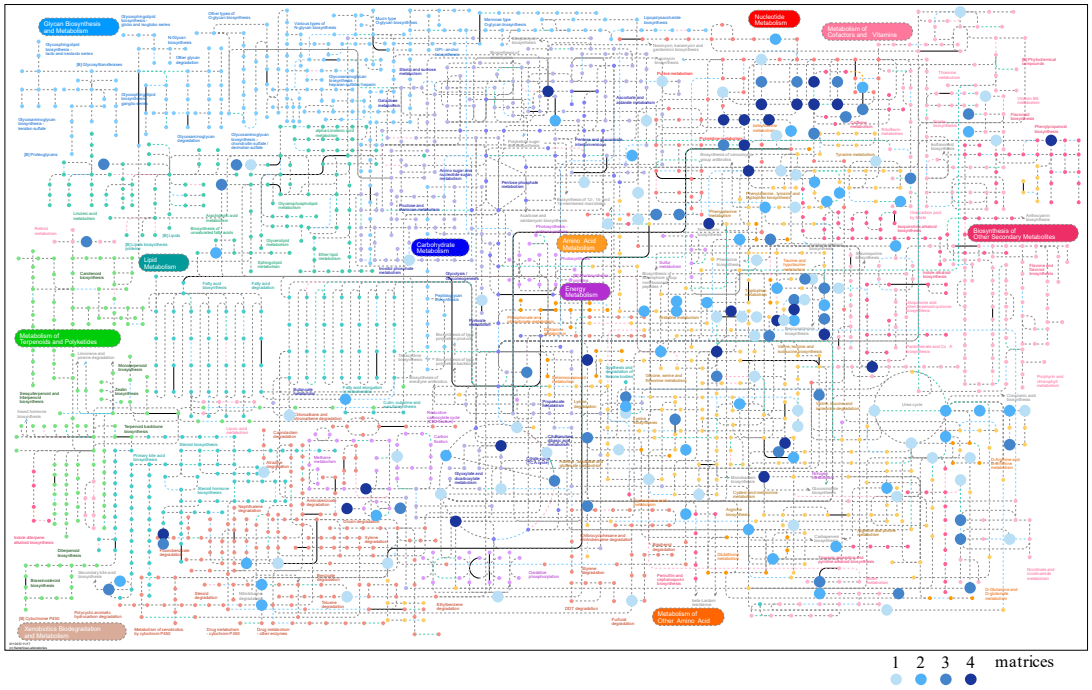


Figure 4. Metabolomic card showing metabolites from urine, blood, saliva, and feces through the metabolomics pathway to illustrate an individual. Metabolites expressed in one, two, three, or four matrices are shown in the blue gradient color.

While lipid metabolism and carbohydrate metabolism were not well described, this discrepancy can be explained by the sample preparation extracting polar metabolites only.

In summary, this study presents the first metabolomics fingerprinting strategy for multimatrix (urine, blood serum, saliva, and feces) analysis using a multiplatform approach ((NMR(^{1}H), RP-UHPLC-MS, and HILIC-UHPLC-MS). We validated a unique preparation for each matrix exploitable with these analytical tools. This targeted multiplatform approach allowed us to detect 201, 143, 136, and 264 metabolites in urine, blood, saliva, and feces, respectively. Currently, in the literature, metabolomics studies of these biological fluids are carried out in untargeted ways [31,37,44]. In 2018, De Paepe et al. [36] proposed a unique multimatrix metabolomics fingerprinting strategy for feces, plasma, and urine using a single analytical platform (RP-UHPLC-MS). They found 9672, 9647, and 6122 features in feces, urine, and plasma, respectively, proving this approach was perfectly fitted to the concept of systems biology.

Other studies have been carried out using a highly targeted approach on dozens of molecules [24,38,39,42]. However, the major limitation of this two-metabolic phenotyping

is the lack of metabolite identification, hindering its use as in disease-related biomarker detection or for potential pathway elucidation. Yet, in order to elucidate intricate metabolic pathways, the identification of the features is needed.

Zhao et al. [24] in 2017 carried out a targeted multimatrix approach in GC-MS and were able to identify 92, 103, and 118 metabolites in serum, urine, and feces, respectively. Nevertheless, analysis of these fluids by other analytical techniques can help to widen human metabolome coverage and better understand the biological system, to provide statistically robust models and help in the discovery of biomarkers [47].

We proposed a multimatrix and multiplatform approach to achieve a high metabolome coverage with structural identification of metabolites (level 2) [48] with a high reproducibility. This identification allowed us to generate a metabolic map, collecting information from different biological compartments. This metabolic phenotype allows for a better characterization of the metabolic status of an individual and opens avenues to better understand the exchanges between different compartments.

3. Materials and Methods

3.1. Sample Preparations

All chemicals were bought from Sigma Aldrich (Saint-Quentin Fallavier, France) unless otherwise specified.

Saliva, feces, serum, and urine samples were obtained from healthy volunteers (with no dietary restrictions or antibiotic treatment prior to sample donation). Volunteers were informed and agreed with the use of their samples for this study.

3.1.1. Urine

Morning-fasted urine were pooled and centrifuged (10 min at 4 °C and 15,000× *g*). Supernatants were collected in 20 aliquots of 200 µL (n = 5 per sample preparation). Three extraction protocols were tested either with solvent extraction (Urine/MeOH(1:8)), (Urine/ACN(1:8)) or urine dilution (Urine/D_2O buffer (1:2) for NMR or Urine/deionized H_2O (1:10) for LC-MS). For extraction protocols, 1600 µL of solvent were added to urine, vortexed for 10 min, and centrifuged (10 min at 4 °C and 15,000× *g*). Supernatants were collected (750 µL for NMR analysis and 375 µL for RP/HILIC analysis), dried under SpeedVac (ThermoFisher, Villebon sur Yvette, France), and stored at −20 °C until use. A graphical presentation of all sample preparation protocols is given in Figure S1.

3.1.2. Blood

A pool of sera was divided into 30 aliquots of 200 µL. Six extraction procedures were tested (n = 5 per procedure): blood/ACN (1:2), blood/ACN (1:8), blood/MeOH (1:2), blood/MeOH (1:8), blood/MeOH:$CHCl_3$ (1:1:1), and blood/MeOH/$CHCl_3$ (1:1.5:2.5). Samples were vortexed, kept at −20 °C for 30 min, then centrifuged (10 min, 4 °C, 15,000× *g*). Supernatants were collected in glass tubes and divided into 3 aliquots (750 µL for RMN, 375 µL for RP-LC-MS, and 375 µL for HILIC-LC-MS) for further solvent evaporation in a SpeedVac (ThermoFisher, Villebon sur Yvette, France). A graphical visualization of all sample preparation protocols is given in Figure S2.

3.1.3. Saliva

Saliva samples were pooled and 20 aliquots of 400 µL (n = 5 per solvent) were collected. For metabolite extraction, four protocols were tested (Saliva/MeOH (1:1) and (1:2), Saliva/ACN (1:1) and (1:2)), vortexed for 10 min, and centrifuged (10 min at 4 °C, 15,000× *g*). Supernatants were collected and divided into 3 aliquots (350 µL for the NMR analysis and 175 µL for the RP/HILIC analysis for the 1:1 ratio and 700 µL for the NMR analysis and 200 µL for the RP/HILIC analysis for the 1:2 ratio). Samples were dried under SpeedVac (ThermoFisher, Villebon sur Yvette, France). A graphical visualization of all sample preparation protocols is given in Figure S3.

3.1.4. Feces

For better reproducibility, fecal samples were freeze-dried (FreeZone®4.5L, Labconco, Kansas City, MO, USA) at −107 °C and 0.2 mbar for 24 h and pooled. Sample aliquots of 6 mg (n = 5 per solvent) were added with 1 mL of the extraction solvent (water(1), MeOH/H_2O (1:1), MeOH/H_2O (4:1), ACN/H_2O (1:1), ACN/H_2O (4:1), MeOH/H_2O/ACN (1:1:1)), vortexed for 10 min, and centrifuged (10 min at 4 °C, 15,000× *g*). Supernatants were collected in 3 aliquots (450 µL for the NMR analysis and 225 µL for the C18/HILIC analysis) and dried under SpeedVac (ThermoFisher, Villebon sur Yvette, France) or freeze-dried for H_2O solvent. A graphical visualization of all sample preparation protocols is given in Figure S4.

Dried residues were dissolved in 100 µL of MeOH/H_2O (1:9) for RP-LC and in 100 µL of ACN/H_2O (75:25) for HILIC. Dried residues were dissolved in 200 µL of a deuterated buffer (0.2 M potassium phosphate buffered deuterium oxide (pH = 7.44 ± 0.5) and 10 µL of deuterium oxide (D_2O) with external reference [3-trimethylsilylpropionic acid] (TSP) at 3.2 mM for the ^{1}H-NMR analysis.

3.2. Data Acquisition

3.2.1. UHPLC-MS

As previously described [49], LC-HRMS analysis was performed on an UPLC Ultimate WPS-3000 system (Dionex, Germany), coupled to a QExactive mass spectrometer (Thermo Fisher Scientific, Bremen, Germany).

The chromatography system was equipped separately with two columns: RP-LC (Reverse Phase Liquid Chromatography) Phenomenex Kinetex® XB-C18 (1.7 µm 100 A 150x 2.1 mm from Phenomenex, CA, USA) kept at a constant temperature of 55 °C. The solvent system comprised mobile phase A (ultrapure water) and mobile phase B (methanol), both acidified with 0.1% formic acid. A constant flow rate of 0.4 mL/min over a run of 24 min was used with multistep gradient. The multistep gradient (followed by 2 min equilibration time) was programmed as follows: 0–6 min, 0.1–25% B; 6–10 min, 25–80% B; 10–12 min, 80–90% B; 12–21 min, 90–99.9% B; 21–23 min, 99.9% B; 23–24 min, 99.9–0.1% B.

HILIC (Hydrophilic Interaction Liquid Chromatography) Waters Cortecs® (unbonded silica; 1.6 µm 100 A 150x2.1mm from Waters, Ireland) kept at a constant temperature of 40 °C. The solvent system comprised mobile phase A (ultrapure water; pH = 3.0) and mobile phase B (acetonitrile added with 5% water to dissolve ammonia formate and prevent precipitation) both contain 10mM of ammonium formate and acidified with 0.5% formic acid. A constant flow rate of 0.3 mL/min over a run of 27 min was used with a multistep gradient. The multistep gradient (followed by 2 min equilibration time) was programmed as follows: 0–8 min, 5–18% A; 8–15 min, 18–25% A; 15–20.5 min, 25–75% A; 20.5–22 min, 75–97% A; 22–23 min, 97–5% A; 23–27 min, 5% A.

The auto sampler temperature (Ultimate WPS-3000 UHPLC system, Dionex, Germany) was set at 4 °C.

HESI (head electrospray ionization) source was used for both chromatography systems with a spray voltage of 3.5 kV, a capillary temperature of 325 °C, a heater temperature of 350 °C, a sheath gas flow of 50 arbitrary units (AU), an auxiliary gas flow of 13 AU, a sweep gas flow of 3 AU, and a tube lens S-Lens RF level 50, operated in positive (ESI+) and negative (ESI−) electrospray ionization modes (one run for each mode).

Detection was performed with a full-scan acquisition at 70,000 resolution (m/z = 200), which ranged from 58.0 to 870.0 m/z, with an automatic gain control (AGC) target of 105 charges and a maximum injection time (IT) of 250 ms.

Xcalibur 2.2 software (Thermo Fisher Scientific, Bremen, Germany) controlled the system.

3.2.2. NMR

^{1}H-NMR spectra were acquired at 298K on an AVANCE III HD 600 MHz system (Bruker Biospin, Karlsruhe, Germany) equipped with a Bruker 5 mm TCI CryoProbe with Z-gradient. ^{1}H-NMR spectra were recorded with the «noesypr1d» pulse sequence with a

relaxation delay of 20 s, on a sweep width of 12 ppm, 64 k data points, an acquisition time of 4.55 s, with 64 transients, and 8 dummy scans.

3.3. Data Processing

3.3.1. UHPLC-MS

Data were analyzed using Workflow4Metabolomics (W4M), an online platform for metabolomics data pre-processing including XCMS Online [50,51]. A pre-processing workflow was created following three global steps: (i) the extraction method used CentWave-chromatographic peak detection with a max tolerated ppm m/z deviation of 15.0 and a minimum difference in m/z of 0.008, (ii) the alignment method was completed by grouping chromatographic peaks within and between samples by PeakDensity with a bandwidth of 10 and by using Obiwarp for the retention time correction [52], and (iii) Camera.annotate was used to annotate isotope peaks, adducts, and fragments in the peak list in order to keep only the monoisotopic peak. Detected peaks were identified by the module bank_inhouse with an internal database made by Mass Spectrometry Metabolite Library of Standards MSMLS® (IROA Technologies™). Only metabolites detected in all repetitions were kept for the further analysis.

3.3.2. ^{1}H-NMR

Spectra were processed using TopSpin 3.6 software (Bruker Biospin, Karlsruhe, Germany) and were reduced to buckets of variable widths using AMIX software (Analysis of MIXture, version 3.2, Bruker Karlsruhe, Germany). Bucket identification was done using ChenomX software (NMR Suite 7.7, ChenomX, free software, Edmonton, Canada), an in-house database, and literature [22].

3.4. Data Fusion

Once the pre-processing of the RP and HILIC MS data were completed, Workflow4Metabolomics generated a compound list for RP-LC-MS and for HILIC-LC-MS. ChenomX generated the NMR data list. The three compound lists were combined thanks to the identification of peaks. When metabolites were detected by several platforms, and to keep the highest data quality, a data cleaning was done according to the best replicate chosen by the lowest coefficient of variation (CV%) to get rid of the redundancy.

3.5. Data Analysis

Data obtained from the different preparation protocols for each matrix and the different analytical tools were analyzed. The number of metabolites, unique metabolites, and the coefficient of variation (CV%) were calculated.

To identify the impacted metabolic pathways using the extracted metabolites, we analyzed the set of metabolites for each preparation by the Network Explorer module [53] implemented in Metaboanalyst4.0, a freely available web-based metabolomics analysis suite (MetaboAnalyst4.0, Quebec, QC, Canada http://www.metaboanalyst.ca/ (accessed on 10 March 2020)) [53]. The Network Explorer module was developed to support a network-based approach for multi-omics data integration and interpretation. This module is used to build the mapping of metabolites onto different types of metabolic pathways. The results are presented in a table that summarizes the most significant metabolite sets with their p-values in the Network Explorer. A cutoff was chosen at a raw p-value of 0.05 to retain the most significantly enriched metabolic pathways.

In this study, for each biofluid analyzed, we selected the extraction method based on the number of total metabolites, their CVs, the number of unique metabolites, and the predicted number of metabolic pathways found by using MetaboAnalyst 4.0.

To visualize the metabolic coverage using the four matrices, we used an interactive pathway explorer (iPath) (https://pathways.embl.de/ accessed on 5 November 2020) [54].

Complementarity of matrices was evaluated using an upset graph (https://bioinfo-fr.net/contrarie-par-les-diagrammes-de-venn-decouvrez-les-diagrammes-upset/ (accessed on 7 April 2020)) [46].

3.6. Intraday Precision

Validation method parameters of untargeted metabolomics [55] are specific but intraday precision needs to be addressed. Precision is a crucial point to validate [56]. This parameter was measured by injecting a series of samples from multiple preparations coming from the same and homogeneous pool. It allows for the evaluation of the closeness of repeated measures of the same sample/analyte [55]. The pools previously prepared for each matrix were thus extracted 16 times to validate the reproducibility of the method. The 16 extractions (named repeated samples) were analyzed in LC-MS and NMR. Previous metabolites detected for each matrix were researched in repeated samples. The variation coefficient was calculated for each metabolite in order to validate the reproducibility of the workflow.

4. Conclusions

This study presents the development of a multimatrix and multiplatform analytical method. This approach makes it possible to see the complementarity between the matrices but also between the platforms. It therefore allows a broad description of the human metabolome. It makes it possible to set up a methodology for obtaining a metabolomic map that describes an individual at a specific time. This methodology can therefore be applied in clinical studies comparing two or more groups with different metabolomes. Moreover, a statistical study of the metabolic interaction between metabolites from different compartments could help to understand exchanges. This approach needs to be implemented with an absolute quantification detection in order to follow an individual along his lifetime.

The aims of this work were to explore and represent in a robust way the whole metabolome in the different compartments of the same individual. The multiplicity of data obtained will provide important information on the state of health of a person, both by comparing their metabolomes over time or with a reference population, and by identifying mechanisms for understanding exchanges in compartments. This multilayer and multiplatform metabolic phenotype has the potential to detect very early metabolic trajectory drifts associated with a pathology.

Supplementary Materials: Figure S1: Schematic presentation of the preparation procedures applied on the pool of urine for the selection of the optimum protocol for the maximum metabolome coverage by LC-MS and 1H-NMR, Figure S2: Schematic presentation of the preparation procedures applied on the pool of sera for the selection of the optimum protocol for the maximum metabolome coverage by UHPLC-MS and 1H-NMR, Figure S3: Schematic presentation of the sample preparation procedure applied on the pool of saliva for the selection of the optimum protocol for the maximum metabolome coverage by UPHLC-MS and 1H-NMR, Figure S4: Schematic presentation of the sample preparation procedures applied in the pool of fecal samples for the selection of the optimum protocol for the maximum metabolome coverage by UHPLC-MS and 1H-NMR, Figure S5: Venn diagram to mention the complementarity in term of metabolite coverage across the different analytical platforms according to the four matrices prepared with the chosen preparation, Table S1: Number of extracted buckets in urine for the different extraction methods in 1H-NMR and UHPLC-MS, Table S2: Number of extracted metabolites and coefficient of validation in urine for the different extraction methods in 1H-NMR and UHPLC-MS, Table S3: Metabolite set coverage highlighted by metabolites extracted (CV < 30%) according to the three sample preparations. The number of hits and the p-values are shown for each preparation by metabolite sets, Table S4: Number of extracted buckets in blood for the different extraction methods in 1H-NMR and UHPLC-MS, Table S5: Number of extracted metabolites and coefficient of validation in blood for the different extraction methods in 1H-NMR and UHPLC-MS, Table S6: Metabolite set coverage highlighted by metabolites extracted in blood (CV < 30%) according to the six sample preparations. The number of hits and the p-values are shown for each preparation by metabolite sets, Table S7: Number of extracted buckets in saliva

for the different extraction methods in 1H-NMR and UHPLC-MS, Table S8: Number of extracted metabolites and coefficient of validation in saliva for the different extraction methods in 1H-NMR and UHPLC-MS, Table S9: Metabolite set coverage highlighted by metabolites extracted in saliva (CV < 30%) according to the four sample preparations. The number of hits and the p-values are shown for each preparation by metabolite sets, Table S10: Number of extracted buckets in feces for the different extraction methods in 1H-NMR and UHPLC-MS, Table S11: Number of extracted metabolites and coefficient of validation in feces for the different extraction methods in 1H-NMR and UHPLC-MS, Table S12: Metabolite set coverage highlighted by metabolites extracted in feces (CV < 30%) according to the six sample preparations. The number of hits and the p-values are shown for each preparation by metabolite sets, Table S13: Metabolite lists linked to the Upset plot.

Author Contributions: Conceptualization, P.E., L.N.-D. and C.M.; methodology, C.M.; software, A.L. and F.M.; validation, S.M., J.G. and D.D.-R.; formal analysis, C.M. and N.B.; investigation, C.M.; resources, L.R., L.S., P.E. and A.L.; writing—original draft preparation, C.M. and L.N.-D.; writing—review and editing, H.B., N.B. and P.E.; visualization, C.M. and N.B.; supervision, P.E. and L.N.-D.; project administration, P.E., INSERM funding acquisition, P.E. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the «Region Centre Val de Loire» project PhenoMet, INSERM U1253 iBrain and the University of Tours. Cécile MARTIAS was the recipient of a PhD fellowship from the Region Centre Val de Loire.

Institutional Review Board Statement: All patients included in studies are informed in writing about the collection of their leftover samples remaining after routine biological analyses for research aims and given the right to refuse such uses. In addition, all patients were informed about the data obtained and about their right to access these data, according to articles L.1121-1 and R1121-2 of the French Public Health Code. Participants had to give written refusal for the use of samples and/or clinical data.

Informed Consent Statement: All samples were pooled before analysis.

Data Availability Statement: Data are contained on the supplementary material.

Acknowledgments: We would like to acknowledge the Metabolomics Core Facility "PST-ASB" of the University of Tours–France and the European Regional Development Fund (FEDER).

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

Sample Availability: Not available.

References

1. Holmes, E.; Wilson, I.D.; Nicholson, J.K. Metabolic Phenotyping in Health and Disease. *Cell* **2008**, *134*, 714–717. [CrossRef]
2. Collino, S.; Martin, F.-P.J.; Rezzi, S. Clinical metabolomics paves the way towards future healthcare strategies. *Br. J. Clin. Pharmacol.* **2013**, *75*, 619–629. [CrossRef] [PubMed]
3. Wishart, D.S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat. Rev. Drug Discov.* **2016**, *15*, 473–484. [CrossRef]
4. Newgard, C.B. Metabolomics and Metabolic Diseases: Where Do We Stand? *Cell Metab.* **2017**, *25*, 43–56. [CrossRef] [PubMed]
5. Pallares-Méndez, R.; Aguilar-Salinas, C.A.; Cruz-Bautista, I.; Del Bosque-Plata, L. Metabolomics in diabetes, a review. *Ann. Med.* **2016**, *48*, 89–102. [CrossRef] [PubMed]
6. Miani, M.; Le Naour, J.; Waeckel-Enee, E.; Verma, S.C.; Straube, M.; Emond, P.; Ryffel, B.; van Endert, P.; Sokol, H.; Diana, J. Gut Microbiota-Stimulated Innate Lymphoid Cells Support beta-Defensin 14 Expression in Pancreatic Endocrine Cells, Preventing Autoimmune Diabetes. *Cell Metab.* **2018**, *28*, 557–572e6. [CrossRef] [PubMed]
7. Urpi-Sarda, M.; Almanza-Aguilera, E.; Llorach, R.; Vázquez-Fresno, R.; Estruch, R.; Corella, D.; Sorli, J.; Carmona, F.; Sanchez-Pla, A.; Salas-Salvadó, J.; et al. Non-targeted metabolomic biomarkers and metabotypes of type 2 diabetes: A cross-sectional study of PREDIMED trial participants. *Diabetes Metab.* **2019**, *45*, 167–174. [CrossRef]
8. Jansson, J.; Willing, B.; Lucio, M.; Fekete, A.; Dicksved, J.; Halfvarson, J.; Tysk, C.; Schmitt-Kopplin, P. Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* **2009**, *4*, e6386. [CrossRef]
9. Scoville, E.A.; Allaman, M.M.; Brown, C.T.; Motley, A.K.; Horst, S.N.; Williams, C.S.; Koyama, T.; Zhao, Z.; Adams, D.W.; Beaulieu, D.B.; et al. Alterations in lipid, amino acid, and energy metabolism distinguish Crohn's disease from ulcerative colitis and control subjects by serum metabolomic profiling. *Metabolomics* **2018**, *14*, 1–12. [CrossRef] [PubMed]
10. Ahmed, I.; Roy, B.C.; Khan, S.A.; Septer, S.; Umar, S. Microbiome, Metabolome and Inflammatory Bowel Disease. *Microorganisms* **2016**, *4*, 20. [CrossRef]

11. Ni, Y.; Xie, G.; Jia, W. Metabonomics of Human Colorectal Cancer: New Approaches for Early Diagnosis and Biomarker Discovery. *J. Proteome Res.* **2014**, *13*, 3857–3870. [CrossRef] [PubMed]
12. Euceda, L.R.; Andersen, M.K.; Tessem, M.-B.; Moestue, S.A.; Grinde, M.T.; Bathen, T.F. NMR-Based Prostate Cancer Metabolomics. *Breast Cancer* **2018**, *1786*, 237–257. [CrossRef]
13. Hoang, G.; Udupa, S.; Le, A. Application of metabolomics technologies toward cancer prognosis and therapy. *Int. Rev. Cell Mol. Biol.* **2019**, *347*, 191–223. [CrossRef] [PubMed]
14. Goonewardena, S.N.; Prevette, L.E.; Desai, A. Metabolomics and Atherosclerosis. *Curr. Atheroscler. Rep.* **2010**, *12*, 267–272. [CrossRef] [PubMed]
15. Zhang, T.; Wu, X.; Ke, C.; Yin, M.; Li, Z.; Fan, L.; Zhang, W.; Zhang, H.; Zhao, F.; Zhou, X.; et al. Identification of Potential Biomarkers for Ovarian Cancer by Urinary Metabolomic Profiling. *J. Proteome Res.* **2013**, *12*, 505–512. [CrossRef] [PubMed]
16. Rauschert, S.; Kirchberg, F.F.; Marchioro, L.; Hellmuth, C.; Uhl, O.; Koletzko, B. Early Programming of Obesity Throughout the Life Course: A Metabolomics Perspective. *Ann. Nutr. Metab.* **2017**, *70*, 201–209. [CrossRef] [PubMed]
17. Bagheri, M.; Farzadfar, F.; Qi, L.; Yekaninejad, M.S.; Chamari, M.; Zeleznik, O.A.; Kalantar, Z.; Ebrahimi, Z.; Sheidaie, A.; Koletzko, B.; et al. Obesity-Related Metabolomic Profiles and Discrimination of Metabolically Unhealthy Obesity. *J. Proteome Res.* **2018**, *17*, 1452–1462. [CrossRef]
18. Orozco, J.S.; Hertz-Picciotto, I.; Abbeduto, L.; Slupsky, C.M. Metabolomics analysis of children with autism, idiopathic-developmental delays, and Down syndrome. *Transl. Psychiatry* **2019**, *9*, 1–15. [CrossRef]
19. Wu, Y.; Li, L. Sample normalization methods in quantitative metabolomics. *J. Chromatogr. A* **2016**, *1430*, 80–95. [CrossRef]
20. Pezzatti, J.; Boccard, J.; Codesido, S.; Gagnebin, Y.; Joshi, A.; Picard, D.; González-Ruiz, V.; Rudaz, S. Implementation of liquid chromatography–high resolution mass spectrometry methods for untargeted metabolomic analyses of biological samples: A tutorial. *Anal. Chim. Acta* **2020**, *1105*, 28–44. [CrossRef]
21. Zhang, A.; Sun, H.; Wang, X. Serum metabolomics as a novel diagnostic approach for disease: A systematic review. *Anal. Bioanal. Chem.* **2012**, *404*, 1239–1245. [CrossRef] [PubMed]
22. Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A.C.; Wilson, M.R.; Knox, C.; Bjorndahl, T.C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; et al. The Human Urine Metabolome. *PLoS ONE* **2013**, *8*, e73076. [CrossRef]
23. Duarte, I.F.; Diaz, S.O.; Gil, A.M. NMR metabolomics of human blood and urine in disease research. *J. Pharm. Biomed. Anal.* **2014**, *93*, 17–26. [CrossRef]
24. Zhao, L.; Ni, Y.; Su, M.; Li, H.; Dong, F.; Chen, W.; Wei, R.; Zhang, L.; Guiraud, S.P.; Martin, F.-P.; et al. High Throughput and Quantitative Measurement of Microbial Metabolome by Gas Chromatography/Mass Spectrometry Using Automated Alkyl Chloroformate Derivatization. *Anal. Chem.* **2017**, *89*, 5565–5577. [CrossRef] [PubMed]
25. Agus, A.; Planchais, J.; Sokol, H. Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease. *Cell Host Microbe* **2018**, *23*, 716–724. [CrossRef] [PubMed]
26. Breban, M.; Tap, J.; Leboime, A.; Said-Nahal, R.; Langella, P.; Chiocchia, G.; Furet, J.-P.; Sokol, H. Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. *Ann. Rheum. Dis.* **2017**, *76*, 1614–1622. [CrossRef]
27. Dunn, W.B.; Bailey, N.J.C.; Johnson, H.E. Measuring the metabolome: Current analytical technologies. *Analyst* **2005**, *130*, 606–625. [CrossRef]
28. Beckonert, O.; Keun, H.C.; Ebbels, T.; Bundy, J.; Holmes, E.; Lindon, J.; Nicholson, J. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692–2703. [CrossRef]
29. Gowda, G.A.N.; Raftery, D. Quantitating Metabolites in Protein Precipitated Serum Using NMR Spectroscopy. *Anal. Chem.* **2014**, *86*, 5433–5440. [CrossRef]
30. Gowda, G.A.N.; Gowda, Y.N.; Raftery, D. Expanding the Limits of Human Blood Metabolite Quantitation Using NMR Spectroscopy. *Anal. Chem.* **2015**, *87*, 706–715. [CrossRef]
31. Rico, E.; González, O.; Blanco, M.E.; Alonso, R.M. Evaluation of human plasma sample preparation protocols for untargeted metabolic profiles analyzed by UHPLC-ESI-TOF-MS. *Anal. Bioanal. Chem.* **2014**, *406*, 7641–7652. [CrossRef]
32. Dunn, W.B. Mass spectrometry in systems biology an introduction. *Methods Enzymol.* **2011**, *500*, 15–35. [PubMed]
33. Stevens, V.L.; Hoover, E.; Wang, Y.; Zanetti, K.A. Pre-Analytical Factors that Affect Metabolite Stability in Human Urine, Plasma, and Serum: A Review. *Metabolites* **2019**, *9*, 156. [CrossRef]
34. Karu, N.; Deng, L.; Slae, M.; Guo, A.C.; Sajed, T.; Huynh, H.; Wine, E.; Wishart, D.S. A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database. *Anal. Chim. Acta* **2018**, *1030*, 1–24. [CrossRef] [PubMed]
35. Johnson, C.H.; Karlsson, E.; Sarda, S.; Iddon, L.; Iqbal, M.; Meng, X.; Harding, J.R.; Stachulski, A.V.; Nicholson, J.; Wilson, I.D.; et al. Integrated HPLC-MS and ^{1}H-NMR spectroscopic studies on acyl migration reaction kinetics of model drug ester glucuronides. *Xenobiotica* **2009**, *40*, 9–23. [CrossRef] [PubMed]
36. De Paepe, E.; Van Meulebroek, L.; Rombouts, C.; Huysman, S.; Verplanken, K.; Lapauw, B.; Wauters, J.; Hemeryck, L.Y.; Vanhaecke, L. A validated multi-matrix platform for metabolomic fingerprinting of human urine, feces and plasma using ultra-high performance liquid-chromatography coupled to hybrid orbitrap high-resolution mass spectrometry. *Anal. Chim. Acta* **2018**, *1033*, 108–118. [CrossRef]

37. Guan, Z.; Wu, J.; Wang, C.; Zhang, F.; Wang, Y.; Wang, M.; Zhao, M.; Zhao, C. Investigation of the preventive effect of Sijunzi decoction on mitomycin C-induced immunotoxicity in rats by ^{1}H NMR and MS-based untargeted metabolomic analysis. *J. Ethnopharmacol.* **2018**, *210*, 179–191. [CrossRef] [PubMed]
38. McHugh, C.E.; Flott, T.L.; Schooff, C.R.; Smiley, Z.; Puskarich, M.; Myers, D.D.; Younger, J.G.; Jones, A.E.; Stringer, K.A. Rapid, Reproducible, Quantifiable NMR Metabolomics: Methanol and Methanol: Chloroform Precipitation for Removal of Macromolecules in Serum and Whole Blood. *Metabolites* **2018**, *8*, 93. [CrossRef]
39. Wallner-Liebmann, S.; Tenori, L.; Mazzoleni, A.; Dieber-Rotheneder, M.; Konrad, M.; Hofmann, P.; Luchinat, C.; Turano, P.; Zatloukal, K. Individual Human Metabolic Phenotype Analyzed by ^{1}H NMR of Saliva Samples. *J. Proteome Res.* **2016**, *15*, 1787–1793. [CrossRef]
40. Schulz, A.; Lang, R.; Behr, J.; Hertel, S.; Reich, M.; Kümmerer, K.; Hannig, M.; Hannig, C.; Hofmann, T. Targeted metabolomics of pellicle and saliva in children with different caries activity. *Sci. Rep.* **2020**, *10*, 1–11. [CrossRef]
41. Deda, O.; Gika, H.G.; Wilson, I.; Theodoridis, G.A. An overview of fecal sample preparation for global metabolic profiling. *J. Pharm. Biomed. Anal.* **2015**, *113*, 137–150. [CrossRef]
42. Romick-Rosendale, L.; Goodpaster, A.M.; Hanwright, P.J.; Patel, N.B.; Wheeler, E.T.; Chona, D.L.; Kennedy, M.A. NMR-based metabonomics analysis of mouse urine and fecal extracts following oral treatment with the broad-spectrum antibiotic enrofloxacin (Baytril). *Magn. Reson. Chem.* **2009**, *47*, S36–S46. [CrossRef]
43. Saric, J.; Wang, Y.; Li, J.; Coen, M.; Utzinger, J.; Marchesi, J.; Keiser, J.; Veselkov, K.; Lindon, J.C.; Nicholson, J.; et al. Species Variation in the Fecal Metabolome Gives Insight into Differential Gastrointestinal Function. *J. Proteome Res.* **2008**, *7*, 352–360. [CrossRef]
44. Xu, W.; Chen, D.; Wang, N.; Zhang, T.; Zhou, R.; Huan, T.; Lu, Y.; Su, X.; Xie, Q.; Li, L.; et al. Development of High-Performance Chemical Isotope Labeling LC–MS for Profiling the Human Fecal Metabolome. *Anal. Chem.* **2014**, *87*, 829–836. [CrossRef] [PubMed]
45. Moosmang, S.; Pitscheider, M.; Sturm, S.; Seger, C.; Tilg, H.; Halabalaki, M.; Stuppner, H. Metabolomic analysis—Addressing NMR and LC-MS related problems in human feces sample preparation. *Clin. Chim. Acta* **2019**, *489*, 169–176. [CrossRef] [PubMed]
46. Lex, A.; Gehlenborg, N.; Strobelt, H.; Vuillemot, R.; Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1983–1992. [CrossRef]
47. Letertre, M.P.M.; Dervilly, G.; Giraudeau, P. Combined Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry Approaches for Metabolomics. *Anal. Chem.* **2021**, *93*, 500–518. [CrossRef] [PubMed]
48. Rathahao-Paris, E.; Alves, S.; Junot, C.; Tabet, J.-C. High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics* **2015**, *12*, 1–15. [CrossRef]
49. Hounoum, B.M.; Blasco, H.; Nadal-Desbarats, L.; Diémé, B.; Montigny, F.; Andres, C.; Emond, P.; Mavel, S. Analytical methodology for metabolomics study of adherent mammalian cells using NMR, GC-MS and LC-HRMS. *Anal. Bioanal. Chem.* **2015**, *407*, 8861–8872. [CrossRef]
50. Guitton, Y.; Tremblay-Franco, M.; Le Corguillé, G.; Martin, J.-F.; Pétéra, M.; Roger-Mele, P.; Delabrière, A.; Goulitquer, S.; Monsoor, M.; Duperier, C.; et al. Create, run, share, publish, and reference your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics. *Int. J. Biochem. Cell Biol.* **2017**, *93*, 89–101. [CrossRef]
51. Giacomoni, F.; Le Corguillé, G.; Monsoor, M.; Landi, M.; Pericard, P.; Pétéra, M.; Duperier, C.; Tremblay-Franco, M.; Martin, J.-F.; Jacob, D.; et al. Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics* **2015**, *31*, 1493–1495. [CrossRef] [PubMed]
52. Smith, C.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78*, 779–787. [CrossRef] [PubMed]
53. Chong, J.; Wishart, D.S.; Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinform.* **2019**, *68*, e86. [CrossRef] [PubMed]
54. Darzi, Y.; Letunic, I.; Bork, P.; Yamada, T. iPath3.0: Interactive pathways explorer v3. *Nucleic Acids Res.* **2018**, *46*, W510–W513. [CrossRef] [PubMed]
55. Dudzik, D.; Barbas-Bernardos, C.; García, A.; Barbas, C. Quality assurance procedures for mass spectrometry untargeted metabolomics. A review. *J. Pharm. Biomed. Anal.* **2018**, *147*, 149–173. [CrossRef]
56. Naz, S.; Vallejo, M.; García, A.; Barbas, C. Method validation strategies involved in non-targeted metabolomics. *J. Chromatogr. A* **2014**, *1353*, 99–105. [CrossRef] [PubMed]

SUPPLEMENTARY DATA

Optimization of sample preparation for metabolomics exploration of urine, feces, blood and saliva in human using combined NMR and UHPLC-HRMS platforms

Cécile Martias[1], Nadine Baroukh[1], Sylvie Mavel[1],Hélène Blasco[1,3], Antoine Lefevre[1], Lea Roch[2], Fréderic Montigny[1], Julie Gatien[2], Laurent Schibler [2], Diane Dufour[1,3], Lydie Nadal-Desbarats[1], Patrick Emond[1,3,]

1 - UMR 1253, iBrain, University of Tours, Inserm, Tours, France

2 – ALLICE, Phenotyping station, Nouzilly, France

3- CHRU Tours, Medical biology center, Tours, France.

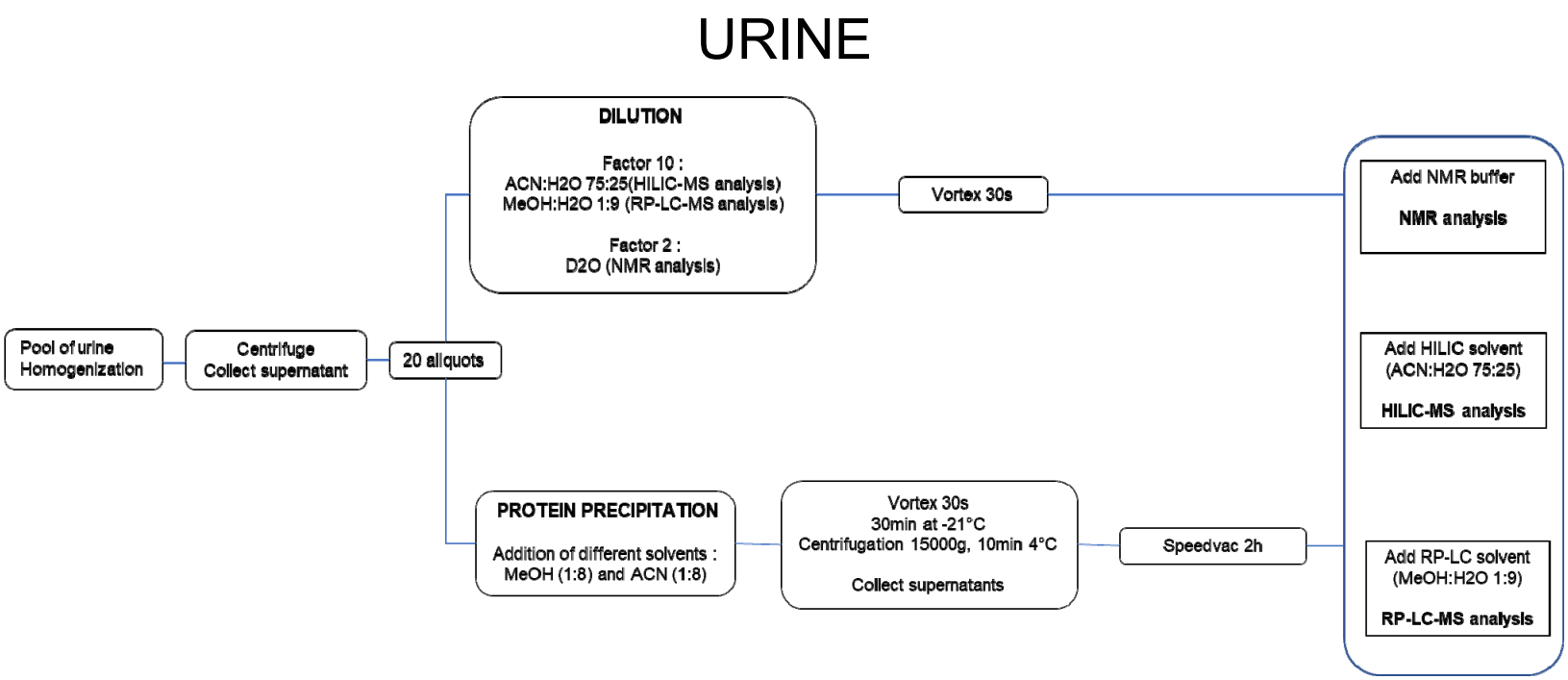


Figure S1: Schematic presentation of the preparation procedures applied on the pool of urine for the selection of the optimum protocol for the maximum metabolome coverage by LC-MS and 1H-NMR.

	Numbers of buckets					CV – standard deviation
	RP-LC POS	**RP-LC NEG**	**HILIC POS**	**HILIC NEG**	**NMR**	
Urine :ACN (1 :8) (n=5)	1976	1053	981	612	44	11% - 7%
Urine :MeOH (1 :8) (n=5)	2335	1533	1030	830	45	9% - 6%
Urine :H_2O (1 :10) (n=5)	2358	1532	1571	1052		9% - 6%
Urine :H_2O (1 :2) (n=5)					46	

Table S1: Number of extracted buckets in urine for the different extraction methods in ^{1}H-NMR and UHPLC-MS

		RP-LC	HILIC	NMR	TOTAL
Urine :ACN (1 :8) (n=5)	Number of metabolites	192	104	37	176
	CV<5%	18	0	9	27
	CV<10%	41	20	35	88
	CV<20%	64	59	36	132
	CV<30%	**72**	**81**	**36**	**145**
	CV % mean (min-max)	20% (1%-158%)	20% (5%-59%)	7% (2%-44%)	17% (1%-158%)
Urine :MeOH (1 :8) (n=5)	Number of metabolites	107	134	37	215
	CV<5%	41	1	23	64
	CV<10%	83	16	36	125
	CV<20%	98	72	36	172
	CV<30%	**102**	**114**	**36**	**201**
	CV % mean (min-max)	8% (1%-76%)	20% (4%-51%)	5% (4%-9%)	12% (1%-76%)
Urine :H_2O (1 :10) (n=5)	Number of metabolites	112	119		205
	CV<5%	47	7		82
	CV<10%	87	47		144
	CV<20%	99	99		186
	CV<30%	**105**	**112**		**197**
	CV % mean (min-max)	9% (0%-77%)	14% (2%-84%)		
Urine :H_2O (1 :2) (n=5)	Number of metabolites			37	
	CV<5%			32	
	CV<10%			35	
	CV<20%			37	
	CV<30%			**37**	
	CV % mean (min-max)			4% (2%-14%)	9% (0%-77%)

Table S2: Number of extracted metabolites and coefficient of validation in urine for the different extraction methods in ^{1}H-NMR and UHPLC-MS

Pathway name	Urine:ACN (1:8)		Urine:MeOH (1:8)		Urine:H_2O	
	Hits	***p-value***	**Hits**	***p*-value**	**Hits**	***p*-value**
Glyoxylate and dicarboxylate metabolism	7	1.86 E^{-3}	12	1.37 E^{-6}	9	4.17 E^{-4}
Phenylalanine metabolism	6	7.16 E^{-6}	6	3.58 E^{-5}	6	3.58 E^{-5}
Glycine, serine and threonine metabolism	7	1.52 E^{-3}	9	3.17 E^{-4}	8	1.65 E^{-3}
Citrate cycle	4	1.31 E^{-2}	6	9.16 E^{-4}	6	9.16 E^{-4}
Alanine, aspartate and glutamate metabolism	/	/	8	1.00 E^{-3}	/	/
Purine metabolism	/	/	13	1.2 E^{-3}	11	1.1 E^{-2}
Phenylalanine, tyrosine and tryptophane biosynthesis	2	2.01 E^{-2}	3	1.85 E^{-3}	3	1.85 E^{-3}
Glycerophospholipid metabolism	/	/	5	2.23 E^{-3}	/	/
Tryptophane metabolism	9	5.05 E^{-4}	9	3.68 E^{-3}	8	1.28 E^{-2}
beta-Alanine metabolism	/	/	6	4.46 E^{-3}	7	7.6 E^{-4}
Pantothenate and CoA biosynthesis	/	/	5	8.3 E^{-3}	/	/
Pyrimidine metabolism	/	/	8	9.46E^{-3}	/	/
Histidine metabolism	/	/	/	/	5	6.25 E^{-3}
Tyrosine metabolism	8	2.81 E^{-3}	/	/	/	/
Sphingolipid metabolism	3	1.38 E^{-2}	/	/	/	/
Synthesis and degradation of ketone bodies	2	3.22 E^{-2}	/	/	/	/
TOTAL	9 pathways		12 pathways		9 pathways	

Table S3: Metabolite set coverage highlighted by metabolites extracted (CV<30%) according to the three sample preparations. The number of hits and the p-values are shown for each preparation by metabolite sets.

BLOOD

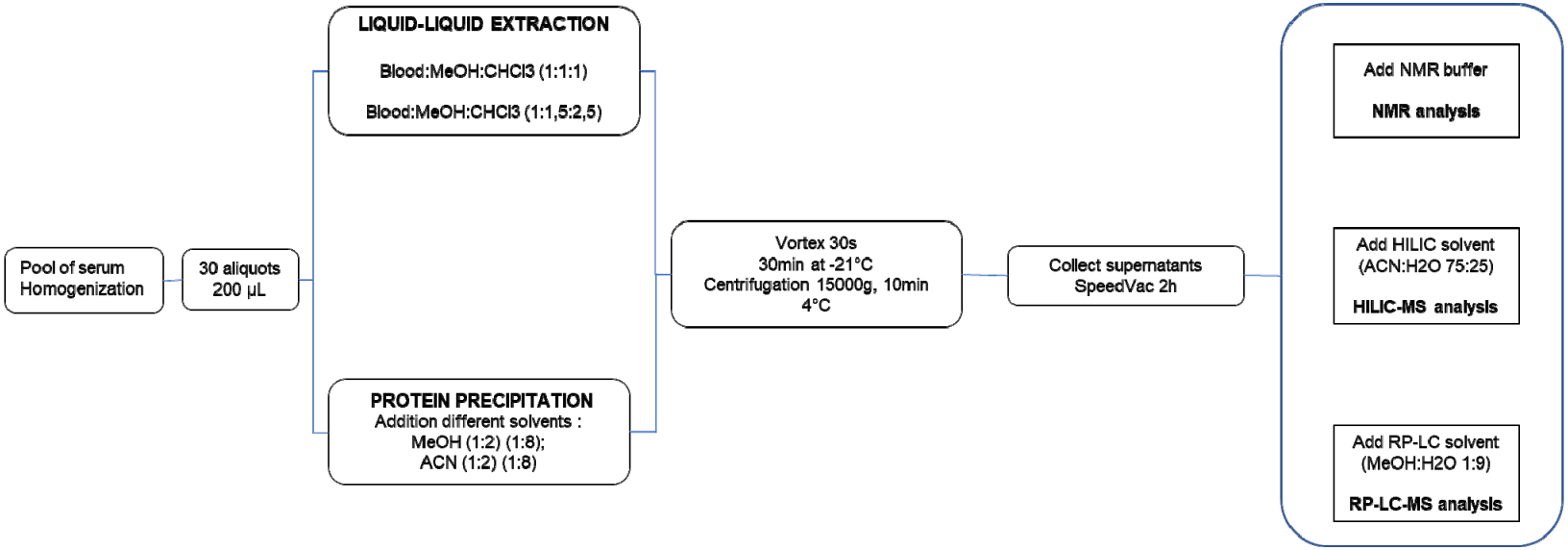


Figure S2: Schematic presentation of the preparation procedures applied on the pool of sera for the selection of the optimum protocol for the maximum metabolome coverage by UHPLC-MS and ^{1}H-NMR.

	Numbers of buckets					CV – standard deviation
	RP-LC POS	**RP-LC NEG**	**HILIC POS**	**HILIC NEG**	**NMR**	
Blood :ACN (1 :2) (n=5)	1233	1137	1624	2345	42	11% - 6%
Blood :ACN (1 :8) (n=5)	1173	997	1907	2373	37	13% - 6%
Blood :MeOH (1 :2) (n=5)	1285	1171	2018	2189	35	14% - 6%
Blood :MeOH (1 :8) (n=5)	1363	1167	2266	1908	42	13% - 5%
Blood: MeOH:$CHCl_3$ (1:1:1) (n=5)	996	1016	1049	1441	43	11% - 6%
Blood: MeOH:$CHCl_3$ (1:1.5:2.5) (n=5)	991	1020	1065	2019	42	14% - 6%

Table S4: Number of extracted buckets in blood for the different extraction methods in ^{1}H-NMR and UHPLC-MS

		RP-LC	HILIC	NMR	Total
Blood :ACN (1 :2) (n=5)	Number of metabolites	75	80	22	139
	CV<5%	45	3	4	51
	CV<10%	66	35	17	96
	CV<20%	71	66	19	122
	CV<30%	**72**	**74**	**20**	**129**
	CV % mean (min-max)	6% (1%-42%)	13%(4%-44%)	11% (4%-76%)	10% (1%-76%)
Blood :ACN (1 :8) (n=5)	Number of metabolites	66	82	22	135
	CV<5%	9	6	0	15
	CV<10%	24	36	4	58
	CV<20%	59	59	10	105
	CV<30%	**62**	**74**	**19**	**123**
	CV % mean (min-max)	12% (1%-39%)	15% (2%-74%)	20% (7%-37%)	14% (1%-74%)
Blood :MeOH (1 :2) (n=5)	Number of metabolites	76	88	22	147
	CV<5%	7	4	0	11
	CV<10%	50	28	4	75
	CV<20%	74	67	19	132
	CV<30%	**76**	**80**	**21**	**141**
	CV % mean (min-max)	9% (1%-29%)	16% (3%-77%)	14% (9%-40%)	12% (1%-77%)
Blood :MeOH (1 :8) (n=5)	Number of metabolites	89	88	22	157
	CV<5%	50	2	12	62
	CV<10%	73	36	19	107
	CV<20%	81	55	21	128
	CV<30%	**84**	**73**	**21**	**143**
	CV % mean (min-max)	7% (1%-60%)	19% (3%-123%)	6% (1%-33%)	11% (1%-72%)
Blood: MeOH:$CHCl_3$ (1:1:1) (n=5)	Number of metabolites	63	74	21	127
	CV<5%	9	5	16	30
	CV<10%	53	18	20	81
	CV<20%	62	52	20	112
	CV<30%	**63**	**61**	**21**	**119**
	CV % mean (min-max)	7% (1%-21%)	21% (1%-109%)	5% (2%-28%)	12% (1%-109%)
Blood: MeOH:$CHCl_3$ (1:1.5:2.5) (n=5)	Number of metabolites	60	76	22	126
	CV<5%	22	3	14	37
	CV<10%	54	26	18	88
	CV<20%	58	52	18	108
	CV<30%	**60**	**65**	**20**	**118**
	CV % mean (min-max)	6% (1%-28%)	17% (2%-56%)	10% (2%-75%)	11% (1%-75%)

Table S5: Number of extracted metabolites and coefficient of validation in blood for the different extraction methods in ^{1}H-NMR and UHPLC-MS

Pathway name	ACN 1:2		ACN 1:8		MeOH 1:2		MeOH 1:8		Blood: MeOH:CHCl3 1:1:1		Blood: MeOH:CHCl3 1:1.5:2.5	
	Hits	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**
Tryptophan metabolism	9	$4.15\ E^{-4}$	9	$1.99\ E^{-4}$	9	$5.55\ E^{-4}$	9	$4.58\ E^{-4}$	/	/	/	/
Phenylalanine, tyrosine and tryptophane biosynthesis	3	$7.63\ E^{-4}$	3	$5.76\ E^{-4}$	3	$8.55\ E^{-4}$	3	$7.93\ E^{-4}$	3	$5.52\ E^{-4}$	3	$5.52\ E^{-4}$
Phenylalanine metabolism	4	$1.81\ E^{-3}$	4	$1.27\ E^{-3}$	4	$2.09\ E^{-3}$	4	$1.19\ E^{-3}$	3	$1.31\ E^{-2}$	3	$1.31\ E^{-2}$
Pantothenate and CoA biosynthesis	5	$2.25\ E^{-3}$	4	$1.08\ E^{-2}$	/	/	5	$2.38\ E^{-3}$	4	$1.02\ E^{-2}$	5	$1.37\ E^{-3}$
Tyrosine metabolism	/	/	/	/	/	/	8	$2.59\ E^{-3}$	/	/	/	/
Valine, leucine and isoleucine metabolism	3	$8.98\ E^{-3}$	3	$6.89\ E^{-3}$	3	$9.99\ E^{-3}$	3	$9.31\ E^{-3}$	3	$6.62\ E^{-3}$	3	$6.62\ E^{-3}$
Purine metabolism	/	/	8	$2.00\ E^{-2}$	9	$1.50\ E^{-2}$	9	$1.28\ E^{-2}$	9	$5.77\ E^{-3}$	9	$5.77\ E^{-3}$
Ubiquinone and other terpenoid-quinone biosynthesis	/	/	/	/	/	/	3	$1.34\ E^{-2}$	/	/	/	/
Pyrimidine metabolism	/	/	/	/	/	/	6	$2.54\ E^{-2}$	/	/	/	/
Glycine, serine and threonine metabolism	7	$1.30\ E^{-3}$	/	/	7	$1.63\ E^{-3}$	/	/	6	$3.88\ E^{-3}$	7	$6.71\ E^{-4}$
Glycerophospholipid metabolism	/	/	4	$2.90\ E^{-2}$	4	$6.17\ E^{-3}$	/	/	4	$3.62\ E^{-3}$	/	/
Glyoxylate and dicarboxylate metabolism	6	$7.90\ E^{-3}$	/	/	4	$9.50\ E^{-3}$	/	/	7	$8.30\ E^{-4}$	6	$4.61\ E^{-3}$
Sphingolipid metabolism	3	$1.29\ E^{-2}$	/	/	3	$1.43\ E^{-2}$	/	/	/	/	/	/
Alanine, aspartate and glutamate metabolism	5	$2.13\ E^{-2}$	/	/	/	/	/	/	/	/	/	/
Pyruvate metabolism	/	/	4	$1.62\ E^{-2}$	/	/	/	/	/	/	4	$1.54\ E^{-2}$
Synthesis and degradation of ketone bodies	/	/	2	$2.57\ E^{-2}$	/	/	/	/	/	/	/	/
Citrate cycle	/	/	/	/	/	/	/	/	4	$8.16\ E^{-3}$	/	/
Biotin metabolism	/	/	/	/	/	/	/	/	/	/	2	$1.55\ E^{-2}$
TOTAL	9 pathways		9 pathways		9 pathways		9 pathways		9 pathways		9 pathways	

Table S6: Metabolite set coverage highlighted by metabolites extracted in blood (CV<30%) according to the six sample preparations. The number of hits and the p-values are shown for each preparation by metabolite sets.

SALIVA

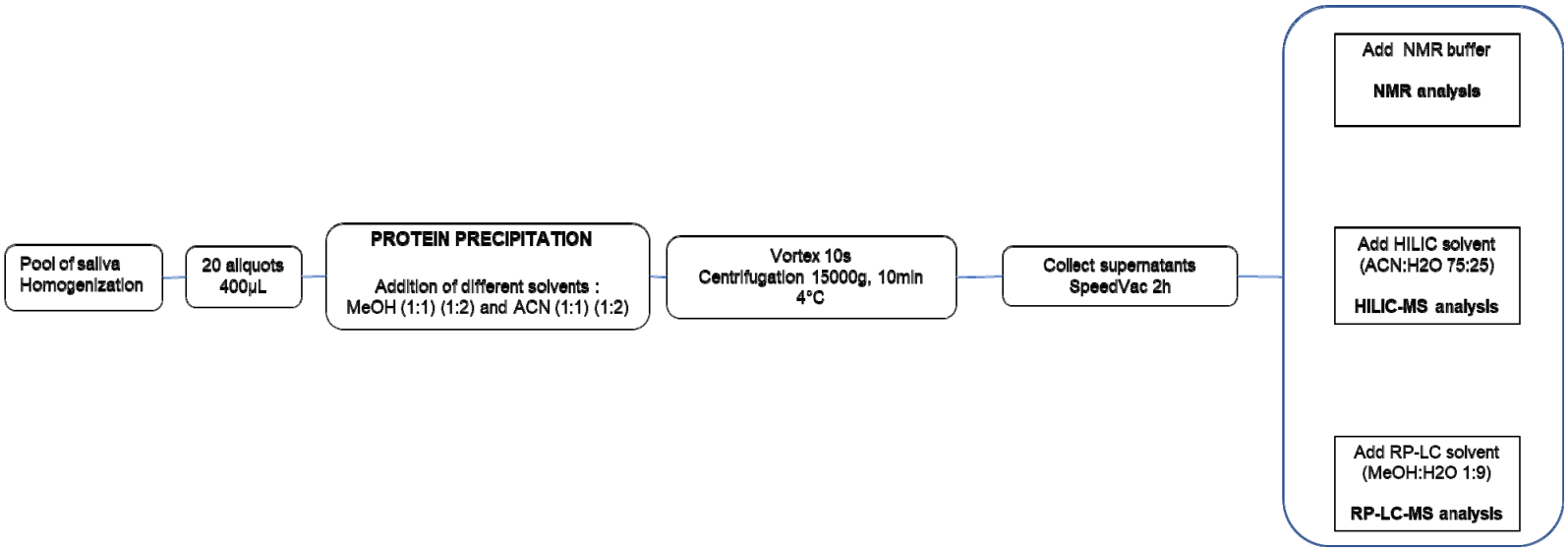


Figure S3: Schematic presentation of the sample preparation procedure applied on the pool of saliva for the selection of the optimum protocol for the maximum metabolome coverage by UPHLC-MS and ^{1}H-NMR

	Numbers of buckets					CV – standard deviation
	RP-LC POS	RP-LC NEG	HILIC POS	HILIC NEG	NMR	
Saliva :ACN (1 :2) (n=5)	987	520	1575	679	43	16% - 6%
Saliva :ACN (1 :1) (n=5)	988	684	1506	684	44	17% - 6%
Saliva :MeOH (1 :2) (n=5)	941	531	1233	725	47	19% - 6%
Saliva :MeOH (1 :1) (n=5)	987	407	1240	736	43	19% - 5%

Table S7: Number of extracted buckets in saliva for the different extraction methods in ^{1}H-NMR and UHPLC-MS

		RP-LC	HILIC	NMR	Total
Saliva :ACN (1 :2) (n=5)	Number of metabolites	53	108	18	147
	CV<5%	11	4	0	15
	CV<10%	38	43	9	77
	CV<20%	46	80	15	120
	CV<30%	**50**	**96**	**16**	**136**
	CV % mean (min-max)	11% (2%- 71%)	17% (3%-75%)	16% (8%-90%)	13% (2%-62%)
Saliva :ACN (1 :1) (n=5)	Number of metabolites	57	110	18	148
	CV<5%	18	3	7	28
	CV<10%	41	22	13	67
	CV<20%	50	63	15	110
	CV<30%	**53**	**98**	**16**	**137**
	CV % mean (min-max)	11% (1%- 78%)	20% (1%-80%)	11% (3%-54%)	14% (1%-80%)
Saliva :MeOH (1 :2) (n=5)	Number of metabolites	53	109	18	146
	CV<5%	13	5	2	20
	CV<10%	36	16	14	61
	CV<20%	46	49	16	96
	CV<30%	**50**	**90**	**18**	**133**
	CV % mean (min-max)	11% (2%- 65%)	21% (2%-80%)	8% (4%-28%)	15% (2%-80%)
Saliva :MeOH (1 :1) (n=5)	Number of metabolites	50	108	18	145
	CV<5%	22	19	0	39
	CV<10%	38	36	0	69
	CV<20%	47	68	14	111
	CV<30%	**49**	**89**	**16**	**133**
	CV % mean (min-max)	8% (1%- 50%)	19% (0%-93%)	22% (16%-52%)	15% (0%-93%)

Table S8: Number of extracted metabolites and coefficient of validation in saliva for the different extraction methods in ^{1}H-NMR and UHPLC-MS

Pathway name	ACN 1:2		ACN 1:1		MeOH 1:2		MeOH 1:1	
	Hits	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**
Purine metabolism	14	1.98 E^{-5}	12	3.57 E^{-4}	14	1.98 E^{-5}	13	9.41 E^{-5}
Phenylalanine metabolism	5	1.62 E^{-4}	5	1.53 E^{-4}	5	1.62 E^{-4}	4	2.19 E^{-3}
Pyrimidine metabolism	9	4.40 E^{-4}	11	9.78 E^{-6}	10	7.72 E^{-5}	10	7.20 E^{-5}
Histidine metabolism	5	2.09 E^{-3}	/	/	6	2.33 E^{-4}	5	2.09 E^{-3}
Pantothenate and CoA biosynthesis	5	2.81 E^{-3}	5	2.66 E^{-3}	5	2.81 E^{-3}	5	2.81 E^{-3}
Glutathione metabolism	5	4.79 E^{-3}	/	/	7	7.64 E^{-5}	6	6.79 E^{-4}
Alanine, aspartate and glutamate metabolism	6	5.99 E^{-3}	7	1.05 E^{-3}	6	5.99 E^{-3}	7	1.13 E^{-3}
Glyoxylate and dicarboxylate metabolism	6	1.01 E^{-2}	7	2.01 E^{-3}	7	2.16 E^{-3}	8	3.87 E^{-4}
Glycerophospholipid metabolism	/	/	4	6.17 E^{-3}	/	/	/	/
Valine, leucine and isoleucine metabolism	/	/	3	9.99 E^{-3}	/	/	/	/
beta-Alanine metabolism	/	/	/	/	5	7.59 E^{-3}	/	/
TOTAL	8 pathways		8 pathways		9 pathways		8 pathways	

Table S9: Metabolite set coverage highlighted by metabolites extracted in saliva (CV<30%) according to the four sample preparations. The number of hits and the p-values are shown for each preparation by metabolite sets.

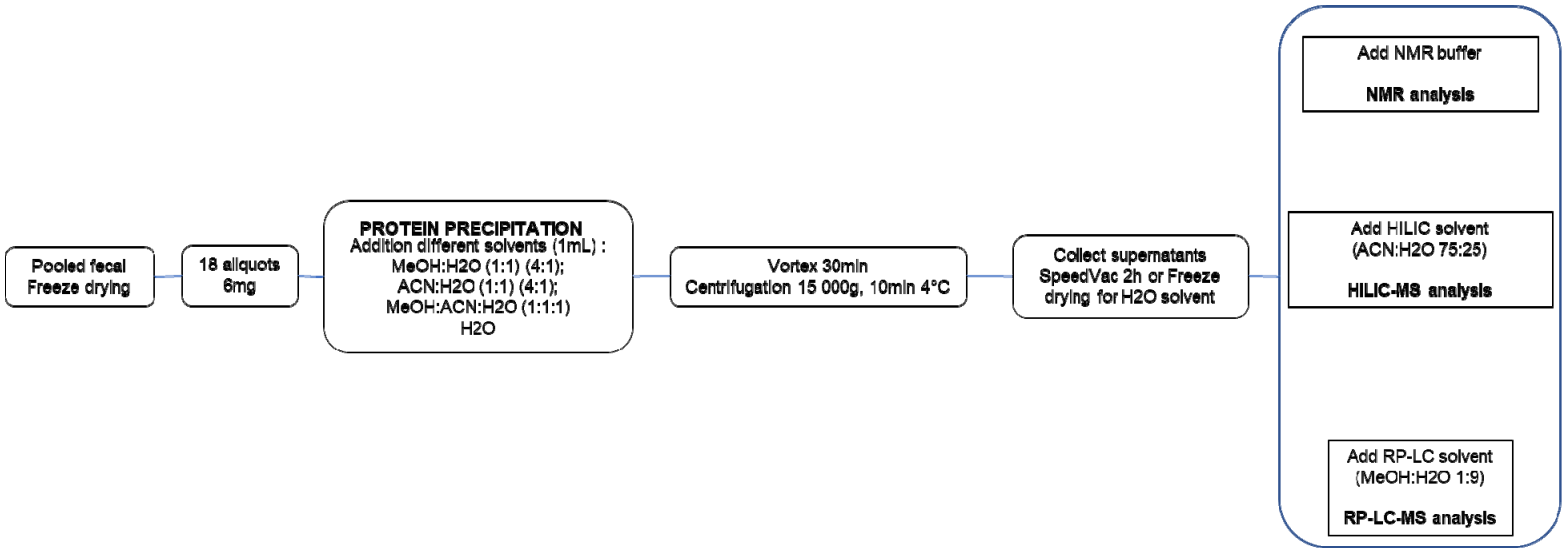


Figure S4: Schematic presentation of the sample preparation procedures applied in the pool of fecal samples for the selection of the optimum protocol for the maximum metabolome coverage by UHPLC-MS and ^{1}H-NMR.

	Numbers of buckets					CV – standard deviation
	RP-LC POS	**RP-LC NEG**	**HILIC POS**	**HILIC NEG**	**NMR**	
ACN:H_2O (1 :1) (n=5)	2013	1963	1948	1251	45	12% - 5%
ACN:H_2O (4 :1) (n=5)	1984	1931	1912	1261	45	10% - 5%
MeOH:H_2O (1:1) (n=5)	1908	1961	1851	1259	45	9% - 4%
MeOH :H_2O (4:1) (n=5)	1991	1953	1865	1187	45	10% - 5%
MeOH:H_2O:ACN (1 :1: 1) (n=5)	1827	1898	1863	1255	45	11% - 5%
H_2O (n=5)	1636	1737	1537	1043	45	12% - 6%

Table S10: Number of extracted buckets in feces for the different extraction methods in ^{1}H-NMR and UHPLC-MS

		RP-LC	HILIC	NMR	TOTAL
ACN:H_2O (1 :1) (n=5)	Number of metabolites	163	203	25	281
	CV<5%	1	28	0	29
	CV<10%	49	119	23	161
	CV<20%	128	174	25	250
	CV<30%	**145**	**192**	**25**	**266**
	CV % mean (min-max)	17% (1%-125%)	12% (1%-74%)	7% (6%-11%)	12% (1%-105%)
ACN:H_2O (4 :1) (n=5)	Number of metabolites	155	199	25	272
	CV<5%	17	30	4	47
	CV<10%	90	118	22	181
	CV<20%	126	172	25	241
	CV<30%	**140**	**189**	**25**	**259**
	CV % mean (min-max)	15% (2%-120%)	11% (2%-78%)	6% (4%-12%)	11% (2%-82%)
MeOH:H_2O (1:1) (n=5)	Number of metabolites	155	201	25	278
	CV<5%	26	41	0	63
	CV<10%	101	121	22	190
	CV<20%	129	171	25	242
	CV<30%	**143**	**190**	**25**	**264**
	CV % mean (min-max)	13% (2%-130%)	11% (2%-76%)	8% (5%-13%)	10% (2%-130%)
MeOH :H_2O (4:1) (n=5)	Number of metabolites	156	196	25	270
	CV<5%	62	24	9	79
	CV<10%	105	97	24	173
	CV<20%	136	166	25	246
	CV<30%	**143**	**184**	**25**	**259**
	CV % mean (min-max)	11% (1%-148%)	13% (2%-62%)	6% (3%-15%)	10% (1%-62%)
MeOH:H_2O:ACN (1 :1: 1) (n=5)	Number of metabolites	160	195	25	275
	CV<5%	19	35	8	55
	CV<10%	83	117	24	176
	CV<20%	127	160	25	236
	CV<30%	**141**	**184**	**25**	**258**
	CV % mean (min-max)	17% (2%-144%)	12% (2%-70%)	6% (3%-11%)	12%(2%-144%)
H_2O (n=5)	Number of metabolites	152	195	25	269
	CV<5%	27	9	4	36
	CV<10%	87	50	21	135
	CV<20%	131	127	25	218
	CV<30%	**140**	**173**	**25**	**254**
	CV % mean (min-max)	12% (2%-65%)	18% (2%-111%)	7% (4%-15%)	13%(2%-69%)

Table S11: Number of extracted metabolites and coefficient of validation in feces for the different extraction methods in ^{1}H-NMR and UHPLC-MS

Name	ACN:H2O 1:1		ACN:H2O 41		MeOH:H2O 1:1		MeOH:H2O 4:1		MeOH:H2O:ACN 1:1:1		H2O	
	Hits	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**	**Hits**	***p*-value**
Histidine metabolism	8	$8.17\ E^{-5}$	9	$9.02\ E^{-6}$	8	$7.35\ E^{-5}$	9	$6.64\ E^{-6}$	7	$4.78\ E^{-4}$	8	$5.91\ E^{-5}$
Tyrosine metabolism	12	$8.52\ E^{-4}$	14	$5.54\ E^{-5}$	13	$1.82\ E^{-4}$	15	$7.08\ E^{-6}$	12	$5.27\ E^{-4}$	13	$1.33\ E^{-4}$
Pyrimidine metabolism	13	$8.95\ E^{-5}$	13	$1.04\ E^{-4}$	13	$7.66\ E^{-5}$	13	$7.08\ E^{-5}$	14	$9.60\ E^{-6}$	13	$5.57\ E^{-5}$
Phenylalanine metabolism	6	$1.89\ E^{-4}$	5	$2.18\ E^{-3}$	5	$1.91\ E^{-3}$	6	$1.67\ E^{-4}$	5	$1.61\ E^{-3}$	5	$1.67\ E^{-3}$
Valine, leucine and isoleucine metabolism	6	$3.03\ E^{-5}$	5	$5.79\ E^{-4}$	6	$2.78\ E^{-5}$	5	$4.88\ E^{-4}$	6	$2.23\ E^{-5}$	6	$2.33\ E^{-5}$
Arginine and proline metabolism	11	$9.80\ E^{-4}$	11	$1.11\ E^{-3}$	11	$8.65\ E^{-4}$	11	$8.12\ E^{-4}$	11	$6.26\ E^{-4}$	11	$6.69\ E^{-4}$
Alanine, aspartate and glutamate metabolism	9	$1.57\ E^{-3}$	9	$1.74\ E^{-3}$	9	$1.41\ E^{-3}$	9	$1.34\ E^{-3}$	9	$1.07\ E^{-3}$	9	$1.14\ E^{-3}$
Tryptophan metabolism	12	$6.69\ E^{-4}$	12	$7.64\ E^{-4}$	12	$5.84\ E^{-4}$	11	$2.08\ E^{-3}$	11	$1.62\ E^{-3}$	12	$4.42\ E^{-4}$
Beta-Alanine metabolism	7	$4.23\ E^{-3}$	7	$4.59\ E^{-3}$	7	$3.90\ E^{-3}$	7	$3.74\ E^{-3}$	/	/	7	$3.28\ E^{-3}$
Phenylalanine, tyrosine and tryptophane biosynthesis	/	/	3	$4.51\ E^{-3}$	/	/	3	$4.06\ E^{-3}$	/	/	/	/
Pantothenate and CoA biosynthesis	6	$5.91\ E^{-3}$	6	$6.35\ E^{-3}$	6	$5.55\ E^{-3}$	6	$5.30\ E^{-3}$	6	$4.56\ E^{-3}$	6	$4.74\ E^{-3}$
Glutathione metabolism	/	/	/	/	/	/	6	$9.75\ E^{-3}$	6	$8.43\ E^{-3}$	6	$8.75\ E^{-3}$
Purine metabolism	14	$6.11\ E^{-3}$	/	/	/	/	/	/	14	$3.78\ E^{-3}$	/	/
Lysine degradation											6	$8.75\ E^{-3}$
TOTAL	11 pathways		11 pathways		10 pathways		12 pathways		11 pathways		12 pathways	

Table S12: Metabolite set coverage highlighted by metabolites extracted in feces (CV<30%) according to the six sample preparations. The number of hits and the p-values are shown for each preparation by metabolite sets.

PLATFORM COMPLEMENTARITY

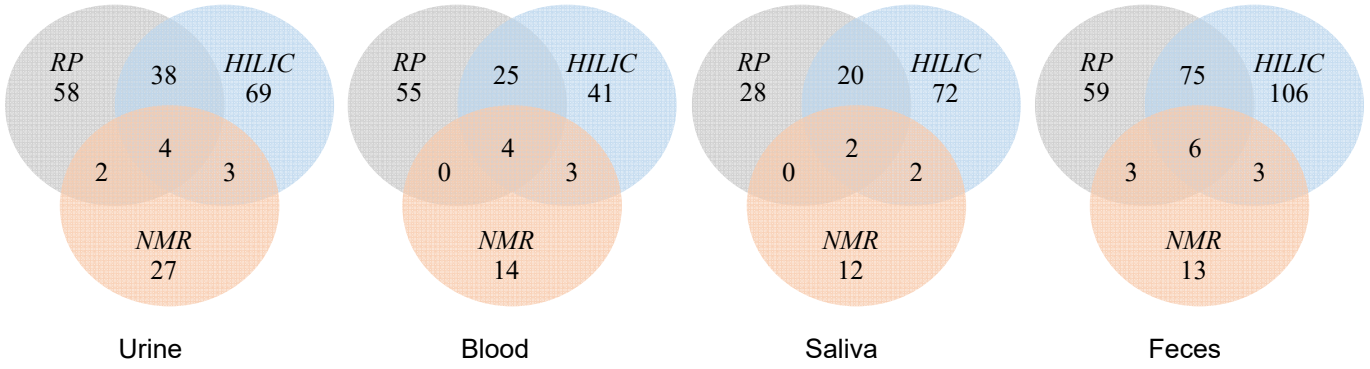


Figure S5: Venn diagram to mention the complementarity in term of metabolite coverage across the different analytical platforms according to the four matrices prepared with the chosen preparation

MATRICES COMPLEMENTARITY

Elements in FECES :		Elements in URINE :	Elements in SALIVA :	Elements in BLOOD :	Elements in FECES URINE :	Elements in BLOOD FECES :	Elements in FECES SALIVA :
66 metabolites		25 metabolites	8 metabolites	6 metabolites	31 metabolites	12 metabolites	7 metabolites
Methylglutaric acid	Imidazoleacetic acid	(2E,4E)-2,4-Hexadienoic acid	2-Ketobutyric acid	1-Methylhistidine	1-Naphthylamine	2-Hydroxy-3-methylbutyric acid	3-Hydroxyphenyllactate
(+/-)-2-Hydroxy-4-(methylthio)butanoic acid	Indole	17a-Estradiol	3-(3-Hydroxyphenyl)propanoic acid	2-Hydroxybutyric acid	2-Aminoisobutyric acid	3,4-Dihydroxybenzeneacetic acid	3-Sulfinoalanine
(R)-Salsolinol	Isovaleric acid	2-Methylglutaric acid	5-Methylcytosine	Dodecanoylcarnitine	3-(2-Hydroxyphenyl)propanoic acid	Glycocholic acid	Cadaverine
1-Aminocyclopropanecarboxylic acid	Ketoleucine	5-Hydroxy-L-tryptophan	Butanal	Propanal	3-Hydroxybenzoic acid	Homogentisic acid	Dethiobiotin
3-Amino-5-hydroxybenzoate	L-Acetylcarnitine	5'-Methylthioadenosine	Deoxyadenosine	Theobromine	3-Methoxytyramine	L-Glutamine	D-Xylose
3-Aminoisobutanoic acid	L-Asparagine	Acetoin	D-Ribose 5-phosphate	Thyroxine	3-Methyldioxyindole	L-Kynurenine	Histamine
4,5-Dihydroorotic acid	L-Dopa	Acetylcysteine	Propionic acid		4-Hydroxybenzoic acid	L-Proline	Malonate
4-Guanidinobutanoic acid	Lithocholic acid	Benzoic acid	Putrescine		CMPciliatine	MG(18:1(9Z)/0:0/0:0)	
4-Hydroxycinnamic acid	L-Leucine	Cinnamaldehyde			Dimethylamine	Protocatechuic acid	
4-Methylcatechol	L-Lysine	Citraconic acid			Ethylmalonic acid	Ribitol	
5-Aminopentanoic acid	Methylmalonic acid	Citric acid			Guanidoacetic acid	Salicyluric acid	
5-Hydroxyindoleacetic acid	Mevalonolactone	Cyclic GMP			Homocysteine thiolactone	Vanylglycol	
6-Mercaptopurine ribonucleoside triphosphate	N-Acetyl-L-methionine	Dimethylglycine			Homovanillic acid		
Adenosine monophosphate	N-Acetylputrescine	Galactaric acid			L-Arginine		
Adipic acid	N-Acetylserine	Glycolate			L-Glutamic acid		
Anserine	N-alpha-Acetyl-L-lysine	Hydroxyphenyllactic acid			L-Histidine		
Ascorbate	Nicotine	Indoxyl sulfate			L-Histidinol		
Benzyl alcohol	Normetanephrine	Itaconic acid			L-Serine		
Betaine	Oleic acid	L-Arabinose			Maleamate		
Butyric acid	Ornithine	Maleic acid			Melatonin		
Catechol	Orotic acid	Mannitol			N-Acetylneuraminic acid		
Citrulline	Palmitic acid	myo-Inositol			N-Methyltryptamine		
Cysteic acid	Phenylethylamine	Nicotinamide ribotide			O-Succinyl-L-homoserine		
Deoxycholic acid	Phosphorylcholine	Ophthalmic acid			Phenylacetaldehyde		
Deoxycytidine	p-Hydroxyphenylacetic acid	Tryptophanol			Pterin		
Diaminopimelic acid	p-Octopamine				Pyrrole-2-carboxylic acid		
Diethyl oxalpropionate	Purine				Shikimic acid		
D-Ribose	Pyridoxamine				Trigonelline		
Fumaric acid	Sarcosine				Ureidopropionic acid		
Gentisic acid	Tartaric acid				Valeric acid		
Heptanoic acid	Thiamine				Xanthurenic acid		
Hydroquinone	trans-Aconitic acid						
Hydroxykynurenine	Tryptamine						

Elements in BLOOD URINE :	Elements in BLOOD SALIVA :	Elements in SALIVA URINE :	Elements in BLOOD FECES URINE :	Elements in FECES SALIVA URINE :	Elements in BLOOD FECES SALIVA :	Elements in BLOOD SALIVA URINE :	Elements in BLOOD FECES SALIVA URINE :
6 metabolites	**2 metabolites**	**2 metabolites**	**26 metabolites**	**19 metabolites**	**10 metabolites**	**5 metabolites**	**37 metabolites**
Creatine	2-Aminophenol	AICAR	4-Acetamidobutanoic acid	1H-Indole-3-acetamide	3-Hydroxycapric acid	Caffeine	1-Methyladenosine
L-Malic acid	Phenylacetic acid	cis-Aconitic acid	4-Pyridoxic acid	3-Methyladenine	3-Methoxytyrosine	Choline	2-Hydroxypyridine
Quinic acid			4-Trimethylammoniobutanoic acid	Acetic acid	Caffeate	Isocitric acid	3-Hydroxyanthranilic acid
S-Adenosylhomocysteine			6-Hydroxynicotinic acid	Adenosine	Diethanolamine	N6-(delta2-Isopentenyl)-adenine	3-Hydroxybutyric acid
Sphinganine			Acetoacetic acid	Biotin	Leukotriene B4	N-Alpha-acetyllysine	3-Hydroxymethylglutaric acid
Trimethylamine N-oxide			All-trans-retinoic acid	Cytidine	L-Valine		Acetylglycine
			Azelaic acid	Deoxyuridine	Pectin		Adenine
			Cholic acid	Glycerophosphocholine	Pipecolic acid		Cortisol
			Cortisone	Guanosine	Theophylline		Creatinine
			Diacetyl	Inosine	Trimethylamine		Cytosine
			Ethanolamine	L-Aspartic acid			D-Glucose
			Indole-3-methyl acetate	N-Acetylglutamic acid			Dihydrouracil
			Indoleacetaldehyde	Nicotinic acid			Formic acid
			Kynurenic acid	Pimelic acid			Glycine
			L-Carnitine	Riboflavin			Guanine
			L-Lactic acid	Serotonin			Hippuric acid
			L-Threonine	Suberic acid			Hypoxanthine
			L-Tryptophan	Thymidine			Indoleacetic acid
			Lumichrome	Thymine			L-Alanine
			Methylguanidine				L-Isoleucine
			N-Acetyl-L-aspartic acid				L-Methionine
			N-acetyltryptophan				L-Phenylalanine
			Niacinamide				L-Tyrosine
			Quinaldic acid				Methyl jasmonate
			Taurine				N-Acetyl-L-alanine
			Uric acid				N-Acetylleucine
							N-Acetyl-L-phenylalanine
							N-Acetylserotonin
							Pantothenic acid
							Pyroglutamic acid
							Succinic acid
							trans-Ferulic acid
							Uracil
							Uridine
							Urocanic acid
							Xanthine
							Xanthosine

Table S13: Metabolites list link to the Upsetplot

B. Chez le modèle animal caprin

1. Contexte et objectif

Nous sommes de plus en plus sensible au bien-être des animaux et à la qualité des produits consommés. Les éleveurs ont de nos jours des dispositifs utilisés en routine tels que la rumination [155], l'activité physique ou la qualité du lait pour juger de l'état de santé de leurs animaux en plus des diagnostics vétérinaires. La métabolomique chez les animaux est un domaine émergeant fournissant de nouveaux biomarqueurs pour un diagnostic précoce, rapide et non invasif chez les animaux d'élevage [156]. Les applications de la métabolomique sur le bétail sont menées dans différents domaines tels que l'évaluation de la santé animale, le diagnostic des maladies [126], la découverte de biomarqueurs pour l'efficacité de l'alimentation [157], le potentiel de croissance et la production laitière. Des maladies telles que la cétose ou la mammite affectent les animaux laitiers et la production de lait. La cétose et la cétose subclinique sont retrouvées chez les vaches laitières après le vêlage généralement liées à un bilan énergétique négatif [158]. L'inflammation de la mammite peut être soit clinique, généralement associée à des signes cliniques, soit subclinique sans signe clinique mais avec un nombre élevé de cellules somatiques affectant la qualité du lait. La mammite est la maladie la plus importante et la plus coûteuse en production caprine laitière [159]. La mammite est une infection du pis, causée par les agents pathogènes les plus courants, les staphylocoques à coagulase négative (SNC) et Staphylococcus aureus (S. aureus). Contrairement à la mammite clinique étant facile à détecter, les animaux souffrant de mammite subclinique sont souvent difficiles à détecter faute de méthodes de diagnostic fiables. Même si certaines études ont relaté des liens entre le nombre de cellules somatiques et la mammite subclinique [160, 161], l'établissement d'un diagnostic précoce et fiable au niveau de la ferme est difficile en raison du délai entre les différentes analyses de contrôle du lait, associé à l'absence d'autres signes clairs.

Dans cette étude, nous nous intéressons aux animaux laitiers, chèvres, qui sont sujettes à ses maladies. Actuellement, il n'existe aucune étude associant des informations métabolomiques provenant de plusieurs matrices utilisant plusieurs plateformes chez les caprins en bonne santé. Ainsi, notre étude décrit une procédure opératoire standardisée (SOP) pour les matrices biologiques (lait, plasma et fèces) analysées par 3 plateformes (RMN, UHPLC(RP)-MS et UHPLC(HILIC)-MS).

2. Principaux résultats et conclusion

Les données obtenues des 3 plateformes sont fusionnées en enlevant les métabolites redondants. Ces données sont évaluées selon deux critères, le nombre de métabolites détectés par méthode d'extraction et la répétabilité de ces métabolites (CV). A l'aide de ces critères, l'extraction biphasique Plasma :MeOH :$CHCl_3$ (1 :1 :1) a été choisi pour le plasma permettant la détection de 198 métabolites. L'extraction Lait :H_2O :MeOH :$CHCl_3$ (1 :3 :4 :6) permet la détection de 221 métabolites dans le lait et 313 métabolites fécaux pour l'extraction MeOH :H_2O (4 :1). La complémentarité de ces matrices a permis de mettre en évidence 130 composés en commun pouvant être des métabolites clés pour la mise en évidence d'échanges compartimentaux. La totalité de ces métabolites détectés a permis la proposition d'un atlas métabolomique caprin qui pourrait ouvrir une nouvelle voie aux éleveurs dans le suivi du bien-être de leur troupeau.

B.1. Publication: Analytical methodology for a metabolome atlas in goat's plasma, milk and feces using ^{1}H-NMR and UHPLC-MS

metabolites

MDPI

Article

Analytical Methodology for a Metabolome Atlas of Goat's Plasma, Milk and Feces Using ^{1}H-NMR and UHPLC-HRMS

Cécile Martias [1], Julie Gatien [2], Léa Roch [2], Nadine Baroukh [1], Sylvie Mavel [1], Antoine Lefèvre [1], Frédéric Montigny [1], Laurent Schibler [2], Patrick Emond [1,3] and Lydie Nadal-Desbarats [1,*]

1 UMR 1253 iBrain, University of Tours, Inserm, 37044 Tours, France; cecile.martias@univ-tours.fr (C.M.); nadine.baroukh@univ-tours.fr (N.B.); sylvie.mavel@univ-tours.fr (S.M.); antoine.lefevre@univ-tours.fr (A.L.); frederic.montigny@univ-tours.fr (F.M.); patrick.emond@univ-tours.fr (P.E.)
2 Allice, Phenotyping Station, 37380 Nouzilly, France; julie.gatien@allice.fr (J.G.); lroch8@gmail.com (L.R.); laurent.schibler@allice.fr (L.S.)
3 CHRU Tours, Medical Biology Center, 37000 Tours, France
* Correspondence: lydie.nadal@univ-tours.fr; Tel.: +33-(0)2-4736-6164

Abstract: Metabolomics has been increasingly used in animal and food sciences. Animal health is one of the most important factor that can also alter animal integrity and welfare. Some studies have already investigated the link between health and metabolic profile of dairy animals. These studies in metabolomics often consider a single type of sample using a single analytical platform (nuclear magnetic resonance or mass spectrometry). Only few studies with multi-platform approaches are also used with a single or a multi type of sample, but they mainly consider dairy cows' metabolome although dairy goats present similar diseases, that it could be interesting to detect early to preserve animal health and milk production. This study aims to create a metabolic atlas of goat plasma, milk and feces, based on healthy animals. Our study describes a standard operating procedure for three goat matrices: blood plasma, milk, and feces using multiple platforms (NMR (^{1}H), UHPLC (RP)-MS and UHPLC (HILIC)-MS) that follows a unique sample preparation procedure for each sample type to be analyzed on multi-platforms basis. Our method was evaluated for its robustness and allowed a better characterization of goat metabolic profile in healthy conditions.

Keywords: metabolic fingerprinting; ^{1}H-NMR; LC-MS; plasma; milk; feces; goat

check for updates

Citation: Martias, C.; Gatien, J.; Roch, L.; Baroukh, N.; Mavel, S.; Lefèvre, A.; Montigny, F.; Schibler, L.; Emond, P.; Nadal-Desbarats, L. Analytical Methodology for a Metabolome Atlas of Goat's Plasma, Milk and Feces Using ^{1}H-NMR and UHPLC-HRMS. *Metabolites* **2021**, *11*, 681. https://doi.org/10.3390/metabo11100681

Academic Editor: Shozo Tomonaga

Received: 31 August 2021
Accepted: 1 October 2021
Published: 4 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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/).

1. Introduction

Over the last decades, animal health and welfare have generated a growing interest in livestock research [1]. The development of tools and identification of important new biomarkers have been a huge concern for the last twenty years to help breeders in the management of their herds. Animal health is a dominant factor influencing animal welfare, quality of products and farms' income at the same time [2]. Nowadays, several parameters, such as rumination [3], activity, milk production and quality [4], are examined in routine by some farmers to monitor health of their animals, in addition to veterinarian diagnosis for proven diseases.

The study of dairy animals' metabolome would allow better targeting biomarkers of diseases. Is nowadays well established that the metabolome represents the ultimate end-point of the biological 'omics' cascade. The metabolome is impacted by the complex interactions between the host and its microbiota and other factors such as diet, stress, gender or age. Metabolomics focuses on the wide variety of low molecular weight metabolites (<1500 Da) present in biological samples (cells, tissues or biological fluids) [5]. Metabolomics is now routinely used in biomedical and nutritional sciences [6]. Livestock metabolomics is an emerging area of metabolomics providing new biomarkers for early, rapid and non-invasive diagnosis in farm animals [7]. Applications of livestock metabolomics is conducted in

different domains such as animal welfare assessment, disease diagnosis [8], biomarker discovery for diet efficiency [9,10], growth potential and milk production.

Diseases such as ketosis or mastitis affect dairy animals and milk production. Among dairy cows after calving, ketosis and subclinical ketosis are usually found related to negative energy balance [11]. Mastitis inflammation can either be clinical and is usually associated with clinical signs or subclinical with no clinical sign but with elevated somatic cell count affecting the milk quality. One of the most important and expensive disease in dairy goat production is mastitis [12]. Mastitis is an infection of the udder, caused by the most common pathogens coagulase-negative staphylococci (CNS) and *Staphylococcus aureus* (*S. aureus*). Mastitis can appear mainly during the first third of lactation. In small ruminants like goats, the prevalence of subclinical mastitis averages 5–30% [13]. Contrary to clinical mastitis which is easy to detect, animals suffering from subclinical mastitis are often difficult to identify due to the lack of reliable diagnostic methods. Even if some studies related links between somatic cells counts and subclinical mastitis [14,15], early and reliable diagnosis establishment at farm level is difficult due to the time between the different milk control analyzes, associated to the absence of others clear signs.

However looking at the literature, very few efforts focused on metabolomics characterization of healthy animals with the aim to identifying baseline metabolite coverage in different biofluids or tissues. These baseline survey values are necessary and very often needed before biomarker studies should be carried on. Nowadays most of the metabolomics studies on dairy animals are usually done on a single biological fluid (mostly blood or milk) and using only one analytical platform (either chromatography-mass spectroscopy (MS) or nuclear magnetic resonance (NMR)). These two platforms are the most widely used technologies. In the literature, milk is one of the most frequently used fluid in dairy animals, as it is easily accessible and produced in a huge amount. Therefore, the search for biomarkers is often carried out by the analysis of milk, for health applications and for milk composition and quality for example [16–18]. However, the information often remained limited to a unique sample type. Some studies showed correlation between milk and blood which help in the detection of others biomarkers [9]. In fact, blood is a matrix commonly used to assess markers of stress in animals. An example is the measurement of cortisol hormone a stress indicator. Cortisol metabolites in blood have been shown to be directly correlated to fecal cortisol metabolites in stressed animals [19]. In most cases, feces can be collected without invasive procedure and in large quantities. Most of the time when multiple fluids are used to describe biological systems, the information remained limited to a unique analytical platform or in contrary when investigating metabolome combining multi-analytical platforms, studies have focused on a single biofluid. However Xu et al. showed that combining data from milk and plasma using ^{1}H-NMR and UHPLC-MS platforms permitted to understand link between energy balance and metabolic profiles [9].

Currently, there is no study associating metabolomics information from multiple matrices using multiple platforms for goat metabolome atlas in healthy condition to get the widest metabolite coverage to describe an individual. Thus the present work describes a unique standard operating procedure by biological matrices (milk, plasma and feces) to be analyzed by different platforms (UHPLC (RP and HILIC)-MS and NMR). The proposal of a metabolomics atlas for goat could open a new path for breeders in monitoring the welfare of their herd.

2. Results

In order to depict as closely as possible the metabolomics coverage of healthy goat, we looked for the best preparation protocol to describe an individual with the largest metabolic coverage meaning using three biological matrices (plasma, milk and feces) easy to harvest and different analytical platforms (^{1}H-NMR, HILIC-MS and RPLC-MS). The aim of this study was to obtain a single preparation procedure able to be explored by ^{1}H-NMR, (ESI+/ESI-)RP-LC-MS and (ESI+/ESI-)HILIC-LC-MS platforms. The metabolic

information collected may have the best reproducibility in the extraction protocol, the largest metabolic coverage without any a priori in the selected metabolites.

2.1. Optimization of Extraction Protocol

2.1.1. Plasma

A total of three preparation protocols were tested with plasma samples (see materials and methods Figure S1a), two biphasic extraction using two ratios of solvents, plasma:MeOH:$CHCl_3$ (1:1:1) and (1:1.5:2.5) ratios and one monophasic extraction with MeOH within 1:8 proportions. All extraction protocols succeeded to provide wide number of metabolites as shown in Figure 1a. Once the chemical redundancy eliminated, the biphasic extraction protocols found 229 compounds all platforms together with 198 metabolites showing a CV <30% in plasma:MeOH:$CHCl_3$ (1:1:1) and 230 compounds were detected in plasma:MeOH:$CHCl_3$ (1:1.5:2.5) including 184 metabolites showing a CV <30%. The monophasic extraction protocol allows to detect 237 compounds all platforms together with 169 metabolites showing a CV <30%. (see Supplementary Data Table S1). When regarding the number of compounds extracted by only one extraction protocol (Figure 1a), 19 compounds were specific to plasma:MeOH:$CHCl_3$ at (1:1:1), nine to plasma: MeOH:$CHCl_3$ at (1:1.5:2.5) and 25 were specific to plasma:MeOH (1:8).

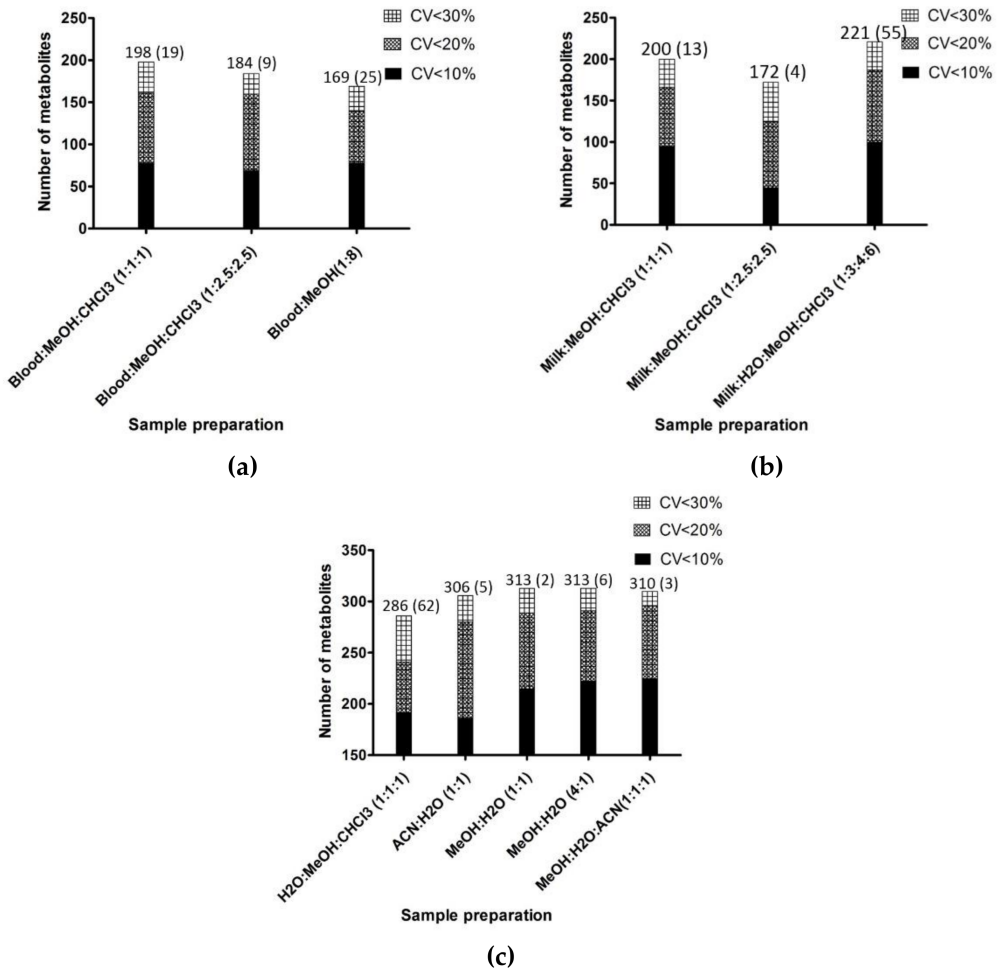


Figure 1. Histograms providing the total number of metabolites extracted (n = 10) for each extraction protocol and respectively their reproducibility evaluated by CV (Coefficient of Variation) in plasma (**a**), milk (**b**) and feces (**c**). In bracket, are mentioned the numbers of metabolites analyzed by only one extraction protocol.

Comparing the three preparations, in term of number of metabolite extracted, the plasma:MeOH:$CHCl_3$ (1:1:1) preparation was the most suitable preparation.

2.1.2. Milk

The sample used for this step of procedure standardization was a pool of 10 females, with no particular features/traits (including lactation phase, parity, milk production for example). The aim was to get the maximum range of metabolites in the different biofluids, as the subsequent studies for biomarkers research will be carried on several diseases occurring at various stages of lactation.

Highly concentrated in lipids, milk was tested with 3 biphasic preparation protocols (see materials and methods Figure S1b). Once the chemical redundancy eliminated, the milk:H_2O:MeOH:$CHCl_3$ (1:3:4:6) preparations gave the largest number of detected metabolites over 263 compounds were detectable all platforms associated including 221 metabolites having a CV <30% (see Supplementary Data Table S2). When considering the number of specific extracted compounds for each preparation protocol (Figure 1b), milk:H_2O:MeOH:$CHCl_3$ (1:3:4:6) preparation gave the largest specific number of annotated metabolites. In fact this preparation gave 55 specific metabolites compared to 13 unique for milk:MeOH:$CHCl_3$ (1:1:1) and four unique for milk:MeOH:$CHCl_3$ (1:1.5:2.5).

Comparing the three preparations, in term of number of metabolites extracted and number of unique compounds, the milk:H_2O:MeOH:$CHCl_3$ (1:3:4:6) preparation was the most suitable one.

2.1.3. Feces

Five extraction protocols (see Materials and Methods Figure S1c) were tested on freeze-dried feces. Once the chemical redundancy eliminated, the MeOH:H_2O (1:1) and (4:1) preparations gave the largest number of detected metabolites over 313 compounds were detectable all platforms associated (see Supplementary Data Table S3). Regarding the number of specific compounds for these two preparations (Figure 1c), the MeOH:H_2O (4:1) preparation gave the largest specific number of detected metabolites. In fact this preparation gave six specific metabolites compared to two for MeOH:H_2O (1:1). Comparing the five preparations, in terms of number of metabolite extracted, the MeOH:H_2O(4:1) preparation was the most suitable preparation.

2.2. Matrix Complementarity

To assess the complementarity of the three matrices a Venn diagram (Figure 2) was created. This diagram provides an efficient way to visualize the numbers and names (Table S4) of metabolites shared by the different matrices. The metabolites detected were separated into seven different chemical classes. A pie chart was obtained with metabolites found specifically in each Venn section. The seven classes were found in milk with a majority of sugars (42%). These seven categories were also found in feces, but aromatic compounds were predominant in feces (34%). In plasma, the category of sugars was not detected specifically and aromatic compounds and amino acids (respectively 36% and 22%) were the two main classes found.

There were between 14 to 46 common metabolites between two matrices (Table S4) and 130 common metabolites to all three. These 130 common compounds are listed in Table S4. Most of them are amino-acids-peptides derivatives (33%) or aromatic compounds-vitamins-amines (24%).

When comparing with the literature (Table 1), we provided with a confidence level 2 [20] a lot of metabolites implicated in metabolic disease as ketosis or mastitis. Our study highlighted 373 robust molecules describing a healthy individual using 3 matrices (Table S4). These 373 metabolites allow to establish a healthy metabo-atlas as a benchmark to follow goat in herd. Table 1 resumed biomarkers in clinical and subclinical mastisis and in ketosis found in the literature according to the explored matrix. In Table 1, we provided the different matrices where we found the biomarkers described in the literature. As mentioned, a

lot of metabolites are present in more than one matrix allowing the health status follow-up simply and easily independently of the biological compartment according to the practice in breeding.

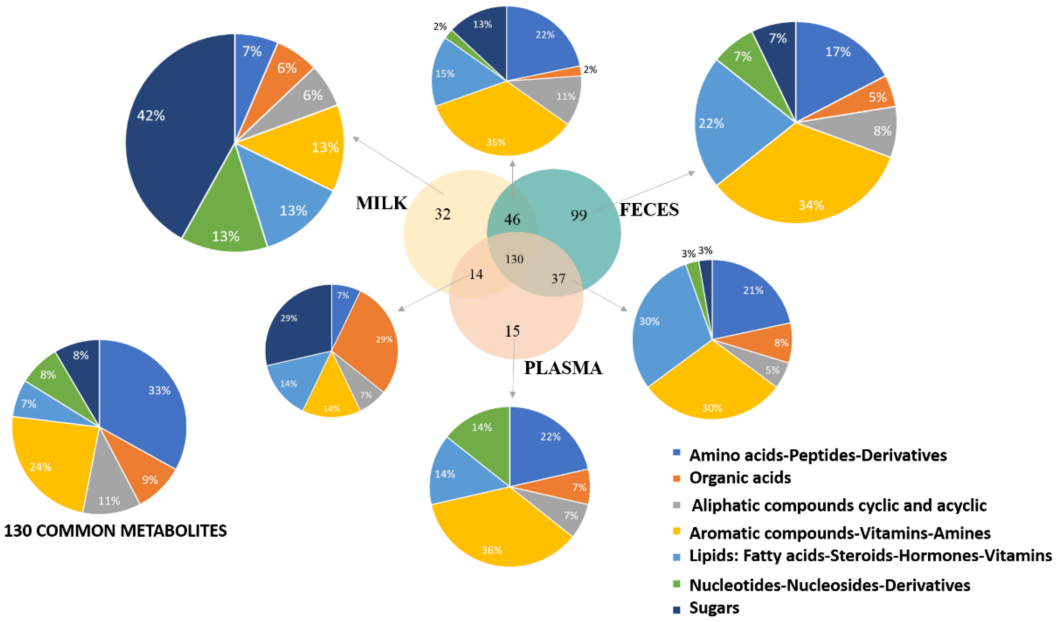


Figure 2. The complementarity of metabolite coverage across the different biological matrices is visualized on a Venn diagram. Pie charts with chemical classes of metabolites found in each Venn section are also displayed.

Table 1. Metabolites of interest differentially present in different matrices according to the literature and found in one, two or the three matrices exploited in our study with a CV <30% according to the best preparation procedure. (↗ increased, ↘ decreased, NF = not found).

Disease/Status	**Mastitis Xi et al. (2017) [8]**			**Ketosis Zhang G. et al. 2021 [11]**		**Mastitis Sundekilde UK et al. 2013 [21]**	**Healthy Martias C. et al. The Present Study**
	Milk Cow			**Urinary Cow**	**Urinary Cow**	**Milk Cow**	**Plasma, Milk and Feces Goat**
	Significantly Different Between			**Metabolite Alteration**		**Strong Association With Stomatic Cell Count (SCC)**	**Matrices Where Metabolites Are Present**
	Healthy and Clinical	**Healthy and Subclinical**	**Subclinical and Clinical**	**Preceding Ketosis**	**During Ketosis**		
3-Hydroxy-butyric acid					↗	↗	Milk, Plasma
4-Hydroxy-phenyl-lactate	↘		↘				Feces, Plasma
4-Hydroxy-phenyl-pyruvate	↘						Feces
5-hydroxy-L-tryptophane		↘					Feces
Acetic acid						↗	Milk, Feces, Plasma
Acetoacetic acid					↗		Milk, Plasma
L-Arginine	↗	↗					Milk, Feces, Plasma
Ascorbic acid				↗	↗		NF
Benzoic acid		↘					Feces
Butyrate						↗	Feces, Plasma

Table 1. *Cont.*

Disease/Status	Mastitis Xi et al. (2017) [8]			Ketosis Zhang G. et al. 2021 [11]		Mastitis Sundekilde UK et al. 2013 [21]	Healthy Martias C. et al. The Present Study
	Milk Cow			Urinary Cow	Urinary Cow	Milk Cow	Plasma, Milk and Feces Goat
	Significantly Different Between			Metabolite Alteration		Strong Association With Stomatic Cell Count (SCC)	Matrices Where Metabolites Are Present
	Healthy and Clinical	Healthy and Sub-clinical	Subclinical and Clinical	Preceding Ketosis	During Ketosis		
L-Carnitine	↘	↘					Milk, Feces, Plasma
Citrate	↘						Milk, Plasma
Dimethylglycine		↘					Plasma
D-Lactose			↘			↘	Milk, Plasma
Dopamine	↗						Feces
Fumarate						↘	Milk, feces
Glucose	↘						Milk, Feces, Plasma
Glucose-1-phosphate	↘	↘	↘				Milk
Guanosine monophos-phate	↘						Plasma
Hippurate	↘					↘	Milk, Feces, Plasma
Isocitric acid				↗			Feces, Plasma
L-Isoleucine	↗					↗	Milk, Feces, Plasma
L-Lactic acid					↗	↗	Milk, Feces, Plasma
Malate	↘						Milk, Feces, Plasma
Oxoglutarate			↘				Milk
Phosphocholine	↘		↘				Milk, Feces, Plasma
L-Proline	↗						Milk, Feces, Plasma
Riboflavin			↘				Milk, Feces
sn-Glycero-3-phosphocholine	↘						Milk, Feces, Plasma
Uridine	↘						Milk, Feces, Plasma
L-Valine	↗						Milk, Feces, Plasma

3. Discussion

For breeders, it's very crucial to monitor the health status of their herd but it is sometimes difficult to identify signs of diseases, and therefore in which biological compartment the metabolic modifications happen. The monitoring of the metabolic status using a metabolomic atlas from multimatrices is currently lacking to anticipate and explain a metabolic dysregulation. This study focused on the metabolomics characterization of healthy goats in aim to describe as exhaustively as possible the baseline metabolite coverage in order to create a metabo-atlas. This baseline survey is necessary and very often needed before biomarker studies should be carried on. Three biological matrices were used (milk, plasma and feces) using the complementarity of three analytical platforms, ^{1}H-NMR and UHPLC(HILIC)-MS and UHPLC(RP)-MS. Our metabolic coverage (metaboatlas) proposes only metabolites with a CV<30% meaning very reliable compounds found in each healthy sample (n = 10) with a validation step of the best operating procedure to prepare them. Moreover, the redundancy of the metabolites in different matrices may provide further information about exchanges between compartments but also to investigate metabolites evolution during pathologies.

Several papers have reported that goat milk is usually prepared with a biphasic protocol [22,23] or with ultrafiltration [16,18]. In our study, testing three biphasic protocol, Milk:H_2O:MeOH:$CHCl_3$ (1:3:4:6) was the most suitable protocol. According to Caboni et al. [22], goat's and sheep's milk have a metabolic profile by untargeted GC-MS dependent on the species. This study is very informative in term of milk origin and quality of transformed products but did not provide the metabolic baseline of a normal milk metabolome to follow the animal health. They identified 38 low molecular weight metabolites mainly amino acids, short chain hydroxylated carboxylic acids, organic acids, polyols and sugars. In our study and according to standardized operating procedure we developed, we identified 221 low molecular weight metabolites with 55 exclusively in this biofluid compared to plasma and feces. Plasma and milk shared (130 + 14) 144 metabolites (Table S4). This observation can be explained by the exchange between vascular compartment and glandular tissue. Metabolomics studies usually show the complementarity between milk and plasma to find prognostic biomarker for risk of ketosis or to evaluate the relationship between energy balance and metabolomics profiles in early lactation [9,16]. In milk, for subclinical ketotic conditions, ketone bodies (β-hydroxybutyrate, acetoacetate and acetone) are already accepted as reliable biomarkers [24,25]. In our study, 3-hydroxybutyric acid and acetoacetate are described in milk and plasma and acetone in plasma. Caboni et al. [22] described increased pyruvate and lactate associated with somatic cell count in milk. In our study pyruvate is also present in plasma and lactic acid in the three matrices explored. Xi et al. [8] showed, in an untargeted milk metabolomics study by UPLC-QTOF, biomarkers to differentiate healthy, clinical and subclinical mastitis. As mentioned in Table 1, most of these metabolites are found in feces and/or plasma. Zhang et al. [11] demonstrated that urinary metabolite can highlight ketosis. However, urine is problematic to collect because goat urinary cycle is unpredictable. Monitoring these metabolites can be done through other fluids (like milk, feces or plasma) (Table 1).

According to the literature, plasma can be prepared using ultrafiltration in order to remove macromolecules [9,16] or with a biphasic method for lipid studies [16], only the lipid phase is analyzed. Plasma can be also prepared by a protein precipitation with cold MeOH [26]. In our study, testing two biphasic protocols and one monophasic extraction, plasma:MeOH:$CHCl_3$ (1:1:1) was the most suitable protocol with 198 molecules with a CV <30%. Enjalbert et al. [25] showed an increased in the ketone bodies in milk and plasma in true positive cows during subclinical ketosis with a good correlation coefficient between blood and milk for acetone (0.96), acetoacetate (0.74) and BHBA (0.66). This observation confirms that it is possible to follow the health status in different biofluids. Xu et al. [9] showed a relation between energy balance and metabolic profiles in plasma and milk. Some metabolites were found in both plasma and milk but some were found correlated

to energy balance only in plasma or in milk. This observation is of importance because it clearly highlights the need to explore several matrices instead of only one matrix.

Sun et al. [26] investigated by GC-TOF/MS the simultaneously responses of four biofluids (serum, milk, rumen fluid and urine) to different forage diets to establish the correlation among biofluids and the forage quality related to milk production and quality in dairy cows. According to the metabolic profiles of the four matrices, authors identified key different metabolic pathways between the two diet groups. Metabolites were identified 165 (rumen fluid), 195 (milk), 218 (serum) and 156 (urine) with 29 common between the four matrices.

In the literature, fecal material could be first freeze-dried and after metabolites could be extracted with EtOH/H_2O [19] or MeOH followed by a SPE step [27]. In our study, testing one biphasic protocol and four monophasic protocols, feces:H_2O (4:1) is the most suitable protocol with 313 molecules with a CV <30%. Currently, very few studies have explored fecal metabolites. Scherpenhuizen et al. [19] explored fecal cortisol metabolite concentrations in sheep for noninvasive quantitative analysis of physiological stress in sheep. Dulude-de Broin et al. [27] evaluated faecal metabolites and cortisol in hair as biomarkers of the hypothalamic–pituitary–adrenal -axis activity in captive mountain goats. The cortisol and some glucocorticoid metabolites as corticosterone were found in our study in feces and in milk matrices.

In this publication, the authors deliberately choose to not detail concentration values of each metabolite found in this goat metabo-atlas. Even if it is possible with NMR and usually done, quantification would have been biased by the fact that samples used for this study have been collected in a single farm. It is clearly known that environmental and farm conditions [11], genetic and breed [18], as the most important effects, could hugely impact values of concentration.

Our study makes it possible to cover a wide range of metabolites. A change in this metabolic profile can allow early detection of metabolic dysregulation that can indicate a disease. The next work will be the establishment of correlation between these metabolic dysregulations and diseases. This step will be of a great interest for livestock production, as it will allow a large screening of the metabolism of the animals for an early detection of diseases. Breeders need tools to help them in the daily management of their herd, especially to monitor health status, as some diseases could not show clear clinical signs (subclinical forms) and provoke negative consequences on animal health and products quality.

4. Materials and Methods

4.1. Sample Collection

For each biofluid/matrix, samples were collected on 10 Alpine goats from a French commercial farm, and pooled, then stored at −80 °C until analysis was performed. Approval by the local ethical committee was not necessary for blood, milk and feces sampling in a commercial farm at the time of the experiment. Samples were collected by staff with skills to conceive and perform experimental procedures from an approved establishment (Allice experimental facilities, n° B-37-175-5), and according to ethical and welfare guidelines usually applied in our experiments.

4.1.1. Plasma

Blood was collected on heparin tubes then centrifuged (15 min, +4 °C, 15,000 g). Pool was formed with plasma supernatant.

4.1.2. Milk

Milk samples were collected in sterilized tubes through manual milking. Pool was constituted with fresh whole milk.

4.1.3. Feces

Fecal samples were collected in sterilized tubes. Samples were individually freeze-dried then pooled and mixed.

4.2. Sample Preparation

Sigma Aldrich (Saint-Quentin Fallavier, France) supplied all chemicals.

4.2.1. Plasma

For each extraction protocol, the extraction was repeated 10 times. Three extraction procedures were tested either on 100 uL of plasma, plasma:MeOH:$CHCl_3$ (1:1:1) and plasma:MeOH:$CHCl_3$ (1:1.5:2.5) or on 200 uL of plasma, plasma:MeOH (1:8). Samples were vortexed after addition of each chemical solvent, kept at −20 °C during 30 min, then centrifuged (10 min, 4 °C, 15,000 g). Supernatant were collected in a glass tube and divided into 2 aliquots (one for the NMR analysis and the other for MS analysis) for further solvent evaporation in a SpeedVac (ThermoFisher, Villebon sur Yvette, France). A graphical visualization of protocols of sample preparation is given in Figure S1a.

4.2.2. Milk

For each extraction protocol, the extraction was repeated 10 times. Three extractions procedures were tested either on 100uL of milk:MeOH:$CHCl_3$ (1:1:1) and milk:MeOH:$CHCl_3$ (1:1.5:2.5) or 200 uL of milk, milk:H_2O:MeOH:$CHCl_3$ (1:3:4:6). Samples were vortexed after addition of each chemical, kept at −20 °C during 30 min, then centrifuged (10 min, 4 °C, 15,000 g). Supernatant were collected in a glass tube and divided into 2 aliquots (one for the NMR analysis and the other for MS analysis) for further solvent evaporation in a SpeedVac (ThermoFisher). A graphical visualization of protocols of sample preparation is given in Figure S1b.

4.2.3. Feces

Fecal samples were freeze-dried (FreeZone®4.5 L, Labconco, Kansan City, MO, USA) at −107 °C, 0.2 mbar for 24 h and then pooled and mixed as previously described by Martias et al. [28].

For each extraction protocol, the extraction was repeated 10 times. Five extraction procedures were tested on 50 mg for each of a pool of dried feces: H_2O:MeOH:$CHCl_3$ (1:1:1), ACN:H_2O (1:1), MeOH:H_2O (1:1), MeOH:H_2O (4:1), MeOH:H_2O:ACN (1:1:1). Samples were vortexed during 10 min and centrifuged (10 min at 4 °C, 15,000 g). Supernatants were collected in 2 aliquots (one for the NMR analysis and the other for MS analysis) for further solvent evaporation in a SpeedVac (ThermoFisher). A graphical visualization of all sample preparation protocols is given in Figure S1c.

For the MS analysis, dried-residues were dissolved in 150 µL of ACN:H_2O (4:1). 75 µL were used for HILIC and the remaining phase were evaporated in a Speedvac (ThermoFisher). Then, dried-residues was dissolved in 75µL of MeOH:H_2O (1:9) for RP-LC. For the ^{1}H-NMR analysis, dried-residues were dissolved in 200 µL of a deuterated buffer (0.2 M potassium phosphate buffered deuterium oxide (pH = 7.44 ± 0.5) and 10 µL of deuterium oxide (D_2O) with external reference [3-trimethylsilylpropionic acid (TSP) at 3.2 mM]).

4.3. Data Acquisition

4.3.1. UHPLC-MS

As previously described [29], LC-HRMS analysis was performed on an UPLC Ultimate WPS-3000 system (Dionex, Sunnyvale, CA, USA), coupled to a QExactive-orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA).

The chromatography system was equipped separately with two columns: Reverse Phase Liquid Chromatography (RP-LC) Kinetex XB-C18 (1.7 µm 100 A 150 × 2.1 mm, Phenomenex, Torrance, CA, USA) and a hydrophilic interaction liquid chromatography

(HILIC) Cortecs (unbonded silica 1.6 µm 100 A 150 × 2.1 mm, Waters, Dublin, Ireland) as previously described [28].

A head electrospray ionization (HESI) source was used for both chromatography system, operated in positive (ESI+) and negative (ESI−) electrospray ionization modes (one run for each mode), as previously described [28]. Detection was performed with a full-scan acquisition at 70,000 resolution (m/z = 200) which ranged from 58.0 to 870.0 m/z, with an automatic gain control (AGC) target of 105 charges and a maximum injection time (IT) of 250 ms as previously described [28]. Xcalibur 2.2 software (Thermo Fisher Scientific, San Jose, CA, USA) controlled the system.

4.3.2. ^{1}H-NMR

As previously described [28], ^{1}H-NMR spectra were acquired at 298 K on an AVANCE III HD 600 MHz system (Bruker Biospin, Karlsruhe, Germany). ^{1}H-NMR spectra were recorded with «noesypr1d» pulse sequence with a relaxation delay of 20 s, on a spectral width of 12 ppm, a time domain of 64 k points, an acquisition time of 4.55 s, with 64 FIDs and 8 dummy scans.

4.4. Data Processing

4.4.1. UHPLC-MS

Data processing was carried out with Xcalibur 2.2 software (Thermo Fisher Scientific, San Jose, CA, USA) as previously described [30]. A library of 495 standard metabolites (Mass Spectroscopy Metabolite Library of Standards MSMLS® (IROA Technologies™), Sea Girt, NJ, USA) was analyzed with the same gradient of mobile phases and in the same conditions. The 495 molecules were divided into 42 pools of 12 molecules dispatched on a 96-well plate. The pools were made with molecules having different masses. Each pool is injected in full-scan acquisition mode then injected a second time in MS^2 (PRM –parallel reaction monitoring- mode). For each metabolite, the high-resolution mass and retention time parameters have been integrated into the "Thermo Xcalibur processing setup" (ThermoFisher Scientific) which is an automatic integration software. Once this list of metabolites has been produced, it is used to find and integrate each of these metabolites in the samples with a precision on the mass of 5 ppm (Mode +) and 10 ppm (Mode −) around the theoretical mass. A window of 12 s relative to the retention time observed for the standards was used in order to select the metabolites to be integrated.

All samples were then processed with this integration method to create a result file using "Thermo Xcalibur Quan browser". All the integrations created in this way were checked and then exported to an .xls file containing the areas of each metabolite. Each peak area was normalized to the total of peak areas of interest.

4.4.2. ^{1}H-NMR

As previously described [28], spectra were processed using TopSpin 3.6 software (Bruker Biospin GmbH, Rheinstetten, Germany) and were reduced to buckets of variable width using AMIX software (Analysis of MIXture, version 3.2, Bruker Biospin GmbH, Rheinstetten, Germany). Bucket identification was done using ChenomX software (NMR Suite 7.7, ChenomX version 8.6, Edmonton, Canada) and literature [31]

4.5. Data Fusion

Once the pre-processing of the spectral data was done, Xcalibur and ChenomX generated respectively a compound list for LC-MS data and for NMR data. Generated names by ChenomX and IROA® database are loaded in metaboanalyst (https://www.metaboanalyst.ca/ accessed on 10 of June 2021) in order to have a generic name for all platforms. Thanks to this homogeneity of names, KEGG and HMDB numbers were attributed for each list of metabolites coming from the different platforms. Thus, NMR list, RP-MS list and HILIC-MS list could be merged. When metabolites were detected by several platforms, redundant metabolites (generic name based on their KEGG, HMDB or PubChem numbers) were

deleted based on their signal variation coefficient (CV) comparing the different platforms. To eliminate the redundancy, the metabolites with the largest CV was deleted and the information coming from the platform having the most reproducible signal was kept.

4.6. Data Analysis

The data generated from each matrix, for each extraction protocol and from each platform were gathered in an .xls file with generic name obtained with metaboanalyst (https://www.metaboanalyst.ca/ accessed on 10 June 2021) as mentioned in data fusion. The number of metabolites, unique metabolite and each coefficient of variation (CV%) were calculated in order to validate the reproducibility of the workflow and to choose the best extraction protocol.

5. Conclusions

This study presents the first metabolomics fingerprinting strategy for multimatrices (plasma, milk and feces) analysis using a multiplatform approach ((NMR (^{1}H), RP and HILIC-MS) on goats. We validate a unique preparation by biological matrix to be exploitable by NMR and LC-MS. This approach makes it possible to see the complementarity between the matrices and to consider the permeation of metabolites across biological compartments. It will therefore allow a broad description of the goat metabolome. Thanks to the metabolic monitoring of a healthy individual, we will be able to demonstrate a dysregulation the goat's health status at an early stage.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/metabo11100681/s1. Figure S1: Schematic presentation of the preparations procedures applied on the pool of plasma (a), milk (b) and feces (c) for the selection of the optimum protocol for the maximum metabolome coverage by UHPLC-MS and 1H-NMR., Table S1: Number of extracted metabolites and coefficient of validation in blood for the different extraction methods using 1H-NMR and UHPLC-MS, Table S2: Number of extracted metabolites and coefficient of validation in milk for the different extraction methods using 1H-NMR and UHPLC-MS, Table S3: Number of extracted metabolites and coefficient of validation in fecal for the different extraction methods using 1H-NMR and UHPLC-MS, Table S4: Metabolites extracted from each biological matrix according to the best preparation procedure, with a CV <30%, Table S5: Metabolites detected by platform in plasma, milk and feces. In bold the platform where the metabolite was annotated with the lowest CV (best reproducibility).

Author Contributions: Conceptualization, P.E., L.N.-D., J.G. and C.M.; methodology, C.M. and L.R.; software, A.L. and F.M.; validation, S.M. and J.G.; formal analysis, C.M., L.R. and N.B.; investigation, C.M.; resources, L.S., P.E., L.N.-D. and A.L.; writing—original draft preparation, C.M. and L.N.-D.; writing—review and editing, N.B. and P.E.; visualization, C.M. and N.B.; supervision, P.E. and L.N.-D.; project administration, P.E. and L.S., INSERM; funding acquisition, P.E. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the "Region Centre Val de Loire" project PhenoMet, INSERM U1253 iBrain and the University of Tours. Cécile MARTIAS was the recipient of a PhD fellowship from the Region Centre Val de Loire.

Institutional Review Board Statement: Ethical review and approval were waived for this study because approval by the local ethical committee was not necessary for blood, milk and feces sampling in a commercial farm at the time of the experiment (meaning before March 2020 in France). The only requirement was that the sample collections should be done by experienced staff from an approved establishment. In this study, samples were collected by staff with skills to conceive and perform experimental procedures from an approved establishment: Allice experimental facilities, n° B-37-175-5, and according to ethical and welfare guidelines usually applied in our experiments.

Informed Consent Statement: Not applicable because this study does not involve humans.

Data Availability Statement: Data are contained on the supplementary material.

Acknowledgments: We would like to acknowledge the Metabolomics Core Facility "PST-ASB" of the University of Tours—France and the European Regional Development Fund 1088 (FEDER). This study was funded by the «Region Centre Val de Loire» project PhenoMet, INSERM U1253 iBrain and the University of Tours.

Conflicts of Interest: Authors declare no conflict of interest.

References

1. Fontanesi, L. Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. *Anim. Front.* **2016**, *6*, 73–79. [CrossRef]
2. Funeshima, N.; Miura, R.; Katoh, T.; Yaginuma, H.; Kitou, T.; Yoshimura, I.; Konda, K.; Hamano, S.; Shirasuna, K. Metabolomic profiles of plasma and uterine luminal fluids from healthy and repeat breeder Holstein cows. *BMC Vet. Res.* **2021**, *17*, 54. [CrossRef]
3. Paudyal, S.; Maunsell, F.P.; Richeson, J.T.; Risco, C.A.; Donovan, D.A.; Pinedo, P.J. Rumination time and monitoring of health disorders during early lactation. *Animal* **2018**, *12*, 1484–1492. [CrossRef]
4. Salama, A.A.K.; Contreras-Jodar, A.; Love, S.; Mehaba, N.; Such, X.; Caja, G. Milk yield, milk composition, and milk metabolomics of dairy goats intramammary-challenged with lipopolysaccharide under heat stress conditions. *Sci. Rep.* **2020**, *10*, 5055. [CrossRef] [PubMed]
5. Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R.N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18*, 1157–1161. [CrossRef] [PubMed]
6. Wishart, D.S. Metabolomics: Applications to food science and nutrition research. *Trends Food Sci. Technol.* **2008**, *19*, 482–493. [CrossRef]
7. Goldansaz, S.A.; Guo, A.C.; Sajed, T.; Steele, M.A.; Plastow, G.S.; Wishart, D.S. Livestock metabolomics and the livestock metabolome: A systematic review. *PLoS ONE* **2017**, *12*, e0177675. [CrossRef] [PubMed]
8. Xi, X.; Kwok, L.Y.; Wang, Y.; Ma, C.; Mi, Z.; Zhang, H. Ultra-performance liquid chromatography-quadrupole-time of flight mass spectrometry MS(E)-based untargeted milk metabolomics in dairy cows with subclinical or clinical mastitis. *J. Dairy Sci.* **2017**, *100*, 4884–4896. [CrossRef]
9. Xu, W.; Vervoort, J.; Saccenti, E.; Kemp, B.; van Hoeij, R.J.; van Knegsel, A.T.M. Relationship between energy balance and metabolic profiles in plasma and milk of dairy cows in early lactation. *J. Dairy Sci.* **2020**, *103*, 4795–4805. [CrossRef]
10. Kim, S.; Kim, J.; Yun, E.J.; Kim, K.H. Food metabolomics: From farm to human. *Curr. Opin. Biotechnol.* **2016**, *37*, 16–23. [CrossRef] [PubMed]
11. Zhang, G.; Mandal, R.; Wishart, D.S.; Ametaj, B.N. A Multi-Platform Metabolomics Approach Identifies Urinary Metabolite Signatures That Differentiate Ketotic from Healthy Dairy Cows. *Front. Vet. Sci.* **2021**, *8*, 595983. [CrossRef]
12. Gabli, Z.; Djerrou, Z.; Gabli, A.E.; Bensalem, M. Prevalence of mastitis in dairy goat farms in Eastern Algeria. *Vet. World* **2019**, *12*, 1563–1572. [CrossRef]
13. Contreras, A.; Sierra, D.; Sánchez, A.; Corrales, J.C.; Marco, J.C.; Paape, M.J.; Gonzalo, C. Mastitis in small ruminants. *Small Rumin. Res.* **2007**, *68*, 145–153. [CrossRef]
14. Hussein, H.A.; Fouad, M.T.; Abd El-Razik, K.A.; Abo El-Maaty, A.M.; D'Ambrosio, C.; Scaloni, A.; Gomaa, A.M. Study on prevalence and bacterial etiology of mastitis, and effects of subclinical mastitis and stage of lactation on SCC in dairy goats in Egypt. *Trop. Anim. Health Prod.* **2020**, *52*, 3091–3097. [CrossRef]
15. Ylva Persson, I.O. Direct and indirect measurement of somatic cell count as indicator of intramammary infection in dairy goats. *Acta Vet. Scand.* **2011**, *53*, 15. [CrossRef]
16. Klein, M.S.; Buttchereit, N.; Miemczyk, S.P.; Immervoll, A.K.; Louis, C.; Wiedemann, S.; Junge, W.; Thaller, G.; Oefner, P.J.; Gronwald, W. NMR metabolomic analysis of dairy cows reveals milk glycerophosphocholine to phosphocholine ratio as prognostic biomarker for risk of ketosis. *J. Proteome Res.* **2012**, *11*, 1373–1381. [CrossRef] [PubMed]
17. O'Callaghan, T.F.; Vázquez-Fresno, R.; Serra-Cayuela, A.; Dong, E.; Mandal, R.; Hennessy, D.; McAuliffe, S.; Dillon, P.; Wishart, D.S.; Stanton, C.; et al. Pasture Feeding Changes the Bovine Rumen and Milk Metabolome. *Metabolites* **2018**, *8*, 27. [CrossRef] [PubMed]
18. Yang, Y.; Zheng, N.; Zhao, X.; Zhang, Y.; Han, R.; Yang, J.; Zhao, S.; Li, S.; Guo, T.; Zang, C.; et al. Metabolomic biomarkers identify differences in milk produced by Holstein cows and other minor dairy animals. *J. Proteom.* **2016**, *136*, 174–182. [CrossRef]
19. Scherpenhuizen, J.M.; Narayan, E.J.; Quinn, J.C. Timed environmental exposure indicates sample stability for reliable noninvasive measurement of fecal cortisol metabolite concentrations in sheep. *Domest. Anim. Endocrinol.* **2020**, *72*, 106423. [CrossRef]
20. Rathahao-Paris, E.; Alves, S.; Junot, C.; Tabet, J.-C. High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics* **2015**, *12*, 10. [CrossRef]
21. Sundekilde, U.K.; Poulsen, N.A.; Larsen, L.B.; Bertram, H.C. Nuclear magnetic resonance metabonomics reveals strong association between milk metabolites and somatic cell count in bovine milk. *J. Dairy Sci.* **2013**, *96*, 290–299. [CrossRef]
22. Caboni, P.; Murgia, A.; Porcu, A.; Manis, C.; Ibba, I.; Contu, M.; Scano, P. A metabolomics comparison between sheep's and goat's milk. *Food Res. Int.* **2019**, *119*, 869–875. [CrossRef]

23. Palma, M.; Hernandez-Castellano, L.E.; Castro, N.; Arguello, A.; Capote, J.; Matzapetakis, M.; de Almeida, A.M. NMR-metabolomics profiling of mammary gland secretory tissue and milk serum in two goat breeds with different levels of tolerance to seasonal weight loss. *Mol. Biosyst.* **2016**, *12*, 2094–2107. [CrossRef] [PubMed]
24. Geishauser, T.; Leslie, K.; Tenhag, J.; Bashiri, A. Evaluation of Eight Cow-Side Ketone Tests in Milk for Detection of Subclinical Ketosis in Dairy Cows. *J. Dairy Sci.* **2000**, *83*, 296–299. [CrossRef]
25. Enjalbert, F.; Nicot, M.C.; Bayourthe, C.; Moncoulon, R. Ketone Bodies in Milk and Blood of Dairy Cows: Relationship between Concentrations and Utilization for Detection of Subclinical Ketosis. *J. Dairy Sci.* **2001**, *84*, 583–589. [CrossRef]
26. Sun, H.Z.; Wang, D.M.; Wang, B.; Wang, J.K.; Liu, H.Y.; Guan, L.L.; Liu, J.X. Metabolomics of four biofluids from dairy cows: Potential biomarkers for milk production and quality. *J. Proteome Res.* **2015**, *14*, 1287–1298. [CrossRef] [PubMed]
27. Dulude-de Broin, F.; Cote, S.D.; Whiteside, D.P.; Mastromonaco, G.F. Faecal metabolites and hair cortisol as biological markers of HPA-axis activity in the Rocky mountain goat. *Gen. Comp. Endocrinol.* **2019**, *280*, 147–157. [CrossRef] [PubMed]
28. Martias, C.; Baroukh, N.; Mavel, S.; Blasco, H.; Lefèvre, A.; Roch, L.; Montigny, F.; Gatien, J.; Schibler, L.; Dufour-Rainfray, D.; et al. Optimization of Sample Preparation for Metabolomics Exploration of Urine, Feces, Blood and Saliva in Humans Using Combined NMR and UHPLC-HRMS Platforms. *Molecules* **2021**, *26*, 4111. [CrossRef]
29. Dieme, B.; Lefevre, A.; Nadal-Desbarats, L.; Galineau, L.; Madji Hounoum, B.; Montigny, F.; Blasco, H.; Andres, C.R.; Emond, P.; Mavel, S. Workflow methodology for rat brain metabolome exploration using NMR, LC-MS and GC-MS analytical platforms. *J. Pharm. Biomed. Anal.* **2017**, *142*, 270–278. [CrossRef]
30. Bitar, T.; Mavel, S.; Emond, P.; Nadal-Desbarats, L.; Lefèvre, A.; Mattar, H.; Soufia, M.; Blasco, H.; Vourc'H, P.; Hleihel, W.; et al. Identification of metabolic pathway disturbances using multimodal metabolomics in autistic disorders in a Middle Eastern population. *J. Pharm. Biomed. Anal.* **2018**, *152*, 57–65. [CrossRef]
31. Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A.C.; Wilson, M.R.; Knox, C.; Bjorndahl, T.C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; et al. The human urine metabolome. *PLoS ONE* **2013**, *8*, e73076. [CrossRef] [PubMed]

SUPPLEMENTARY DATA

Analytical methodology for a metabolome atlas of goat's plasma, milk and feces using 1H-NMR and UHPLC-HRMS

Cécile Martias[1], Julie Gatien[2], Lea Roch[2], Nadine Baroukh[1], Sylvie Mavel[1], Antoine Lefevre[1], Fréderic Montigny[1], Laurent Schibler[2], Patrick Emond[1,3] ,Lydie Nadal-Desbarats[1]

1 - UMR 1253, iBrain, University of Tours, Inserm, Tours, France
2 – ALLICE, Phenotyping station, Nouzilly, France
3-CHRU Tours, Biochemistry and molecular biology department, Tours, France

SAMPLE PREPARATION

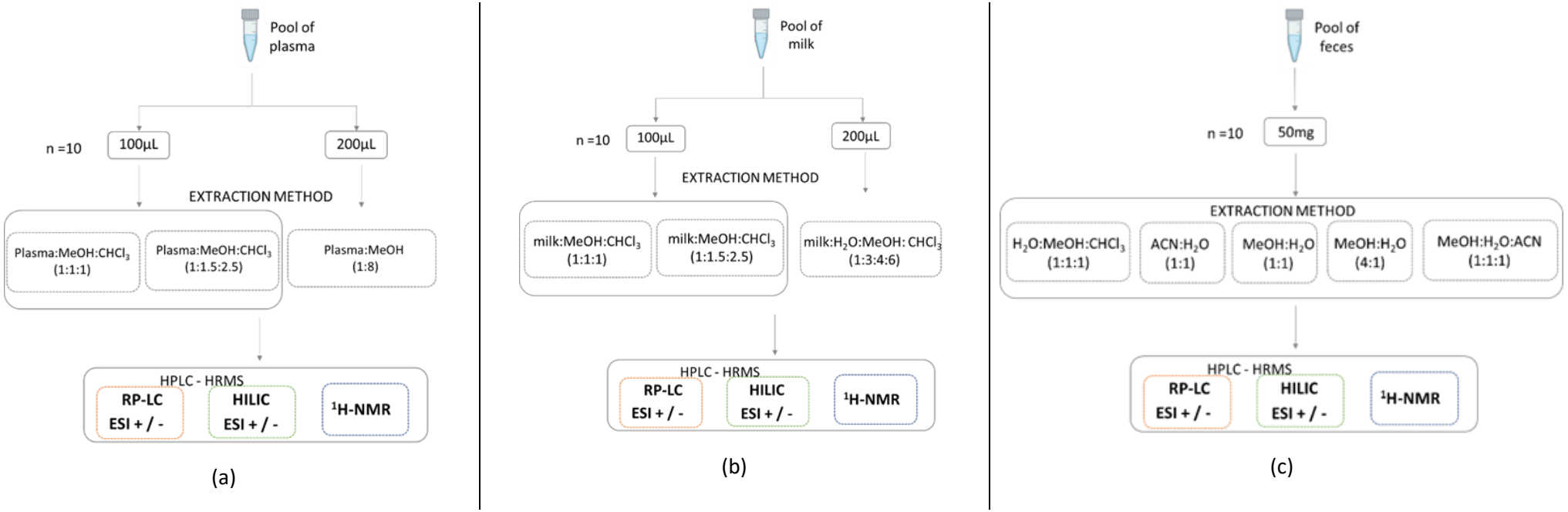


Figure S1. Schematic presentation of the preparation procedures applied on the pool of plasma (a), milk (b) and feces (c) for the selection of the optimum protocol for the maximum metabolome coverage by UHPLC-MS and ^{1}H-NMR.

PLASMA

		RP-LC	HILIC	NMR	Total
Plasma:MeOH :$CHCl_3$ (1 :1 :1) (n=10)	Number of metabolites :	137	131	39	229
	CV<30%	108	97	39	198
	CV<20%	81	73	39	161
	CV<10%	32	31	29	78
	CV % mean (min-max)	22% (4%-93%)	27% (3%-135%)	8% (4%-15%)	18% (3%-93%)
Plasma:MeOH :$CHCl_3$ (1 :2.5 :1) (n=10)	Number of metabolites :	145	129	36	230
	CV<30%	123	84	36	184
	CV<20%	103	65	35	160
	CV<10%	40	21	24	69
	CV % mean (min-max)	18% (4%-92%)	26% (3%-135%)	8% (3%-20%)	19% (3%-92%)
Plasma:MeOH (1 :8) (n=10)	Number of metabolites :	136	153	38	237
	CV<30%	120	61	38	169
	CV<20%	98	33	38	140
	CV<10%	41	15	30	78
	CV % mean (min-max)	16% (1%-50%)	41% (2%-120%)	6% (1%-19%)	24% (1%-120%)

Table S1: Number of extracted metabolites and coefficient of validation in blood for the different extraction methods using ^{1}H-NMR and UHPLC-MS

MILK

		RP-LC	HILIC	NMR	Total
Milk:MeOH :$CHCl_3$ (1 :1 :1) (n=10)	Number of metabolites :	142	131	35	234
	CV<30%	126	97	35	200
	CV<20%	111	58	35	166
	CV<10%	68	13	28	95
	CV % mean (min-max)	14% (1%-102%)	25% (1%-107%)	7% (3%-17%)	17% (1%-84%)
Milk:MeOH :$CHCl_3$ (1 :2.5 :2.5) (n=10)	Number of metabolites :	144	131	35	237
	CV<30%	95	90	35	172
	CV<20%	69	49	34	125
	CV<10%	19	8	23	45
	CV % mean (min-max)	30% (4%-175%)	27% (1%-110%)	9% (7%-20%)	25% (1%-122%)
Milk:H_2O:MeOH :$CHCl_3$ (1 :3 :4 :6) (n=10)	Number of metabolites :	139	151	33	263
	CV<30%	121	109	33	221
	CV<20%	107	107	32	187
	CV<10%	65	35	21	100
	CV % mean (min-max)	15% (2%-78%)	20% (2%-84%)	8% (2%-22%)	17% (2%-84%)

Table S2: Number of extracted metabolites and coefficient of validation in milk for the different extraction methods using ^{1}H-NMR and UHPLC-MS

FECES

		RP-LC	HILIC	NMR	Total
H_2O:MeOH :$CHCl_3$ (1 :1 :1) (n=10)	Number of metabolites :	236	250	28	347
	CV<30%	194	199	21	286
	CV<20%	175	143	17	241
	CV<10%	141	109	7	192
	CV % mean (min-max)	15% (2%-92%)	21% (3%-168%)	21% (5%-53%)	17% (2%-168%)
ACN:H_2O 1 :1 (n=10)	Number of metabolites :	236	256	28	348
	CV<30%	195	220	28	306
	CV<20%	169	196	28	280
	CV<10%	111	124	8	186
	CV % mean (min-max)	17% (2%-103%)	17% (3%-149%)	11% (6%-17%)	14% (2%-97%)
MeOH:H_2O 1 :1 (n=10)	Number of metabolites :	238	262	28	355
	CV<30%	198	222	28	313
	CV<20%	187	191	26	289
	CV<10%	137	132	13	215
	CV % mean (min-max)	15% (2%-98%)	16% (2%-157%)	11% (5%-24%)	14% (2%-157%)
MeOH:H_2O 4:1 (n=10)	Number of metabolites :	240	261	28	353
	CV<30%	201	221	25	313
	CV<20%	188	204	22	291
	CV<10%	138	140	19	222
	CV % mean (min-max)	15% (2%-99%)	16% (3%-119%)	13% (4%-54%)	13% (2%-119%)
MeOH:H_2O:ACN 1:1: (n=10)	Number of metabolites :	236	254	28	348
	CV<30%	201	219	28	310
	CV<20%	185	203	28	296
	CV<10%	149	138	16	225
	CV % mean (min-max)	14% (2%-95%)	15% (1%-119%)	9% (5%-15%)	13% (1%-119%)

Table S3: Number of extracted metabolites and coefficient of validation in fecal for the different extraction methods using ^{1}H-NMR and UHPLC-MS

List of metabolites found in each biological matrix or shared between two or three matrices

Table S4: Metabolites extracted from each biological matrix according to the best preparation procedure, with a CV<30%.

Only in plasma				
Metabolite	**Chemical class**	**HMDB**	**PubChem**	**KEGG**
Dihydroxyacetone phosphate	Acides phosphoriques	HMDB0001473	668	C00111
Acetone	Aliphatic compounds cyclic and acyclic	HMDB0001659	180	C00207
Dimethylglycine	Amino acids-Peptides-Derivatives	HMDB0000092	673	C01026
L-Cystine	Amino acids-Peptides-Derivatives	HMDB0000192	67678	C00491
Thyroxine	Amino acids-Peptides-Derivatives	HMDB0000248	5819	C01829
1-Hydroxy-2-naphthoate	Aromatic compounds-Vitamins-Amines	METPA0374		C03203
4-Hydroxycinnamic acid	Aromatic compounds-Vitamins-Amines	HMDB0002035	637542	C00811
Bilirubin	Aromatic compounds-Vitamins-Amines	HMDB0000054	21252250	C00486
Phenylacetic acid	Aromatic compounds-Vitamins-Amines	HMDB0000209	999	C07086
Phenylpyruvic acid	Aromatic compounds-Vitamins-Amines	HMDB0000205	997	C00166
(S)-3-Hydroxyisobutyric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000023	87	C06001
Methyl jasmonate	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0036583	5367719	C11512
Guanosine monophosphate	Nucleotides-Nucleosides-Derivatives	HMDB0001397	6804	C00144
Xanthosine	Nucleotides-Nucleosides-Derivatives	HMDB0000299	64959	C01762
Pyruvic acid	Organic acids	HMDB0000243	1060	C00022

Only in Milk				
Metabolite	**Chemical class**	**HMDB**	**PubChem**	**KEGG**
4-Methylcatechol	Aromatic compounds-Vitamins-Amines	HMDB0000873	9958	C06730
Adenosine monophosphate	Nucleotides-Nucleosides-Derivatives	HMDB0000045	6083	C00020
ADP-glucose	Nucleotides-Nucleosides-Derivatives	HMDB0006557	16500	C00498
Butanal	Aliphatic compounds cyclic and acyclic	HMDB0003543	261	C01412
Cellobiose	Sugars	HMDB00055	10712	C00185
D-Fructose	Sugars	HMDB00660	5984	C00095
D-Tagatose	Sugars	HMDB03418	439312	C00795
D-Xylitol	Sugars	HMDB02917	6912	C00379
Dethiobiotin	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0003581	445027	C01909
Dimethyl sulfone	Organic acids	HMDB0004983	6213	C11142
Galactaric acid	Sugars	HMDB0000639	3037582	C00879
Galactose 1-phosphate	Sugars	HMDB0000645	123912	C00446
Glucaric acid	Sugars	HMDB0000663	33037	C00818
Glycerol 3-phosphate	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000126	439162	C00093
Heptanoic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000666	8094	C17714
Hypotaurine	Organic acids	HMDB0000965	107812	C00519
L-Fucose	Sugars	HMDB0000174	3034656	C01019
L(-)-Nicotine	Aromatic compounds-Vitamins-Amines	HMDB0001934	89594	C00745
Melatonin	Aromatic compounds-Vitamins-Amines	HMDB0001389	896	C01598
N-Methylhydantoin	Aliphatic compounds cyclic and acyclic	HMDB03646	69217	C02565

Oxoglutaric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000208	51	C00026
Palatinose	Sugars	HMDB0256083	52994241	C01742
Phenylacetylglycine	Amino acids-Peptides-Derivatives	HMDB00821	68144	C05598
Phosphocreatine	Amino acids-Peptides-Derivatives	HMDB0001511	587	C02305
Phosphocreatine	Amino acids-Peptides-Derivatives	HMDB0001511	587	C02305
Sorbitol	Sugars	HMDB00247	5780	C00794
Sucrose	Sugars	HMDB0000258	5988	C00089
Theophylline	Aromatic compounds-Vitamins-Amines	HMDB0001889	2153	C07130
Uridine-5-diphospho-n-acetylglucosamine	Nucleotides-Nucleosides-Derivatives	HMDB00290	445675	C00043
Uridine diphosphate-N-acetylgalactosamine	Nucleotides-Nucleosides-Derivatives	HMDB0000304	1167	C00203
Uridine diphosphate glucose	Sugars	HMDB0000286	53477679	C00029
Uridine diphosphategalactose	Sugars	HMDB0000302	18068	C00052

Milk Plasma				
Metabolite	**Chemical class**	**HMDB**	**PubChem**	**KEGG**
Glycolaldehyde dimer	Aliphatic compounds cyclic and acyclic		186078	
Gamma-Aminobutyric acid	Amino acids-Peptides-Derivatives	HMDB0000112	223130	C00334
Dimethylamine	Aromatic compounds-Vitamins-Amines	HMDB0000087	674	C00543
Theobromine	Aromatic compounds-Vitamins-Amines	HMDB0002825	5429	C07480
3-Hydroxybutyric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000357	441	C01089
Lithocholyltaurine	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000722	53477716	C02592
3-Hydroxymethylglutaric acid	Organic acids	HMDB0000355	1662	C03761
Acetoacetic acid	Organic acids	HMDB0000060	96	C00164
Citric acid	Organic acids	HMDB0000094	311	C00158
Methylmalonic acid	Organic acids	HMDB0000202	487	C02170
Allose	Sugars	HMDB0001151	12285879	C01487
Alpha-Lactose	Sugars	HMDB0000186	84571	C00243
D-Galactose	Sugars	HMDB0000143	439357	C00984
L-Sorbose	Sugars	HMDB0001266	441484	C08356

Milk Feces				
Metabolite	**Chemical class**	**HMDB**	**PubChem**	**KEGG**
1-Aminocyclopropanecarboxylic acid	Amino acids-Peptides-Derivatives	HMDB0036458	535	C01234
2-Hydroxypyridine	Aromatic compounds-Vitamins-Amines	HMDB0013751	8871	C02502
3-Aminoisobutanoic acid	Amino acids-Peptides-Derivatives	HMDB0003911	64956	C05145
3-Methyladenine	Aromatic compounds-Vitamins-Amines	HMDB0011600	1673	C00913
4-Hydroxybenzaldehyde	Aromatic compounds-Vitamins-Amines	HMDB0011718	126	C00633
5-Hydroxylysine	Amino acids-Peptides-Derivatives	HMDB0000450	4433	C16741
Acetamide	Aromatic compounds-Vitamins-Amines	HMDB0031645	178	C06244
Anserine	Amino acids-Peptides-Derivatives	HMDB0000194	112072	C01262
Cholic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000619	221493	C00695
Citraconic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000634	643798	C02226
Corticosterone	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0001547	5753	C02140
Cortisol	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000063	657311	C00735
Cortisol-21-acetate	Lipids: Fatty acids-Steroids-Hormones-Vitamins		5744	C02821
Cyclic AMP	Nucleotides-Nucleosides-Derivatives	HMDB0000058	6076	C00575
Diacetyl	Aliphatic compounds cyclic and acyclic	HMDB0003407	650	C00741

Epinephrine	Aromatic compounds-Vitamins-Amines	HMDB0000068	5816	C00788
Ethanolamine	Aliphatic compounds cyclic and acyclic	HMDB0000149	700	C00189
Fumaric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000134	444972	C00122
Glucosamine 6-phosphate	Sugars	HMDB0001254	439217	C00352
Glucose 6-phosphate	Sugars	HMDB0001401	5958	C00092
Glyceraldehyde-3-phosphatediethylacetal	Aliphatic compounds cyclic and acyclic		3334922	
Imidazoleacetic acid	Aromatic compounds-Vitamins-Amines	HMDB0002024	96215	C02835
Indolepyruvate	Aromatic compounds-Vitamins-Amines	HMDB0060484	803	C00331
Kynurenic acid	Amino acids-Peptides-Derivatives	HMDB0000715	3845	C01717
L-Acetylcarnitine	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000201	7045767	C02571
L-Allothreonine	Amino acids-Peptides-Derivatives	HMDB0004041	99289	C05519
L-Dihydroorotic acid	Aliphatic compounds cyclic and acyclic	HMDB0003349	439216	C00337
Mannitol	Sugars	HMDB0000765	6251	C00392
N-Acetyl-D-glucosamine	Sugars	HMDB0000215	439174	C00140
N-Acetylgalactosamine	Sugars	HMDB0000212	84265	C01074
N-Acetyl-L-methionine	Amino acids-Peptides-Derivatives	HMDB0011745	6180	C02712
N-Acetylmannosamine	Sugars	HMDB0001129	11096158	C00645
Niacinamide	Aromatic compounds-Vitamins-Amines	HMDB0001406	936	C00153
Nicotinic acid	Aromatic compounds-Vitamins-Amines	HMDB0001488	938	C00253

N-Methyltryptamine	Aromatic compounds-Vitamins-Amines	HMDB0004370	6088	C06213
N-Trimethyllysine	Amino acids-Peptides-Derivatives		440121	C03793
Oxfenicine	Amino acids-Peptides-Derivatives	HMDB0246152	36143	D05292
Pantolactone	Aliphatic compounds cyclic and acyclic	HMDB0059876	439368	C01012
Pterin	Aromatic compounds-Vitamins-Amines	HMDB0000802	73000	C00715
Pyrrole-2-carboxylic acid	Amino acids-Peptides-Derivatives	HMDB0004230	12473	C05942
Riboflavin	Aromatic compounds-Vitamins-Amines	HMDB0000244	493570	C00255
Salicyluric acid	Aromatic compounds-Vitamins-Amines	HMDB0000840	10253	C07588
Thymine	Aromatic compounds-Vitamins-Amines	HMDB0000262	1135	C00178
trans-Aconitic acid	Organic acids	HMDB0000958	444212	C02341
trans-Ferulic acid	Aromatic compounds-Vitamins-Amines	HMDB0000954	445858	C01494
Tryptophanol	Aromatic compounds-Vitamins-Amines	HMDB0003447	10685	C00955

Milk Feces Plasma				
Metabolite	Chemical class	HMDB	PubChem	KEGG
Acetylcholine	Aliphatic compounds cyclic and acyclic	HMDB0000895	6060	C01996
Allantoin	Aliphatic compounds cyclic and acyclic	HMDB0000462	204	C01551
Biotin	Aliphatic compounds cyclic and acyclic	HMDB0000030	171548	C00120
Choline	Aliphatic compounds cyclic and acyclic	HMDB0000097	305	C00114
Creatinine	Aliphatic compounds cyclic and acyclic	HMDB0000562	588	C00791
Cyclohexane-1,2-Diol	Aliphatic compounds cyclic and acyclic		13601	C03739
L-Carnitine	Aliphatic compounds cyclic and acyclic	HMDB0000062	2724480	C00318
Methylguanidine	Aliphatic compounds cyclic and acyclic	HMDB0001522	10111	C02294
O-Phosphoethanolamine	Aliphatic compounds cyclic and acyclic	HMDB0000224	1015	C00346
Pantothenic acid	Aliphatic compounds cyclic and acyclic	HMDB0000210	6613	C00864
Phosphorylcholine	Aliphatic compounds cyclic and acyclic	HMDB0001565	8691	C00588
Propanal	Aliphatic compounds cyclic and acyclic	HMDB0003366	527	C00479
Trimethylamine	Aliphatic compounds cyclic and acyclic	HMDB0000906	1146	C00565
Ureidopropionic acid	Aliphatic compounds cyclic and acyclic	HMDB0000026	111	C02642
2,4-Dihydroxypyrimidine-5-carboxylic acid	Amino acids-Peptides-Derivatives		90301	C03030
3-Methylhistidine	Amino acids-Peptides-Derivatives	HMDB0000479	64969	C01152
4-Hydroxy-L-proline	Amino acids-Peptides-Derivatives	HMDB0006055	69248	C01015
5-Aminolevulinic acid	Amino acids-Peptides-Derivatives	HMDB0001149	137	C00430

Acetylglycine	Amino acids-Peptides-Derivatives	HMDB0000532	10972	
Beta-Alanine	Amino acids-Peptides-Derivatives	HMDB0000056	239	C00099
Betaine	Amino acids-Peptides-Derivatives	HMDB0000043	247	C00719
Citrulline	Amino acids-Peptides-Derivatives	HMDB0000904	9750	C00327
Creatine	Amino acids-Peptides-Derivatives	HMDB0000064	586	C00300
Formyl-L-methionylpeptide	Amino acids-Peptides-Derivatives		13611	C11439
Glycine	Amino acids-Peptides-Derivatives	HMDB0000123	750	C00037
Guanidoacetic acid	Amino acids-Peptides-Derivatives	HMDB0000128	763	C00581
Hippuric acid	Amino acids-Peptides-Derivatives	HMDB0000714	464	C01586
L-Arginine	Amino acids-Peptides-Derivatives	HMDB0000517	6322	C00062
L-Asparagine	Amino acids-Peptides-Derivatives	HMDB0000168	6267	C00152
L-Aspartic acid	Amino acids-Peptides-Derivatives	HMDB0000191	5960	C00049
L-Glutamic acid	Amino acids-Peptides-Derivatives	HMDB0000148	33032	C00025
L-Glutamine	Amino acids-Peptides-Derivatives	HMDB0000641	5961	C00064
L-Histidine	Amino acids-Peptides-Derivatives	HMDB0000177	6274	C00135
L-Homoserine	Amino acids-Peptides-Derivatives	HMDB0000719	12647	C00263
L-Isoleucine	Amino acids-Peptides-Derivatives	HMDB0000172	6306	C00407
L-Kynurenine	Amino acids-Peptides-Derivatives	HMDB0000684	161166	C00328
L-Leucine	Amino acids-Peptides-Derivatives	HMDB0000687	6106	C00123
L-Lysine	Amino acids-Peptides-Derivatives	HMDB0000182	5962	C00047
L-Methionine	Amino acids-Peptides-Derivatives	HMDB0000696	6137	C00073

L-Norleucine	Amino acids-Peptides-Derivatives	HMDB0001645	21236	C01933
L-Phenylalanine	Amino acids-Peptides-Derivatives	HMDB0000159	6140	C00079
L-Proline	Amino acids-Peptides-Derivatives	HMDB0000162	145742	C00148
L-Serine	Amino acids-Peptides-Derivatives	HMDB0000187	5951	C00065
L-Threonine	Amino acids-Peptides-Derivatives	HMDB0000167	6288	C00188
L-Tryptophan	Amino acids-Peptides-Derivatives	HMDB0000929	6305	C00078
L-Tyrosine	Amino acids-Peptides-Derivatives	HMDB0000158	6057	C00082
L-Valine	Amino acids-Peptides-Derivatives	HMDB0000883	6287	C00183
N-Acetyl-L-alanine	Amino acids-Peptides-Derivatives	HMDB0000766	88064	
N-Acetyl-L-phenylalanine	Amino acids-Peptides-Derivatives	HMDB0000512	74839	C03519
N-Acetylleucine	Amino acids-Peptides-Derivatives	HMDB0011756	70912	C02710
N-Acetylserine	Amino acids-Peptides-Derivatives	HMDB0002931	65249	
N-acetyltryptophan	Amino acids-Peptides-Derivatives	HMDB0013713	700653	
Ornithine	Amino acids-Peptides-Derivatives	HMDB0000214	6262	C00077
Orotic acid	Amino acids-Peptides-Derivatives	HMDB0000226	967	C00295
Pipecolic acid	Amino acids-Peptides-Derivatives	HMDB0000070	849	C00408
Pyroglutamic acid	Amino acids-Peptides-Derivatives	HMDB0000267	7405	C01879
Sarcosine	Amino acids-Peptides-Derivatives	HMDB0000271	1088	C00213
2-Aminophenol	Aromatic compounds-Vitamins-Amines	METPA0236		C01987
3-(2-Hydroxyphenyl)propanoic acid	Aromatic compounds-Vitamins-Amines	HMDB0033752	873	C01198
3-Hydroxyphenylacetic acid	Aromatic compounds-Vitamins-Amines	HMDB0000440	12122	C05593

3-Methyldioxyindole	Aromatic compounds-Vitamins-Amines	HMDB0004186	151066	C05834
3,4-Dihydroxybenzeneacetic acid	Aromatic compounds-Vitamins-Amines	HMDB0001336	547	C01161
5-Hydroxyindoleacetic acid	Aromatic compounds-Vitamins-Amines	HMDB0000763	1826	C05635
Adenine	Aromatic compounds-Vitamins-Amines	HMDB0000034	190	C00147
Benzyl alcohol	Aromatic compounds-Vitamins-Amines	HMDB0003119	244	C03485
Cytosine	Aromatic compounds-Vitamins-Amines	HMDB0000630	597	C00380
Gentisic acid	Aromatic compounds-Vitamins-Amines	HMDB0000152	3469	C00628
Guanine	Aromatic compounds-Vitamins-Amines	HMDB0000132	764	C00242
Homogentisic acid	Aromatic compounds-Vitamins-Amines	HMDB0000130	780	C00544
Homovanillic acid	Aromatic compounds-Vitamins-Amines	HMDB0000118	1738	C05582
Hypoxanthine	Aromatic compounds-Vitamins-Amines	HMDB0000157	790	C00262
Indoleacetic acid	Aromatic compounds-Vitamins-Amines	HMDB0000197	802	C00954
Indoxyl sulfate	Aromatic compounds-Vitamins-Amines	HMDB0000682	10258	
Lumichrome	Aromatic compounds-Vitamins-Amines	METPA0216		C01727
N-Acetylserotonin	Aromatic compounds-Vitamins-Amines	HMDB0001238	903	C00978
N,N-Dimethylphenethylamine	Aromatic compounds-Vitamins-Amines	HMDB0032243	86536	
Ortho-Hydroxyphenylacetic acid	Aromatic compounds-Vitamins-Amines	HMDB0000669	11970	C05852
p-Hydroxyphenylacetic acid	Aromatic compounds-Vitamins-Amines	HMDB0000020	127	C00642
Paraxanthine	Aromatic compounds-Vitamins-Amines	HMDB0001860	4687	C13747
Protocatechuic acid	Aromatic compounds-Vitamins-Amines	HMDB0001856	72	C00230
Pyrocatechol	Aromatic compounds-Vitamins-Amines	HMDB0000957	3390	C15571

trans-Cinnamic acid	Aromatic compounds-Vitamins-Amines	HMDB0000930	444539	C10438
Trigonelline	Aromatic compounds-Vitamins-Amines	HMDB0000875	5570	C01004
Uracil	Aromatic compounds-Vitamins-Amines	HMDB0000300	1174	C00106
Uric acid	Aromatic compounds-Vitamins-Amines	HMDB0000289	1175	C00366
Urocanic acid	Aromatic compounds-Vitamins-Amines	HMDB0000301	736715	C00785
Xanthine	Aromatic compounds-Vitamins-Amines	HMDB0000292	1188	C00385
Thymine	Aromatic compounds-Vitamins-Amines	HMDB0000262	1135	C00178
12-Hydroxydodecanoic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0002059	79034	C08317
2-Hydroxybutyric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000008	11266	C05984
4-Trimethylammoniobutanoic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0001161	134	C01181
Caproic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000535	8892	C01585
Ethylmalonic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000622	11756	
Galactitol	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000107	11850	C01697
Glycerophosphocholine	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000086	71920	C00670
Glycocholic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000138	23617285	C01921
Ketoleucine	Lipids: Fatty acids-Steroids-Hormones-	HMDB0000695	70	C00233

Vitamins				
1-Methyladenosine	Nucleotides-Nucleosides-Derivatives	HMDB0003331	27476	C02494
5'-Methylthioadenosine	Nucleotides-Nucleosides-Derivatives	HMDB0001173	439176	C00170
Adenosine	Nucleotides-Nucleosides-Derivatives	HMDB0000050	60961	C00212
Deoxycytidine	Nucleotides-Nucleosides-Derivatives	HMDB0000014	13711	C00881
Deoxyguanosine	Nucleotides-Nucleosides-Derivatives	HMDB0000085	187790	C00330
Guanosine	Nucleotides-Nucleosides-Derivatives	HMDB0000133	6802	C00387
Inosine	Nucleotides-Nucleosides-Derivatives	HMDB0000195	6021	C00294
N6-(delta2-Isopentenyl)-adenine	Nucleotides-Nucleosides-Derivatives	METPA1684		C04083
Thymidine	Nucleotides-Nucleosides-Derivatives	HMDB0000273	5789	C00214
Uridine	Nucleotides-Nucleosides-Derivatives	HMDB0000296	6029	C00299
3-Hydroxycapric acid	Organic acids	HMDB0002203	26612	
4-Acetamidobutanoic acid	Organic acids	HMDB0003681	18189	C02946
Acetic acid	Organic acids	HMDB0000042	176	C00033
Azelaic acid	Organic acids	HMDB0000784	2266	C08261
Citramalic acid	Organic acids	HMDB0000426	441696	C00815
Formic acid	Organic acids	HMDB0000142	284	C00058
L-Lactic acid	Organic acids	HMDB0000190	61503	C00186
L-Malic acid	Organic acids	HMDB0000156	222656	C00149
N-Acetylputrescine	Organic acids	HMDB0002064	122356	C02714
Suberic acid	Organic acids	HMDB0000893	10457	C08278

Succinic acid	Organic acids	HMDB0000254	1110	C00042
Taurine	Organic acids	HMDB0000251	1123	C00245
2-Deoxy-D-Glucose	Sugars		3865	C00586
Alpha-D-Glucose	Sugars	HMDB0003345	79025	C00267
D-Galactosamine	Sugars	METPA0263		C02262
D-Mannose	Sugars	HMDB0000169	18950	C00936
Gluconic acid	Sugars	HMDB0000625	10690	C00257
Glyceric acid	Sugars	HMDB0000139	439194	C00258
L-Arabitol	Sugars	HMDB0001851	94154	C00532
Methyl beta-D-glucopyranoside	Sugars	HMDB0029965	94214	
myo-Inositol	Sugars	HMDB0000211		C00137
N-Acetylneuraminate	Sugars		3568	C00270
Ribitol	Sugars	HMDB0000508		C00474

Feces				
Metabolite	**Chemical class**	**HMDB**	**PubChem**	**KEGG**
Ciliatine	Acides phosphoriques	HMDB0011747	339	C03557
3-Dehydroshikimate	Aliphatic compounds cyclic and acyclic	METPA0312		C02637
Acetoin	Aliphatic compounds cyclic and acyclic	HMDB0003243	4068	C00466
Agmatine	Aliphatic compounds cyclic and acyclic	HMDB0001432	199	C00179
Cadaverine	Aliphatic compounds cyclic and acyclic	HMDB0002322	273	C01672
Quinic acid	Aliphatic compounds cyclic and acyclic	HMDB0003072	6508	C00296
Valerolactone	Aliphatic compounds cyclic and acyclic	NA	10953	C02240
Shikimic acid	Aliphatic compounds cyclic and acyclic	HMDB0003070	8742	C00493
Spermidine	Aliphatic compounds cyclic and acyclic	HMDB0001257	1102	C00315
3-Sulfinoalanine	Amino acids-Peptides-Derivatives	HMDB0000996	439270	C00606
L-Alanine	Amino acids-Peptides-Derivatives	HMDB0000161	5950	C00041
L-Cystathionine	Amino acids-Peptides-Derivatives	HMDB0000099	439258	C02291
N-Formylglycine	Amino acids-Peptides-Derivatives		75606	
Picolinic acid	Amino acids-Peptides-Derivatives	HMDB0002243	1018	C10164
Quinaldic acid	Amino acids-Peptides-Derivatives	HMDB0000842	7124	C06325
5-Hydroxy-L-tryptophan	Amino acids-Peptides-Derivatives	HMDB0000472	144	C00643
L-Dopa	Amino acids-Peptides-Derivatives	HMDB0000181	6047	C00355
DL-O-Phosphoserine	Amino acids-Peptides-Derivatives	HMDB0001721	106	C01005

Homocysteine	Amino acids-Peptides-Derivatives	HMDB0000742	778	C00155
S-Adenosylhomocysteine	Amino acids-Peptides-Derivatives	HMDB0000939	439155	C00021
2-Aminoisobutyric acid	Amino acids-Peptides-Derivatives	HMDB0001906	6119	C03665
3-Methoxytyrosine	Amino acids-Peptides-Derivatives	HMDB0001434	1670	
Diaminopimelic acid	Amino acids-Peptides-Derivatives	HMDB0001370	99290	C00666
3-Nitrotyrosine	Amino acids-Peptides-Derivatives	HMDB0001904	65124	
Cysteic acid	Amino acids-Peptides-Derivatives	HMDB0002757	25701	C00506
Hydroxykynurenine	Amino acids-Peptides-Derivatives	HMDB0000732	89	C02794
2-Hydroxyphenethylamine	Aromatic compounds-Vitamins-Amines	HMDB0001065	1000	C02735
3-Methoxytyramine	Aromatic compounds-Vitamins-Amines	HMDB0000022	1669	C05587
4-Hydroxyphenylpyruvic acid	Aromatic compounds-Vitamins-Amines	HMDB0000707	979	C01179
Dihydrouracil	Aromatic compounds-Vitamins-Amines	HMDB0000076	649	C00429
Isovaleric acid	Aromatic compounds-Vitamins-Amines	HMDB0000718	10430	C08262
L-Histidinol	Aromatic compounds-Vitamins-Amines	HMDB0003431	165271	C00860
p-Aminobenzoic acid	Aromatic compounds-Vitamins-Amines	HMDB0001392	978	C00568
Purine	Aromatic compounds-Vitamins-Amines	HMDB0001366	1044	C15587
Pyridoxine	Aromatic compounds-Vitamins-Amines	HMDB0000239	1054	C00314
Resorcinol monoacetate	Aromatic compounds-Vitamins-Amines		5055	D02393
Tyramine	Aromatic compounds-Vitamins-Amines	HMDB0000306	5610	C00483
Uracil-5-carboxylic acid	Aromatic compounds-Vitamins-Amines		90301	C03030
Vanillylmandelic acid	Aromatic compounds-Vitamins-Amines	HMDB0000291	736172	C05584

2,6-Dihydroxypyridine	Aromatic compounds-Vitamins-Amines	METPA0354		C03056
3-Hydroxybenzaldehyde	Aromatic compounds-Vitamins-Amines		101	C03067
5-Methylcytosine	Aromatic compounds-Vitamins-Amines	HMDB0002894	65040	C02376
Biliverdin	Aromatic compounds-Vitamins-Amines	HMDB0001008	5353439	C00500
Norepinephrine	Aromatic compounds-Vitamins-Amines	HMDB0000216	439260	C00547
4-Hydroxybenzoic acid	Aromatic compounds-Vitamins-Amines	HMDB0000500	135	C00156
Benzoic acid	Aromatic compounds-Vitamins-Amines	HMDB0001870	243	C00539
3-Amino-4-hydroxybenzoic acid	Aromatic compounds-Vitamins-Amines		65083	C12115
3-Amino-5-hydroxybenzoic acid	Aromatic compounds-Vitamins-Amines		127115	C12107
3-Hydroxyanthranilic acid	Aromatic compounds-Vitamins-Amines	HMDB0001476	86	C00632
3,4-Dihydroxymandelicacid	Aromatic compounds-Vitamins-Amines		85782	C05580
Ethyl-3-indole acetate	Aromatic compounds-Vitamins-Amines			C00954
Hydroquinone	Aromatic compounds-Vitamins-Amines	HMDB0002434	785	C00530
Vanylglycol	Aromatic compounds-Vitamins-Amines	HMDB0001490	10805	C05594
1H-Indole-3-acetamide	Aromatic compounds-Vitamins-Amines	HMDB0029739	397	C02693
Dimethylbenzimidazole	Aromatic compounds-Vitamins-Amines	HMDB0003701	675	C03114
Indoleacetaldehyde	Aromatic compounds-Vitamins-Amines	HMDB0001190	800	C00637
1-Naphthylamine	Aromatic compounds-Vitamins-Amines	METPA1135		C14790
5-Hydroxymethyluracil	Aromatic compounds-Vitamins-Amines	HMDB0000469	78168	C03088
Serotonin	Aromatic compounds-Vitamins-Amines	HMDB0000259	5202	C00780
Caprylic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000482	379	C06423

Ergocalciferol	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000900	5280793	C05441
Lithocholic Acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000761	11740284	C03990
Maleic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000176	444972	C01384
Mono-methylglutarate	Lipids: Fatty acids-Steroids-Hormones-Vitamins		6993194	
Oleic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000207	445639	C00712
Rosmarinic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0003572	5281792	C01850
Sphinganine	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000269	91486	C00836
Vitamin A	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000305	445354	C17276
Petroselinic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0002080	5461010	C08363
All-trans-retinoic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0001852	444795	C00777
Chenodeoxycholic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000518	10133	C02528
Cholesteryl acetate	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0003822	6427285	C02530
2-Methylglutaric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000422	12046	
7-Dehydrocholesterol	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000032	439423	C01164
Glycerol	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000131	753	C00116
Alpha-Tocopherol	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0001893	14985	C02477
Elaidic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000573	637517	C01712
Cortexolone	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000015	53477676	C05488
Deoxycorticosterone acetate	Lipids: Fatty acids-Steroids-Hormones-Vitamins		5952	C14554
Valeric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000892	7991	C00803
5'-Deoxyadenosine	Nucleotides-Nucleosides-Derivatives	HMDB0001983	439182	C05198

dCMP	Nucleotides-Nucleosides-Derivatives	HMDB0001202	13945	C00239
Cytidine2',3'-cyclicphosphate	Nucleotides-Nucleosides-Derivatives		5401	C02354
5'UMP	Nucleotides-Nucleosides-Derivatives		3405	C00105
Cytidine	Nucleotides-Nucleosides-Derivatives	HMDB0000089	6253	C00475
Cyclic GMP	Nucleotides-Nucleosides-Derivatives	HMDB0001314	24316	C00942
Deoxyuridine	Nucleotides-Nucleosides-Derivatives	HMDB0000012	13712	C00526
Dimethylsulfone	Organic acids	HMDB0004983	6213	C11142
Malonic acid	Organic acids	HMDB0000691	867	C04025
Pimelic acid	Organic acids	HMDB0000857	385	C02656
4-Heptenal diethyl acetal	Organic acids	HMDB0032306	97750	C04067
Isobutyric acid	Organic acids	HMDB0001873	6590	C02632
D-Maltose	Sugars	HMDB0000163	10991489	C00208
Rhamnose	Sugars	HMDB0000849	25310	C00507
D-Lyxose	Sugars		3759	C00476
2-Acetamido-2-deoxy-beta-glucosylamine	Sugars			
Glucosamine	Sugars	HMDB0001514	439213	C00329
D-Ribose	Sugars	HMDB0000283	5779	C00121
2-Amino-2-deoxy-D-gluconate	Sugars	METPA0433		C03752

Feces Plasma				
Metabolite	Chemical Class	HMDB	PubChem	KEGG
Diethanolamine	Aliphatic compounds cyclic and acyclic	HMDB0004437	8113	C06772
Putrescine	Aliphatic compounds cyclic and acyclic	HMDB0001414	1045	C00134
4-Guanidinobutanoic acid	Amino acids-Peptides-Derivatives	HMDB0003464	500	C01035
5-Aminopentanoic acid	Amino acids-Peptides-Derivatives	HMDB0003355	138	C00431
Aminoadipic acid	Amino acids-Peptides-Derivatives	HMDB0000510	469	C00956
Carnosine	Amino acids-Peptides-Derivatives	HMDB0000033	439224	C00386
L-Norvaline	Amino acids-Peptides-Derivatives		4949	C01826
N-Acetyl-L-aspartic acid	Amino acids-Peptides-Derivatives	HMDB0000812	65065	C01042
N-Acetylglutamic acid	Amino acids-Peptides-Derivatives	HMDB0001138	185	C00624
N-Alpha-acetyllysine	Amino acids-Peptides-Derivatives	HMDB0000446	192590	C12989
2',4'-Dihydroxyacetophenone	Aromatic compounds-Vitamins-Amines	HMDB0029659	6990	C03663
2-Aminobenzoic acid	Aromatic compounds-Vitamins-Amines	HMDB0001123	227	C00108
3-Hydroxybenzoic acid	Aromatic compounds-Vitamins-Amines	HMDB0002466	7420	C00587
Caffeic acid	Aromatic compounds-Vitamins-Amines	HMDB0001964	1549111	C01481
Histamine	Aromatic compounds-Vitamins-Amines	HMDB0000870	774	C00388
Hydroxyphenyllactic acid	Aromatic compounds-Vitamins-Amines	HMDB0000755	9378	C03672
Indole-3-methyl acetate	Aromatic compounds-Vitamins-Amines	HMDB0029738	74706	

Phenylacetaldehyde	Aromatic compounds-Vitamins-Amines	HMDB0006236	998	C00601
Pyridoxamine	Aromatic compounds-Vitamins-Amines	HMDB0001431	1052	C00534
Salicylamide	Aromatic compounds-Vitamins-Amines	HMDB0015687	5147	NA
Tryptamine	Aromatic compounds-Vitamins-Amines	HMDB0000303	1150	C00398
17a-Estradiol	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000429	68570	C02537
2,3-Dihydroxypropyl tetradecanoate	Lipids: Fatty acids-Steroids-Hormones-Vitamins		79050	
21-Tetrahydroxy-5-alpha-pregnan-20-one	Lipids: Fatty acids-Steroids-Hormones-Vitamins		5864	
3-Methylglutaric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000752	12284	
Butyric acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000039	264	C00246
Cortisone	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0002802	225609	C00762
Deoxycholic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0000626	222528	C04483
Dihydroxy-beta-cholanate	Lipids: Fatty acids-Steroids-Hormones-Vitamins			
Gamma-Linolenic acid	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0003073	5280933	C06426
Glyceryl monoacetate	Lipids: Fatty acids-Steroids-Hormones-Vitamins		33510	
Leukotriene B4	Lipids: Fatty acids-Steroids-Hormones-Vitamins	HMDB0001085	5283128	C02165
Deoxyadenosine	Nucleotides-Nucleosides-Derivatives	HMDB0000101	13730	C00559
Adipic acid	Organic acids	HMDB0000448	196	C06104
Isocitric acid	Organic acids	HMDB0000193	1198	C00311
Propionic acid	Organic acids	HMDB0000237	1032	C00163
D-Glucuronic acid	Sugars	HMDB0000127	94715	C00191

List of metabolites according to the platform

Table S5: Metabolites detected by platform in plasma, milk and feces. In bold the platform where the metabolite was annotated with the lowest CV (best reproducibility)

Plasma (Blood :MeOH : $CHCl_3$ 1 :1 :1)	
Metabolite	**Analytical platform**
3-Hydroxycapric acid	RP+
1-hydroxy-2-naphthoate	HILIC-
1-Methyladenosine	RP- **HILIC+**
Glyceryl monoacetate	HILIC+
2',4'-Dihydroxyacetophenone	RP+
2,4-Dihydroxypyrimidine-5-carboxylic acid	HILIC-
Gentisic acid	HILIC-
2-Aminophenol	HILIC+
Deoxyadenosine	NMR
2-Deoxy-D-Glucose	RP+
Deoxyguanosine	HILIC-
2-Hydroxybutyric acid	RP- **NMR**
Ortho-Hydroxyphenylacetic acid	HILIC-
Ketoleucine	RP- **NMR**

Protocatechuic acid	HILIC-
3,4-Dihydroxybenzeneacetic acid	HILIC-
3-(2-Hydroxyphenyl)propanoic acid	**RP-** HILIC-
Hydroxyphenyllactic acid	HILIC-
21-Tetrahydroxy-5-alpha-pregnan-20-one	HILIC-
Dihydroxy-beta-cholanate	HILIC-
3-Hydroxymethylglutaric acid	HILIC-
3-Hydroxybenzoic acid	HILIC-
3-Hydroxybutyric acid	NMR
(S)-3-Hydroxyisobutyric acid	HILIC- **NMR**
3-Hydroxyphenylacetic acid	HILIC-
3-Methylglutaric acid	RP-
3-Methyldioxyindole	RP-
Ureidopropionic acid	RP-
4-Acetamidobutanoic acid	RP-
Gamma-Aminobutyric acid	RP+
4-Hydroxycinnamic acid	HILIC+
4-Guanidinobutanoic acid	RP+
4-Hydroxy-L-proline	RP+
p-Hydroxyphenylacetic acid	HILIC-
5-Aminolevulinic acid	RP+
5-Aminopentanoic acid	RP+

5-Hydroxyindoleacetic acid	**RP+ HILIC-**
5'-Methylthioadenosine	RP+ **HILIC+**
Pyroglutamic acid	RP+
Acetic acid	NMR
Acetoacetic acid	HILIC-
Acetone	NMR
N-Acetyl-L-alanine	RP+
N-Acetyl-L-aspartic acid	HILIC+
N-Acetylglutamic acid	RP+
Acetylglycine	RP-
N-Acetylleucine	RP+
N-Acetylneuraminate	RP-
N-Acetyl-L-phenylalanine	**RP-** HILIC-
N-Acetylputrescine	HILIC+
N-Acetylserine	RP-
N-Acetylserotonin	HILIC+
N-acetyltryptophan	RP-
Adenine	RP+
Adenosine	HILIC-
Adipic acid	RP+
Allantoin	**RP-** HILIC+ NMR
Allose	HILIC-

Alpha-D-Glucose	RP- HILIC- **NMR**
Aminoadipic acid	RP-
2-Aminobenzoic acid	HILIC+
L-Arabitol	HILIC-
L-Arginine	RP-
L-Asparagine	**RP-** HILIC+
L-Aspartic acid	RP-
Azelaic acid	RP-
Benzyl alcohol	RP- **HILIC-**
Beta-Alanine	RP+ **NMR**
Betaine	RP+
Bilirubin	HILIC+
Biotin	RP+
Butyric acid	NMR
Caffeic acid	HILIC-
L-Carnitine	RP+
Carnosine	RP- **NMR**
Pyrocatechol	RP-
Choline	NMR
trans-Cinnamic acid	RP+ **HILIC-**
Citramalic acid	RP-
Citric acid	NMR

Citrulline	RP-
Cortisone	HILIC-
Creatine	**RP- NMR**
Creatinine	NMR
Cyclohexane-1,2-Diol	HILIC+
L-Cystine	RP-
Cytosine	RP+
4-Trimethylammoniobutanoic acid	RP+
Deoxycholic acid	HILIC-
Deoxycytidine	HILIC-
Diethanolamine	HILIC+
Dihydroxyacetone phosphate	HILIC-
N,N-Dimethylphenethylamine	HILIC+
Dimethylamine	NMR
Dimethylglycine	NMR
17a-Estradiol	HILIC+
O-Phosphoethanolamine	RP-
Ethylmalonic acid	RP-
Formic acid	NMR
Formyl-L-methionylpeptide	RP-
Galactitol	HILIC-
D-Galactosamine	RP+

D-Galactose	HILIC-
D-Glucuronic acid	RP-
L-Glutamic acid	RP-
L-Glutamine	RP+
Glyceric acid	RP-
Glycerophosphocholine	RP+
Glycine	RP+ **NMR**
Glycocholic acid	RP-
Glycolaldehyde dimer	HILIC+
Guanidoacetic acid	RP+ **HILIC+**
Guanine	HILIC-
Guanosine	HILIC+
Guanosine monophosphate	RP+
Caproic acid	HILIC+
Hippuric acid	RP- **HILIC-** NMR
Histamine	HILIC+
L-Histidine	RP- **NMR**
Homogentisic acid	HILIC-
L-Homoserine	RP-
Homovanillic acid	HILIC-
Hypoxanthine	RP-
Indoleacetic acid	**RP+** HILIC-

Indoxyl sulfate	**RP-** HILIC-
Inosine	RP+ **HILIC+**
Isocitric acid	RP-
L-Isoleucine	NMR
N6-(delta2-Isopentenyl)-adenine	RP-
L-Kynurenine	RP+
L-Lactic acid	RP- **NMR**
Alpha-Lactose	NMR
L-Leucine	NMR
L-Norleucine	RP+
Leukotriene B4	RP+
Gamma-Linolenic acid	HILIC+
Lumichrome	**RP+** HILIC+
L-Lysine	RP+ **NMR**
L-Malic acid	RP-
D-Mannose	**HILIC+** NMR
L-Methionine	**RP-** NMR
Methyl beta-D-glucopyranoside	HILIC-
Indole-3-methyl acetate	RP+ **HILIC+**
Methyl jasmonate	HILIC+
Methylguanidine	HILIC+
3-Methylhistidine	NMR

Methylmalonic acid	HILIC-
myo-Inositol	HILIC-
N-Alpha-acetyllysine	RP+
L-Norvaline	RP+
Acetylcholine	NMR
12-Hydroxydodecanoic acid	RP- **HILIC-**
Ornithine	RP+
Orotic acid	HILIC-
Pantothenic acid	HILIC+
Paraxanthine	RP+ **HILIC+**
Phenylacetaldehyde	HILIC-
Phenylacetic acid	HILIC-
L-Phenylalanine	RP+ **HILIC+**
Phenylpyruvic acid	HILIC+
Phosphorylcholine	NMR
Pipecolic acid	RP- **HILIC+**
L-Proline	RP+
Propanal	HILIC+
Propionic acid	NMR
Putrescine	RP+
Pyridoxamine	RP+
Pyruvic acid	NMR

2,3-Dihydroxypropyl tetradecanoate	RP+ **HILIC+**
Ribitol	HILIC-
Salicylamide	HILIC+
Sarcosine	NMR
L-Serine	RP+
Sorbose	HILIC-
Suberic acid	RP-
Succinic acid	RP-
Taurine	RP-
Lithocholyltaurine	HILIC-
Theobromine	HILIC+
L-Threonine	RP+ **NMR**
Thymidine	**HILIC+** HILIC-
Thyroxine	RP-
Trigonelline	HILIC+
Trimethylamine	HILIC+
Tryptamine	RP+
L-Tryptophan	**RP-** HILIC+ NMR
L-Tyrosine	RP- **HILIC+** NMR
Uracil	**RP+** HILIC+
Uric acid	**RP+** HILIC-
Uridine	**RP-** HILIC+

Urocanic acid	RP+
L-Valine	RP+ **HILIC+** NMR
Xanthine	**RP+** HILIC-
Xanthosine	HILIC-

Milk (Milk :H_2O :MeOH :$CHCl_3$ 1 :3 :4 :6)	
Metabolites	**Analytical platform**
3-Hydroxycapric acid	**RP-** HILIC+
1-Aminocyclopropanecarboxylic acid	RP+
1-Methyladenosine	HILIC+
2,4-Dihydroxypyrimidine-5-carboxylic acid	HILIC-
Gentisic acid	HILIC-
2-Aminophenol	HILIC+
2-Deoxy-D-Glucose	**RP+ HILIC-**
Deoxyguanosine	HILIC-
2-Hydroxybutyric acid	RP-
Ortho-Hydroxyphenylacetic acid	HILIC-
2-Hydroxypyridine	HILIC+
Citraconic acid	HILIC-
Oxoglutaric acid	NMR
Protocatechuic acid	HILIC-

3,4-Dihydroxybenzeneacetic acid	HILIC-
Cyclic AMP	HILIC+
3-(2-Hydroxyphenyl)propanoic acid	**RP-** HILIC-
3-Aminoisobutanoic acid	HILIC-
3-Hydroxymethylglutaric acid	HILIC-
3-Hydroxybutyric acid	NMR
3-Hydroxyphenylacetic acid	HILIC-
3-Methyldioxyindole	RP+
3-Methyladenine	RP+ **HILIC-**
Ureidopropionic acid	RP-
4-Acetamidobutanoic acid	**RP-** HILIC+
Gamma-Aminobutyric acid	**RP+** HILIC+
Kynurenic acid	RP+
4-Hydroxybenzaldehyde	HILIC-
Oxfenicine	RP+
4-Hydroxy-L-proline	**RP+** HILIC+
p-Hydroxyphenylacetic acid	**RP-** HILIC-
Imidazoleacetic acid	HILIC-
Ketoleucine	**RP-** HILIC-
4-Methylcatechol	RP-
5-Aminolevulinic acid	**RP+** HILIC+
5-Hydroxyindoleacetic acid	HILIC+

5-Hydroxylysine	RP+
5'-Methylthioadenosine	**RP+** HILIC+
Pyroglutamic acid	RP-
L-Fucose	HILIC-
Acetamide	NMR
Acetic acid	NMR
Acetoacetic acid	**HILIC-** NMR
N-Acetyl-L-alanine	RP+
L-Acetylcarnitine	**HILIC+** NMR
N-Acetylgalactosamine	HILIC-
N-Acetyl-D-glucosamine	HILIC+
Acetylglycine	RP+
N-Acetylleucine	HILIC-
N-Acetylmannosamine	RP+ **HILIC-**
N-Acetyl-L-methionine	RP-
N-Acetylneuraminate	RP-
N-Acetyl-L-phenylalanine	**RP-** HILIC-
N-Acetylputrescine	HILIC+
N-Acetylserine	RP-
N-Acetylserotonin	HILIC+
N-acetyltryptophan	RP-
Adenine	HILIC+

Adenosine	**RP+** HILIC-
ADP-glucose	RP-
Adenosine monophosphate	RP-
Allantoin	RP- **HILIC+**
Allose	HILIC-
L-Allothreonine	HILIC+
Galactose 1-phosphate	HILIC+
Alpha-D-Glucose	**RP-** HILIC- NMR
Anserine	RP+
L-Arabitol	RP- **HILIC-**
L-Arginine	**RP+** HILIC-
L-Asparagine	**RP+** HILIC-
L-Aspartic acid	RP-
Azelaic acid	RP-
Benzyl alcohol	HILIC-
Beta-Alanine	**RP+** HILIC+ NMR
Betaine	**RP+** HILIC+
Biotin	**RP-** HILIC+
Butanal	HILIC+
L-Carnitine	**RP+** HILIC+
Pyrocatechol	RP-
Cellobiose	HILIC-

Cholic acid	HILIC-
Choline	RP+ **NMR**
trans-Cinnamic acid	RP+
Citramalic acid	HILIC+
Citric acid	NMR
Citrulline	**RP-** HILIC-
Corticosterone	RP+
Cortisol	RP+
Cortisol-21-acetate	RP+
Creatine	RP+ HILIC+ **NMR**
Phosphocreatine	NMR
Creatinine	**RP-** HILIC+
Cyclohexane-1,2-Diol	HILIC+
Cytosine	RP+ **HILIC+**
4-Trimethylammoniobutanoic acid	**RP+** HILIC+
Deoxycytidine	HILIC+
Dethiobiotin	RP+
Diacetyl	HILIC+
L-Dihydroorotic acid	RP+
Dimethyl sulfone	NMR
N,N-Dimethylphenethylamine	HILIC+
Dimethylamine	NMR

Epinephrine	RP+
Ethanolamine	HILIC+
O-Phosphoethanolamine	RP-
Ethylmalonic acid	HILIC-
trans-Ferulic acid	**RP-** HILIC+
Formic acid	NMR
Formyl-L-methionylpeptide	RP-
D-Fructose	HILIC-
Fumaric acid	NMR
Galactaric acid	RP-
Galactitol	HILIC-
D-Galactosamine	RP+
D-Galactose	HILIC-
Gluconic acid	RP-
Glucosamine 6-phosphate	RP-
Glucose 6-phosphate	RP-
L-Glutamic acid	**RP+** HILIC-
L-Glutamine	**RP+** HILIC+
Glyceraldehyde-3-phosphatediethylacetal	RP+
Glyceric acid	RP-
Glycerol 3-phosphate	HILIC+
Glycerophosphocholine	**RP+** HILIC+

Glycine	**RP+** HILIC+
Glycocholic acid	RP-
Glycolaldehyde dimer	HILIC+
Guanidoacetic acid	RP+ **HILIC+**
Guanine	HILIC+
Guanosine	**RP+** HILIC+
Heptanoic acid	HILIC+
Caproic acid	HILIC+
Hippuric acid	**RP-** HILIC+ NMR
L-Histidine	RP-
Homogentisic acid	HILIC+
L-Homoserine	**RP-** HILIC-
Homovanillic acid	HILIC-
Hypotaurine	HILIC-
Hypoxanthine	RP- **HILIC+**
Indoleacetic acid	RP+
Tryptophanol	RP+ **HILIC-**
Indolepyruvate	HILIC+
Indoxyl sulfate	**RP-** HILIC-
Inosine	**RP-** HILIC+ NMR
L-Isoleucine	HILIC+
N6-(delta2-Isopentenyl)-adenine	RP-

L-Kynurenine	RP+
L-Lactic acid	RP- HILIC- **NMR**
Alpha-Lactose	NMR
L-Leucine	HILIC+
Lumichrome	**RP+** HILIC-
L-Lysine	RP+
L-Malic acid	RP-
Mannitol	**RP-** HILIC-
D-Mannose	HILIC+
Melatonin	HILIC+
L-Methionine	RP-
Methyl beta-D-glucopyranoside	HILIC-
Methylguanidine	HILIC+
3-Methylhistidine	RP+
Methylmalonic acid	RP+
N-Methyltryptamine	**RP+** HILIC+
myo-Inositol	RP+ HILIC- **NMR**
N-Trimethyllysine	**RP+** HILIC+
Niacinamide	**RP+** HILIC+
Nicotinic acid	RP-
L(-)-Nicotine	HILIC+
N-Methylhydantoin	NMR

L-Norleucine	HILIC+
Phenylacetylglycine	NMR
Acetylcholine	NMR
12-Hydroxydodecanoic acid	RP-
Ornithine	RP+ **HILIC-**
Orotic acid	HILIC- **NMR**
Palatinose	HILIC-
Pantolactone	HILIC-
Pantothenic acid	**RP-** HILIC+
Paraxanthine	HILIC+
L-Phenylalanine	**RP-** HILIC+
Phosphorylcholine	RP+ **NMR**
Pipecolic acid	HILIC+
L-Proline	RP+ **HILIC+**
Propanal	HILIC+
Pterin	RP+
Pyrrole-2-carboxylic acid	HILIC+
Ribitol	HILIC-
Riboflavin	HILIC+
Glucaric acid	RP-
Salicyluric acid	HILIC+
Sarcosine	HILIC+

L-Serine	**RP-** HILIC+
Sorbitol	HILIC-
L-Sorbose	HILIC-
Suberic acid	RP-
Succinic acid	RP- HILIC+ **NMR**
Sucrose	HILIC-
D-Tagatose	HILIC-
Taurine	RP- **HILIC+** NMR
Lithocholyltaurine	HILIC-
Theobromine	HILIC+
Theophylline	HILIC+
L-Threonine	**RP+** HILIC-
Thymidine	HILIC-
Thymine	**RP+** HILIC+
trans-Aconitic acid	RP- **HILIC-** NMR
Trigonelline	**RP+** HILIC+
Trimethylamine	HILIC+
L-Tryptophan	**RP-** HILIC+
L-Tyrosine	**RP+** HILIC+
Uracil	**RP-** HILIC+
Uric acid	**RP+** HILIC-
Uridine	**RP-** NMR

Uridine diphosphategalactose	HILIC- **NMR**
Uridine diphosphate glucose	HILIC- **NMR**
Uridine diphosphate-N-acetylgalactosamine	HILIC-
Uridine-5-diphospho-n-acetylglucosamine	HILIC- **NMR**
Urocanic acid	**RP+** HILIC-
L-Valine	RP+
Xanthine	**RP-** HILIC-
D-Xylitol	HILIC-

FECES (MeOH : H_2O 4 :1)	
Metabolites	**Analytical platform**
3-Hydroxycapric acid	RP+
1-Aminocyclopropanecarboxylic acid	**RP+** HILIC +
1-Methyladenosine	**RP+** HILIC +
1-Naphthylamine	HILIC+
Glyceryl monoacetate	HILIC+
2',4'-Dihydroxyacetophenone	HILIC-
2,4-Dihydroxypyrimidine-5-carboxylic acid	HILIC-
Gentisic acid	**RP- HILIC-**
Diaminopimelic acid	RP+

2,6-Dihydroxypyridine	HILIC+
2-Acetamido-2-deoxy-beta-glucosylamine	**RP+** HILIC-
2-Aminoisobutyric acid	HILIC+
Ciliatine	RP-
2-Aminophenol	**RP+** HILIC-
Deoxyadenosine	RP- **HILIC+**
dCMP	RP-
2-Deoxy-D-Glucose	**RP+** HILIC-
Deoxyguanosine	HILIC-
2-Hydroxybutyric acid	RP-
Ortho-Hydroxyphenylacetic acid	HILIC- NMR
2-Hydroxypyridine	HILIC+
2-Methylglutaric acid	HILIC-
Citraconic acid	HILIC-
Quinaldic acid	HILIC-
Protocatechuic acid	HILIC-
3,4-Dihydroxybenzeneacetic acid	HILIC-
L-Dopa	**RP+** HILIC+
3-(2-Hydroxyphenyl)propanoic acid	**RP-** HILIC-
Hydroxyphenyllactic acid	RP-
4-Hydroxyphenylpyruvic acid	RP+

21-Tetrahydroxy-5-alpha-pregnan-20-one	**RP+** HILIC-
Dihydroxy-beta-cholanate	HILIC-
Lithocholic Acid	RP-
3-Amino-4-hydroxybenzoic acid	HILIC+
3-Amino-5-hydroxybenzoic acid	RP+
3-Aminoisobutanoic acid	HILIC-
3-Dehydroshikimate	RP-
3-Hydroxyanthranilic acid	**RP+** HILIC+
3-Hydroxybenzaldehyde	HILIC-
3-Hydroxybenzoic acid	RP- **HILIC-**
Hydroxykynurenine	RP+
3-Hydroxyphenylacetic acid	**RP-** HILIC-
Vanillylmandelic acid	**RP-** HILIC+
3-Methoxytyrosine	HILIC+
3-Methoxytyramine	RP+ **HILIC-**
3-Methylglutaric acid	**RP- HILIC-**
3-Methyldioxyindole	RP- **HILIC+**
3-Methyladenine	RP+ **HILIC+**
3-Nitrotyrosine	RP-
3-Sulfinoalanine	RP-
Ureidopropionic acid	**RP+** HILIC+

4-Acetamidobutanoic acid	HILIC-
p-Aminobenzoic acid	HILIC-
4-Guanidinobutanoic acid	**RP+** HILIC+
Kynurenic acid	**RP+** HILIC+
Vanylglycol	HILIC+
4-Hydroxybenzaldehyde	HILIC-
4-Hydroxybenzoic acid	HILIC-
Oxfenicine	RP+
4-Hydroxy-L-proline	**RP+** HILIC+
p-Hydroxyphenylacetic acid	**RP+** HILIC-
Imidazoleacetic acid	RP+ **HILIC+**
Ketoleucine	RP- **HILIC-**
Dihydrouracil	**RP+** HILIC+
5-Aminolevulinic acid	RP+
5-Aminopentanoic acid	RP+ **HILIC+**
5'-Deoxyadenosine	HILIC+
5-Hydroxyindoleacetic acid	**RP+** HILIC-
5-Hydroxylysine	RP+
5-Hydroxymethyluracil	RP-
5-Hydroxy-L-tryptophan	RP-
5-Methylcytosine	**RP+ HILIC+**

5'-Methylthioadenosine	RP+ **HILIC+**
Pyroglutamic acid	RP- **HILIC-**
Valerolactone	HILIC+
Pimelic acid	HILIC-
Acetamide	NMR
Acetic acid	NMR
Acetoin	HILIC-
N-Acetyl-L-alanine	RP+
N-Acetyl-L-aspartic acid	RP-
L-Acetylcarnitine	**HILIC+** NMR
N-Acetylgalactosamine	HILIC-
N-Acetyl-D-glucosamine	**RP-** HILIC+
N-Acetylglutamic acid	RP+
Acetylglycine	RP-
N-Acetylleucine	RP+
N-Acetylmannosamine	**RP+ HILIC+**
N-Acetyl-L-methionine	RP- **HILIC+**
N-Acetylneuraminate	RP+
N-Acetyl-L-phenylalanine	RP+ **HILIC+**
N-Acetylputrescine	RP+ **HILIC+**
N-Acetylserine	RP+

N-Acetylserotonin	**RP-** HILIC+
N-acetyltryptophan	RP+
Adenine	**RP+** HILIC+
Adenosine	RP+ **HILIC+**
Cyclic AMP	HILIC-
S-Adenosylhomocysteine	HILIC+
Adipic acid	**RP-** HILIC-
Agmatine	RP+
L-Alanine	HILIC+
Allantoin	RP-
L-Allothreonine	HILIC+
Alpha-D-Glucose	HILIC-
Alpha-Tocopherol	HILIC+
Aminoadipic acid	**RP+** HILIC+
Anserine	HILIC+
2-Aminobenzoic acid	HILIC-
L-Arabitol	RP-
L-Arginine	**RP+ HILIC+**
L-Asparagine	**RP+** HILIC-
L-Aspartic acid	RP+
Azelaic acid	**RP-** HILIC-

Benzoic acid	HILIC-
Benzyl alcohol	HILIC-
Beta-Alanine	**RP+** HILIC+ NMR
Betaine	**RP+** HILIC+ NMR
Biliverdin	HILIC+
Biotin	RP+ **HILIC+**
Butyric acid	HILIC- **NMR**
Cadaverine	RP+ **HILIC+**
Caffeic acid	RP-
Caprylic acid	HILIC-
L-Carnitine	**RP+** HILIC+
Carnosine	RP+ **HILIC+**
Pyrocatechol	**RP-** HILIC-
Chenodeoxycholic acid	HILIC-
Cholic acid	RP+
7-Dehydrocholesterol	HILIC+
Cholesteryl acetate	HILIC+
Choline	RP+ **NMR**
trans-Cinnamic acid	**RP+** HILIC-
Citramalic acid	**RP-** HILIC+
Citrulline	**RP+** HILIC+

Corticosterone	HILIC+
Cortisol	RP+
Cortisol-21-acetate	RP-
Cortisone	HILIC-
Creatine	RP+
Creatinine	**RP+** HILIC+
Cyclohexane-1,2-Diol	HILIC-
L-Cystathionine	RP-
Cysteic acid	**RP-** HILIC-
Cytidine	**RP+** HILIC+
Cytidine2',3'-cyclicphosphate	RP-
Cytosine	RP+ **HILIC+**
4-Trimethylammoniobutanoic acid	**RP+** HILIC+
Deoxycholic acid	HILIC-
Deoxycorticosterone acetate	RP+
Deoxycytidine	RP- **HILIC+**
Deoxyuridine	HILIC+
Diacetyl	HILIC+
Diethanolamine	**RP+** HILIC+
4-Heptenal diethyl acetal	HILIC-
L-Dihydroorotic acid	RP-

3,4-Dihydroxymandelicacid	RP-
N,N-Dimethylphenethylamine	HILIC - **NMR**
Dimethylbenzimidazole	RP+
Elaidic acid	HILIC-
Epinephrine	RP+
17a-Estradiol	HILIC+
Ethanolamine	HILIC+
O-Phosphoethanolamine	RP+
Ethyl-3-indole acetate	HILIC-
Ethylmalonic acid	**RP-** HILIC-
trans-Ferulic acid	HILIC-
Formic acid	NMR
N-Formylglycine	RP+
Formyl-L-methionylpeptide	RP+
Fumaric acid	RP-
Galactitol	RP+
D-Galactosamine	RP+
Gluconic acid	RP-
2-Amino-2-deoxy-D-gluconate	RP-
Glucosamine	RP- **HILIC+**
Glucosamine 6-phosphate	RP-

Glucose 6-phosphate	RP-
D-Glucuronic acid	RP-
L-Glutamic acid	**RP+** HILIC+
L-Glutamine	**RP+** HILIC-
Glyceraldehyde-3-phosphatediethylacetal	RP-
Glyceric acid	RP-
Glycerol	NMR
Glycerophosphocholine	RP+ **HILIC+**
Glycine	**RP+** HILIC+ NMR
Glycocholic acid	RP-
Guanidoacetic acid	HILIC+
Guanine	RP- **HILIC+**
Guanosine	RP+ **HILIC+**
Cyclic GMP	RP-
Caproic acid	HILIC-
Hippuric acid	RP- **HILIC+**
Histamine	RP+ **HILIC+**
L-Histidine	**RP+** HILIC+
L-Histidinol	HILIC+
Homocysteine	RP+
Homogentisic acid	RP+

L-Homoserine	RP-
Homovanillic acid	HILIC-
Hydroquinone	HILIC-
Hypoxanthine	**RP+** HILIC+
Indoleacetaldehyde	HILIC+
1H-Indole-3-acetamide	RP+ **HILIC+**
Indoleacetic acid	RP+ **HILIC-**
Tryptophanol	HILIC+
Indolepyruvate	HILIC-
Indoxyl sulfate	HILIC-
Inosine	**RP-** HILIC-
Isobutyric acid	HILIC- **NMR**
Isocitric acid	RP-
L-Isoleucine	HILIC+ **NMR**
N6-(delta2-Isopentenyl)-adenine	RP- **HILIC+**
Isovaleric acid	NMR
L-Kynurenine	RP-
L-Lactic acid	RP- HILIC- **NMR**
L-Leucine	HILIC+
L-Norleucine	RP+
Leukotriene B4	**RP+** HILIC-

Gamma-Linolenic acid	HILIC+
Lumichrome	**RP+** HILIC+
L-Lysine	**RP+** HILIC+
D-Lyxose	**RP+** HILIC-
L-Malic acid	RP-
Maleic acid	HILIC-
Malonic acid	RP-
D-Maltose	RP-
Mannitol	RP-
D-Mannose	HILIC+
L-Methionine	**RP+** HILIC+
Methyl beta-D-glucopyranoside	HILIC-
Methylguanidine	HILIC+
3-Methylhistidine	RP+
Indole-3-methyl acetate	**RP+** HILIC+
N-Methyltryptamine	**RP+** HILIC+
Mono-methylglutarate	HILIC-
myo-Inositol	HILIC-
N-Alpha-acetyllysine	**RP-** HILIC+
N-Trimethyllysine	**RP+** HILIC+
Niacinamide	**RP+** HILIC+

Nicotinic acid	**RP+** HILIC+
Norepinephrine	RP+
L-Norvaline	RP+ **HILIC+**
Acetylcholine	NMR
Oleic acid	RP+ **HILIC+**
12-Hydroxydodecanoic acid	**RP+** HILIC-
Ornithine	RP+ **HILIC+**
Orotic acid	RP+ **HILIC-**
Pantolactone	HILIC-
Pantothenic acid	**RP+ HILIC+**
Paraxanthine	HILIC+
Valeric acid	NMR
Petroselinic acid	HILIC-
Phenylacetaldehyde	HILIC-
L-Phenylalanine	**RP+** HILIC+ NMR
2-Hydroxyphenethylamine	HILIC+
Phosphorylcholine	NMR
DL-O-Phosphoserine	RP-
Picolinic acid	HILIC+
Pipecolic acid	RP+
L-Proline	**RP+** HILIC+

Propanal	HILIC+
Propionic acid	NMR
Pterin	**RP+** HILIC+
Purine	RP+ **HILIC+**
Putrescine	RP+
Pyridoxamine	RP+
Pyridoxine	RP+
Pyrrole-2-carboxylic acid	**RP+** HILIC-
Quinic acid	RP-
2,3-Dihydroxypropyl tetradecanoate	RP+
Cortexolone	HILIC+
Resorcinol monoacetate	HILIC-
All-trans-retinoic acid	RP+ **HILIC+**
Vitamin A	HILIC+
Rhamnose	HILIC-
Ribitol	HILIC-
Riboflavin	RP+ **HILIC+**
D-Ribose	HILIC-
Rosmarinic acid	HILIC-
Salicylamide	**RP- HILIC-**
Salicyluric acid	HILIC+

Sarcosine	HILIC-
L-Serine	**RP- HILIC+**
Serotonin	RP+ **HILIC+**
Shikimic acid	HILIC+
Spermidine	RP+ **HILIC+**
Sphinganine	RP+
Suberic acid	**RP-** HILIC-
Succinic acid	**RP-** HILIC+ NMR
Taurine	**RP- HILIC-**
L-Threonine	**RP+** HILIC+ NMR
Thymidine	RP- **HILIC-**
Thymine	RP+ **HILIC-**
trans-Aconitic acid	RP-
Trigonelline	**RP+** HILIC+
Trimethylamine	HILIC+
Tryptamine	HILIC+
L-Tryptophan	**RP+** HILIC+
Tyramine	HILIC+
L-Tyrosine	**RP+** HILIC- NMR
Uracil	RP+ **HILIC-**
Uracil-5-carboxylic acid	RP+

Uric acid	RP+
Uridine	**RP+** HILIC+
5'UMP	HILIC+
Urocanic acid	RP+ **HILIC+**
L-Valine	**RP+ HILIC+ NMR**
Ergocalciferol	HILIC+
Xanthine	**RP+** HILIC-

II. Conclusion générale et perspectives

Ces deux études présentent le développement d'une méthode analytique standardisée multi-matrice et multiplateforme permettant la description large d'un métabolome. Nos études permettent de couvrir une large gamme de métabolites. Cette stratégie d'empreinte métabolomique permet de valider une préparation unique pour chaque matrice et a permis de voir la complémentarité entre matrices et plateformes. Cette approche permet d'avoir le métabolome type d'un individu sain pouvant être comparé à deux ou plusieurs groupes ayant des métabolomes différents. Ce métabolome est visualisable grâce à une carte métabolomique chez l'humain et sous forme d'atlas chez le caprin. Une modification de ce profil métabolique peut permettre une détection précoce d'un dérèglement métabolique pouvant indiquer une maladie. Cette approche peut être complétée par la mise en œuvre d'une méthode quantitative afin de suivre un individu tout au long de sa vie. En effet, actuellement les comparaisons entre deux groupes d'individus se fait par quantification relative nécessitant de travailler en campagne d'échantillon. Le développement de quantification absolue facilitera l'utilisation de ces méthodes en routine clinique et en élevage.

Chapitre 3 : Quantification de métabolites par dérivation chimique

I. Introduction

1. Intérêt de la dérivation chimique en spectrométrie de masse

Les approches actuelles de la métabolomique basée sur la LC-MS peuvent généralement être classées en métabolomique ciblée et non ciblée. Lors d'études ciblées, un nombre défini de métabolites est sélectionné à l'avance connaissant leurs ions précurseurs et leurs schémas de fragmentation. Cette approche, en combinaison avec l'utilisation d'étalons internes isotopiques, est couramment utilisée pour la quantification absolue précise de métabolites. En revanche, la métabolomique non ciblée ne nécessite aucune information préalable, elle est utilisée pour explorer de manière globale le métabolome. Les approches non ciblées sont plus proches du but de la métabolomique, c'est-à-dire fournir en même temps des informations qualitatives et quantitatives pour le plus de métabolites possibles. Cependant, la réalisation de cet objectif est très difficile. Cette tâche est entravée par l'extrême complexité du métabolome. Premièrement, les métabolites aux propriétés physiques et chimiques diverses coexistent dans les échantillons biologiques, leur détection peut être limitée par un seul instrument d'analyse. Ainsi les combinaisons de plateformes et/ou de méthodes instrumentales sont nécessaires pour le profilage complet du métabolome (figure 22 (A)). Cela introduit un temps et un coût d'analyses supplémentaires.

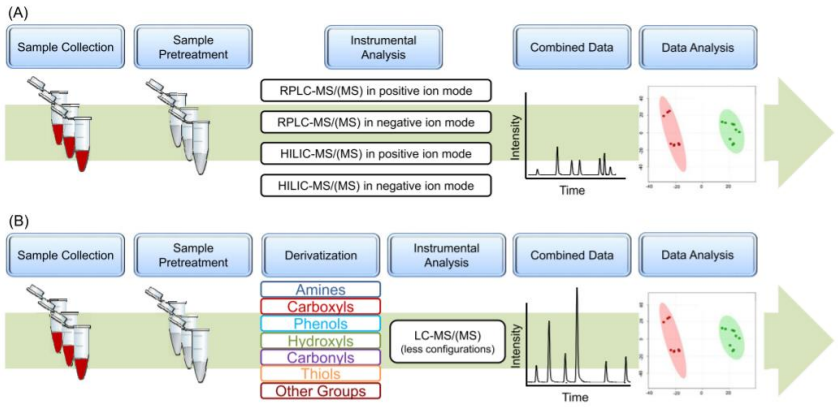


Figure 22 : Workflow général (A) utilisé lors d'étude métabolomique par une analyse LC-MS et par dérivation chimique en LC-MS (B) [124]

Deuxièmement, une large gamme de concentrations de métabolites (du pM au mM) dans les échantillons complexifie la détection et la quantification de tous les métabolites simultanément. La détection et la quantification des métabolites peu abondants sont très difficiles, surtout si les métabolites ont de faibles réponses lors de leur analyse. Troisièmement, les échantillons biologiques sont généralement très précieux (qu'ils proviennent de modèles animaux ou de volontaires/patients en cliniques). Pour plusieurs types d'échantillons, la quantité d'échantillon peut être très limitée. Le liquide céphalorachidien (LCR) est l'un des meilleurs échantillons reflétant les activités cérébrales. Cependant, la collecte du LCR nécessite des opérations minutieuses et professionnelles. Le volume qui peut être collecté est très faible chez l'animal (environ 10µl chez une souris) nécessitant des méthodes très sensibles pour son analyse. Enfin, bien que de nombreuses bases de données de métabolites ont été construites [9], l'identification et la quantification rapide des métabolites connus et inconnus en métabolomique sont encore difficiles en raison de la disponibilité limitée d'étalon et de la complexité des matrices. La dérivation chimique a été développée pour contrer cette difficulté. Pour se faire, l'ensemble des métabolites contenus dans le métabolome sont classés en plusieurs groupes en fonction de leurs groupements chimiques (carboxyle, carbonyle, amine, phénol, alcool, thiol ...). Différentes étiquettes chimiques sont utilisées pour cibler spécifiquement différents sous-métabolomes et ainsi fournir une meilleure couverture de l'ensemble de la gamme de composés d'intérêt (figure 22 (B)) [124].

Ce type d'approche permet d'aider à l'identification de métabolites inconnus grâce à leur caractérisation spectrale en MS [142, 162]. L'utilisation en simultanée d'étiquettes chimiques s'avère efficace pour le profilage du métabolome de différents groupes de métabolites. Par exemple, l'étude multi étiquettes, chlorure de Dansyle, DmPABr et dansyle hydrazine, de Zhao et al. [163] a montré une approche performante pour cibler les métabolites ayant des fonctions amine/phénol, acide carboxylique, carbonyle et alcool. Cette étude a permis la détection d'un total de 7431 pics avec 6109 pics uniques marqués dans le plasma humain. Parmi eux, 670, soit 9%, ont pu être identifiés formellement, 6256 soit 84.3% ont été identifiés de façon putative par les bases de données connues. Ces résultats permettent de montrer que l'analyse de 4 sous-métabolomes permet de détecter une grande quantité de métabolites. La dérivation chimique permet donc d'améliorer la couverture du métabolome et donc de permettre la découverte de nouveaux biomarqueurs lors d'approches non ciblées.

2. A quoi sert la dérivation chimique ?

La dérivation chimique est utilisée en chromatographie gazeuse (GC) dans le but de diminuer la polarité des analytes les plus polaires et d'augmenter leur volatilité. En chromatographie liquide (LC), la dérivation chimique a été introduite historiquement pour améliorer les performances de détection ultraviolet-visible et fluorescence par l'introduction de groupes fonctionnels tels que les chromophores et fluorophores, respectivement [164, 165]. L'introduction des spectromètres de masse a entrainé de nouveau challenge quant à son couplage avec les techniques séparatives plus particulièrement pour la LC. En effet, les étiquettes chimiques préalablement utilisées ne sont pas toutes compatibles avec les phases mobiles HPLC qui contiennent généralement des sels. Ainsi, la plupart des analyses LC-MS sont aujourd'hui réalisées sans dérivation chimique. Toutefois, les analyses avec dérivation chimique présentent plusieurs avantages analytiques par rapport à une

analyse LC-MS sans dérivation. Elle permet d'améliorer la séparation, la détection, la stabilité et la quantification des composés lors d'approches LC-MS.

2.1. Séparation LC

La suppression d'ions est un obstacle à l'analyse de mélanges complexes utilisant la spectrométrie de masse, en particulier avec la source électro spray (ESI)-MS. Les espèces chimiques se font concurrence pour atteindre la surface des gouttelettes ESI et s'ioniser. Ce processus de compétition peut altérer l'efficacité d'ionisation de nombreux métabolites dans des matrices complexes. La chromatographie en phase liquide est généralement utilisée pour l'analyse du métabolome dans le but de réduire l'effet de suppression d'ions. Généralement les chromatographies liquides en phase inverse (RPLC) et à interaction hydrophile (HILIC) sont utilisées seules ou combinées (figure 22 (A)). Cependant, chaque approche apporte des limites quant à la séparation des composés, par exemple la RPLC ne permet pas la séparation des composés très polaires ou ioniques. La stratégie de la dérivation s'avère efficace pour surmonter ces problèmes et améliorer la capacité de séparation en introduisant un fragment chimique au métabolite. Ce fragment chimique appelé étiquette chimique peut contenir une partie hydrophobe telle qu'un cycle aromatique (par exemple, chlorure de dansyle [166, 167]) et permet de changer la polarité des groupements chimiques du métabolite et son hydrophobie. Cette modification entraine une meilleure rétention et forme de pic sur une colonne RPLC. L'équipe de Lkhagva [168] a mené une étude comparative de marquage de 7 acides aminés par 5 étiquettes chimiques ayant un caractère hydrophobe plus grand que les acides aminés. Cette étude a montré qu'en phase inverse la formation de produits dérivés avec une plus grande hydrophobie aide à séparer les analytes cibles permettant ainsi d'avoir des temps de rétention plus grands et distinctifs pour chaque acide aminé. De cette façon, les métabolites polaires au sein d'un échantillon biologique peuvent être séparés efficacement même si seule la configuration RPLC est utilisée. Ainsi, dans certaines méthodes de dérivation, la

RPLC est généralement suffisante pour analyser le métabolome presque complet (figure 22 (B)). L'amélioration de la séparation HILIC après dérivation a été également étudié. Bawazeer et al. a observé qu'une dérivation des monosaccharides par la d_5-aniline permet une meilleure séparation et forme de pic en HILIC [169]. Avec ce réactif de marquage, ils ont pu effectuer la séparation du fructose, du glucose, galactose et mannose avec une meilleure résolution et forme de pic offrant ainsi une amélioration significative dans l'analyse des monosaccharides.

2.2. Détection

La détection d'espèces chimiques en spectrométrie de masse repose sur l'ionisation de molécules dans la source d'un spectromètre de masse. Une approche métabolomique en LC-MS combine classiquement l'utilisation du mode positif et négatif afin de fournir des informations complémentaires pour le profilage des métabolomes d'un échantillon biologique (figure 22 (A)). L'utilisation des deux modes d'ionisation pour la détection double également le temps d'analyse. La dérivation améliore la détection des métabolites par plusieurs aspects. Tout d'abord, comme décrit dans le paragraphe 2.1 (Séparation LC), l'introduction d'un fragment hydrophobe dans la structure des métabolites augmente leur rétention sur la colonne RPLC. Les métabolites marqués seront alors élués à une composition plus élevée de phase organique, conduisant à une désolvatation plus efficace dans le processus d'ionisation ESI. Deuxièmement, l'ajout d'un groupe de marquage ou d'une étiquette augmente la masse de l'analyte, évitant ainsi les interférences dans la zone des m/z faibles causées par les impuretés et d'autres molécules contenues dans les solvants. Ceci fournit une amélioration significative de la détection des métabolites de faible poids moléculaire, tels que les acides aminés. Enfin et surtout, de nombreuses réactions de dérivation introduisent un fragment chargé en permanence ou facilement ionisable améliorant ainsi l'ionisation des molécules. Pour la détection en mode positif, les fragments contenant des atomes d'azote sont souvent conçus comme des groupes facilement ionisables, tels que la structure

diméthylamino présente dans le chlorure de dansyle (DnsCl), le DMED ou encore le DmPABr. Par exemple, dans le cas du chlorure de dansyle, ce groupement permet d'augmenter la capacité de détection des métabolites amines/phénols de 10 à 1000 fois [170].

2.3. Stabilité

Certaines catégories de métabolites dans les échantillons biologiques telles que ceux contenant des thiols sont réactives et non stables au cours de la préparation d'échantillon et de l'analyse [171]. En effet, le groupe sulfhydrile est très facilement oxydable et peut avoir des interactions avec différentes enzymes et d'autres types de molécules. De plus, un échantillon est considéré stable si sa détection en MS est améliorée. La fragmentation en source des ions entraine une incertitude quant à la détection des métabolites. Les fragments d'ions produits peuvent être considérés à tort comme les ions moléculaires de métabolites. La dérivation peut améliorer la stabilité des ions, réduisant leur chance de former des fragments d'ions pendant le transport des ions de la source d'ionisation à l'analyseur de masse.

2.4. Quantification

La dérivation chimique permet d'avoir une approche en quantification absolue (figure 23).

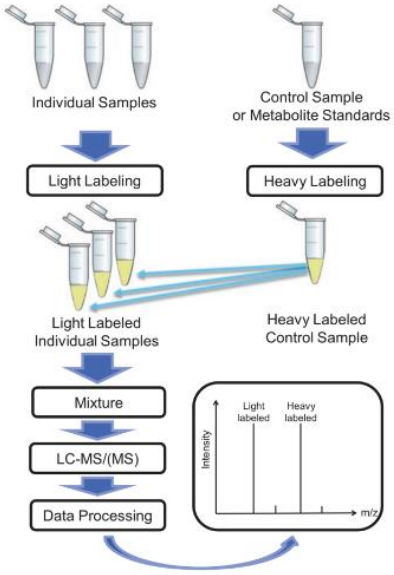


Figure 23 : Workflow général de la dérivation chimique utilisée en LC-MS lors d'approche métabolomique [172]

Les échantillons sont dérivés avec une étiquette légère et des standards ou un échantillon de groupe contrôle sont dérivées avec une étiquette lourde. Ensuite, les deux échantillons dérivés sont mélangés et injectés en LC-MS pour analyse. Étant donné que le dérivé léger et le dérivé lourd ont des propriétés physico-chimiques presque identiques, ils sont élués au même temps de rétention. Les dérivés lourds sont utilisés comme étalon interne pour effectuer la quantification des métabolites contenus dans l'échantillon biologique. Ce type d'approche a permis d'identifier des biomarqueurs [167, 173].

II. Dérivation des fonctions chimiques

1. Fonction acide carboxylique

Les composés contenant une fonction acide carboxylique représentent un grand nombre de biomolécules telles que les acides organiques, les acides gras qu'ils soient à chaîne courte, moyenne ou longue et certains métabolites tels que les acides aminés ou les stéroïdes. De plus ces fonctions sont contenues dans divers

composés pharmaceutiques comme les analgésiques, mais également dans les aliments et les boissons tels que les acides hydroxy-benzoïques et les acides gras. La détection d'acides carboxyliques non dérivés par LC-MS est généralement menée en mode ions négatifs, qui a une sensibilité inférieure et plus d'interférences que le mode ion positif, en particulier dans des matrices complexes comme les échantillons biologiques. De plus, il est souvent nécessaire d'utiliser deux types de colonnes LC pour les séparer efficacement. L'analyse en phase inverse permet de séparer tous les acides apolaires par exemple les acides gras, mais ne permet pas de séparer les molécules polaires tels que les acides aminés ou les dérivés d'acide benzoïque. Pour une meilleure séparation de ces composés il est recommandé de les dériver avec un groupement apolaire. A ce jour, il existe de nombreuses étiquettes chimiques et stratégies pour la détection des fonctions acides en spectrométrie de masse qui sont décrit dans des revues [124, 174].

La dérivation chimique permet d'améliorer la séparation des métabolites en phase inverse et leur ionisation en utilisant notamment l'ESI positive. Ainsi, divers agents de dérivation ont été utilisés pour former davantage de dérivés ionisés et hydrophobes avec une sensibilité grandement améliorée en mode d'ionisation positive [175].

1.1. Dérivés de pipérazine et amine : couplage en deux étapes

Les dérivés pipérazine et amine dérivent une fonction acide carboxylique en formant une liaison amide. La formation d'une liaison amide s'effectue en deux étapes. La première étape de cette réaction est l'activation du groupement carboxyle par un agent de couplage. La deuxième étape consiste à l'ajout de l'étiquette chimique par réaction nucléophile.

1.1.1. Première étape : Activation du groupement carboxyle

Le couplage entre une fonction acide carboxylique et une amine implique une déshydratation qui n'a pas lieu spontanément à température ambiante. Toutefois,

par une réaction acido-basique, l'amine peut se protoner en présence de l'acide et en l'absence de tout autre réactif dans le milieu réactionnel (voir figure 24 (A)). C'est pourquoi il est nécessaire d'activer la fonction acide carboxylique par conversion du -OH en un groupement activateur (composé X). L'activation de l'extrémité du groupement carboxyle du métabolite (composé X) par un agent de couplage permet à l'extrémité N terminale de l'étiquette chimique de faire une attaque nucléophile (figure 24 (B)).

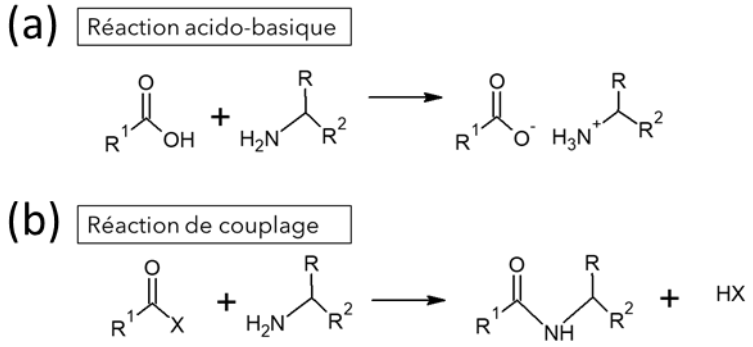


Figure 24 : Réactions entre la fonction acide carboxylique et l'étiquette chimique lors d'une réaction acido-basique (A) et lors de la formation d'une liaison amide (B).

Cette étape d'activation est souvent la plus critique car elle dépend de l'agent de couplage utilisé. Dans la littérature, le couplage en deux étapes du marquage d'une fonction acide carboxylique est décrit pour la synthèse peptidique. Toutefois, ce type de protocole est utilisé lors d'études métabolomiques [43, 44, 176, 177]. Le tableau 1 décrit de façon non exhaustive quelques agents de couplage. Il existe un très grand nombre d'agents de couplage décrits dans la littérature. De nombreuses revues ont été écrites à ce jour [178, 179] .

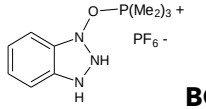
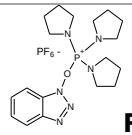
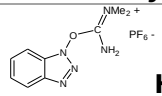
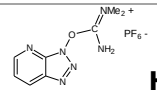
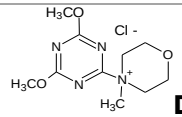
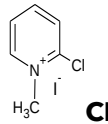
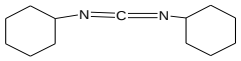
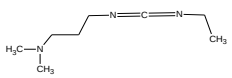
Catégorie	Agent de couplage	Conditions réactionnelles	Commentaires	Agent anti-épimérisation (AAE)
Sels d'phosphonium	 **BOP**	Température ambiante 1 équivalent pour 2 équivalents de base [180]	-Soluble dans tous les solvants -Sous-produit toxique -Cyclisation des acides aminés	Ajout d'AAE pour accélérer la réaction
	 **PyBOP**		-Version non toxique du BOP -Cyclisation des acides aminés	
Sels d'uronium	 **HBTU**	1 équivalent pour 1 équivalent de base [181]	Réactivité similaire que HATU	-Structure chimique semblable aux AAE -Pas besoin d'AAE
	 **HATU**		Plus efficace que BOP	
Autres	 **DMT-MM**	Température ambiante 5min sans base [44]	Utilisé lors de l'étude métabolomique [44]	Non renseigné
	 **CMPI**	1équivalent pour 2 équivalents de base 40°C à 60min [177]	Utilisé lors de l'étude métabolomique [177]	Non renseigné
Carbodiimide	 **DCC**	Sur colonne [182]	Phase solide	Besoin d'AAE pour éviter épimérisation
	 **EDC**	1 équivalent et 1équivalent d'AAE 15sec à température ambiante [183]	Soluble dans l'eau	

Tableau 6 : Résumé des principaux agents de couplage utilisés lors du marquage en deux étapes des fonctions acides carboxyliques

Les **sels de phosphonium** ont montré leur efficacité pour l'activation de la fonction acide lors d'un couplage peptidique.

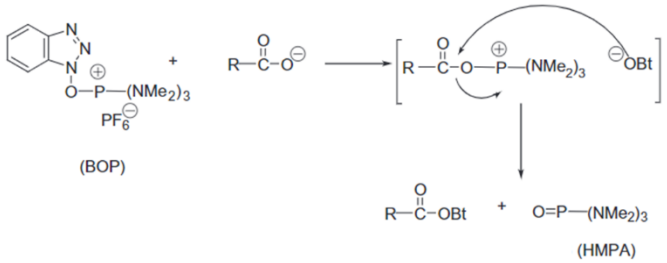


Figure 25 : Schéma réactionnel d'activation d'une fonction acide carboxylique par un sel de phosphonium (BOP) [184]

Le BOP (Benzotriazol-1-yloxy-tris(dimethylamino)-phosphonium hexafluorophosphate) est le premier d'une large gamme d'agents de couplages de sels de phosphonium. Il a été introduit par B. Castro et son équipe en 1975 [182]. Il est un bon agent de couplage et a une bonne solubilité dans la plupart des solvants communs. Toutefois il engendre un sous-produit toxique, le HMPA (hexamethylphosphoramide) (figure 25). Ainsi, en 1990 Merck KGaA a introduit une version non toxique du BOP le PyBOP (Benzotriazol-1-yloxy-tripyrrolidino-phosphonium hexafluorophosphate) dans lequel le motif diméthylamine a été remplacé par une pyrrolidine. Le couplage est réalisé à température ambiante pendant 10 à 60 min à l'aide d'une résine dans laquelle est ajouté 1 équivalent de sels de phosphonium pour 2 équivalents de base (par exemple la N, N-diisopropyléthylamine DIPEA) [180]. La présence dans le milieu d'un excès en HOBt permet d'accélérer la réaction de couplage et de réduire la perte de configuration de l'acide [185]. Les agents de couplage à base de sels de phosphonium sont préférablement utilisés pour des réactions de cyclisation de peptide lorsqu'ils sont présents en excès dans le milieu [186].

D'après la revue bibliographique des sels de phosphonium, cet agent de couplage n'est pas adapté à nos objectifs car nous ne souhaitons pas cycliser nos molécules.

Les **sels d'uronium** sont des agents de couplage couramment utilisés et permettent la réaction de couplage entre l'acide et l'amine d'après le schéma réactionnel décrit en figure 26.

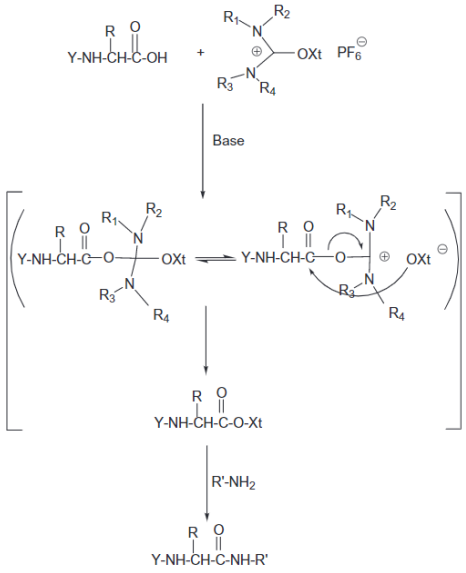


Figure 26 : Schéma réactionnel d'activation d'une fonction acide carboxylique contenu dans un acide aminé par un sel d'uronium (HATU ou HBTU) [184]

Le premier sel d'uronium HBTU (2-(1H-Benzotriazol-1-yl)N,N,N,N-tetramethylaminium hexafluorophosphate) est découvert par Dourtoglou en 1978 [187]. Le HATU (2-(7-Aza-1H-benzotriazol-1-yl)-N,N,N',N'- tetramethylaminium tetrafluoroborate/ hexafluorophosphate) ayant un azote en plus dans son cycle par rapport au HBTU est utilisé pour le couplage d'acide N-méthyl-amine [188]. L'utilisation des sels d'uronium est répandue en phase solide mais également en solution. Leurs sous-produits résultants (structure proche de l'urée) (voir figure 26) sont solubles dans l'eau et dans les solvants organiques standards. Pour une réaction de couplage favorable il faut un équivalent de sels d'uronium dans le milieu pour un équivalent de base par exemple DIPEA [181].

D'après une étude de Ehrlich le HATU est plus efficace que le sel de phosphonium BOP [189]. Cette observation est confirmée par Albericio et Kates [179] qui conclut de son étude que les sels d'uronium sont plus stables que les sels de phosphonium en absence de base.

Ainsi pour notre étude comme nous souhaitons marquer la fonction acide des métabolites nous choisirons le sel d'uronium HATU plutôt que les sels de phosphonium pour tester son efficacité par rapport aux autres agents de couplage.

Il existe également **d'autres types d'agents de couplage** qui sont des « sels d'halogénure » tels que le DMT-MM et le CMPI.

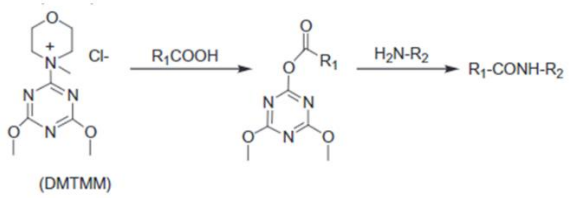


Figure 27 : Réaction d'activation d'une fonction acide carboxylique par le DMT-MM [184]

Le DMT-MM (sel de 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)- 4-methylmorpholinium) est un agent de couplage issu du groupe des triazines qui a été développé par Kaminski et son équipe [190]. Son schéma réactionnel est décrit figure 27. Il est connu pour sa grande efficacité permettant un bon rendement de couplage mais également par son bas niveau de racémisation. Il a été utilisé lors d'une approche métabolomique pour l'activation de la fonction acide pour marquer des acides gras contenus dans la salive [44]. Les conditions réactionnelles se font à température ambiante pendant 5 min, le DMT-MM est dissout dans l'éthanol sans présence de base. Le CMPI (2-Chloro-1-methylpyridinium iodide) est utilisé par Mukaiyama en 1979 [191] comme activateur des fonctions carboxyliques afin de synthétiser des esters, des lactones et des amides. Il est également utilisé de nos jours dans des études métabolomiques notamment pour activer le marquage de la fonction acide des acides gras dans des études faites sur le foie, cœur et cerveau des rats [177]. Les conditions réactionnelles se font à 40°C pendant 1h et 2 équivalents de base (TEA) sont ajoutés pour 1 équivalent de CMPI.

Les recherches bibliographiques actuelles ne permettent pas de privilégier l'utilisation de l'un par rapport à l'autre. Ces deux agents de couplage ont montré leur efficacité lors d'études métabolomiques. Dans le but d'explorer et d'avoir une

première approche de marquage en deux étapes de la fonction acide, nous avons décidé arbitrairement de choisir le CMPI pour la suite de notre étude.

Une autre classe d'agent de couplage sont les **carbodiimides.** Un carbodiimide est un groupe fonctionnel de formule RN =C= NR utilisé comme activateur de fonction depuis des décennies en phase solide et en solution pour la synthèse peptidique.

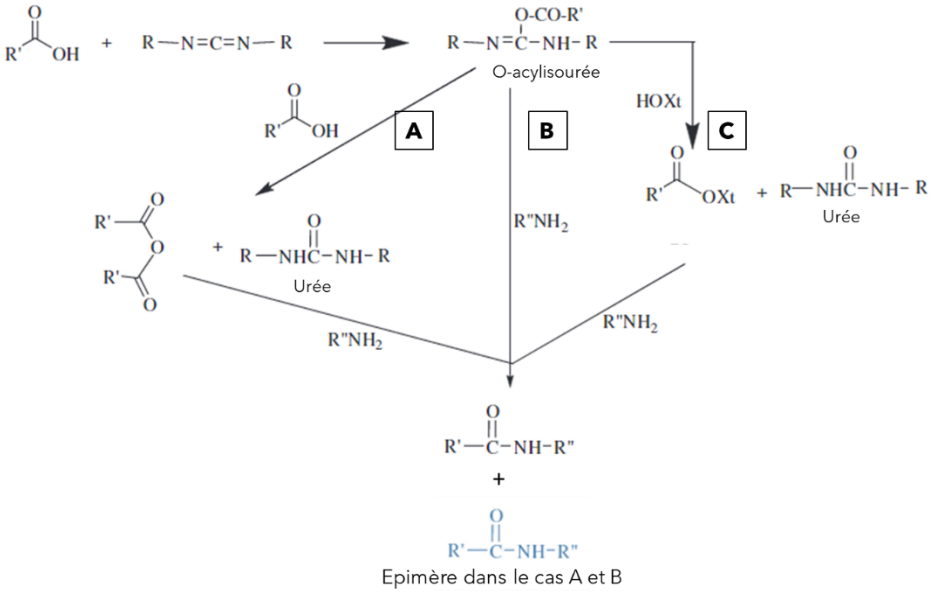


Figure 28 : Schéma réactionnel d'activation d'une fonction acide carboxylique par un carbodiimide lors d'une épimérisation (A), sans AAE (B) avec un AAE (C). Schéma inspiré par [184]

Le DCC (Dicyclohexylcarbodiimide) est utilisé depuis 1955 en phase solide pour la synthèse peptidique [192]. Contrairement au DCC, l'EDC (N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide) est un carbodiimide soluble dans l'eau qui a été désigné spécialement pour les couplages en milieu aqueux [193]. La réaction entre l'acide et le carbodiimide génère une espèce activée appelée O-acylisourée (figure 28 B) qui réagit avec l'étiquette chimique (R-NH_2) pour former la liaison amide et l'urée comme sous-produit. Cependant, d'autres réactions peuvent avoir lieu après la formation de la O-acylisourée. Celle-ci peut réagir avec une seconde molécule d'acide (figure 28 A) pour former l'anhydride symétrique qui réagit ensuite avec l'étiquette chimique pour conduire à la molécule marquée. Cette réaction consomme un équivalent d'acide et peut provoquer une baisse de

rendement de marquage et/ou une épimérisation de l'acide utilisé. Afin de minimiser ces réactions secondaires, ces agents de couplage sont utilisés en parallèle avec un « additif » nommé ici agent anti-épimérisation (AAE) (figure 28 C). La présence d'un AAE de type N-hydroxybenzotriazoles (HOXt) limite les réactions secondaires. Il protone la molécule d'O-acylisourée, empêchant ainsi l'épimérisation et la réaction avec un autre équivalent d'acide. Le rôle des AAE est décrit en 1.1.2. (Problème d'épimérisation/racémisation (ci-dessous)).

1.1.2. Problème d'épimérisation/racémisation

Dans le cas des molécules non planes, par exemple les α-hydroxy acides ou les acides aminés, il y a des risques d'épimérisation. En effet, le fait d'activer l'extrémité possédant la fonction carboxyle peut entrainer une épimérisation. Le proton porté par le carbone asymétrique devient alors plus acide et l'inversion de configuration du carbone α est facilitée (Figure 29). L'ajout d'« additif » appelé ici agent anti-épimérisation (AAE) permet de réduire ce phénomène.

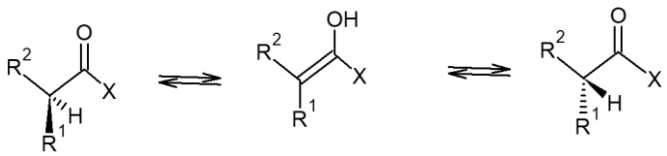
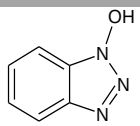
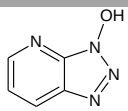
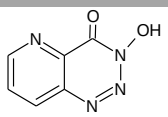
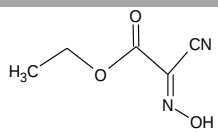


Figure 29 : Schéma d'épimérisation dans le cas d'un acide carboxylique ayant une fonction en position α

La présence d'AAE dans le milieu réactionnel dans le cas des carbodiimides et des sels d'uronium permet d'améliorer l'efficacité du couplage mais aussi de réduire la formation d'épimères dans le milieu.

Tableau 7 : Agents anti-épimérisation (AAE) les plus utilisés lors d'études métabolomiques et de synthèse peptidique

Agent anti-épimérisation (AAE)			
 **HOBt**	 **HOAt**	 **HODhat**	 **Oxyma Pure**
-Explosif en milieu aqueux	-Explosif en milieu aqueux -Plus efficace que HOBt	-Explosif en milieu aqueux -Meilleur rendement que HOBt	-Pas explosif -Réactivité similaire à HOBt et HOAt

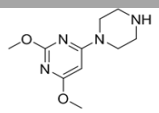
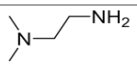
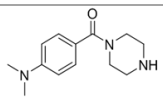
En 1970, König et Geiger [194] découvre le **HOBt** (1-Hydroxybenzotriazole). Cette molécule est l'agent anti-épimérisation (AAE) le plus populaire et reste à ce jour le plus efficace pour la suppression de mélange racémique dû à l'intermédiaire de réaction carbodiimide-acide (O-acylisourée). Le principal désavantage de cet AAE est son caractère explosif en milieu aqueux. L'AAE **HOAt** (1-Hydroxy-7-aza-1H-benzotriazole) créé en 1993 [188] par Carpino, est plus efficace du fait de l'assistance de l'azote pyridinique mais il est beaucoup plus cher à synthétiser. Au début des années 2000, le **HODhat** (Hydroxy-3,4-dihydro-4-oxo-1,2,3- aza-benzotriazine) a été synthétisé, il est beaucoup plus réactif que HOBt et donne généralement de meilleurs rendements et un taux de racémisation inférieur lorsqu'il est couplé avec le EDC [195]. En effet, l'étude comparative de Carpino et al. a montré que le couple EDC/HODhat permet d'avoir un rendement de marquage de 98,6% et un taux de racémisation (LDL-isomère) inférieur à 0,1% or le couple EDC/HOBt a un rendement de 90% et un taux de racémisation de 0,2%. Récemment, **l'Oxyma Pure®** (Ethyl 2-cyano-2-(hydroximino)acetate) (marque déposée par Luxembourg Bio Technologies Ltd) a été développé en tant qu'AAE avec une réactivité similaire au HOBt et HOAt mais en enlevant l'aspect explosif [196]. Ainsi, le fait de rajouter dans le milieu réactionnel des dérivés de type benzotriazole et azabenzotriazole tels que HOBt ou HOAt [188] minimise les réactions secondaires et augmente ainsi l'efficacité du couplage.

L'agent anti-épimérisation (AAE) HOAt est décrit lors d'une étude métabolomique dérivant la fonction acide des acides gras en deux étapes avec l'EDC [183]. De plus l'HOAt, d'après la littérature, est plus réactif que l'HOBt. Ainsi pour la suite de notre étude nous privilégierons l'utilisation de l'HOAt.

1.1.3. Deuxième étape : ajout de l'étiquette chimique

Le groupe amino secondaire de la pipérazine se condense avec les acides carboxyliques pour donner l'amide secondaire correspondant. Ainsi, des étiquettes chimiques ont été conçues pour la dérivation des acides carboxyliques.

Tableau 8 : Etiquettes de dérivation chimique pour le marquage des fonctions acides carboxylique en deux étapes

Etiquettes chimiques	Conditions réactionnelles	Applications
 **DMPP**	Agent de couplage EDC et l'AAE HOAt (voir tableau 1)	Acide gras dans des tissus humains [183]
 **DMED**	Agent de couplage DMT-MM (voir tableau 1)	Acides gras dans la salive [44]
 **DAPPZ**	Agent de couplage CMPI et la base TEA (voir tableau 1)	Acide gras dans le sérum [177] et dans le plasma [176]

La réaction du **DMPP (**2,4-diméthoxy-6-pipérazin-1-yl pyrimidine) avec le groupe carboxylique est très rapide et les produits ont une hydrophobie élevée ce qui améliore la détection en ESI-MS. Le DMPP a été développé pour l'analyse des acides carboxyliques par LC / ESI-MS / MS. Lors d'une approche en infusion directe, le DMPP réagit avec la fonction acide carboxylique des acides gras en 15 sec en présence de l'HOAt et de l'agent de couplage EDC [183] (figure 30).

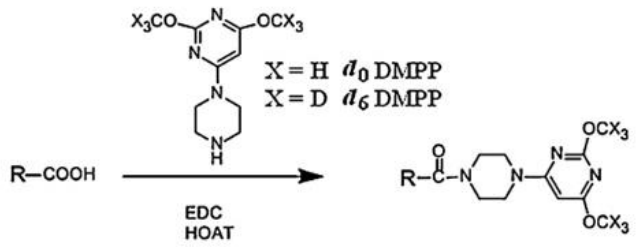


Figure 30 : Schéma réactionnel du DMPP avec une fonction acide [43]

Ainsi, cette étiquette a permis le profilage des acides gras contenus dans des échantillons complexes en améliorant le signal des acides dérivés par le DMPP en ESI-MS. Avec cette approche, des niveaux accrus de 16 acides gras ont été détectés dans les tissus cancéreux par rapport aux tissus sains [183].

La 1 – ((4-diméthylaminophényl) carbonyle) pipérazine (**DAPPZ**) est un autre réactif pipérazine qui est utilisé pour la dérivation chimique d'acides gras (l'acide eicosapentaénoïque salivaire et de l'acide arachidonique) par une analyse LC-MS/MS [44]. Son schéma réactionnel utilisant l'agent de couplage DMT-MM est représenté figure 31.

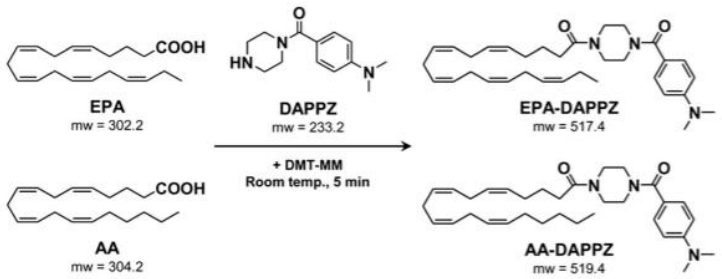


Figure 31 : Schéma réactionnel du DAPPZ avec une fonction acide appartenant à des acides gras [44]

Le DAPPZ a amélioré respectivement la détectabilité de l'acide eicosapentaénoïque et de l'acide arachidonique d'un facteur 100 à 300 par rapport à ces mêmes composés non dérivés [44].

Le **DMED** a été utilisé pour la dérivation d'une série de 19 eicosanoïdes, y compris les acides hydroxy, dihydroxy et époxy eicosatétraénoïque pour leur quantification dans des échantillons de sérums humains provenant de patients diabétiques en

utilisant l'UPLC-MS/MS. Par rapport à une étude sans dérivation des acides gras, cette étiquette a permis d'avoir un effet matrice diminué et une sensibilité améliorée (5 à 138 fois) [177]. Cette étiquette chimique réagit avec la fonction acide carboxylique selon le schéma réactionnel ci-dessous (figure 32).

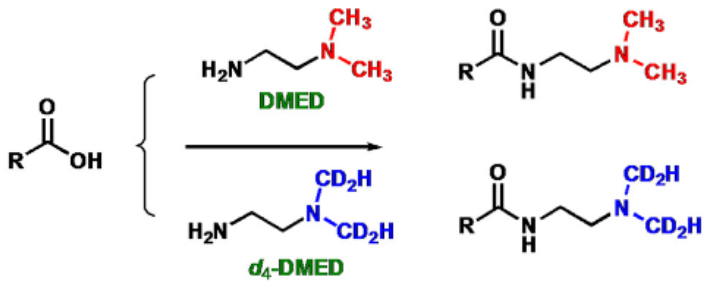


Figure 32 : Schéma réactionnel du DMED avec une fonction acide [177]

Le DMED et son isotope deutéré ont également été utilisés pour profiler des métabolites contenant des groupes carboxyliques dans des échantillons biologiques provenant de fumeurs et de non-fumeurs. Un total de 269 métabolites contenant des groupes carboxyliques a été découvert en utilisant cette méthodologie avec 88 métabolites ayant une différence de quantité significative entre les fumeurs et les non-fumeurs [176].

Ces dérivés pipérazine et amine ont été utilisés lors d'approches métabolomiques ciblant les acides gras. Toutefois, la littérature ne décrit aucune étude comparative permettant de choisir celle qui améliore au mieux la détection des acides gras. Ainsi, pour notre étude nous avons décidé de tester les trois.

1.2. Dérivé bromé : couplage en une étape

Le DmPABr est un dérivé bromé qui forme une liaison ester une fois couplé à la fonction acide carboxylique (voir figure 33). Il réagit avec l'acide selon une substitution nucléophile.

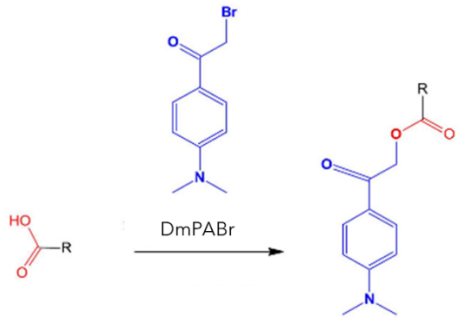
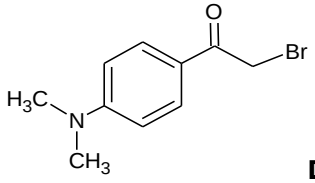


Figure 33 : Schéma réactionnel du DmPABr et d'une fonction acide carboxylique [197]

Tableau 9 : Structure chimique, conditions réactionnelles et utilisation de l'étiquette DmPABr pour le marquage en une étape des fonctions acides carboxyliques

Etiquette chimique	Conditions réactionnelles	Applications
 **DmPABr**	Ajout de TEA 65°C pendant 1h	Acides organiques urine [197, 198] Acides aminés dans l'urine [199]

Le 3-dimethylaminophenacyl bromide est un réactif utilisé pour la première fois en 2013 par Peng et Li [198] pour améliorer la détection du sous-métabolome des acides organiques tels que l'acide formique, acétique, glutarique contenus dans l'urine. Cette étude a permis un profilage de 2500 métabolites d'acides organiques en faisant abstraction des interférences des fonctions amines. En 2019, Awad et al [197] ont également utilisé la même méthode que Peng et Li pour la quantification de 17 métabolites ayant une fonction carboxylique contenue dans l'urine humaine. Toujours sur ce même fluide, Willacey et al [199] a démontré en 2019 que le DmPABr réagit avec les fonctions acides mais également avec les amines, thiols des métabolites. Ce réactif n'est donc pas spécifique à la dérivation d'une fonction acide.

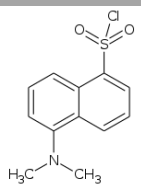
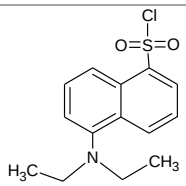
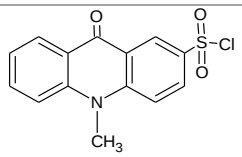
2. Fonctions amine et phénol

Les composés aminés tels que les acides aminés ont un rôle crucial dans les fonctions biologiques (ADN, neurotransmetteurs). Le groupement amine dans les métabolites est relativement facile à protoner en phase mobile acide et à ioniser en la source ESI. Toutefois, l'analyse du sous-métabolome amine sans la dérivation est problématique. Le problème vient de la polarité élevée et de la solubilité dans l'eau de ces composés, entrainant une mauvaise rétention sur la colonne RPLC. La dérivation de la fonction amine augmente l'hydrophobie de la molécule et permet une meilleure séparation des composés marqués. L'identification et la détermination des amines dans différentes matrices à l'aide de la dérivation chimique ont été largement étudiées [200]. Les composés phénoliques sont impliqués dans les processus cellulaires de base ou le métabolisme des macronutriments et des micronutriments [201]. Ils possèdent une fonction aromatique ionisable et un groupe hydroxyle qui pourraient être détectés par LC-MS, en particulier dans le mode d'ionisation négative [202]. Cependant, leur sensibilité n'est pas satisfaisante. Ainsi, ils sont dérivés pour augmenter leur détectabilité. De nombreuses étiquettes chimiques d'amines conviennent également pour la dérivation des métabolites contenant une fonction phénol.

2.1. Dérivés du Dansyle

La dansylation est l'une des procédures de dérivation la plus pratique pour l'analyse LC / ESI-MS des amines et des phénols. Le tableau 5 résume 3 dérivés de Dansyle utilisés de nos jours.

Tableau 10 : Structure chimique, conditions réactionnelles et utilisations des étiquettes chimiques utilisées pour une réaction de dansylation

Etiquettes chimiques	Conditions réactionnelles	Applications
 **Chlorure de Dansyle**	60°C 60min Solution tampon de carbonate/bicarbonate de sodium (pH=9,4)	-Identification de métabolites amines et phénols dans l'urine [203], le salive [204], les fèces [205, 206]..
 **DENSCl**	60°C 60min Solution tampon de carbonate/bicarbonate de sodium (pH=9,4)	-Plus ionisable que chlorure de Dansyle -Quantification de 20 acides aminés dans l'urine [207]
 **MASCl**	Tampon borate (pH=8) 5min à 60°C	-Quantification absolue par ajout de standards d'acide aminé dans le sérum -Réaction rapide 5min [208]

Le **chlorure de Dansyle** (DNSCl) réagit rapidement avec le groupement amine des métabolites en milieu alcalin pour former les sulfamides. Il réagit également avec les phénols pour former les esters de sulfone. L'étiquette chlorure de Dansyle introduit à la fois un groupe amino tertiaire basique et un fragment naphtalène aux métabolites. Le premier est facilement ionisé dans la phase mobile LC acide et le second augmente l'hydrophobicité et la rétention sur la phase inverse. Ainsi, la dansylation permet d'augmenter la réponse en ionisation positive des amines, acides aminés et phénols. Cette étiquette chimique réagit avec les amines et phénols selon le schéma réactionnel figure 34.

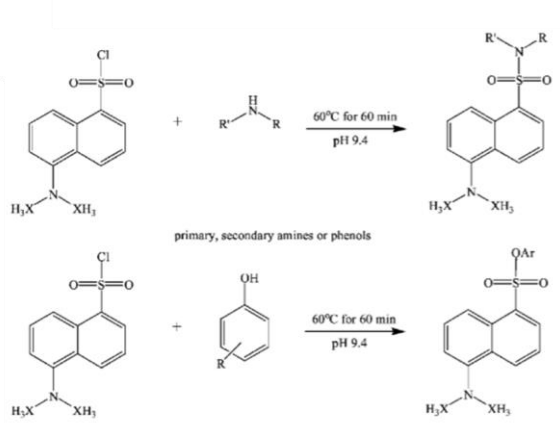


Figure 34 : Schéma réactionnel du marquage des fonctions amine et phénol par le chlorure de Dansyle. X= ^{12}C ou ^{13}C [203]

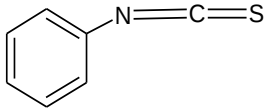
L'application de $^{13}C_0$ / $^{13}C_2$-DNSCl en UHPLC / ESI-FTICR-MS pour la quantification des amines, acides aminés et phénols a été publiée pour la première fois en 2009 dans les urines [203] . Depuis, la dansylation a été réalisée dans différentes études métabolomiques utilisant la salive [204], l'urine [205, 209, 210], la sueur [210] et les fèces [205, 206]. Le réactif 5-diethylaminonaphtalene-1-sulfonyl chloride (DENSCl) a été introduit pour améliorer le chlorure de Dansyle. Le groupement éthyl du **DENSCl,** comparé au groupement méthyl du chlorure de Dansyle, augmente l'ionisation de cette étiquette, résultant de sa plus grande affinité pour les protons. De plus, l'hydrophobicité accrue du groupe éthyle améliore également la rétention chromatographique en phase inverse de la molécule dérivée au DENSCl. Une étude métabolomique sur l'urine humaine a été réalisée en utilisant le DENSCl comme étiquette chimique. Cette étude utilisant le DENSCl et son isotope ^{13}C permet la quantification de 20 acides aminés dans l'urine [207]. Un autre dérivé sulfonyl chloride, le 10 méthyl-acridone-2-sulfonyl chloride (**MASCl)** a été développé en 2015 par Song et son équipe [208]. MASCl a été testé lors d'une étude métabolomique pour dériver les acides aminés contenus dans le sérum [208]. Par rapport au chlorure de Dansyle et au DENSCl, cette étiquette chimique permet de dériver les acides aminés dans des conditions plus douces (pH=8) et plus rapides, 5min de temps de réaction.

Le chlorure de Dansyle est l'étiquette chimique la plus utilisée lors d'étude métabolomique. Elle peut être considérée comme un « gold standard » pour le marquage des fonctions amine et phénol. Ainsi pour notre étude, nous utiliserons le chlorure de Dansyle pour cibler les fonctions amine et phénol.

2.2. Dérivé isothiocyanate

L'isothiocyanate de phényle (PITC) réagit avec un groupe amino N-terminal non chargé, dans des conditions légèrement alcalines, pour former un dérivé cyclique de phénylthiocarbamide.

Tableau 11 : Structure chimique, conditions réactionnelles et utilisation de l'étiquette PITC dans le kit de dosage Biocrates

Etiquette chimique	Conditions réactionnelles	Application
 **PITC**	Eau :Ethanol :Pyridine (1 :1 :1) 20min à température ambiante	-Kit de dosage Biocrates ® [211]

Originellement jusqu'en 1986, les acides aminés sont détectés par la technique de résine échangeuse d'ion. Cette technique est limitée par sa vitesse d'acquisition longue et sa sensibilité faible. Ainsi, une méthode de détection des acides aminés en LC a été développée. Toutefois, l'analyse chromatographique en phase liquide des acides aminés souffre d'inconvénients tels que leurs mauvaises stabilités et quantification dues à l'interférence entre les pics. En 1978, Sigrist et son équipe sont les premiers à décrire l'affinité du PITC avec les fonctions amines en 1978 [212]. En 1986, Lawrence et son équipe ont montré que le PITC peut être utilisé comme étiquette chimique ciblant les fonctions amines lors d'une détection UV [213]. L'étiquette PITC est donc utilisée pour améliorer la détection des acides aminés en UV avec une longueur d'onde de 254nm. Cette méthode de marquage des acides aminés est utilisée par Battaglia et son équipe en 1999 pour leur détection dans des extraits de tumeurs cérébrales humaines [214]. Les conditions de couplage entre le

PITC et les acides aminés sont de 20min à température ambiante. Le PITC est ajouté à l'échantillon selon un mélange MeOH :TEA :Eau (7 :1 :1). Cette solution de dérivation est stable pendant 2 jours à 4°C. En effet, seul 5% de l'échantillon est perdu au bout du deuxième jour. A cette époque, la dansylation est également utilisée pour la détection des acides aminés mais cette étiquette chimique crée des interférences avec l'éluant entrainant des biais de détection. La molécule de PITC sert donc d'alternative afin d'éviter la formation de produits secondaires [215].

De nos jours en spectrométrie de masse les propriétés de marquage du PITC sont utilisés par le Kit de dosage Biocrates ® pour la détection et la quantification de 630 métabolites issus de 36 classes chimiques (les acides aminés, amines, acylcarnitines, phospholipides et la phosphatidylcholine les sphingomyélines) [211]. Les conditions de marquage sont similaires à celles utilisées lors de l'étude de Battaglia. Le PITC est ajouté à l'échantillon dans un mélange Eau :Ethanol :Pyridine (1 :1 :1) pendant 20min à température ambiante. Ce kit permet la quantification métabolomique de divers fluides biologiques (par exemple plasma, sérum, urine, lait [216])

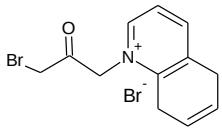
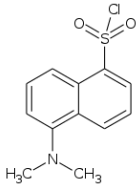
Le PITC est utilisé pour améliorer l'ionisation et la séparation des molécules. Pour une approche quantitative absolue, ce sont les composés qui seront marqués isotopiquement et non le PITC. Cette étiquette permet donc une approche quantitative différente des autres étiquettes décrites dans ce paragraphe.

3. Fonctions alcool et thiol

Les composés contenant des groupements hydroxyles représentent un grand groupe de composés ayant une importance biologique comme les réactions de déshydratation à la base de l'élongation des chaines carbonées qui lient les molécules biologiques simples en longues chaines. Cette déshydratation permet la jonction d'un acide gras au glycérol pour former un triacylglycérol en éliminant le -OH de l'extrémité carboxyle de l'acide gras. Ces composés sont difficilement

ionisables en mode négatif et retenus sur les colonnes RPLC. Ils ont donc une mauvaise réponse en spectrométrie de masse. Par conséquent, la dérivation est nécessaire pour la détection de ces composés. Les métabolites contenant des thiols, telle que la cystéine, ont un rôle important dans le contrôle de l'homéostasie redox et dans les fonctions des protéines. Les thiols existent également dans l'environnement en raison du métabolisme microbien et de la décomposition biologique/non biologique de substances organiques [217]. La détermination des composés thiols dans des spécimens biologiques ou des échantillons environnementaux est difficile en raison de leur faible abondance ainsi que des effets matrice. Ainsi, la dérivation chimique est apparue comme une technique prometteuse pour améliorer la sensibilité, la robustesse et la spécificité des dosages de ces composés.

Tableau 12 : Structure chimique, conditions réactionnelles et applications des étiquettes chimiques utilisées pour le marquage des thiols et alcools

Etiquettes chimiques	Conditions réactionnelles	Applications
 **BQB**	Solution tampon à pH=3,5 3min sous microondes	Dérivation de thiols contenus dans l'urine [218]
 **Chlorure de Dansyle**	3 équivalents DMAP 60°C pendant 60min	Quantification absolue de 85 composés alcools dans les urines et identification putative en non ciblée de 8021 métabolites [166]

Le bromure de bromoacétonylquinolinium (BQB) réagit sélectivement avec les thiols, tels que la cystéine, l'homocystéine et la N-acétylcystéine (Nac), en milieu acide pour produire des dérivés chargés [218]. Lors de cette étude, le BQB est utilisé pour la détection de thiols dans l'urine, la dérivation des composés se fait pendant 3min à l'aide d'un four à microondes.

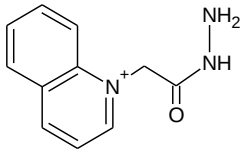
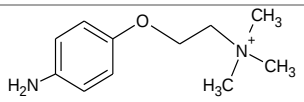
L'étiquette de chlorure de Dansyle permet le marquage des composés hydroxylés les rendant visibles en LC-UV et LC-MS. Cette étiquette décrite pour le marquage des fonctions amine et phénol marque aussi les fonctions alcools dans des conditions basiques à 60°C pendant 1h. En effet, l'étude de Zhao et al. [166] montre la quantification des alcools sur une base de 85 standards hydroxylés en LC-UV (347nm) et l'identification putative de 8021 métabolites contenus dans l'urine à l'aide de base de données métabolomiques.

Ces deux étiquettes chimiques sont utilisées en métabolomique pour cibler les fonctions thiols (BQB) et alcools (Chlorure de Dansyle).

4. Fonctions carbonyles

Contrairement aux groupements amine ou carboxyle, le carbonyle est une fonction neutre qui a une faible efficacité d'ionisation en ESI. De nombreuses cétones ou aldéhydes ont une faible abondance (exemple les hormones) et/ou faible rétention sur colonne RPLC (exemple les glucides). La dérivation par l'introduction d'un fragment hydrophobe permet d'augmenter leur sensibilité de détection et leur rétention. Des étiquettes chimiques ont été développées pour la dérivation chimique des composés carbonylés lors d'études métabolomiques.

Tableau 13 : Structure chimique, conditions réactionnelles et utilisations des étiquettes chimiques utilisées pour le marquage des carbonyles

Etiquettes chimiques	Conditions réactionnelles	Applications
 **HIQB**	50°C pendant 3h dans une solution de MeOH acidifiée à 1% d'acide acétique	Détection de 156 composés carbonylés dans le sérum humain [219]
 **APC**	15°C pendant 3h avec une solution tampon d'acétate d'ammonium (pH= 5,7)	Détection de 21 composés aldéhydes dans le sérum humain [220]

Plusieurs réactifs à base d'hydrazine ont contribué de manière significative à l'analyse des composés carbonylés dans diverses matrices [221]. Par exemple, le bromure de 2-(2-hydrazinyl-2-oxoéthyl)isoquinoléine-2-ium (HIQB) est utilisé pour profiler les composés carbonylés dans le sérum humain [219]. Cette dérivation est effectuée pendant 3 heures à 50°C dans une solution de MeOH contenant 1% d'acide acétique. Une telle dérivation couplée à une analyse MS a permis la détection de 156 composés carbonylés dans le sérum humain.

Les réactifs à base d'hydrazine sont généralement instables, inflammables, irritants pour la peau et toxiques pour les voies respiratoires [222]. Ainsi, de nombreux réactifs n'appartenant pas au groupe à base d'hydrazine ont été développés et adoptés pour la dérivation chimique des composés carbonylés. Par exemple, l'halogénure de 4-(2-(triméthylammonio)éthoxy)benzénamine (APC) est utilisé pour la dérivation des aldéhydes. L'APC a été utilisé pour le profilage complet des aldéhydes dans l'urine humaine et le vin blanc, permettant la détection de 21 et 23 aldéhydes, respectivement [220]. Les conditions réactionnelles de marquage des fonctions aldéhyde par l'APC nécessitent un pH de 5,7 à 15°C pendant 3h.

Le marquage des fonctions carbonyles demande des conditions réactionnelles un peu plus complexes que dans le cas des autres groupements comme par exemple le maintien de la température à 15°C. Ainsi, si le métabolite contient d'autres fonctions dans sa structure il sera préférable de privilégier le marquage de ces autres fonctions plutôt que celle du groupement carbonyle.

III. Plan expérimental

Les études utilisant la détection des fonctions acides sont très répandues lors d'approches métabolomiques. Ce groupement est présent dans de nombreuses structures de métabolites telles que les acides aminés, les acides organiques, les acides gras et les α-hydroxy acides. Toutefois, leur détection en RPLC-MS en mode négatif est très difficile à cause de leur faible sensibilité et hydrophobie. La dérivation chimique de ces fonctions est donc nécessaire. D'après l'étude de la

littérature, cette dérivation peut être effectuée en deux étapes ou en une étape. Le marquage des fonctions acides en deux étapes nécessite l'utilisation d'agents de couplage. L'étude de la littérature nous a permis de choisir trois types d'agents de couplage un carbodiimide, un sel d'uronium et une iodide. L'efficacité de marquage des trois étiquettes chimiques décrites lors d'études métabolomiques sera testée (DMED, DAPPZ, DMPP). En parallèle de cette étude, le marquage des fonctions acides en une étape est également testé. Pour ce faire, nous allons utiliser l'étiquette DmPABr décrite par la littérature comme étiquette se fixant sur les acides carboxyliques. Toutefois, des études ont montré que le DmPABr peut également cibler les fonctions amines. Ainsi, cette étiquette sera comparée à deux étiquettes dites de « référence » pour le marquage des fonctions amines : le chlorure de Dansyle et le PITC. Ainsi, un panel de standards de métabolites représentant au mieux la diversité biologique et chimique des fonctions acides et amines est utilisé. L'objectif est de déterminer l'étiquette chimique qui permettra de marquer le plus efficacement possible les fonctions acides carboxyliques des métabolites.

IV. Matériels et méthodes

1. Produits chimiques et réactifs

1.1. Standards de métabolites

Une liste de 20 métabolites appartenant à 6 catégories chimiques, couramment trouvés dans les fluides biologiques et décrits dans différents types de pathologies, sont testés (voir figure 35). Tous les métabolites ont été achetés chez Sigma-Aldrich (St-Quentin-Fallavier, France) et serviront de standards pour évaluer les différents protocoles de marquage.

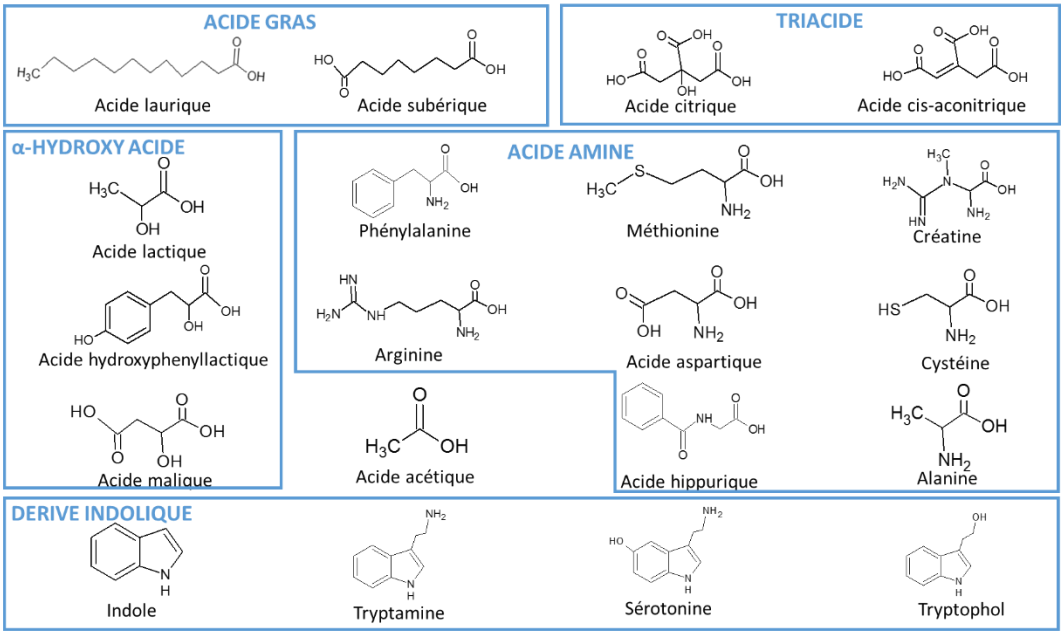


Figure 35 : Liste des standards utilisés dans notre étude selon leurs classes chimiques

Ces composés ont été choisis dans le but de couvrir la diversité de groupements chimiques pouvant être retrouvée dans les fluides biologiques (groupement acide carboxylique, amine primaire, amine secondaire, hydroxyle, phénol, thiol). De part un recueil facile, l'urine et le sang sont par exemple les fluides biologiques où le dosage de composés tels que les acides organiques fait l'objet de nombreuses études. Par exemple, l'acide lactique produit par les muscles peut-être dosé dans le sang et signer d'une activité physique intense. L'acide hippurique est le biomarqueur le plus fréquemment utilisé dans le suivi biologique d'exposition professionnelle au toluène. L'acide subérique est présent dans l'urine des patients présentant des troubles de l'oxydation des acides gras. L'acide acétique se trouve dans tous les organismes, des bactéries aux plantes en passant par les humains. Dans le corps humain, l'acide acétique est normalement produit après la consommation d'alcool. L'éthanol est converti en acétaldéhyde qui est alors converti en acide acétique sous l'action d'enzymes. Les acides aminés ont un rôle dans les processus biologiques et font l'objet de nombreuses études. En effet, ils permettent la synthèse protéique mais sont également les précurseurs de composés biologiquement actifs tels que les bases de l'ADN et les

neurotransmetteurs. De plus, ces métabolites sont présents dans les voies métaboliques telles que le cycle de Krebs (acides citrique, cis-aconitique, malique) ou bien d'autres voies métaboliques telles que celle du tryptophane plus précisément le métabolisme de la sérotonine (sérotonine, tryptamine) et le métabolisme des indoles (indole, tryptophol).

1.2. Etiquettes chimiques, agents de couplage et bases

Les étiquettes DMPP,DAPPZ, DMPT et DmPABr ont été synthétisées par l'ICOA (Institut de Chimique Organique et Analytique, Orléans, France). Le DMED, le chlorure de Dansyle, le PITC, les agents de couplage CMPI, HATU, HOAt, EDC et les bases TEA (triéthylamine), TEOA (triéthanolamine) ont été commandés sur le site Sigma Aldrich (St-Quentin-Fallavier, France).

Tableau 14 :Structures chimiques, formules brutes et m/z des étiquettes chimiques DMED, DMPP, DMPT, DAPPZ, chlorure de Dansyle et PITC utilisées dans notre étude

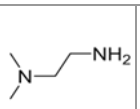
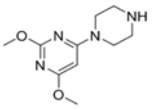
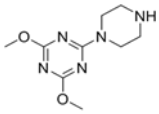
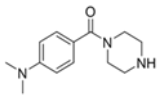
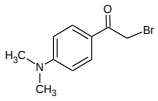
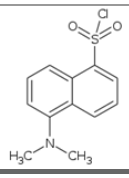
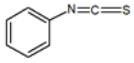
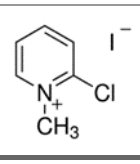
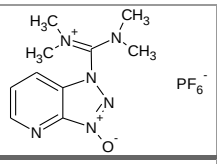
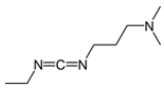
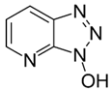
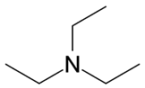
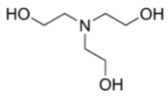
							
Nom	**DMED**	**DMPP**	**DMPT**	**DAPPZ**	**DmPABr**	**Chlorure de Dansyle**	**PITC**
Formule brute	$C_4N_2H_{12}$	$C_{10}N_4O_2H_{16}$	$C_9N_5O_2H_{15}$	$C_{13}N_3OH_{19}$	$C_{10}H_{12}ONBr$	$C_{12}H_{12}NO_2SCl$	C_7H_5NS
m/z $[M+H]^+$	89.1073	226.1298	225.1346	234.1600	242,0171	270.0350	136.0215

Tableau 15 : Structures chimiques, formules brutes et m/z des agents de couplages, AAE et bases utilisés dans notre étude

						
Nom	**CMPI**	**HATU**	**EDC**	**HOAt**	**TEA**	**TEOA**
Formule brute	C_6H_7NClI	$C_{10}N_6OH_{15}(PF_6^-)$	$C_8N_3H_{17}$	$C_5N_4OH_4$	C_6NH_{15}	$C_6O_3NH_{15}$
m/z	M^+ 128.0261	M^+ 235.1301	$[M+H]^+$ 156.1495	$[M+H]^+$ 137.0457	$[M+H]^+$ 102.1277	$[M+H]^+$ 150.1124

2. Protocole de marquage des fonctions acides

2.1. Marquage en deux étapes avec les dérivés pipérazine et amine

D'après la partie II-1 (Fonction acide carboxylique) plusieurs types d'agents de couplage peuvent être utilisés pour dériver les fonctions acides en amide à l'aide d'amine. Ces agents de couplage peuvent être des carbodiimides, sels de phoshonium, sels d'uronium ou autres. Nous avons décidé d'expérimenter un agent de couplage par catégorie en excluant les sels de phosphonium qui sont à privilégier lors de cyclisation.

2.1.1. Protocole CMPI

Le protocole suivi pour le marquage de fonctions acides par l'agent de couplage CMPI selon le schéma réactionnel figure 36 est inspiré de Zhu *et al*. [177].

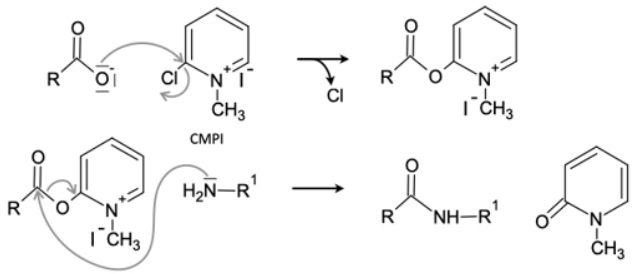


Figure 36 : Schéma réactionnel de la formation d'une liaison peptidique à l'aide de l'agent de couplage CMPI

Le nombre de moles utilisé pour le marquage des fonctions acides avec l'agent de couplage CMPI est résumé ci-dessous (tableau 16) :

Tableau 16 : Nombre de moles des standards,base TEA, agent de couplage CMPI et étiquettes chimiques (DMED, DAPPZ, DMPT et DMPP) utilisé lors du protocole CMPI

Nombre de moles			
Standards	TEA	CMPI	Etiquette
0.44 nmol (1 équivalent)	1.1nmol (2.5 équivalents)	1.1 nmol (2.5 équivalents)	3nmol (6 équivalents)

1mg de chaque standard a été pesé et dissout dans 1mL l'eau miliQ. Cette solution est diluée à 10µg/mL et un volume est prélevé (Tableau S1) de façon à ce que le volume total soit de 50µL (0.44nmol). 15µL d'une solution de TEA à 72µM (1.1nmol) est ajoutée ainsi que 28µL d'une solution de CMPI à 39µM(1.1nmol). La solution est mélangée à l'aide d'un vortex pendant 2min. 50µL d'une solution contenant l'étiquette chimique (DMED, DAPPZ, DMPT ou DMPP) à 3nmol est ajouté. Le mélange est incubé pendant 60min à 40°C dans un bloc chauffant sec, puis analysé en RP-LC (colonne C18) en détection en mode positif par MS.

2.1.2. Protocole HOAt/EDC

Le protocole de marquage de fonction acide avec les agents de couplage HOAt/EDC selon le schéma réactionnel figure 37 est inspiré de l'équipe de Leng [43].

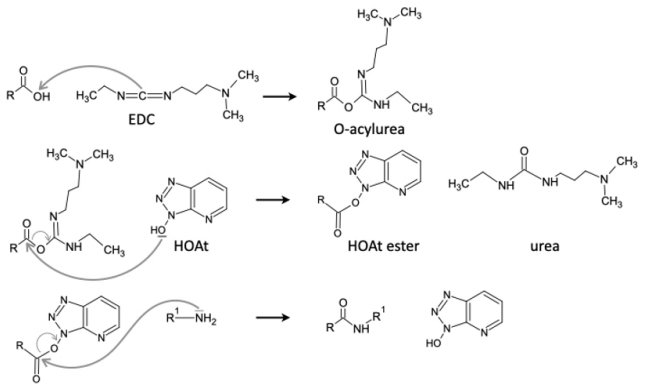


Figure 37 : Schéma réactionnel lors de la formation d'une liaison peptidique à l'aide de l'agent de couplage HOAt/EDC inspiré de Leng et al. [43]

Le nombre de moles utilisé pour le marquage des fonctions acides avec l'agent de couplage EDC et l'AAE HOAt est résumé ci-dessous (tableau 17) :

Tableau 17 :Nombre de moles des standards, agent de couplage EDC, AAE HOAt et étiquettes chimiques DMED, DAPPZ, DMPT et DMPP utilisé lors du protocole EDC/HOAt

Nombre de moles			
Standards	EDC	HOAt	Etiquettes chimiques
0.44 nmol (1 équivalent)	1.1nmol (2.5 équivalents)	1.1 nmol (2.5 équivalents)	3nmol (6 équivalents)

1mg de chaque standard a été pesé et dissout dans 1mL l'eau miliQ. Cette solution est diluée à 10µg/mL et un volume est prélevé (Tableau S1) de façon à ce que le volume total soit de 50µL (0.44nmol). 17µL d'une solution de EDC à 64µM (1.1nmol) est ajouté et 15µL d'une solution de HOAt à 74µM (1.1nmol). La solution est mélangée à l'aide d'un vortex pendant 2min. 50µL d'une solution contenant l'étiquette chimique (DMED, DAPPZ, DMPT ou DMPP) à 3nmol est ajouté. Le mélange est agité pendant 5min à l'aide d'un agitateur vibrant, puis analysé en RP-LC (colonne C18) en détection en mode positif par MS.

2.1.3. Protocole HATU

Dans le cadre de dérivation chimique, le protocole défini pour l'agent de couplage HATU n'a pas été trouvé dans la littérature. Le schéma réactionnel a donc été inspiré d'une N-acylation utilisant l'HATU (figure 38). Le protocole a été créé à partir de ce schéma réactionnel et du protocole utilisant le CMPI car de même que l'HATU cet agent de couplage ne nécessite pas la présence d'AAE.

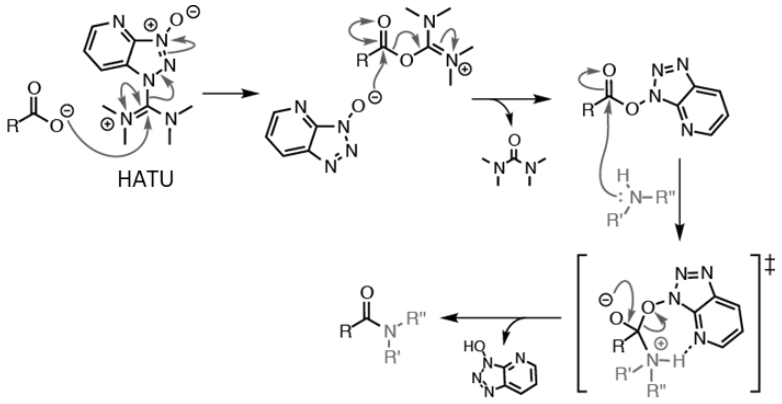


Figure 38 : Schéma réactionnel d'une N-acylation utilisant l'agent de couplage HATU

Le nombre de moles utilisé pour le marquage des fonctions acides avec l'agent de couplage HATU est résumé ci-dessous (tableau 18) :

Tableau 18 : Nombre de moles des standards, base TEA, agent de couplage HATU et étiquettes chimiques DMED, DAPPZ, DMPT et DMPP utilisé lors du protocole HATU

Nombre de moles			
Standards	TEA	HATU	Etiquettes chimiques
0.44 nmol (1 équivalent)	1.1nmol (2.5 équivalents)	1.1 nmol (2.5 équivalents)	3nmol (6 équivalents)

1mg de chaque standard a été pesé et dissout dans 1mL l'eau miliQ. Cette solution est diluée à 10µg/mL et un volume est prélevé (Tableau S1) de façon à ce que le volume total soit de 50µL (0.44nmol). 15µL d'une solution de TEA à 72µM (1.1nmol) est ajouté et 21µL d'une solution de HATU à 52µM (1.1nmol). La solution est mélangée à l'aide d'un vortex pendant 2min. 50µL d'une solution (3nmol) contenant l'étiquette chimique (DMED, DAPPZ, DMPT ou DMPP) à 3nmol est ajouté. Le mélange est agité pendant 30min à l'aide d'un agitateur vibrant à température ambiante, puis analysé en RP-LC (colonne C18) en détection en mode positif par MS.

2.2. Marquage en une étape avec le DmPABr

Le protocole de marquage de la fonction acide par le DmPABr selon le schéma réactionnel figure 39 a été inspiré de Willacey et al. [199]

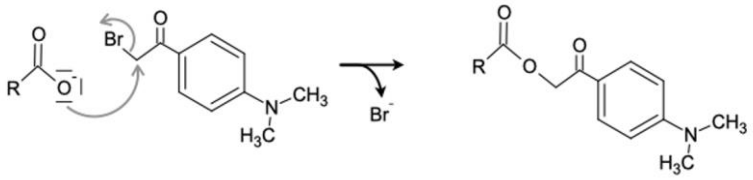


Figure 39 : Schéma réactionnel de la formation d'un ester à partir du DmPABr

Le nombre de moles utilisé lors de ce protocole est résumé ci-dessous (tableau 19) :

Tableau 19 : Nombre de moles des standards,base TEOA et étiquette chimique DmPABr utilisé lors du protocole de marquage des fonctions acides par le DmPABr

Nombre de mol		
Standards	TEOA	Etiquette DmPABr
0.44 nmol (1 équivalent)	370 µmol (large excès)	32µmol (large excès)

1mg de chaque standard a été pesé et dissout dans 1mL l'eau miliQ. Cette solution est diluée à 10µg/mL et un volume est prélevé (Tableau S1) de façon à ce que le volume total soit de 50µL (0.44 nmol). 50µL de TEOA à 370 µmol et 50µL de DmPABr à 0,2µmol sont ajoutés. La solution est vortexée pendant 1 min puis incubée dans un bain chauffant à sec pendant 60min à 60°C. Afin de stopper toute réaction chimique 50µL (32µmol) d'acide formique à 651mM est ajouté et la solution est incubée pendant 15min à 60°C. Elle est ensuite analysée en LC-MS.

3. Protocole de marquage des fonctions amine et phénol par les étiquettes chlorure de Dansyle et PITC

3.1. Protocole chlorure de Dansyle

Le protocole de marquage des fonctions amine et phénol par le chlorure de Dansyle selon le schéma réactionnel figure 40 a été inspiré de l'équipe de Xu [206]. Nous avons décidé de diluer les proportions d'un facteur 20 afin d'éviter toute saturation du détecteur MS et de la colonne LC.

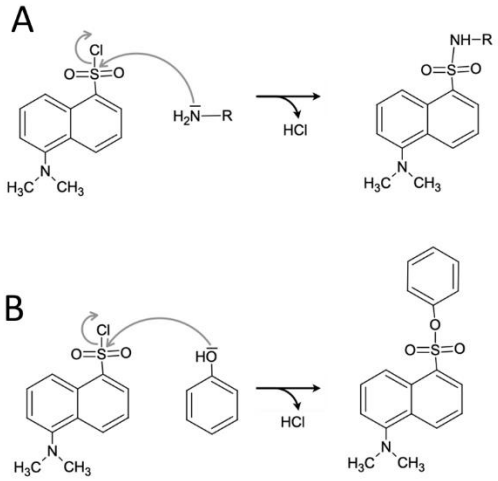


Figure 40 : Schéma réactionnel d'une dansylation réalisée pour une amine primaire (A) et pour un phénol (B) à l'aide du chlorure de Dansyle

Le nombre de moles utilisé lors de ce protocole est résumé ci-dessous (tableau 20) :

Tableau 20 : Nombre de moles des standards et étiquette chimique chlorure de Dansyle utilisé lors du protocole de marquage des fonctions amine/phénol par le chlorure de Dansyle

Nombre de moles	
Standards	Etiquette chlorure de Dansyle
Entre 30µmol -0.1mmol (1 équivalent)	13.5mmol (large excès)

50µL d'une solution de standard à 10µg/mL est prélevé. 25µL d'une solution tamponnée à pH 9.4 de carbonate de sodium/bicarbonate de sodium à 25mM et 25µL d'ACN sont ajoutés. 50µL d'une solution de chlorure de Dansyle diluée dans

l'acétone (1mg/mL) est ajoutée. Le mélange est incubé pendant 60min à 60°C dans un bloc chauffant sec, puis 10µL d'une solution de NaOH à 12,5mM est ajouté. Le mélange est à nouveau incubé pendant 10min à 60°C puis 50µL d'une solution d'acide formique à 21,2mM est ajoutée. La solution est ensuite analysée en RP-LC (colonne C18) en détection en mode positif par MS.

3.2. Protocole PITC

Le protocole de marquage de fonction amine secondaire par le PITC selon le schéma réactionnel figure 41 a été inspiré par la notice d'utilisation du kit biocrates (Note d'application : « Targeted Metabolomics Using the UPLC/MS-based AbsoluteIDQ p180 Kit »).

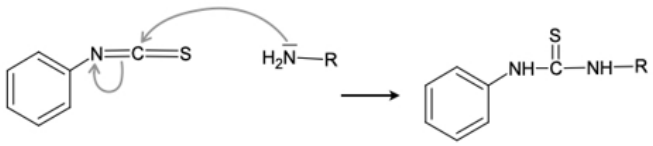


Figure 41 : Schéma réactionnel de dérivation du PITC avec une fonction amine primaire

Le nombre de moles utilisé lors de ce protocole est résumé ci-dessous (tableau 21) :

Tableau 21 : Nombre de moles des standards, base pyridine et étiquette chimique PITC utilisé lors du protocole de marquage des fonctions amine/phénol par le PITC

Nombre de moles		
Standards	Pyridine	PITC
Entre 60µmol – 0.2mmol (1 équivalent)	23.6mmol (large excès)	40mmol (large excès)

Dans une plaque 96 puits, 100µL de solution de standard (10µg/ML) est ajouté. La plaque est ensuite placée sous séchage N_2 à température ambiante pendant 10min. 50µL d'une solution dite de dérivation composée de 1900µL d'éthanol, 1900µL de pyridine (23.6mmol), 1900µL eau miliQ et 300µL PITC (solution à 99%- 40mmol) est ajoutée. La plaque est couverte par un film aluminium et mise sous agitation à température ambiante pendant 1h à 300 rpm. La plaque est ensuite séchée sous N_2

à température ambiante pendant 35min. Le résidu sec est repris dans 100µL de MeOH/H_2O 10/90 et analysé par RP-LC (colonne C18) par détection en mode positif en MS.

V. Acquisition des données

1. Paramètres LC

Les analyses sont réalisées à l'aide de UPLC Ultimate WPS-3000 system (Dionex, Germany) en phase inverse à l'aide d'une colonne Phenomenex Kinetex XB-C18 (1,7µm 100 A 150x2,1mm) à une température de 55°C. La phase mobile est composée de H_2O en phase A et de MeOH en phase B, les deux phases sont acidifiées à l'aide de 0,1% d'acide formique. Un gradient de 24min avec un débit de 0,4mL/min. Après 2min d'équilibre de la colonne à 0,1% de phase B le gradient démarre de 0 à 6min : 0,1 à 25% de phase B, 6 à 10min : 25 à 80% de phase B, 10 à 12min : 80 à 90% de phase B, 12 à 21min : 90 à 99,9% de phase B et de 21 à 24min 99,9% de B.

2. Paramètres MS

L'acquisition des données MS est réalisée en détection positive par un spectromètre de masse QExactive Orbitrap (Thermo Fisher Scientific, Bremen, Germany) et avec une source électrospray HESI (Head electrospray ionization). Les paramètres de l'analyseur et de la source sont décrits respectivement dans le tableau 22 (a) et (b).

Tableau 22 : Paramètres utilisés pour la détection positive en full-scan (a) et paramètres de la source ESI (b)

	Mode +/-
Scan range	60 à 900 m/z
Fragmentation	None
Résolution	70.000
Polarité	Positif
Microscans	1
Lock masses	Off
AGC target	1e6
Maximum inject time	250 ms

(a)

	Source ESI
Sheath gas flow rate	35
Aux gas flow rate	10
Spray voltage (kV)	3
Capillary temp (°C)	325
S-Lens RF level	60
Heater temp (°C)	325

(b)

Pour cette étude, deux types d'acquisitions sont réalisées. La première étape de l'étude est une analyse non ciblée consistant à une analyse en full-scan suivi d'une acquisition DDA en mode Top10. MS^2. Cette acquisition permet de déterminer si la réaction de marquage a fonctionné et quelles sont les adduits de molécules formées ($[M+H]^+$,$[M+Na]^+$…). La deuxième étape est une analyse ciblée en mode PRM, l'énergie de collision est fixée à 35 et le premier quadripôle qui sert de filtre a une fenêtre de masse de 1 m/z. Les m/z de chaque ion identifié lors de la première étape sont soumis à une MS^2 afin d'identifier le schéma de fragmentation. Les modes d'acquisitions sont détaillés chapitre 1.

VI. Résultats et discussion

1. Marquage en deux étapes par les dérivés pipérazine et amine de la fonction acide carboxylique

Les dérivés pipérazine et amine permettent le marquage de la fonction acide en deux étapes. La première étape consiste à l'activation de la fonction acide par un agent de couplage. Nous allons comparer les agents de couplage CMPI, EDC/HOAt et HATU.

Le but ici est de déterminer l'étiquette chimique et l'agent de couplage adaptés à nos acides.

1.1. Cas des acides gras et triacide

Le rapport m/z des acides gras et triacides couplées aux étiquettes DMED, DAPPZ, DMPT et DMPP est résumé tableau 23.

Tableau 23 : m/z attendus pour le marquage par les dérivés pipérazines et amine de la fonction acide des acides gras et triacides sous la forme [M+H]+

		m/z $[M+H]^+$			
	Marqué	**DMED**	**DAPPZ**	**DMPT**	**DMPP**
Acide laurique	1	271.2744	416.3272	407.3017	409.3047
Acide citrique	1	263.1237	408.1765	399.1510	400.1446
	2	333.2132	623.3187	605.2678	607.2582
	3	403.3027	838.4610	811.3845	814.3703
Acide cis-aconitique	1	245.1132	390.1660	381.1405	383.1435
	2	315.2027	605.3082	587.2573	590.2571
	3	385.2921	820.4505	793.3741	797.3692
Acide acétique	1	131.1179	277.1784	267.1452	269.1482
Acide subérique	1	245.1859	390.2387	381.2132	383.2162
	2	315.2754	605.3809	587.3300	590.3298

La figure 42 décrit les résultats obtenus dans le cas du marquage de la fonction acide des acides gras et triacides. Les résultats sont présentés sous forme d'histogrammes avec en abscisse les 4 étiquettes chimiques marquant une fois la molécule ou deux fois. Par exemple dans le cas de l'acide subérique DMED 2 représente la molécule marquée deux fois par l'étiquette DMED. L'ordonnée représente l'aire brute des acides marqués mesurée à partir de leur m/z (tableau 23).

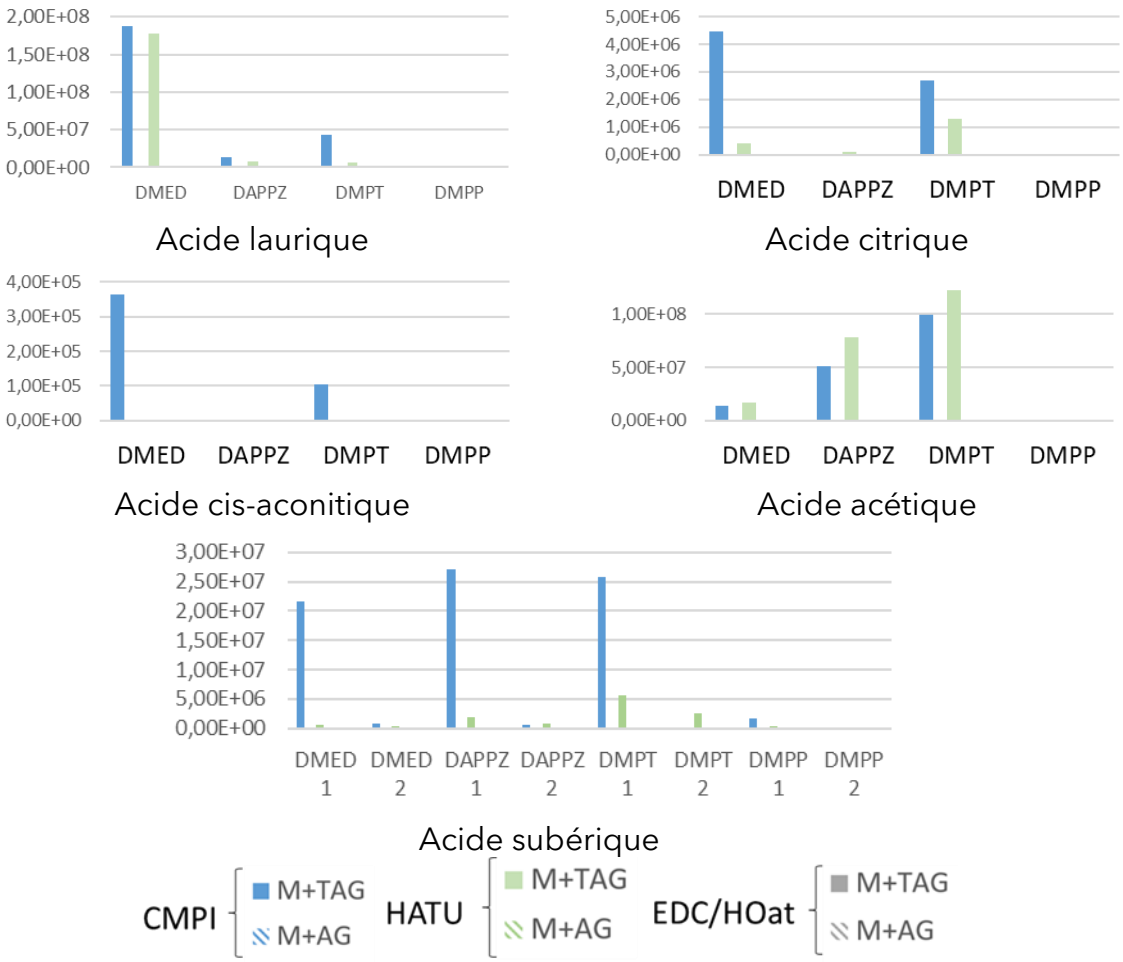


Figure 42: Résultats de marquage des acides gras et triacides par les dérivés pipérazine et amine. L'ordonnée représente l'aire brute des pics et l'abscisse l'étiquette chimique considérée.AG= agent de couplage, CMPI, HATU, EDC ou HOAt et TAG=DMED, DAPPZ, DMPT ou DMPP

Globalement, les étiquettes chimiques DMED, DAPPZ et DMPT se fixent sur la fonction acide des composés avec les agents de couplage CMPI et HATU. Les conditions utilisant le carbodiimide EDC et l'AAE HOAt ne permettent pas de marquer la fonction acide. Ceci peut s'expliquer par deux raisons : soit les quantités de molécules dérivées sont insuffisantes pour une détection MS soit les conditions de dérivation sont à revoir. Les conditions de marquage utilisant l'agent de couplage CMPI semblent plus efficace que celles avec l'agent HATU. Toutefois dans le cas de l'acide laurique dérivé avec le DMED, l'aire M+TAG obtenue est similaire dans les deux cas, de même pour l'acide acétique. L'acide acétique est le standard

le mieux dérivé par les 3 étiquettes avec une aire de pic allant jusqu'à 10^8 dans le cas du DMPT. Ceci peut s'expliquer par le fait que l'acide acétique est l'acide carboxylique le plus simple avec une chaine courte (C_2). En présence de l'agent de couplage CMPI, les acides gras (acides subérique et laurique) se lient mieux aux étiquettes chimiques par rapport à l'agent HATU. L'acide subérique est un diacide qui est marquée deux fois par les étiquettes toutefois, la forme marquée une fois est majoritairement détectée dans le milieu.

L'observation des spectres MS^2 permet de comprendre le schéma de fragmentation des trois étiquettes chimiques, DMED, DAPPZ, DMPT une fois couplée aux acides gras. La figure 43 représente les MS^2 obtenus pour l'acide subérique couplé 1 fois avec le DMED avec les agents de couplage CMPI et HATU (a), l'acide laurique couplé au DMED avec l'agent de couplage CMPI (b), l'acide acétique couplé au DAPPZ (c) et au DMPT (d) dans les conditions CMPI et HATU.

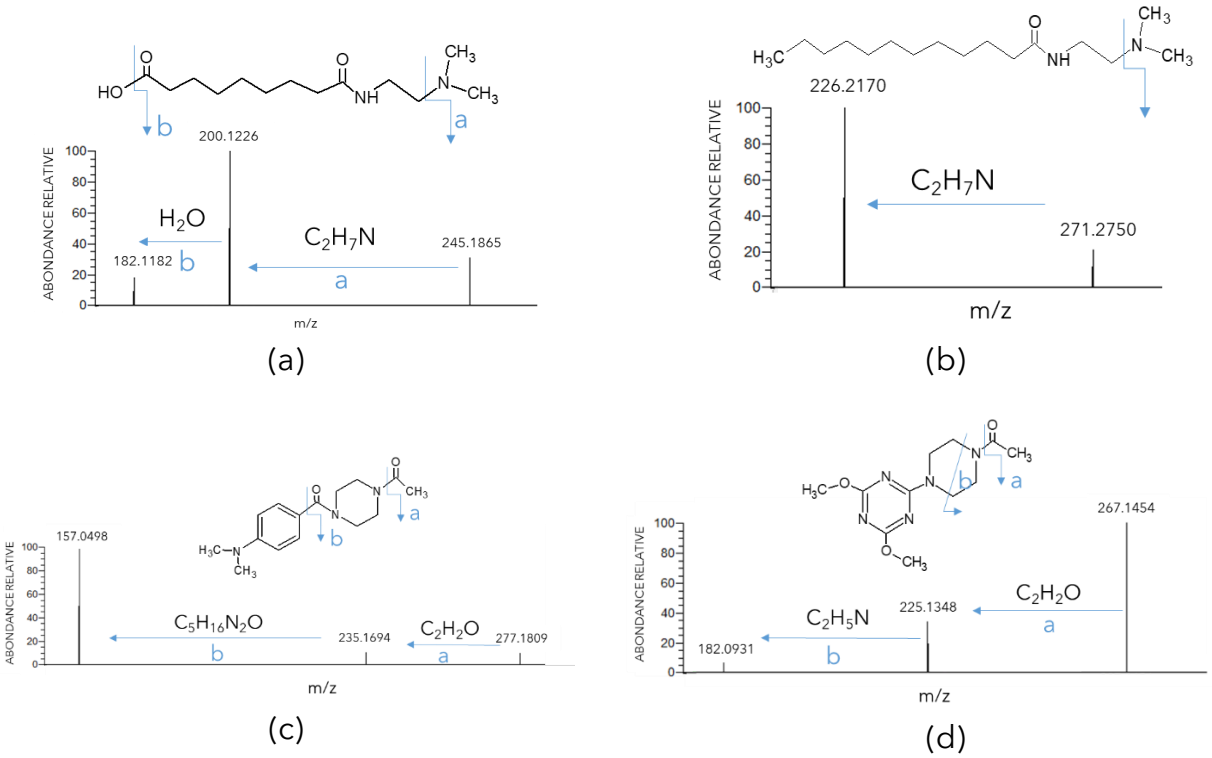


Figure 43 : spectres MS^2 de l'acide subérique marqué une fois par le DMED (a), l'acide laurique marqué avec le DMED (b), l'acide acétique marqué avec le DAPPZ (c) et l'acide acétique marqué avec le DMPT (d)

La fragmentation des acides subérique et laurique (figure 43 (a) et (b)) couplées avec le DMED produit un ion fragment majoritaire qui correspond à une perte de C_2H_7N. Cette perte correspond au groupement diméthylamine du DMED, qui d'après ces deux MS², est le premier groupement à être fragmenté. Dans le cas de l'acide laurique, la fragmentation s'arrête à cette perte. Dans le cas de l'acide subérique qui a deux fonctions acides carboxyliques, nous observons une déshydratation qui correspond à la perte de la fonction hydroxyle de la fonction carboxyle. Les figures 43 (c) et (d) représentent les spectres MS² de l'acide acétique couplée, respectivement, aux étiquettes DAPPZ et DMPT. La première fragmentation dans ces deux cas correspond à la perte du groupement C_2H_2O qui correspond à la fraction de l'acide acétique couplé aux étiquettes. Ensuite, dans le cas de l'étiquette DAPPZ il y a une perte du fragment $C_5H_{16}N_2O$ qui correspond au cycle pipérazine lié à la fonction cétone. Et dans le cas du DMPT, il y a une perte de C_2H_5N qui correspond à la moitié du cycle pipérazine de cette étiquette. Ainsi d'après ces MS², il semble que chaque étiquette chimique a un fragment spécifique. La perte de C_2H_7N est un fragment spécifique au couplage de la molécule avec le DMED, $C_5H_{16}ON_2$ pour l'étiquette DAPPZ et C_2H_5N pour le DMPT.

Les triacides (acides citrique et cis-aconitique) se lient également aux étiquettes chimiques. Les étiquettes DMED et DMPT avec les conditions de couplage CMPI semblent plus efficaces que la DAPPZ et la condition HATU. Toutefois ces deux acides ne se marquent qu'une fois alors qu'il y a la possibilité d'un triple marquage. Ceci peut s'expliquer par deux hypothèses. Soit par un manque de sensibilité du détecteur, car la quantité de molécule dérivée deux fois est trop faible, soit par l'encombrement stérique qui rend inaccessible les autres fonctions acides une fois l'acide marqué une première fois.

1.2. Cas des acides aminés

Le tableau 24 répertorie le rapport m/z attendus dans le cas de chaque acide aminé lorsqu'ils sont marqués par les étiquettes DMED, DAPPZ, DMPT et DMPP.

Tableau 24 : m/z attendus pour le marquage du DMED, DAPPZ, DMPT et DMPP de la fonction acide des acides aminés sous la forme $[M+H]^+$

		m/z $[M+H]^+$			
	Marqué	**DMED**	**DAPPZ**	**DMPT**	**DMPP**
Phénylalanine	1	236.1757	381.2285	372.2030	374.2060
Alanine	1	160.1444	305.1972	296.1717	298.1747
Arginine	1	245.2084	390.2611	381.2357	383.2387
Méthionine	1	220.1478	365.2006	356.1751	358.1781
Créatine	1	202.1662	347.2190	338.1935	340.1965
Cystéine	1	192.1165	337.1693	328.1438	330.1468
Acide hippurique	1	250.1550	395.2078	386.1823	388.1853
Acide aspartique	1	204.1342	349.1870	340.1615	342.1646
	2	274.2237	564.3292	546.2783	549.2782

La recherche de ces m/z dans les données LC-MS a permis de créer la figure 44 qui décrit l'aire brute des molécules marquées par les étiquettes chimiques [M+TAG] et par les agents de couplage [M+AG].

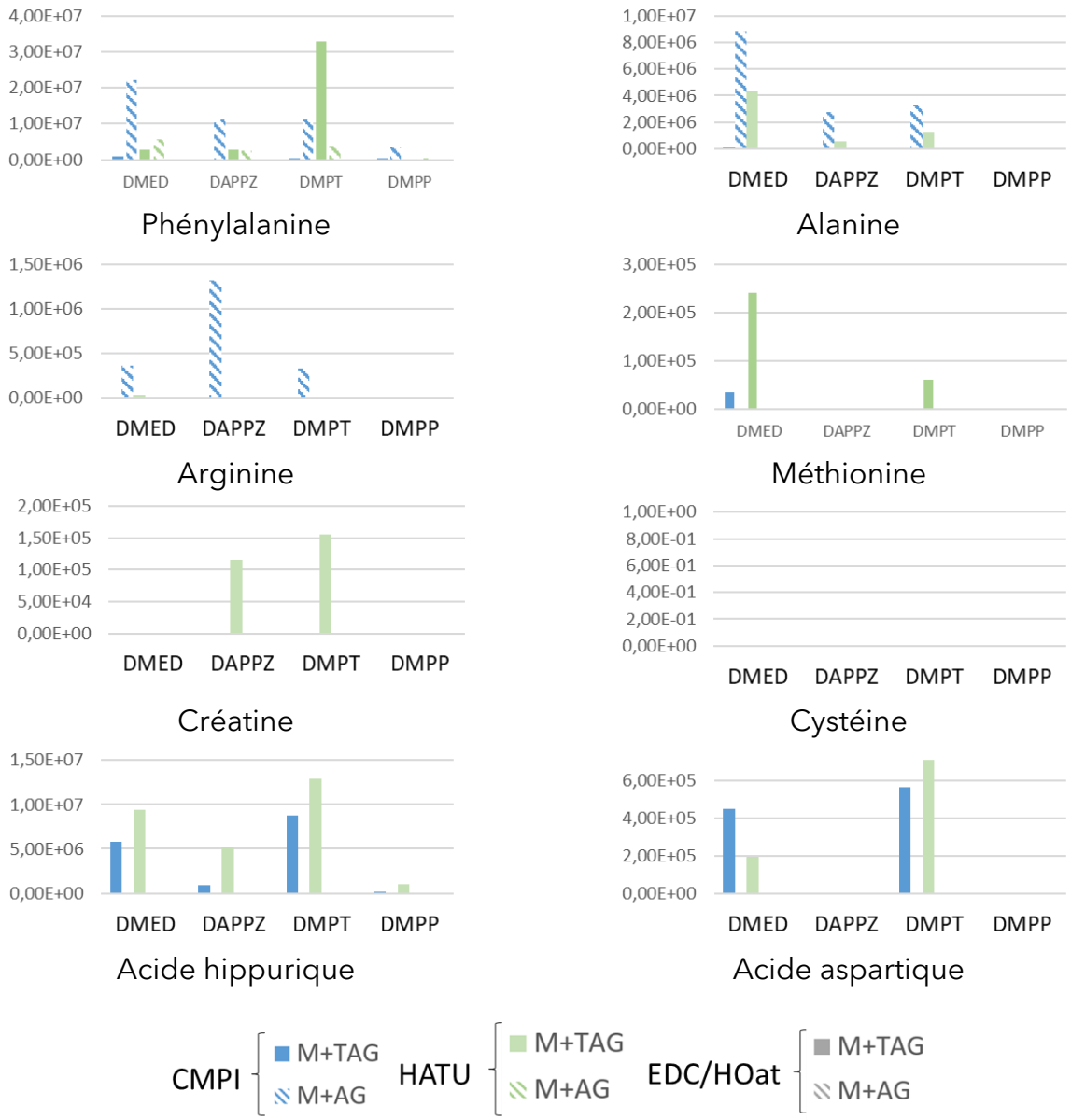


Figure 44 : Résultats de marquage des acides aminés par les dérivés pipérazine et amine. L'ordonnée représente l'aire brute des pics et l'abscisse l'étiquette chimique considérée. AG= agent de couplage, CMPI, HATU, EDC ou HOAt et TAG=DMED, DAPPZ, DMPT ou DMPP

Comme dans le cas précédent, les étiquettes chimiques DMED, DAPPZ et DMPT permettent de marquer les fonctions avec les agents de couplage CMPI et HATU. Les conditions utilisant le carbodiimide EDC et l'AAE HOAt ne permettent pas de dériver la fonction acide de ces composés. La cystéine n'est marquée par aucune des étiquettes chimiques. Les conditions de marquage ne sont donc pas adaptées à cet acide aminé.

Dans cette catégorie chimique de molécules, l'étiquette DMPT avec l'agent de couplage CMPI permet de marquer en plus grande proportion la phénylalanine et l'acide hippurique (aire brute de 10^7). Les analyses MS² sont réalisées en mode DDA et l'intensité du pic de l'acide hippurique couplé au DMPT est assez grande pour être sujette à une MS² par notre spectromètre de masse (figure 45).

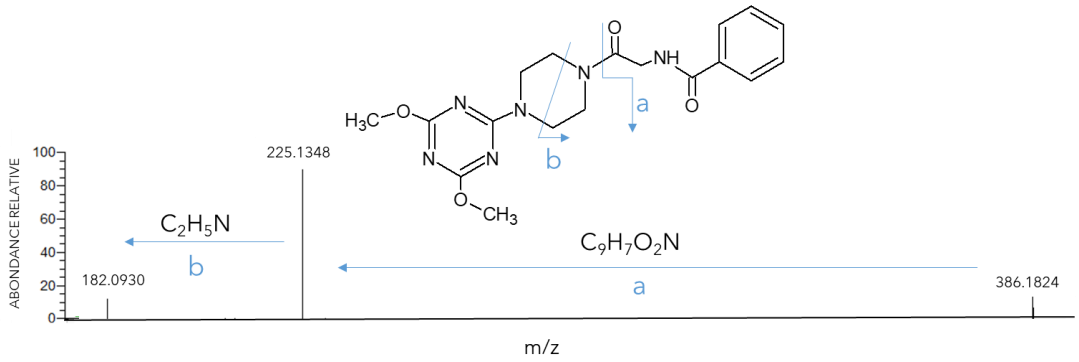


Figure 45 : Spectre MS² de l'acide hippurique couplé au DMPT sous les conditions HATU

L'ion moléculaire 386.182 (tableau 24) de l'acide hippurique couplé au DMPT est fragmenté avec une perte de $C_9H_7O_2N$. Cette perte de 161.047 correspond à la perte de l'acide hippurique couplé au DMPT. Le pic 225.134 ainsi obtenu est celui du DMPT, puis ce fragment perd C_2H_5N, qui correspond à la moitié du cycle pipérazine, pour donner l'ion 182.093. Ce schéma de fragmentation est similaire à celui de l'acide acétique couplé au DMPT figure 43 (d). Les fragments 225.234 et 182.093 sont retrouvés dans ces deux spectres MS².

Selon les agents de couplage utilisés, la réactivité des acides aminés est différente, entrainant des réactions de couplage non totales. En effet, l'intermédiaire réactionnel avec l'agent de couplage CMPI est détecté [M+AG] (AG=agent de couplage) dans le milieu dans le cas de la phénylalanine, l'alanine et l'arginine. Les m/z attendus pour le marquage des acides aminés sont répertoriés tableau 25.

Tableau 25 : m/z attendus des réactions secondaires de la fonction acide des acides aminés avec les agents de couplage (CMPI, HATU et EDC) et l'AAE (HOAt)

		m/z $[M]^+$	m/z $[M+H]^+$		
	Marqué	**CMPI**	**HATU**	**EDC**	**HOAt**
Phénylalanine	1	257.1284	284.1142	321.2285	285.1220
Alanine	1	181.0971	208.0829	245.1972	209.0907
Arginine	1	266.1611	293.1468	330.2611	294.1547
Méthionine	1	241.1005	268.0863	305.2006	269.0941
Créatine	1	223.1189	250.1047	287.2190	251.1125
Cystéine	1	213.0692	240.0550	277.1693	241.0628
Acide hippurique	1	271.1077	298.0935	335.2078	299.1013
Acide aspartique	1	225.0869	252.0727	289.1870	253.0805
	2	317.1370	370.1006	442.3136	404.1061

Le CMPI est une molécule halogénée qui comparée aux deux autres agents de couplage a une réactivité plus grande. En effet, le CMPI réagit selon une substitution nucléophile contrairement à l'HATU et l'HOAt/EDC qui réagissent selon une réaction acide/base. De plus, cette molécule est initialement chargée et conserve sa charge positive une fois couplée à l'acide. Cet intermédiaire réactionnel initialement chargé positivement est alors plus sensiblement détectable en MS que les deux autres agents de couplage utilisés dans cette étude. De plus, l'observation des spectres MS^2 de ses trois acides aminés couplés au CMPI nous permettent de remarquer qu'il y a une réaction « parasite ». En effet, le CMPI se fixe à la fonction amine des acides aminés. Les spectres MS^2 sont obtenus en mode DDA, l'intensité du pic de l'arginine couplée au CMPI n'est pas suffisante pour être fragmentée. Ainsi, la figure 46 représente les spectres MS^2 dans le cas de l'alanine (a) et la phénylalanine (b) couplées au CMPI. Dans les deux cas, l'ion moléculaire est fragmenté avec une perte de 46.007 associé à une formule brute de CH_2O_2. Cette perte correspond à la fonction acide de la structure chimique de l'alanine, et la phénylalanine. L'observation de cette perte à un m/z correspondant à [M+CMPI] confirme qu'il est impossible que le CMPI se soit fixé à la fonction acide. Il est donc forcément fixé à la fonction amine empêchant la suite de la réaction de couplage.

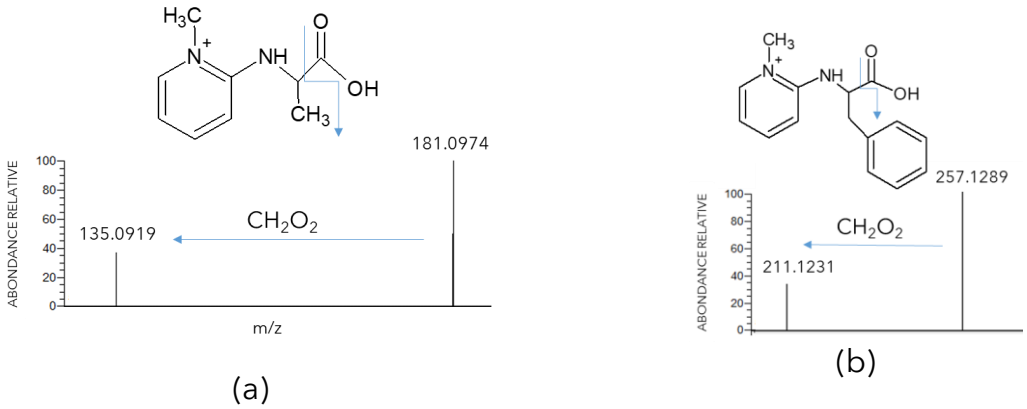


Figure 46 : Spectre MS² de l'alanine (a) et de la phénylalanine (b) couplées au CMPI

Ainsi dans le milieu réactionnel, il y a une concurrence de réactivité, le CMPI a le choix de se fixer sur les deux fonctions (acide et amine). Il faut savoir qu'une fonction amine primaire NH_2 est plus nucléophile qu'une fonction acide. Le CMPI se fixe donc préférentiellement sur la fonction amine, ne permettant pas l'activation de la fonction acide pour un couplage avec l'étiquette chimique. Cette observation explique donc pourquoi nous retrouvons en très faible proportion ces acides aminés couplés aux étiquettes chimiques pipérazine et amine.

Ce couplage M+AG n'est pas retrouvé dans le cas de la méthionine, la créatine, l'acide hippurique, l'acide aspartique. La méthionine a pourtant une structure chimique proche de celle de l'alanine ou la phénylalanine. Les formes de résonnances dans la structure de la créatine induites par les fonctions amines conjuguées peuvent expliquer la difficulté du CMPI à se fixer sur celles-ci. Dans le cas de l'acide hippurique, le groupement benzoyle joue le rôle de groupement protecteur de l'amine la rendant ainsi moins réactive qu'une fonction amine primaire. Ce groupement cétone aromatique permet ainsi de favoriser la réaction du CMPI avec la fonction acide et donc le couplage avec les étiquettes chimiques. Dans le cas de l'acide aspartique, il y a la possibilité qu'une liaison hydrogène se forme entre la fonction acide et la fonction amine (voir figure 47).

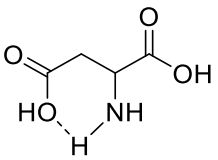


Figure 47 : Schéma de la liaison hydrogène dans la structure de l'acide aspartique

Cette liaison hydrogène rend la fonction amine moins réactive donc moins accessible pour un couplage avec le CMPI. La deuxième fonction acide de l'acide aspartique est libre d'accès pour un couplage avec le CMPI puis par l'étiquette chimique.

Dans le cas des acides aminés de notre étude, les étiquettes DMED, DMPT et DAPPZ permettent la dérivation de la fonction acide avec l'agent de couplage HATU. Les conditions utilisant l'agent de couplage CMPI ne sont pas adaptées car créent des réactions « parasites » stoppant le schéma réactionnel au niveau du produit intermédiaire.

1.3. Cas des α-hydroxy acides

La figure 48 décrit les résultats obtenus pour la dérivation de la fonction acide portée par les α-hydroxy-acides à partir de la mesure des m/z (tableau 26).

Tableau 26 : m/z attendus pour le marquage par le DMED, DAPPZ, DMPT et DMPP de la fonction acide des α-hydroxy acides sous la forme $[M+H]^+$

		m/z $[M+H]^+$			
	Marqué	**DMED**	**DAPPZ**	**DMPT**	**DMPP**
Acide lactique	1	161.1284	306.1812	297.1557	299.1588
Acide hydroxyphenyllactique	1	253.1546	398.2073	389.1819	391.1849
Acide malique	1	205.1183	350.1711	341.1456	343.1486
	2	275.2078	565.3133	547.2624	550.2622

De même que dans le cas des triacides, l'acide malique contient plusieurs fonctions acides (exactement deux fonctions) accessibles pour le marquage avec les étiquettes chimiques DMED, DAPPZ, DMPT et DMPP. Par exemple, les abscisses DMED 1 et DMED 2 représentent, respectivement, l'acide malique marqué une et deux fois par l'étiquette DMED.

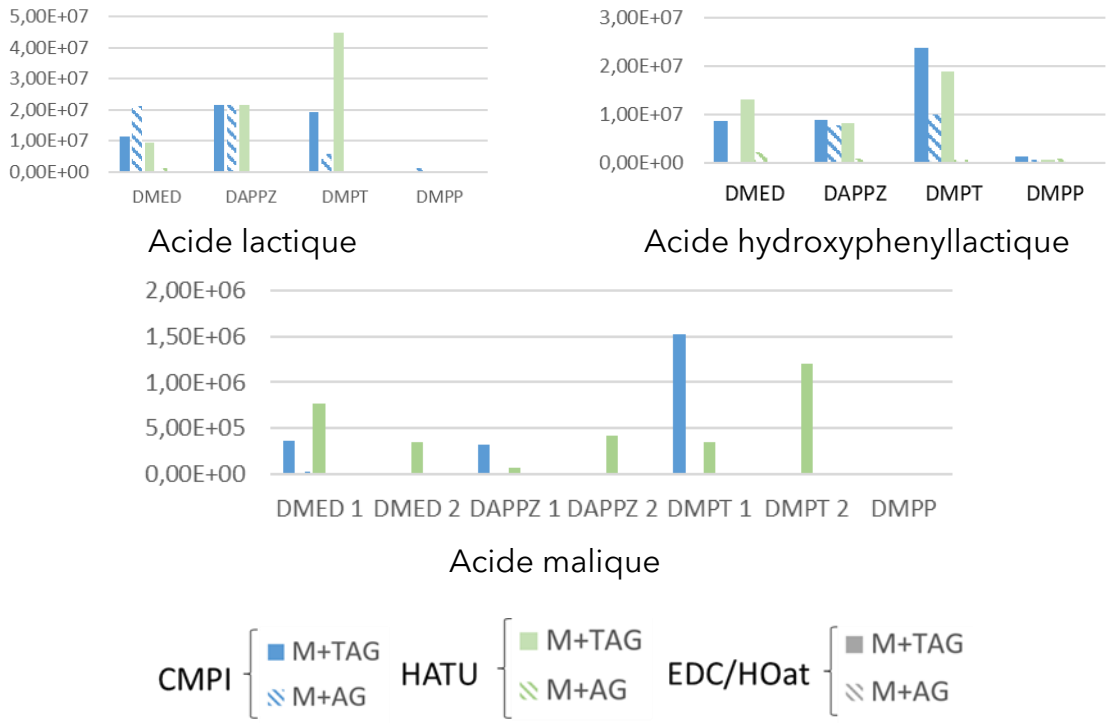


Figure 48: Résultats de marquage des α-hydroxy acides par les dérivés pipérazine et amine. L'ordonnée représente l'aire brute des pics et l'abscisse l'étiquette chimique considérée AG= agent de couplage, CMPI, HATU, EDC ou HOAt et TAG=DMED, DAPPZ, DMPT ou DMPP

Les α-hydroxy acides sont des composés possédant une ou plusieurs fonctions acides avec une fonction alcool en position alpha de l'acide. Le même constat est fait que dans le cas des acides aminés, triacides et acides gras c'est-à-dire que l'étiquette DMPP ne permet pas ou peu de marquer les acides. De même que l'agent de couplage EDC/HOAt ne permet pas une dérivation efficace de la fonction acide. Les trois étiquettes DMED, DAPPZ et DMPT permettent de marquer la fonction acide avec l'agent de couplage CMPI ou HATU dans des proportions similaires. L'HATU semble plus adapté que le CMPI pour marquer le diacide (acide malique) de plus les réactions sont totales ne laissant pas dans le milieu de forme M+AG (observée avec le CMPI). En effet, sous ces conditions il est possible d'observer l'acide malique marqué deux fois avec un ordre de grandeur plus grand avec l'étiquette DMPT.

Le spectre MS^2 de l'étiquette DMPT couplée à l'acide lactique sous les conditions de HATU est représenté figure 49.

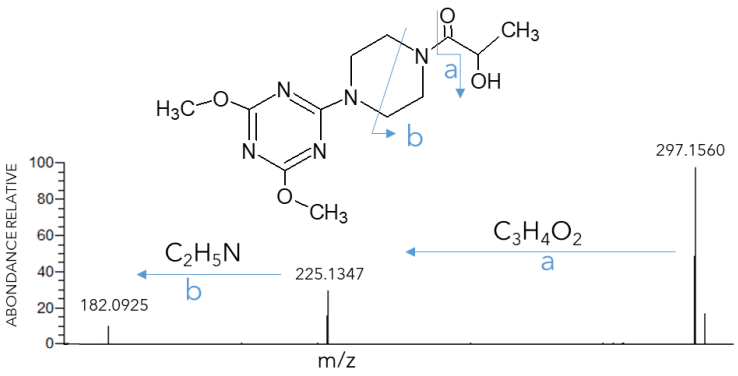


Figure 49 : Spectre MS² de l'acide lactique couplé au DMPT sous les conditions HATU

L'étude de ce spectre MS² permet de mettre en évidence une perte de 72.021 soit $C_3H_4O_2$ qui correspond à la perte de l'acide lactique. Puis une perte de 43.042 soit C_2H_5N qui correspond à la moitié de la pipérazine composant l'étiquette DMPT. Ce schéma de fragmentation MS² confirme bien le couplage de l'acide lactique avec l'étiquette DMPT.

Semblable aux acides aminés, l'intermédiaire de réaction [M+CMPI] est retrouvé dans le milieu pour l'acide lactique et l'acide hydroxyphenyllactique. Le tableau 27 résume les m/z attendus pour chaque α-hydroxy acide couplé avec les agents de couplage CMPI, HATU, EDC et HOAt.

Tableau 27 : m/z attendus des réactions secondaires de la fonction acide des α-hydroxy acides avec les agents de couplage (CMPI, HATU et EDC) et l'AAE (HOAt)

		m/z $[M]^+$	m/z $[M+H]^+$		
	Marqué	**CMPI**	**HATU**	**EDC**	**HOAt**
Acide lactique	1	182.0811	209.0669	246.1812	210.0747
Acide hydroxyphenyllactique	1	274.1073	301.0931	338.2074	302.1009
Acide malique	1	226.0710	253.0568	290.1711	254.0646
	2	318.1211	371.0547	443.2977	405.0902

Les spectres MS² de l'acide lactique et de l'acide hydroxyphenyllactique fixés au CMPI sont représentés, respectivement, figure 50 (a) et (b). Ces spectres ont le fragment 110.060 en commun qui correspond à la molécule de CMPI avec un hydroxyle à la place du chlore.

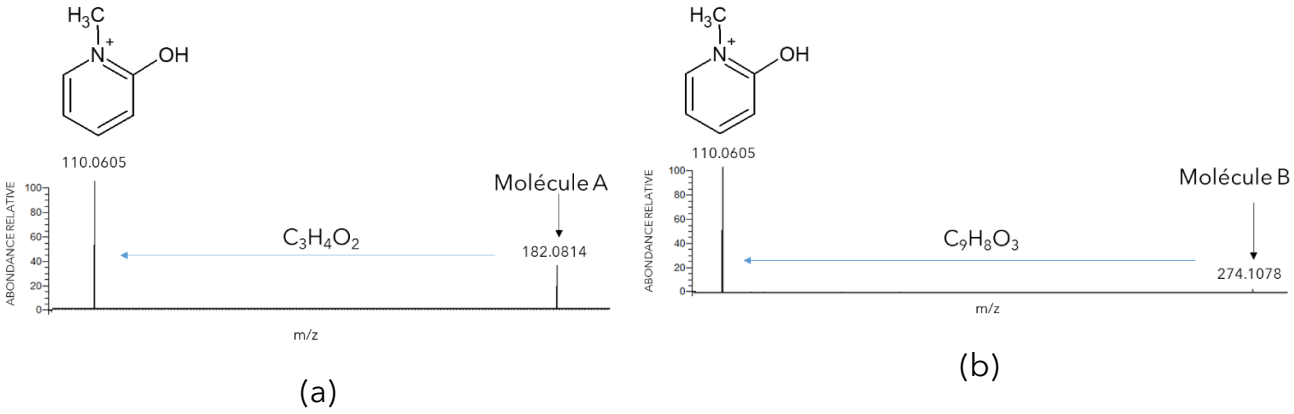


Figure 50 : Spectre MS² du couplage avec le CMPI dans le cas de l'acide lactique (a) et l'acide hydroxyphenyllactique (b)

Dans le cas de l'acide lactique, la fragmentation de l'ion 182.081 (molécule A) entraine une perte de 72.020 correspondant à une formule brute de $C_3H_4O_2$. Cette formule brute correspond à la perte de l'acide lactique déshydraté. La fragmentation de l'acide hydroxyphenyllactique couplé au CMPI entraine également une perte de cet acide déshydraté. En effet, le pic moléculaire de 274.107 (molécule B) est fragmenté avec une perte de 164.047 correspondant à une formule brute $C_9H_8O_3$ qui est l'acide hydroxyphenyllactique déshydraté. Toutefois, la perte de ces deux acides sous forme déshydratée ne permet pas de déterminer sur quelle fonction le CMPI s'est fixé préférentiellement. D'après la figure 51, il y a deux possibilités de structures chimiques correspondant au m/z du [M+CMPI] et à la perte de forme déshydraté de ces deux α-hydroxy acides. Le CMPI s'est fixé soit sur la fonction alcool de ces deux composés, soit sur la fonction acide carboxylique.

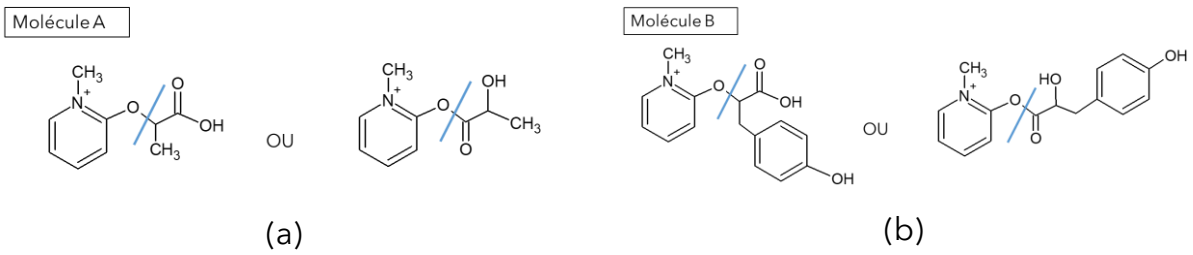


Figure 51 : Structures chimiques de l'acide lactique (a) et l'acide hydroxyphenyllactique (b) couplés au CMPI sur leurs fonctions hydroxyle et carboxylique

Il n'est donc pas possible de vérifier si le CMPI s'est fixé sur la fonction alcool ou sur la fonction acide carboxylique. Toutefois, l'analyse des chromatogrammes pour l'acide lactique (figure 51 (a)) et l'acide hydroxyphenyllactique (figure 51 (b)) permet l'observation de deux temps de rétentions pour un même m/z. En effet, pour l'acide lactique le m/z 182.081 ([M+CMPI]) est détecté à 0.88min et 1.79min et l'ion 274.107 ([M+CMPI]) est détecté à 0.91min et 5.69min pour l'acide hydroxyphenyllactique. Les pics en début de chromatogrammes sont plus importants. Cela montre qu'il y a une forme (figure 52) plus prépondérante que l'autre.

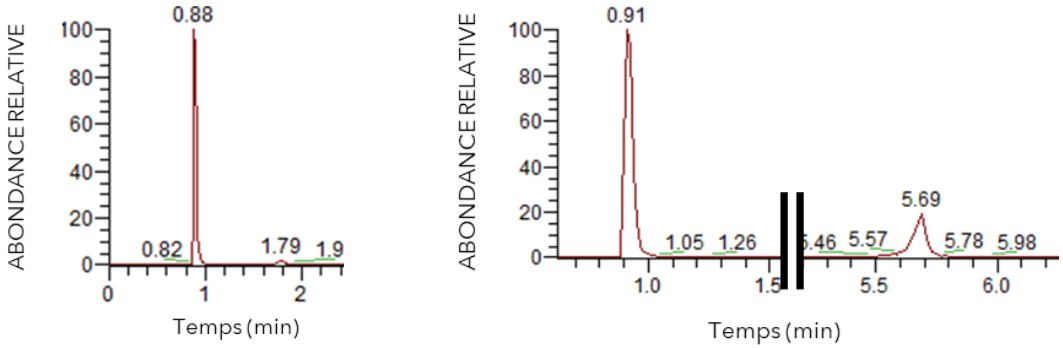


Figure 52 : Chromatogrammes de l'acide lactique pour un m/z de 182.081 (a) et de l'acide hydroxyphenyllactique pour un m/z de 274.107 (b) couplés au CMPI

L'observation des temps de rétention montre bien qu'il y a deux composés qui ont le même m/z mais deux temps de rétention distincts. Les deux structures chimiques des intermédiaires réactionnels (figure 51) sont donc retrouvées dans le milieu dans des proportions différentes. Deux explications peuvent être faites. Premièrement le CMPI reste fixé à la fonction acide carboxylique en formant un intermédiaire stable empêchant l'étape de couplage avec l'étiquette chimique de se faire. Deuxièmement, le CMPI se fixe à la fonction hydroxyle de l'acide lactique et l'acide hydroxyphenyllactique stoppant la réaction de couplage à l'intermédiaire réactionnel. Nous pouvons donc supposer qu'il y a une concurrence réactionnelle entre les fonctions acide et alcool de ces deux α-hydroxy acides. Dans le cas de l'acide malique, l'intermédiaire réactionnel avec le CMPI n'est pas retrouvé. Ceci est surement dû à la formation d'une liaison hydrogène entre la fonction alcool et la

fonction acide de la molécule (comme dans le cas de l'acide aspartique voir figure 47). La fonction alcool est donc moins accessible rendant ainsi l'activation et la dérivation de la fonction acide possible.

Les étiquettes pipérazines et amine permettent la dérivation de la fonction acide des α-hydroxy acides avec l'agent de couplage HATU. Les conditions utilisant l'agent de couplage CMPI ne sont pas adaptées aux α-hydroxy acides.

1.4. Bilan du marquage en deux étapes

Parmi les 4 dérivés pipérazines et amine considérés dans cette étude, les étiquettes DMED et DMPT permettent d'avoir les meilleurs résultats de dérivation. Dans le cas de l'étiquette DMED cela peut s'expliquer par sa structure chimique qui est différente des trois autres. En effet, cette étiquette chimique est linéaire et a une masse moléculaire plus petite ce qui lui permet d'accéder plus facilement à la fonction acide activée lors du couplage. De plus, son site de fixation est une amine secondaire comparé aux trois autres étiquettes qui portent une amine primaire. Or, une fonction amine primaire a une réactivité plus grande qu'une fonction amine secondaire, le DMED est donc plus réactif que le DAPPZ, DMPT et DMPP. Les meilleurs résultats de marquage obtenus par l'étiquette DMPT peuvent provenir de sa détection en MS. En effet, par rapport à l'étiquette DMPP, l'étiquette DMPT a un azote supplémentaire dans son cycle aromatique lui procurant de meilleurs propriétés d'ionisation la rendant ainsi plus ionisable que les autres étiquettes. La dérivation de la fonction acide par l'étiquette DMPP est peu observée même dans les conditions utilisant l'EDC et l'HOAt qui sont celles décrites par la littérature [43]. Deux hypothèses peuvent expliquer cette observation. La première est que le DMPP est moins ionisable que les autres étiquettes rendant ainsi la forme [M+TAG] moins visible par la MS. La deuxième est qu'il y a une fragmentation en source du DMPP. Le m/z recherché lors de l'étude n'est donc plus valable ou le fragment généré est neutre donc pas détectable en MS.

L'analyse des spectres MS^2 a permis de mettre en évidence des schémas de fragmentation retrouvés pour chaque étiquette. Ce schéma peut être la perte d'un fragment spécifique ou la présence dans le spectre MS^2 d'un ion fragment. Dans le cas du DMED, la perte du fragment C_2H_5N correspondant au groupement diméthylamine est spécifique à l'étiquette DMED. Cette fragmentation est également décrite par Quan-Fei Zhu et son équipe [177]. Dans le cas de l'étiquette DAPPZ, le fragment $C_5H_{16}N_2O$ correspondant au groupement pipérazine lié à une cétone est une perte spécifique à cette étiquette qui est observé par Ogawa et son équipe [44]. Deux ions fragments le 225.148 et le 182.092 sont retrouvés dans les spectres MS^2 des molécules couplées avec le DMPT. Seul le fragment à 225.148 est décrit par la littérature lors d'un couplage du DMPT avec des acides gras [43]. Nos résultats sont donc en accord avec ceux de la littérature.

Parmi les 5 catégories d'acides considérées dans cette étude, les étiquettes chimiques permettent de marquer plus efficacement les acides gras et triacides. Cette observation est en lien avec la littérature car les étiquettes ont été développées pour le marquage des acides gras contenus dans l'urine [43], le sérum [177] ou encore la salive [44]. La réactivité d'un acide gras et des autres types d'acides utilisés dans cette étude est différente. En effet, on a pu observer que la présence de groupement en position alpha de l'acide tel qu'un hydroxyle ou une amine complexifie la formation de l'amide. La formation de l'amide a été évaluée selon 3 protocoles utilisant 3 types d'agents de couplage. Plusieurs observations sont faites. L'agent de couplage CMPI permet de former un complexe M+AG très stable empêchant la réaction de cet intermédiaire avec l'étiquette chimique. L'agent HATU forme également un complexe M+AG mais moins visible et ne permet pas d'améliorer le couplage de l'acide avec l'étiquette chimique. L'agent de couplage HATU n'est donc pas assez stable et/ou réactif. Les conditions utilisant l'EDC en duo avec l'HOAt ne permettent d'observer aucun intermédiaire de réaction et aucun couplage d'acide avec les étiquettes chimiques. Ceci peut s'expliquer par un pH trop acide ne permettant pas à l'acide d'être sous forme ionisé réduisant ainsi le

couplage de l'acide avec l'agent de couplage EDC. Cette différence de réactivité des agents de couplage entraine un marquage de la fonction acide à la hauteur de l'efficacité de l'agent de couplage utilisé.

De par la présence des fonctions en alpha de l'acide et la nécessité d'agent de couplage, le couplage en deux étapes par les dérivés pipérazine et amine n'est pas adapté à la dérivation de fonction acide carboxylique. De plus, un transfert de méthode en clinique avec des échantillons biologiques riches en familles de molécules engendrerait de nombreuses réactions parasites entre les métabolites, les agents de couplage et les étiquettes chimiques. Le mélange final serait trop complexe et difficile à exploiter. Nous considérons donc que ces étiquettes ne sont pas adaptées pour le marquage des fonctions acides dans un milieu biologique complexe. Afin de pallier aux nombreuses réactions parasites et aux intermédiaires réactionnels, un marquage en une étape semblerait plus adapté.

2. Marquage en une étape par le DmPABr de la fonction acide carboxylique

La molécule de DmPABr permet le marquage de la fonction acide en réalisant une substitution nucléophile avec la fonction acide préalablement déprotonée [199].

Le but ici est de déterminer si un marquage en une étape des composés précédents est plus efficace qu'un marquage en deux étapes.

2.1. Cas des acides gras et triacides

Les acides gras sont des acides carboxyliques à chaîne aliphatique possédant une structure chimique simple et les triacides possèdent 3 fonctions acides carboxyliques (figure 35).

Tableau 28 :m/z des standards marquées 1,2,3 ou 4 fois par l'étiquette DmPABr (TAG) en détection positif sous forme $[M+H]^+$

		1TAG	2TAG	3TAG	4TAG
Métabolites	Nombre de site	m/z	m/z	m/z	m/z

Acide acétique	1	222,112			
Acide cis-aconitique	3	336,108	497,191	658,276	
Acide citrique	4	354,118	515,202	676,286	837,370
Acide laurique	1	362,269			
Acide subérique	2	336,181	497,264		

Les résultats obtenus pour le marquage de ces composés par le DmPABr sont représentés figure 53 à partir de la mesure de leur m/z (tableau 28).

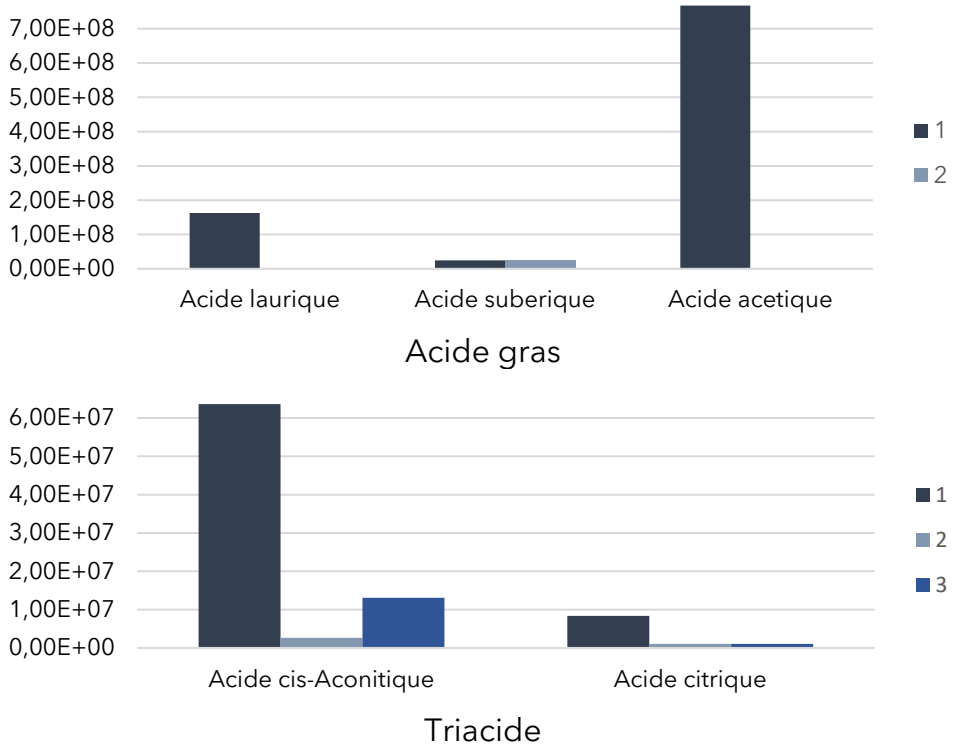


Figure 53 : Histogrammes des résultats de marquage par l'étiquette DmPABr dans le cas des acides gras et triacides. L'axe des ordonnées représente l'aire brute sous la courbe de chaque molécule marquée. Le code couleur 1,2 et3 représente le nombre de fois où la molécule est marquée par le DmPABr.

L'étiquette DmPABr permet de marquer les 3 acides gras avec une aire de pic très importante pour l'acide acétique. Cet acide est l'acide carboxylique ayant la structure chimique la plus simple avec une chaîne carbonée composée de 2 atomes. Avec une chaîne carbonée de 12 atomes, l'acide laurique est également marqué par le DmPABr. Le diacide, acide subérique, est retrouvé sous forme marqué 1 fois mais également 2 fois par le DmPABr. Cette observation est similaire pour les triacides (acides cis-aconitique et citrique) qui sont marqués 1,2 et 3 fois par le DmPABr. Toutefois la forme marquée 1 fois est majoritairement présente dans le milieu. Le marquage une fois majoritaire dans le milieu aurait pu s'expliquer par un encombrement stérique dû à la liaison de l'acide avec le DmPABr, entrainant une gêne quant à l'ajout d'une deuxième ou troisième étiquette sur la molécule une fois marquée une première fois. Mais la forme marquée 3 fois de l'acide cis-aconitique avec le DmPABr est plus importante que celle marquée deux fois. Cette hypothèse n'est donc pas valable.

2.2. Cas des acides aminés et α-hydroxy acides

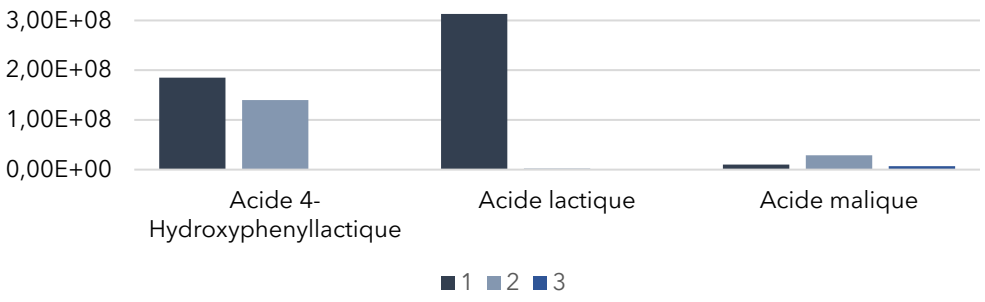
Les acides aminés et les α-hydroxy acides sont des composés qui ont en alpha de leur fonction acide carboxylique une fonction amine ou hydroxyle. Le rapport m/z de ces composés marqués avec l'étiquette DmPABr est résumé tableau 29.

Tableau 29 : m/z des standards marquées 1,2,3 ou4 fois par l'étiquette DmPABr (TAG) en détection positif sous forme $[M+H]^+$

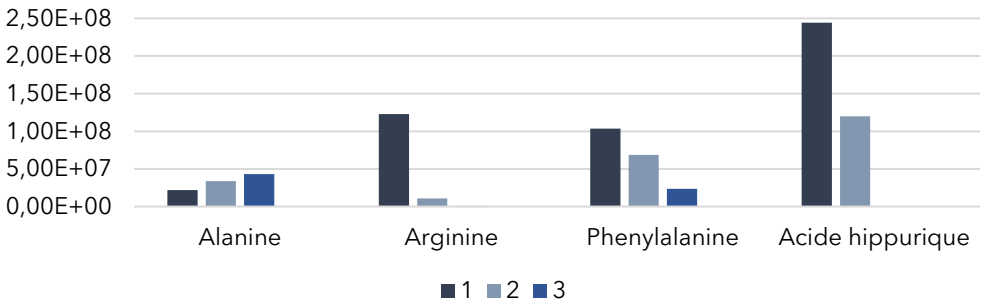
		1TAG	2TAG	3TAG	4TAG
Métabolites	Nombre de site	m/z	m/z	m/z	m/z
Acide hydroxyphenyllactique	3	344,149	505,233	666,317	
Alanine	3	251,139	412,223	573,307	
Arginine	3	336,203	497,287	658,371	
Acide aspartique	4	295,129	456,212	617,297	778,381
Créatine	3	293,161	454,244	615,329	

Cystéine	4	283,111	444,195	605,279	766,363
Acide hippurique	2	341,15	502,233		
Acide lactique	2	252,123	413,207		
Acide malique	3	296,113	457,196	618,281	
Méthionine	3	311,142	472,226	633,311	
Phénylalanine	3	327,170	488,254	649,338	

Les résultats obtenus pour le marquage de ces composés par le DmPABr sont représentés figure 54.



α-Hydroxy-acide



Acide aminé

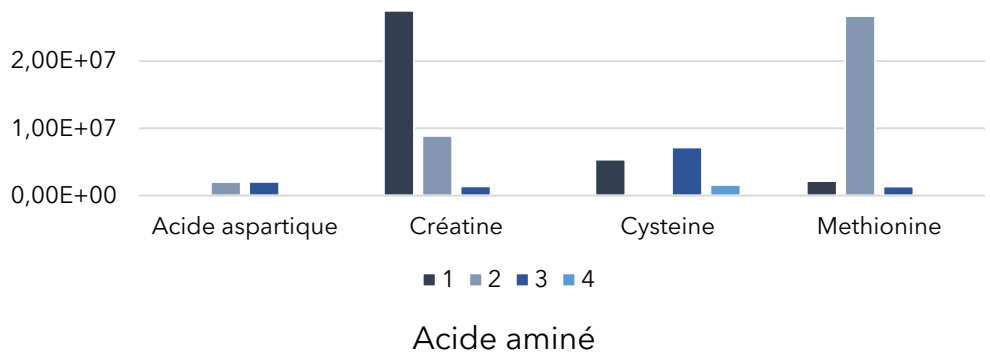


Figure 54 : Histogrammes des résultats de l'étiquette DmPABr dans le cas des α-hydroxy acides et acides aminés. L'axe des ordonnées représente l'aire brute sous la courbe de chaque molécule marquée. Le code couleur 1,2,3 et 4 représente le nombre de fois la molécule est marquée par le DmPABr.

D'après nos observations, ces composés peuvent être marqués jusqu'à 4 fois par le DmPABr. L'alanine est détectée jusqu'à 3 fois marquée par le DmPABr. La cystéine a une structure chimique proche de celle de l'alanine avec une fonction thiol en plus en bout de chaine. D'après l'observation des pics obtenus dans le cas de la cystéine, nous remarquons qu'elle est marquée 4 fois avec le DmPABr. En comparant ce résultat à celui de l'alanine, nous pouvons émettre l'hypothèse que la fonction thiol de la cystéine est également marquée par le DmPABr. Ainsi, ces observations mettent en évidence que le DmPABr bien que décrit initialement pour marquer la fonction acide carboxylique se fixe également sur les fonctions amine, alcool et thiol. Parmi les 11 composés considérés ici, 6 ont la forme marquée une fois présente majoritairement dans le milieu. Dans les autres cas comme celui de l'alanine et de la cystéine, la forme marquée 3 fois est majoritaire et la forme marquée 2 fois est majoritaire dans le cas de la méthionine. Ceci est peut-être notamment dû à l'ionisation de ces formes qui est meilleure grâce à la présence du DmPABr en plus grande quantité sur la molécule.

Le DmPABr permet donc de marquer les acides aminés et les α-hydroxy acides. Toutefois, ces composés sont marqués entre 1 et 4 fois par le DmPABr. Cette étiquette peut donc se fixer à la fonction acide carboxylique mais également sur les fonctions amine, alcool et thiol. Le groupement marqué préférentiellement par le DmPABr peut être déterminé à l'aide d'une analyse MS^2. Cette analyse permet de

connaitre le schéma de fragmentation du DmPABr et d'identifier d'éventuels fragments types du couplage du DmPABr avec une fonction acide carboxylique, amine ou alcool

2.3. Acquisition MS²

Afin de confirmer, les résultats précédents et vérifier le site de fixation du DmPABr avec les métabolites, une analyse ciblée est réalisée pour les m/z de chaque molécule marquée 1 à 4 fois. Des analyses MS² en détection positive sont réalisées par une méthode MRM (décrite chapitre 1). Cette technique consiste à fragmenter tous les m/z définis.

2.3.1. Cas des acides gras triacides

Les résultats obtenus en full-scan pour la catégorie chimique des acides gras et triacides a permis de confirmer que l'étiquette DmPABr se fixe sur les fonctions acides carboxyliques. Toutefois, dans le cas de l'acide citrique qui est marqué jusqu'à 3 fois avec le DmPABr, la question est de savoir si sa fonction hydroxyle est également marquée. Le tableau 30 résume les fragments majoritaires retrouvés dans chaque spectre MS².

Tableau 30 : Tableau du fragment MS² majoritaire détecté après fragmentation en mode MRM des acides gras et triacides couplés au DmPABr (TAG) 1 à 3 fois. RT = temps de rétention

	1 TAG		2 TAG		3 TAG	
Métabolites	RT (min)	mz	RT (min)	m/z	RT (min)	m/z
Acide acétique	9.74	180.103				
Acide cis-aconitique	9.52 9.80 10.10	180.103 134.097 180.103	10.22 10.88	134.097 134.097	11.52	134.097
Acide citrique	10.92	148.075	9.36 10.21	134.097 134.097	11.58	349.153
Acide laurique	13.50	180.103				
Acide subérique	11.68	180.103	10.60	134.097		

L'analyse des spectres MS² met en évidence deux types de fragments majoritaires dans chaque MS², le 180.10 et le 134.09, pour les acides acétique, cis-aconitique,

laurique et subérique. Le fragment 180.10 correspond à l'étiquette DmPABr oxydé et le fragment 134.09 à un réarrangement de son cycle aromatique (tableau 32). Les m/z attendus pour l'acide cis-aconitique marqué une fois et deux fois ont plusieurs temps de rétention. Par exemple, la forme marquée une fois a trois temps de rétention assez proches (9.52min, 9.80min et 10.10min). Cette observation permet d'estimer que le DmPABr s'est fixé aux trois fonctions acides distinctement. De même que dans le cas de la forme marquée deux fois, il a deux types de structure chimique possibles d'où l'observation de deux temps de rétention (10.22min et 10.88min). Bien que l'acide citrique ait une structure chimique proche de l'acide cis-aconitique, ses spectres MS² contiennent des fragments différents (figure 55).

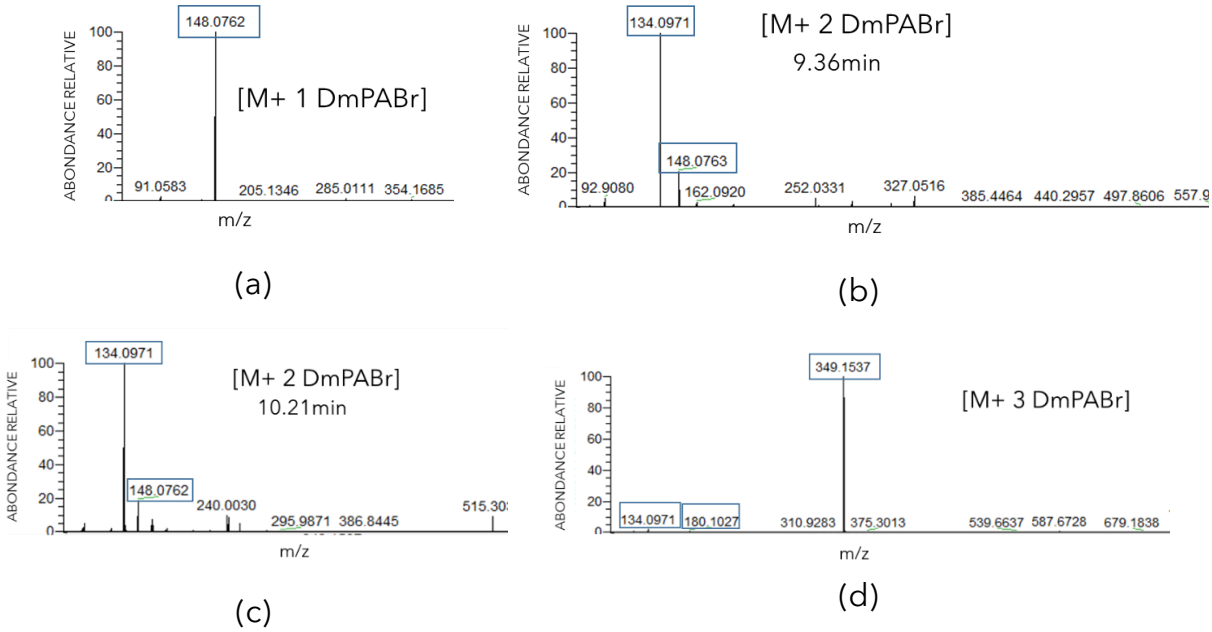


Figure 55 : spectres MS² de l'acide citrique marqué par l'étiquette chimique DmPABr, une fois (a) , deux fois (b) et (c) et trois fois (d)

Ces 4 spectres MS² ne contiennent pas l'ion moléculaire de l'acide citrique marqué avec le DmPABr, cela peut s'expliquer par une énergie de collision qui est trop forte pour cette molécule entrainant la fragmentation complète de l'ion moléculaire. La forme marquée une fois par le DmPABr est retrouvée à un seul temps de rétention avec majoritairement le fragment 148.076. Ce fragment est peut-être issu de la liaison de la fonction hydroxyle de l'acide citrique avec le DmPABr. D'après notre tolérance en masse de 10ppm, un m/z de 148.075 correspond à une formule brute $C_9H_{10}ON$ ou $C_7H_8N_4$. Nous savons que la formule brute du DmPABr est $C_{10}H_{12}NOBr$, ainsi la première formule brute pourrait correspondre à un fragment du DmPABr avec une perte d'un carbone et de deux hydrogènes (CH_2). La perte d'un fragment CH_2 peut provenir de la fragmentation du groupement diméthylamine mais ne permet pas de distinguer le site de fixation du DmPABr sur l'acide citrique. L'acide citrique marquée 2 fois à des spectres MS² ayant majoritairement le fragment 134.097 (figure 55 (b) et (c)). Notons également, la présence de ce même fragment 148.076 retrouvé dans le premier spectre (figure 55 (a)). Le spectre MS² de la forme marquée 3 fois (figure 55 (d)) a majoritairement l'ion fragment 349.153. Cet ion peut correspondre à 3 formules brutes $C_5H_{21}O_8N_{10}$, $C_{19}H_{19}O_2N_5$ ou $C_{21}H_{21}O_3N_2$. Cependant, à ce jour nous n'avons pas pu déterminer à quoi ce fragment correspond. Il est peut-être issu d'une molécule présente dans le milieu réactionnel ayant un m/z proche de l'acide citrique [M + 3 DmPABr]. En effet, le quadrupôle utilisé pour le mode MRM à une fenêtre de filtration de +/- 1 m/z. Ce qui entraine que des m/z proches peuvent être fragmentés en produisant des ions fragments de forte abondance. Toutefois, dans ce même spectre nous pouvons noter la présence des fragments 134.097 et 180.103 identifiés précédemment comme appartenant au DmPABr. Mais ces fragments ne permettent pas de différencier le site de fixation du DmPABr. Ainsi l'acide citrique est bien marqué jusqu'à 3 fois par le DmPABr sur ses fonctions acides carboxylique et/ou hydroxyle.

L'analyse des spectres MS² des acides gras et triacides a permis de mettre en évidence deux fragments spécifiques communs à la plupart des MS² mais n'a pas

permis de différencier le marquage d'une fonction acide carboxylique par rapport à une fonction hydroxyle.

2.3.2. Cas des acides aminés et α-hydroxy acides

L'analyse full-scan des catégories chimiques acides aminés et α-hydroxy acides a montré que l'étiquette DmPABr se fixe à la fonction acide carboxylique mais également aux fonctions amines primaire, secondaire, hydroxyle et thiol. Les spectres MS² en mode MRM sont effectués afin de différencier le site de fixation sur lequel le DmPABr réagit (tableau 31). Toutefois, de même que dans le cas des acides gras et triacides, deux fragments sont majoritairement retrouvés le 180.103 et le 134.097 (tableau 32) ne permettant pas de différencier le site de fixation du DmPABr sur les acides aminés et α-hydroxy acides.

Tableau 31 : Tableau du fragment MS² majoritaire détecté après fragmentation en mode MRM des acides aminés et α-hydroxy acides couplés au DmPABr (TAG) 1 à 4 fois. RT = temps de rétention

	1 TAG		2 TAG		3 TAG		4 TAG	
Métabolites	RT (min)	mz	RT (min)	m/z	RT (min)	m/z	RT (min)	m/z
Acide hydroxyphenyllactique	10.01	180.103	10.21	134.097	10.20	134.097		
Acide aspartique	/	/	10.20	134.097	10.22	134.097	/	/
Acide hippurique	10.21	180.103	9.25 9.96 10.64	134.097 161.090 134.097				
Acide lactique	9.58	180.103	9.36	134.097				
Acide malique	9.16	180.103	10.63	180.103	10.21	134.097		
Alanine	7.74 9.57	180.103 180.103	9.90	134.097	11.08	134.097		
Arginine	5.91	180.103	10.21	134.097	/	/		
Créatine	11.70	180.103	10.22	134.097	10.21	148.075		
Cystéine	10.49 11.53	134.097 134.097	/	/	13.29	134.097	12.51	134.097
Méthionine	8.85	180.103	9.04 9.23	134.097 134.097	10.19	134.097		
Phénylalanine	9.31 10.26	180.103 134.097	9.73	134.097	12.16	134.097		

Nous observons que pour un même m/z plusieurs temps de rétention sont détectés. Cette observation peut laisser supposer que le DmPABr se fixe sur différentes fonctions chimiques d'une même molécule entrainant des propriétés physico-chimiques et des temps de rétention différents. Par exemple, l'acide hipuurique marquée deux fois par le DmPABr à 3 temps de rétention (figure 56) avec comme un ion fragment majoritaire 134.097 à 9.25min et 10.64min et l'ion 161.090 à 9.96min (figure 57).

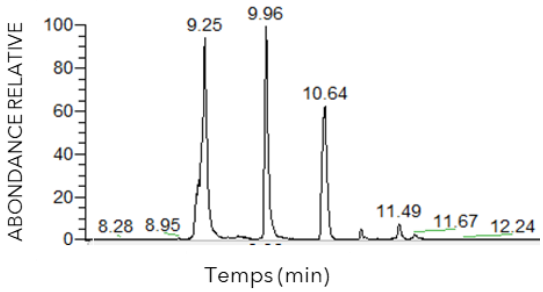
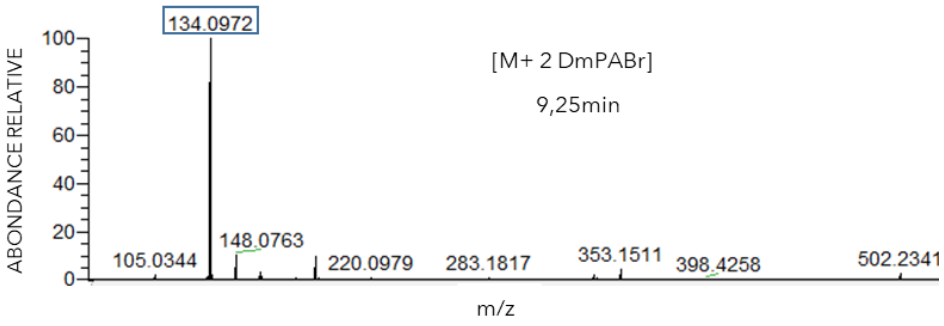


Figure 56 :Pics chromatographiques correspondant à l'acide hippurique marqué deux fois par le DmPABr (m/z de 502.2)

Or d'après sa structure chimique, l'acide hippurique peut se coupler deux fois avec le DmPABr sur sa fonction acide et sur sa fonction amine secondaire. Nous pouvons donc supposer que l'acide hippurique marqué deux fois n'a qu'un seul temps de rétention. La présence de trois temps de rétention au m/z de 502.233 [M+ 2 DmPABr] correspondant à l'acide hippurique marqué deux fois par le DmPABr peut être expliquée par la présence d'impureté.



(a)

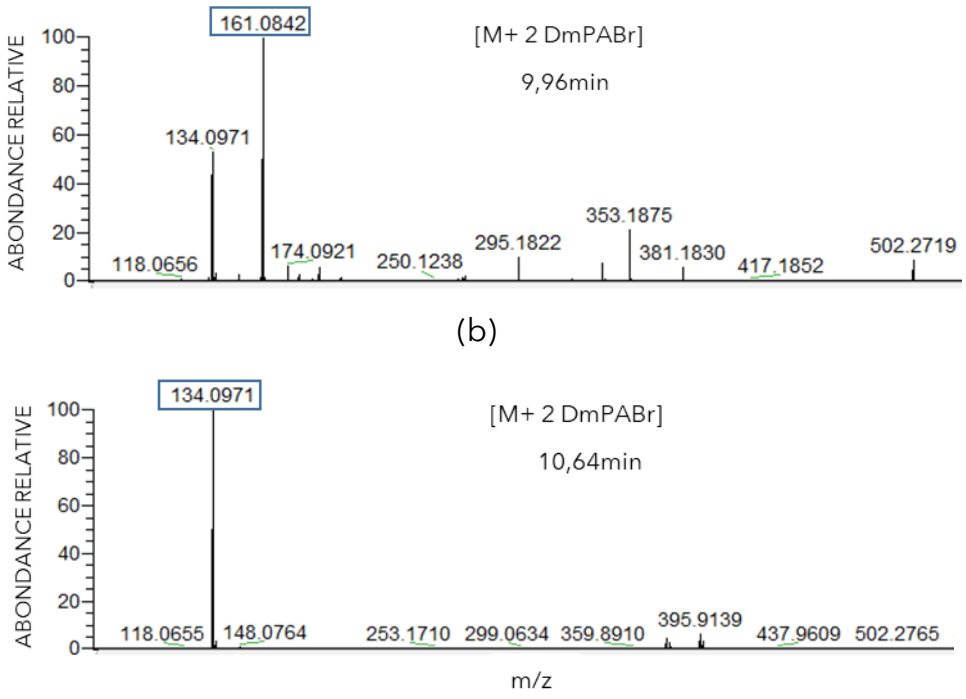
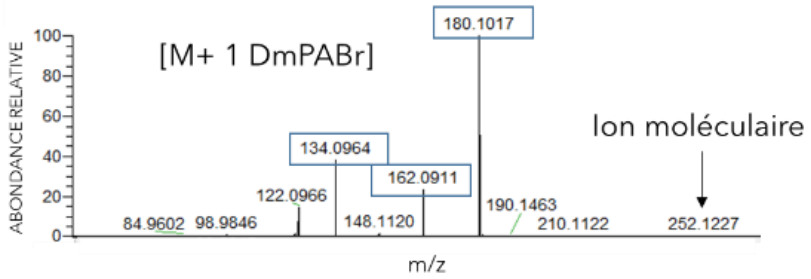


Figure 57 : Spectre MS² de l'acide hippurique marqué deux fois par le DmPABr au temps de rétention 9.25min (a), 9.96min (b) et 10.64min (c)

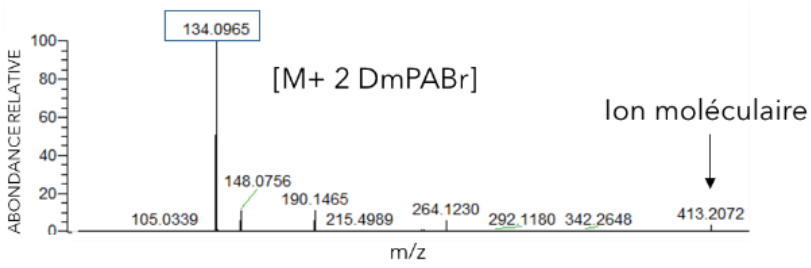
En effet, dans un premier temps, l'ion moléculaire de m/z 502.233 [M+2 DmPABr] n'est pas retrouvé dans les 3 spectres MS². De plus ; l'ion fragment 161.090, retrouvé majoritairement au temps 9.96min a pour formule brute $C_{10}H_{13}ON$. Cette formule brute est proche de celle de l'ion fragment 180.103 déshydraté, mais n'ai présent dans aucun autre spectre de cette étude. L'obtention de cet ion fragment provient peut-être d'une impureté qui a un m/z proche de 502.233. De plus, l'analyse des spectres MS² aux deux autres temps de rétention ne permet pas de les différencier tout en sachant qu'ils ont majoritairement l'ion 134.097. La présence de ces 3 pics chromatographiques au m/z correspondant à l'acide hippurique marqué 2 fois est peut être due à des impuretés de m/z proche.

L'acide lactique est le composé des α-hydroxy acides ayant la structure chimique la plus simple ($C_3H_6O_3$). L'analyse de son spectre MS² peut aider à distinguer le marquage de l'étiquette DmPABr sur une fonction acide carboxylique ou alcool. Cependant, ses spectres (figure 58 (a) et (b)) ont majoritairement les fragments

180.103 et 134.097 ne permettant pas de connaitre le site de fixation du DmPABr sur l'acide lactique. Le fragment 162.091 (figure 58 (a)) est lié au fragment 180.10 qui a perdu son groupement hydroxyle de bout de chaine.



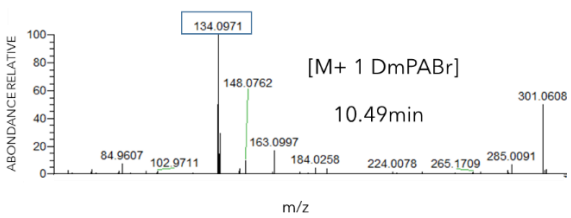
(a)



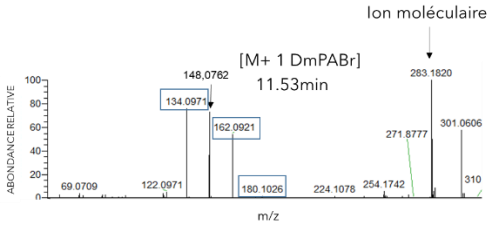
(b)

Figure 58 : Spectre MS² de l'acide lactique couplé au DmPABr une fois (a) et deux fois (b)

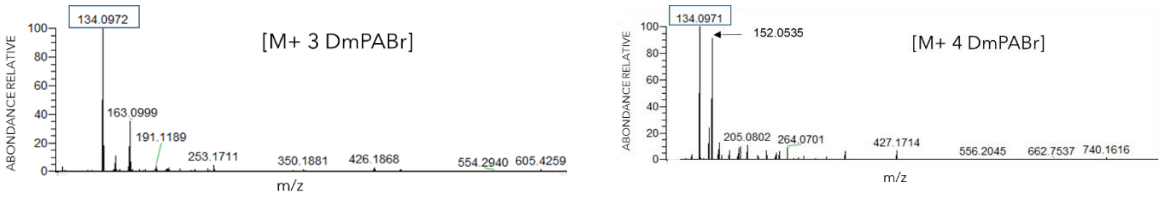
Parmi les acides aminés considérés dans cette étude, la cystéine a la structure chimique qui contient trois types de groupement chimique différents, une amine primaire, un acide carboxylique et un thiol. L'analyse de ses spectres MS² peuvent aider à la compréhension du marquage du DmPABr sur ces trois fonctions chimiques. La figure 59 représente les spectres MS² de la cystéine marquée une fois (figure 59 (a) et (b)), trois fois (figure 59 (c)) et quatre fois (figure 59 (d)) par le DmPABr.



(a)



(b)



(c) (d)

Figure 59 : Spectre MS² de la cystéine couplé au DmPABr une fois (a) et (b), trois fois (c) et quatre fois (d)

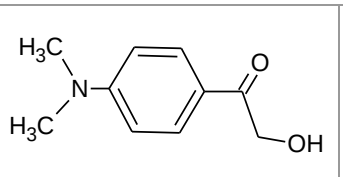
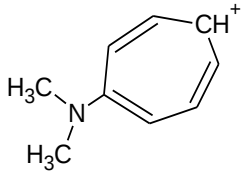
Dans ces quatre spectres MS², seul le spectre de la cystéine marquée 1 fois par le DmPABr à 11.53min contient le pic de l'ion moléculaire. Cette absence de l'ion moléculaire dans le spectre MS² peut être dû à une énergie de collision trop forte pour ces molécules entrainant la fragmentation complète de l'ion moléculaire. Ou bien l'ion recherché n'est pas présent dans le milieu. Toutefois, l'analyse de ces spectres MS² permet de mettre en évidence la présence de l'ion de m/z 134.097 qui correspond au réarrangement du cycle aromatique du DmPABr. De par la présence de ce fragment nous pouvons confirmer que la cystéine est bien marquée jusqu'à 4 fois avec le DmPABr. Toutefois, l'analyse des spectres MS² ne permet pas de déterminer le groupement chimique sur lequel le DmPABr est fixé.

L'analyse des spectres MS² des acides aminés et α-hydroxy acides a permis de mettre en évidence deux fragments spécifiques communs à la plupart des spectres MS² mais n'a pas permis de différencier le marquage d'une fonction acide carboxylique par rapport à une fonction amine, alcool ou thiol.

2.3.3. Bilan

L'analyse des MS² permet de mettre en évidence 2 fragments spécifiques au couplage avec le DmPABr (tableau 32).

Tableau 32 : principaux fragments retrouvés lors de la MS² du DmPABr couplé aux composés

		
Formule brute	$C_{10}H_{14}O_2N$	$C_9H_{12}N$

m/z	180.103	134.097

Un fragment avec un m/z de 180.10 représente le DmPABr oxydé. Ce fragment est retrouvé, par exemple, dans le spectre MS^2 de l'acide acétique et de l'acide hydroxyphenyllactique. Il pourrait donc être défini comme un fragment spécifique du marquage des fonctions acides carboxylique ou hydroxyle. En effet, il possède un hydroxyle au niveau du site de fixation de l'étiquette qui pourrait provenir d'une précédente liaison avec un atome d'oxygène. (Cette hypothèse sera infirmée par la suite). Le fragment de m/z 134.09 est retrouvé dans une grande majorité des spectres MS^2, il est issu d'un réarrangement du cycle aromatique du DmPABr. Il peut être considéré comme un fragment de suivi pour identifier un couplage du DmPABr. La découverte des fragments 180 et 134 est également décrite par l'étude de Willacey en 2019 [199]. Cette étude cible le marquage d'acides aminés et organiques. Parmi les 64 composés utilisés dans cette étude, 10 composés sont communs avec notre étude. L'analyse de leur spectre MS^2 a permis d'observer que 30 composés ont le fragment 180 ou 134 dans leurs MS^2.

2.4. Bilan du marquage en une étape

Le DmPABr est une étiquette chimique qui permet d'améliorer la séparation des composés en chromatographie liquide inverse et la détection des composés acides en mode d'ionisation positive. Le DmPABr permet de marquer la fonction acide en une étape des acides gras, triacides, acides aminés et α-hydroxy acides. La plupart des études retrouvées dans la littérature décrivent cette étiquette pour le marquage des acides organiques [223]. Par exemple, l'étude de Peng [198] décrit le marquage de la fonction acide carboxylique des acides organiques contenus dans l'urine. Grâce à l'utilisation de bases de données ils ont pu identifier des métabolites à partir des m/z détectés. A un même m/z peut correspondre plusieurs molécules. Ainsi, cette étude de Peng *et al.* a permis l'identification putative de l'acide acétique, acide subérique, acide citrique, cis aconitique, l'acide lactique, malique marqués une fois

par le DmPABr. Les formes marquées plusieurs fois ne sont pas recherchées dans cette étude. Ce constat est également fait dans l'étude de Guo et son équipe [173]. Ils ont créé une base de données de 113 standards d'acides organiques en ne se focalisant que sur les formes marquées une fois par le DmPA. Dans cette base de données, nous retrouvons les acides malique, citrique, hippurique, créatine et lactique qui sont dans notre étude marqué entre 1 à 3 fois. L'équipe de Willacey en 2019 [199] est la première à montrer que les acides organiques tels que l'acide malique peuvent être marqués deux fois. Il a également montré que les acides aminés sont marqués plus d'une fois par le DmPABr. Par exemple, l'alanine, l'arginine, la méthionine, la phénylalanine et la cystéine sont marquées 3 fois, et l'acide aspartique 4 fois ce qui est en accord avec nos résultats.

L'analyse des spectres MS^2 a permis de mettre en évidence 3 fragments spécifiques au couplage du DmPABr qui pourront plus tard aider lors d'analyses non ciblées dans les fluides biologiques. En effet, il sera possible d'identifier de nouveaux métabolites par le suivi de ces 3 fragments afin de faire des identifications putatives aux m/z associés à ses fragments.

Une récente étude datant d'août 2020 [224], réalisée en même temps que nos expériences, a mis en évidence le marquage des fonctions amines primaires, secondaires et thiols par le DmPABr. Cette étiquette originellement utilisée pour le marquage du sous-métabolome des acides carboxyliques se relève être également une étiquette pouvant marquer d'autres fonctions. Dans le cas des fonctions amine et alcool la question suivante est de savoir si cette étiquette est aussi efficace que des étiquettes de référence.

3. Comparaison du DmPABr avec des étiquettes chimiques connues : chlorure de Dansyl et PITC

D'après les résultats précédents, nous avons constaté que l'étiquette DmPABr se fixe à d'autres fonctions que la fonction acide carboxylique. Nous avons donc émis deux hypothèses. La première est de confirmer que le DmPABr se fixe bien aux

fonctions amine et alcool. La deuxième est de comparer le rendement de marquage du DmPABr à deux étiquettes chimiques étant des références quant au marquage des sous-métabolomes amine et phénol soit : le chlorure de dansyl et le PITC.

3.1. L'étiquette DmPABr marque-t-elle les fonctions amine et alcool ?

Afin de répondre à la première hypothèse, nous avons décidé de tester le protocole de marquage du DmPABr dans le cas des fonctions amine et alcool en utilisant des dérivés indoliques (figure 35). Ces composés contiennent des fonctions amines primaire, secondaire et alcool et sont également impliqués dans de nombreuses voies métaboliques. Dans cette catégorie nous avons choisi l'indole qui est un composé organique aromatique hétérocyclique. La structure de l'indole est présente dans les autres composés choisis : la tryptamine, la sérotonine et le tryptophol. Grâce à ce panel de composés, nous pouvons tester la réactivité de l'étiquette DmPABr sur les fonctions amine primaire, secondaire, phénol et alcool.

La figure 60 représente l'aire brute des dérivés indoliques mesurée après marquage avec le DmPABr à partir du calcul de leur m/z (tableau 33).

Tableau 33 : m/z des standards de dérivés indoliques marqués 1,2,3 ou 4 fois par l'étiquette DmPABr (TAG) en détection positive sous forme $[M+H]^+$

		1TAG	2TAG	3TAG	4TAG
Métabolites	Nombre de site	m/z	m/z	m/z	m/z
Indole	1	279,149			
Sérotonine	4	338,186	499,3	660,354	821,4384
Tryptamine	3	322,191	483,3	644,36	
Tryptophol	2	323,175	484,3		

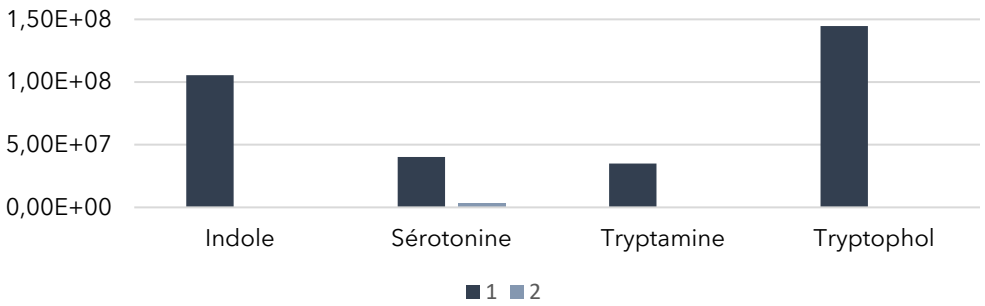


Figure 60 : Histogramme des résultats du marquage par l'étiquette DmPABr dans le cas des dérivés indoliques. L'axe des ordonnées représente l'aire brute sous la courbe de chaque molécule marquée. Le code couleur 1et 2 représente le nombre de fois où la molécule est marquée par le DmPABr.

Nous remarquons que les 4 composés sont marqués une fois par le DmPABr. Une faible quantité de sérotonine marquée deux fois est détectée. Le tryptophol contient deux fonctions, alcool et amine, pouvant réagir avec le DmPABr. Nous savons qu'une amine secondaire est moins nucléophile qu'un hydroxyle particulièrement si elle est cyclisée. Nous pouvons donc supposer que le tryptophol a été marqué sur sa fonction alcool. De plus, on observe que son aire brute marquée au DmPABr est plus important que celle des 3 autres composés. L'analyse des spectres MS^2 (tableau 34) ne permet pas de déterminer sur quelle fonctionne le DmPABr s'est fixé préférentiellement.

Tableau 34 : Tableau du fragment MS^2 majoritaire détecté après fragmentation en mode MRM des dérivés indoliques couplés au DmPABr (TAG) 1 à 2 fois

	1 TAG		2 TAG	
Molécules	RT (min)	mz	RT (min)	m/z
Indole	11.41	180.103		
Sérotonine	12.0	180.103	9.4	134.097
Tryptamine	8.86	134.097		
Tryptophol	11.31	148.075		

En effet, de même que pour les catégories d'acides gras, triacides, acides aminés et α-hydroxy acides, les fragments 180.103 et 134.097 sont majoritairement retrouvés. L'ion fragment 148.075 ($C_9H_{10}ON$) retrouvé dans le cas de l'acide citrique est également majoritaire dans le spectre MS^2 du tryptophol. Il peut provenir d'une impureté ou de la fragmentation du DmPABr ($C_{10}H_{12}NOBr$) (perte de CH_2 et Br).

L'hypothèse émise plus haut supposant que le fragment 180.103 est spécifique au marquage des fonctions acides est fausse. En effet, il est également retrouvé dans le spectre MS^2 de l'indole qui ne présente aucun atome d'oxygène dans sa structure. Donc il n'y a aucune possibilité que le DmPABr se soit fixé à une fonction hydroxyle ou carboxylique dans ce cas.

La découverte du DmPABr comme étiquette se fixant sur les fonctions amines est très récente (2020) [224] . La littérature ne décrit pas le marquage des dérivés indoliques par le DmPABr. Nous avons donc décidé de confronter nos observations à deux étiquettes chimiques largement utilisées pour le marquage des fonctions amines et phénol, le chlorure de Dansyl et le PITC.

3.2. L'étiquette DmPABr est-elle plus efficace que le chlorure de Dansyl et le PITC ?

Les étiquettes chimiques chlorure de Dansyle et PITC ont été largement décrites dans la littérature pour marquer respectivement les fonctions amine primaire/phénol et amine primaire. Parmi les 6 catégories chimiques de composés choisies dans cette étude, il y a 3 catégories communes aux étiquettes DmPABr, chlorure de Dansyle et PITC (acide aminé, dérivé indolique et α-hydroxy acide). Parmi les 8 acides aminés considérés dans cette étude (voir figure 35), l'acide hippurique ne peut être marqué ni par le chlorure de Dansyle ni par le PITC car il possède une fonction amine secondaire. Dans le cas des α-hydroxy acides, seul l'acide hydroxyphenyllactique peut être dérivé par le chlorure de Dansyle car il possède une fonction phénol. Le PITC ne peut se coupler à aucun des α-hydroxy

acides. Cette absence de fonction cible a pour conséquence que dans le cas des dérivés indoliques, l'indole et le tryptophol ne peuvent être dérivés par ses étiquettes. Ainsi pour la suite de notre comparaison, nous ne considérons que l'alanine, la créatine, l'arginine, la cystéine, la phénylalanine, la méthionine, l'acide aspartique, l'acide hydroxyphenyllactique, la sérotonine et la tryptamine.

3.2.1. Le pseudo-rendement de marquage

Afin de comparer les 3 étiquettes chimiques, nous avons décidé de calculer un pseudo-rendement de marquage. Ce n'est pas un rendement comme défini en chimie organique car il met en jeu le rapport des aires des molécules et non les quantités de matière. Il est défini par le rapport suivant :

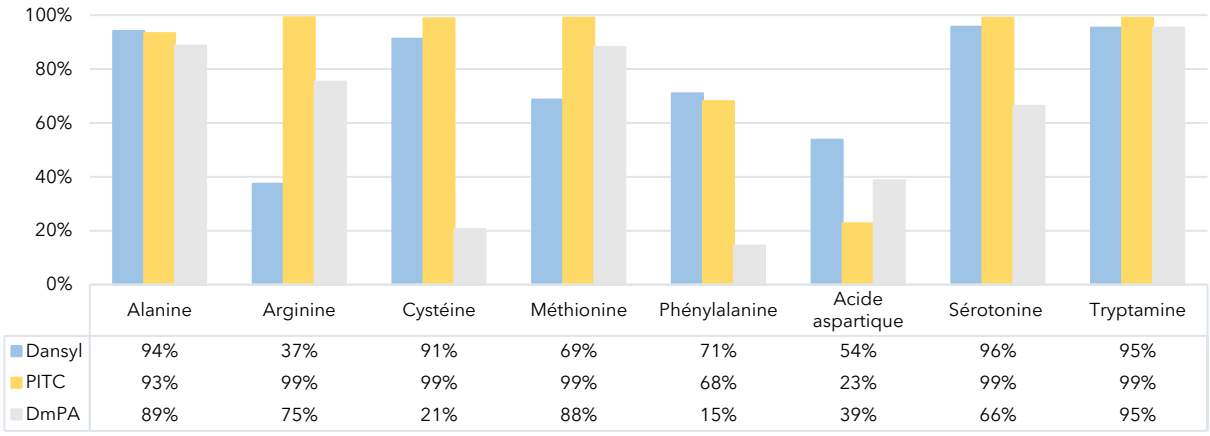
$$\frac{Aire\ [M + TAG]}{(\sum Aire[M + TAG] + Aire\ [Molécule\ restante]}$$

Aire[M+TAG] correspond à l'aire de la molécule marquée par l'étiquette chimique sous sa forme la plus abondante. Aire [Molécule restante] correspond à l'aire de la molécule non marquée par l'étiquette restante dans le milieu réactionnel non marquée. Et ∑Aire[M+TAG] correspond à la somme des autres aires de la molécule marquée 1 à 4 fois selon la molécule considérée. Par exemple l'alanine est marquée majoritairement 3 fois avec l'étiquette DmPABr. Ainsi son pseudo rendement est exprimé comme suit :

$$\frac{Aire\ [Alanine + 3DmPA]}{(Aire[Alanine + 1DmPA] + Aire\ [Alanine + 2DmPA] + Aire\ [Alanine\ restante]}$$

Ce rapport calculé pour chaque composé et chaque étiquette chimique a permis de créer la figure 61. D'après les résultats obtenus précédemment pour le DmPABr, les formes majoritaires ont été déterminées pour chaque composé. Ainsi pour l'alanine et la cystéine la forme marquée 3 fois a été considéré, la forme marquée 2 fois est considéré pour la méthionine et l'acide aspartique et celle marquée 1 fois pour l'arginine, la phénylalanine, la sérotonine et la tryptamine. Le chlorure de Dansyle et le PITC ne marquent qu'une fois les métabolites.

L'intégration des pics chromatographiques dans le cas du chlorure de Dansyle et PITC a été réalisée selon la valeur des m/z définis tableau S3.



	Alanine	Arginine	Cystéine	Méthionine	Phénylalanine	Acide aspartique	Sérotonine	Tryptamine
Dansyl	94%	37%	91%	69%	71%	54%	96%	95%
PITC	93%	99%	99%	99%	68%	23%	99%	99%
DmPA	89%	75%	21%	88%	15%	39%	66%	95%

Figure 61 : figure représentant le pseudo-rendement de marquage de chaque molécule en fonction du DmPABr, chlorure de Dansyle et PITC

Ce pseudo rendement n'a pas pu être calculé pour 2 composés. L'acide hydroxyphenyllactique qui n'est pas ionisable en mode positif. La créatine seule, non marquée par les étiquettes n'a pas été détectée au m/z attendu. Cette difficulté de détection peut être due à deux facteurs : soit la créatine se fragmente en source, soit les formes de résonnances de la créatine, contenant une imine et des amines, délocalisent la charge positive de la molécule complexifiant sa détection. Cette observation permet de confirmer la nécessité de dériver les molécules afin d'augmenter leur sensibilité de détection dans les fluides biologiques.

Globalement, le pseudo-rendement de marquage du DmPABr est toujours inférieur à celui du chlorure de Dansyle et du PITC. En calculant le pseudo-rendement moyen de chaque étiquette chimique, on retrouve 61%, 76% et 85%, respectivement, pour le DmPABr, chlorure de Dansyle et PITC. Ce taux inférieur du DmPABr par rapport aux deux autres étiquettes peut être dû à plusieurs facteurs. Premièrement, les fonctions ciblées du DmPABr ne sont pas forcément les mêmes que le chlorure de Dansyle ou le PITC. En effet, le DmPABr marque les fonctions acides mais également les amines et les alcools entrainant ainsi une différence de réactivité entre chaque

fonction. Deuxièmement, les conditions de marquage du DmPABr ne sont pas optimales, le protocole est donc à revoir en faisant varier les paramètres critiques tels que la température de chauffage, le temps d'agitation ou le milieu réactionnel. Troisièment, une autre hypothèse est liée à la structure du DmPABr, la partie dédiée à la détection du DmPABr par la MS est peut-être moins « ionisable » que le cas du chlorure de Dansyle et le PITC.

3.3. Bilan

L'étiquette DmPABr est utilisée pour marquer les fonctions acides. Nous avons montré (paragraphe 2) que cette étiquette permet aussi le marquage des fonctions amines, primaire et secondaire, alcool et thiol. Nous avons donc comparé le DmPABr à deux étiquettes de références pour le marquage des fonctions amine et phénol. Cependant comparée aux 2 étiquettes de référence soit le chlorure de Dansyle et le PITC, il semblerait que l'étiquette DmPABr ait une efficacité de marquage inférieure notamment dû à son pseudo-rendement de marquage pour la plupart de nos standards à l'exception de l'arginine et de la méthionine. Ainsi dans le cas du marquage des fonctions amine primaire/phénol et amine primaire, le chlorure de Dansyle et le PITC sont à privilégier par rapport au DmPABr.

Dans la littérature, l'étiquette DmPABr est décrite pour marquer le sous-métabolome des composés possédant une fonction acide carboxylique. Cette étiquette n'est pas comparée au chlorure de Dansyle ou au PITC mais est utilisée en parallèle. En effet, des études « multi-étiquettes » utilisant le DmPABr et le chlorure de Dansyle sont faites dans le but d'isoler et de quantifier deux sous-métabolomes. Le sous-métabolome des phénol et amine primaire est ciblé par le chlorure de Dansyle et celui des acides carboxyliques avec le DmPABr. Par exemple l'étude de Jia et al. en 2019 [223] a permis d'établir une base de données urinaire avec des informations MS et MS/MS pour plus de 800 composés soit 671 potentiels biomarqueurs (en enlevant la redondance) dans des pools d'urine (505 phénols et

166 acides carboxyliques). Par cette même approche, l'équipe de Luo *et al*. [225] a créé en 2016 une base de données sur les levures de 275 métabolites dont 120 acides aminés identifiés grâce au chlorure de Dansyle et 33 acides organiques identifiés grâce au DmPABr. Ces deux études utilisent le chlorure de Dansyle et le DmPABr en parallèle. Un mélange dans le milieu réactionnel de ces deux étiquettes chimiques entrainerait une compétitivité de marquage car de même que le chlorure de Dansyle, le DmPABr réagit avec les fonctions amines.

Les études actuelles avec l'étiquette PITC sont réalisées à l'aide du kit Biocrates (AbsoluteIDQ p180 Kit). Ce kit permet de dériver différentes catégories de molécules telles que les acides aminés ou les nucléotides [226]. Dans ce kit, la molécule de PITC est utilisée uniquement pour sa capacité à améliorer la détection et la séparation des composés. Ce sont les composés marqués isotopiquement qui permettent leur quantification. Dans la littérature il y a peu d'articles, actuels ou anciens, mentionnant l'utilisation de l'isotope du PITC lors d'une approche métabolomique. En 1986 Miller *et al*. [227] a montré qu'enrichir le site de fixation du PITC en ^{13}C permet de distinguer les groupements amines des protéines lors d'une analyse en ^{13}C-RMN. Ce marquage permet notamment de différencier les groupements amines présents sur la chaine de l'insuline. Cependant, ce type d'études n'a pas été poursuivi et pourrait être une idée de perspectives. En effet, d'après nos résultats le PITC est aussi efficace que le chlorure de Dansyle pour le marquage des fonctions amines secondaires. Ainsi, l'isotope du PITC pourrait lui permettre d'être utilisé comme le chlorure de Dansyle et le DmPABr pour une quantification absolue des métabolites.

Dans le cas du DmPABr, des études peuvent être envisagées dans le but de le rendre aussi efficace que le chlorure de Dansyle et le PITC. Une amélioration des conditions de marquage du DmPABr et/ou celle de sa structure chimique pourrait être envisageable.

VII. Conclusion

La dérivation chimique utilisant des étiquettes chimiques permet de marquer les fonctions des métabolites et d'améliorer la détectabilité de ces composés. En effet, elle permet de détecter les acides en détection positive en MS ce qui permet de réduire le nombre d'acquisitions LC-MS et d'avoir un maximum d'informations en une injection. La dérivation de la fonction acide en deux étapes est envisageable pour les acides gras ou triacides mais n'est pas adaptée pour les acides aminés et les α-hydroxy acides, notamment dû à la nécessité d'agents de couplage qui complexifient le milieu réactionnel provocant des réactions « parasites ». L'utilisation d'agent de couplage dans un milieu biologique est à proscrire. Ainsi le marquage de la fonction acide carboxylique en deux étapes est un protocole inadapté pour nos études en clinique ou en préclinique. Le marquage de la fonction acide en une étape par le DmPABr a donné de meilleur résultat et a permis de dériver les acides aminés, les α-hydroxy acides, les acide gras et triacides. Toutefois ce composé est non spécifique à la fonction acide et marque notamment les fonctions amine, alcool et thiol. Ces réactions secondaires entrainent une multitude de marquage, diversifiant ainsi l'étiquette sur plusieurs types de fonctions. De plus, l'étude comparative du DmPABr face à deux « gold standard » du marquage des fonctions amine et phénol a montré que cette dernière est moins efficace. Cependant, la versatilité du DmPABr ouvre d'autres perspectives dans le fait de développer une étiquette chimique qui pourrait marquer plusieurs sous-métabolomes en même temps. Celle-ci permettra de réduire le temps et le coût de la phase pré-analytique par rapport aux études actuelles multi-étiquettes. Les perspectives sont donc de trouver une étiquette chimique réalisant une réaction nucléophile avec les métabolites mais ayant un site de fixation plus réactif que le DmPABr et une structure mieux ionisable.

Conclusion générale

L'ensemble de ces travaux montrent l'étendue des possibilités offertes par la métabolomique multi-matrices et multiplateforme à l'heure actuelle et dans un futur proche. Le développement de méthodes analytiques standardisées décrivant le métabolome le plus large possible permet d'avoir une approche applicable par d'autres laboratoires. En effet, de nos jours il n'existe pas de consensus quant à la préparation des échantillons lors d'études métabolomiques limitant ainsi une comparaison des données inter-sites et réduisant la proposition de biomarqueurs aux études utilisant le même protocole de préparation d'échantillons. Ainsi le transfert vers une utilisation en routine clinique ou en santé animale reste limité par ce manque de standardisation. Lors de notre étude nous avons montré que l'étape pré-analytique de préparation d'échantillon était cruciale dans la description d'une matrice biologique tant sur le nombre de métabolites visibles que sur la reproductibilité de l'analyse. Cette standardisation des protocoles de préparation est nécessaire pour décrire un échantillon biologique, pour estimer la complémentarité entre matrices et plateformes et pour décrire le métabolome type d'un individu à un instant *t*. Ce profil métabolique ainsi standardisé peut-être utilisé lors d'étude clinique ou en élevage pour la comparaison à d'autres profils métaboliques ou pour un suivi ou cours du temps. Ainsi, une modification de ce profil peut permettre une détection précoce d'un dérèglement métabolique pouvant indiquer une maladie. La découverte de biomarqueurs notamment en clinique ou en élevage pourra être facilité par la comparaison de ces métabolomes. La comparaison de ces métabolomes est de nos jours réalisée grâce à une quantification relative, nécessitant l'acquisition des données en campagne. Il est donc difficile de comparer des études entre elles lorsqu'elles ne sont pas acquises en même temps et/ou avec un pré analytique différent. Ainsi le deuxième levier pour une utilisation de la métabolomique en routine clinique ou en santé animale est la quantification absolue des molécules identifiées soit pour un suivi longitudinal des individus et ainsi permettre de mettre en évidence très précocement une dérive

métabolique soit pour permettre la quantification de molécules identifiées comme biomarqueurs et pouvoir les doser au fil de l'eau. Ainsi, le développement de méthode de quantification absolue facilitera l'utilisation de ces méthodes en routine clinique et en élevage, lors d'acquisition à des temps différents et/ou par d'autres laboratoires. Les approches métabolomiques sont des études nécessitant des temps d'analyse et de traitement de données assez long comparées à d'autres approches. La méthode de quantification absolue doit être relativement rapide et permettre une quantification la plus large possible du métabolome. Ainsi la dérivation chimique semble être la plus appropriée en terme de temps et de coût pour une quantification absolue des métabolites à grande échelle. Comparé aux autres méthodes développées à ce jour, elle a le meilleur compromis en terme de couverture du métabolome et du temps à réaliser cette étape. La dérivation chimique des fonctions en plusieurs étapes n'est pas adaptée à nos échantillons car les échantillons biologiques sont des milieux complexes. L'ajout de plusieurs réactifs dans ces milieux entraine de nombreuses réactions parasites limitant ainsi l'efficacité de marquage des étiquettes chimiques et la détection des métabolites marqués. Dans la perspective d'une utilisation en routine clinique ou en santé animale c'est-à-dire sur des échantillons complexes, il serait préférable d'utiliser une méthode mono étape. Lors de notre étude, le DmPaBr est un bon candidat qui marque les fonctions acides mais aussi les fonctions amine et thiol. Cependant une optimisation de sa structure est nécessaire pour que son efficacité d'ionisation avec de bonnes propriétés séparatives soit aussi optimale que certaines étiquettes moins versatiles utilisées actuellement.

ANNEXE

Métabolites	Nombre de moles (mol)	Masse molaire (g/mol)	Volume prélevé de solution mère (µL)	Volume de solvant ajouté (µL)
Acide acétique	4,40E-10	60,052	2,6	47
Acide aspartique	4,40E-10	133,1027	5,8	44
Acide cis-aconitique	4,40E-10	174,108	7,6	42
Acide citrique	4,40E-10	192,1235	8,4	42
Acide hippurique	4,40E-10	179,17	7,8	42
Acide hydroxyphenyllactique	4,40E-10	182,173	8,0	42
Acide lactique	4,40E-10	90,077	3,9	46
Acide laurique	4,40E-10	200,317	8,8	41
Acide malique	4,40E-10	134,087	5,9	44
Acide subérique	4,40E-10	118,09	5,2	45
Alanine	4,40E-10	89,093	3,9	46
Arginine	4,40E-10	174,201	7,6	42
Créatine	4,40E-10	131,1332	5,7	44
Cystéine	4,40E-10	121,16	5,3	45
Indole	4,40E-10	117,15	5,2	45
Méthionine	4,40E-10	149,211	6,5	43
Phénylalanine	4,40E-10	165,189	7,3	43
Sérotonine	4,40E-10	176,215	7,8	42
Tryptamine	4,40E-10	160,216	7,1	43
Tryptophol	4,40E-10	161,2	7,1	43

Tableau S1 : Nombre de mol et volume ajouté pour chaque standard avec une solution mère à 10µg/mL pour la dérivation avec les étiquettes DMED, DAPPZ, DMPT,DMPP et DmPABr

Nom	Nombre de sites de marquage			
	DMED DAPPZ DMPT DMPP	DNSCl	PITC	DmPABr
Acide acétique	1	0	0	1
Acide aspartique	2	1	1	4
Acide cis-aconitique	3	0	0	3
Acide citrique	3	0	0	4
Acide hippurique	1	0	0	2
Acide hydroxy-phenyllactique	1	1	0	3
Acide lactique	1	0	0	2
Acide laurique	1	0	0	1
Acide malique	2	0	0	3
Acide subérique	2	0	0	2
Alanine	1	1	1	3
Arginine	1	2	2	3
Créatine	1	2	2	3
Cystéine	1	1	1	4
Indole	0	1	1	1
Méthionine	1	1	1	3
Phénylalanine	1	1	1	3
Sérotonine	0	2	1	4
Tryptamine	0	1	2	3
Tryptophol	0	0	0	2

Tableau S2 : Nombre de fois où les étiquettes chimiques (DMED, DAPPZ, DMPT, DMPP, DNSCl, PITC et DmPABr) ont la possibilité de marquer les fonctions chimiques des standards sélectionnés

	Formule brute	[M+H]	[M+DANSYL+H]	[M+PITC+H]
Alanine	$C_3H_7NO_2$	90,054	323,106	225,069
Arginine	$C_6H_{14}N_4O_2$	175,118	408,170	310,133
Cystéine	$C_3H_7NO_2S$	122,027	355,078	257,041
Méthionine	$C_5H_{11}NO_2S$	150,058	383,109	285,072
Phénylalanine	$C_9H_{11}NO_2S$	166,086	399,137	301,100
Acide aspartique	$C_4H_7NO_4$	134,044	367,095	269,059
Indole	C_8H_7N	118,065	351,116	253,079
Tryptamine	$C_{10}H_{12}N_2$	161,107	394,158	296,121
Tryptophol	$C_{10}H_{11}NO$	162,091	395,142	297,105
Sérotonine	$C_{10}H_{12}N_2O$	177,102	410,153	312,116
DANSYL	$C_{12}H_{12}ClNO_2S$	270,035		
PITC	C_7H_5NS	136,021		

Tableau S3 : m/z des standards marqués par les étiquettes de chlorure de Dansyle et PITC en détection positive sous forme $[M+H]^+$

REFERENCES

1. Horgan, R. and L.C. Kenny, *'Omic'technologies: genomics, transcriptomics, proteomics and metabolomics.* The Obstetrician & Gynaecologist, 2011. **13**: p. 189–195.

2. Williams, R.J., *Normal young men.* Perspect Biol Med, 1957. **1**(1): p. 97-104.

3. Pauling, L., et al., *Quantitative analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography.* Medical Sciences, 1971. **68**: p. 2374-2376.

4. Horning, E.C. and M.G. Horning, *Metabolic Profiles: Gas-Phase Methods for Analysis of Metabolites.* Clinical chemistry, 1971. **17**: p. 802-809.

5. Nicholson, J.K., et al., *Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects.* Biochem, 1984. **217**: p. 365-375.

6. Oliver, S.G., et al., *Systematic functional analysis of the yeast genome.* Trends Biotechnol, 1998. **16**: p. 373–378.

7. Fiehn, O., et al., *Metabolite profiling for plant functional genomics.* Nature biotechnology, 2000. **18**: p. 1157–1161.

8. Holmes, E., et al., *Detection of urinary drug metabolite (Xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (nmr) spectroscopy.* Anal Chem, 2007. **79**: p. 2629-2640.

9. Wishart, D.S., et al., *HMDB 4.0: the human metabolome database for 2018.* Nucleic Acids Res, 2018. **46**(D1): p. D608-D617.

10. Damiani, C., et al., *Systems metabolomics: from metabolomic snapshots to design principles.* Curr Opin Biotechnol, 2020. **63**: p. 190-199.

11. Wishart, D.S., *Emerging applications of metabolomics in drug discovery and precision medicine.* Nat Rev Drug Discov, 2016. **15**(7): p. 473-84.

12. Newgard, C.B., *Metabolomics and Metabolic Diseases: Where Do We Stand?* Cell Metab, 2017. **25**(1): p. 43-56.

13. Amato, P., et al., *Metatranscriptomic exploration of microbial functioning in clouds.* Sci Rep, 2019. **9**(1): p. 4383.

14. Segers, K., et al., *Analytical techniques for metabolomic studies: a review.* Bioanalysis, 2019. **11 (24)**: p. 2297-2318.

15. Barton, R.H., et al., *The influence of EDTA and citrate anticoagulant addition to human plasma on information recovery from NMR-based metabolic profiling studies.* Mol Biosyst, 2010. **6**(1): p. 215-24.

16. Gratton, J., et al., *Optimized Sample Handling Strategy for Metabolic Profiling of Human Feces.* Anal Chem, 2016. **88**(9): p. 4661-8.

17. Ulaszewska, M.M., et al., *Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies.* Mol Nutr Food Res, 2019. **63**(1): p. e1800384.

18. Vaught, J.B., *Blood collection, shipment, processing, and storage.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(9): p. 1582-4.

19. Pinto, J., et al., *Human plasma stability during handling and storage: impact on NMR metabolomics.* Analyst, 2014. **139**(5): p. 1168-77.

20. Jobard, E., et al., *A Systematic Evaluation of Blood Serum and Plasma Pre-Analytics for Metabolomics Cohort Studies.* Int J Mol Sci, 2016. **17**(12).

21. Roux, A., et al., *Impact of collection conditions on the metabolite content of human urine samples as analyzed by liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance spectroscopy.* Metabolomics, 2015. **11**: p. 1095-1105.

22. Gika, H.G., et al., *Within-Day Reproducibility of an HPLC-MS-Based Method for Metabonomic Analysis: Application to Human Urine.* Journal of Proteome Research 2007. **6**: p. 3291-3303.

23. Semrena, T.Ž., et al., *Gas chromatographic-mass spectrometric analysis of urinary volatile organic metabolites: Optimization of the HS-SPME procedure and sample storage conditions.* Talanta, 2018. **176**: p. 537-543.

24. Phua, L.C., et al., *Global gas chromatography/time-of-flight mass spectrometry (GC/TOFMS)-based metabonomic profiling of lyophilized human feces.* J Chromatogr B Analyt Technol Biomed Life Sci, 2013. **937**: p. 103-13.

25. Dieme, B., et al., *Workflow methodology for rat brain metabolome exploration using NMR, LC-MS and GC-MS analytical platforms.* J Pharm Biomed Anal, 2017. **142**: p. 270-278.

26. Karu, N., et al., *A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database.* Anal Chim Acta, 2018. **1030**: p. 1-24.

27. Saric, J., et al., *Metabolic profiling of an Echinostoma caproni infection in the mouse for biomarker discovery.* PLoS Negl Trop Dis, 2008. **2**(7): p. e254.

28. Gardner, A., et al., *Developing and Standardizing a Protocol for Quantitative Proton Nuclear Magnetic Resonance ((1)H NMR) Spectroscopy of Saliva.* J Proteome Res, 2018. **17**(4): p. 1521-1531.

29. Wallner-Liebmann, S., et al., *Individual Human Metabolic Phenotype Analyzed by (1)H NMR of Saliva Samples.* J Proteome Res, 2016. **15**(6): p. 1787-93.

30. De Paepe, E., et al., *A validated multi-matrix platform for metabolomic fingerprinting of human urine, feces and plasma using ultra-high performance liquid-chromatography coupled to hybrid orbitrap high-resolution mass spectrometry.* Anal Chim Acta, 2018. **1033**: p. 108-118.

31. Dill, K.A., *Denatured states of proteins.* Biochem, 1991. **60**: p. 795-825.

32. Nagana Gowda, G.A. and D. Raftery, *Quantitating metabolites in protein precipitated serum using NMR spectroscopy.* Anal Chem, 2014. **86**(11): p. 5433-40.

33. Schulz, A., et al., *Targeted metabolomics of pellicle and saliva in children with different caries activity.* Sci Rep, 2020. **10**(1): p. 697.

34. Folch, J., M. Lees, and G.H.S. Stanley, *A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues.* Journal of Biological Chemistry, 1957. **226**(1): p. 497-509.

35. Bligh, E.G. and W.J. Dyer, *A rapid method of total lipid extraction and purification.* Can. J. Biochem. Physiol., 1959. **37**.

36. Matyash, V., et al., *Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics.* J Lipid Res, 2008. **49**(5): p. 1137-46.

37. Iverson, S.J., S.L.C. Lang, and M.H. Cooper, *Comparison of the Bligh and Dyer and Folch Methods for Total Lipid Determination in a Broad Range of Marine Tissue.* Lipids, 2001. **36**.

38. Ulmer, C.Z., et al., *Optimization of Folch, Bligh-Dyer, and Matyash sample-to-extraction solvent ratios for human plasma-based lipidomics studies.* Anal Chim Acta, 2018. **1037**: p. 351-357.

39. McHugh, C.E., et al., *Rapid, Reproducible, Quantifiable NMR Metabolomics: Methanol and Methanol: Chloroform Precipitation for Removal of Macromolecules in Serum and Whole Blood.* Metabolites, 2018. **8**(4).

40. Lamichhane, S., et al., *Strategy for Nuclear-Magnetic-Resonance-Based Metabolomics of Human Feces.* Anal Chem, 2015. **87**(12): p. 5930-7.

41. Chapuis, F., V. Pichon, and M.C. Hennion, *Méthode de préconcentration par extraction en phase solide : principe et application aux industries environnementales et pétrolières.* Oil & Gas Science and Technology, 2006. **60**(6): p. 899-912.

42. Rico, E., et al., *Evaluation of human plasma sample preparation protocols for untargeted metabolic profiles analyzed by UHPLC-ESI-TOF-MS.* Anal Bioanal Chem, 2014. **406**(29): p. 7641-52.

43. Leng, J., et al., *A highly sensitive isotope-coded derivatization method and its application for the mass spectrometric analysis of analytes containing the carboxyl group.* Analytica Chimica Acta, 2013. **758**: p. 114-121.

44. Ogawa, S., et al., *LC/ESI-MS/MS method for determination of salivary eicosapentaenoic acid concentration to arachidonic acid concentration ratio.* Biomed. Chromatogr., 2015: p. 6.

45. Garner, C.E.S., et al., *Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease.* FASEB J, 2007. **21**(8): p. 1675-88.

46. Agin, A., et al., *Metabolomics – an overview. From basic principles to potential biomarkers (part 1).* Médecine Nucléaire, 2016. **40**(1): p. 4-10.

47. Lane, A.N. and T.W. Fan, *NMR-based Stable Isotope Resolved Metabolomics in systems biochemistry.* Arch Biochem Biophys, 2017. **628**: p. 123-131.

48. Gonzalez-Dominguez, R., A. Sayago, and A. Fernandez-Recamales, *Metabolomics in Alzheimer's disease: The need of complementary analytical platforms for the identification of biomarkers to unravel the underlying pathology.* J Chromatogr B Analyt Technol Biomed Life Sci, 2017. **1071**: p. 75-92.

49. Murgia, F., et al., *Metabolomics As a Tool for the Characterization of Drug-Resistant Epilepsy.* Front Neurol, 2017. **8**: p. 459.

50. Puig-Castellvi, F., I. Alfonso, and R. Tauler, *Untargeted assignment and automatic integration of (1)H NMR metabolomic datasets using a multivariate curve resolution approach.* Anal Chim Acta, 2017. **964**: p. 55-66.

51. Le Guennec, A., F. Tayyari, and A.S. Edison, *Alternatives to Nuclear Overhauser Enhancement Spectroscopy Presat and Carr-Purcell-Meiboom-Gill Presat for NMR-Based Metabolomics.* Anal Chem, 2017. **89**(17): p. 8582-8588.

52. Meiboom, S. and D. Gill, *Modified Spin-Echo Method for Measuring Nuclear Relaxation Times.* Review of Scientific Instruments, 1958. **29**(8): p. 688-691.

53. Xi, Y., et al., *Improved identification of metabolites in complex mixtures using HSQC NMR spectroscopy.* Anal Chim Acta, 2008. **614**(2): p. 127-33.

54. Van, Q.N., et al., *Comparison of 1D and 2D NMR Spectroscopy for Metabolic profiling.* Journal of proteome research, 2008. **7**: p. 630-639.

55. Issaq, H.J., et al., *Analytical and statistical approaches to metabolomics research.* J Sep Sci, 2009. **32**(13): p. 2183-99.

56. Dumas, M.E., et al., *Metabonomic assessment of physiological disruptions using 1H-13C HMBC-NMR spectroscopy combined with pattern recognition procedures performed on filtered variables.* Anal Chim Acta, 2002. **74**: p. 2261-2273.

57. Dumas, M.-E., et al., *Homeostatic signature of anabolic steroids in cattle using 1H-13C HMBC NMR Metabonomics.* Journal of proteome research, 2005. **4**: p. 1493-1502.

58. Kim, H.K., Y.H. Choi, and R. Verpoorte, *NMR-based plant metabolomics: where do we stand, where do we go?* Trends Biotechnol, 2011. **29**(6): p. 267-75.

59. Martineau, E., et al., *Absolute quantification of metabolites in breast cancer cell extracts by quantitative 2D (1) H INADEQUATE NMR.* NMR Biomed, 2012. **25**(8): p. 985-92.

60. Le Guennec, A., et al., *Resolution-enhanced 2D NMR of complex mixtures by non-uniform sampling.* Magn Reson Chem, 2015. **53**(11): p. 913-20.

61. Smolinska, A., et al., *NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review.* Anal Chim Acta, 2012. **750**: p. 82-97.

62. Ravanbakhsh, S., et al., *Accurate, fully-automated NMR spectral profiling for metabolomics.* PLoS One, 2015. **10**(5): p. e0124219.

63. Anderson, P.E., et al., *Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data.* Metabolomics, 2010. **7**(2): p. 179-190.

64. Savorani, F., G. Tomasi, and S.B. Engelsen, *icoshift: A versatile tool for the rapid alignment of 1D NMR spectra.* J Magn Reson, 2010. **202**(2): p. 190-202.

65. Ivan, K.S.Y., et al., *Metabonomic and Microbiological Analysis of the Dynamic Effect of Vancomycin-Induced Gut Microbiota Modification in the Mouse.* Journal of Proteome Research, 2008. **7**: p. 3718-3728.

66. Kohl, S.M., et al., *State-of-the art data normalization methods improve NMR-based metabolomic analysis.* Metabolomics, 2012. **8**(Suppl 1): p. 146-160.

67. Frank Dieterle, A.R., G. Schlotterbeck, and H. Senn, *Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics.* Anal. Chem, 2006. **78**: p. 4281-4290.

68. Euceda, L.R., G.F. Giskeodegard, and T.F. Bathen, *Preprocessing of NMR metabolomics data.* scandinavian Journal of Clinical & Laboratory Investigation, 2015. **15**: p. 193-203.

69. Donarski, J.A., D.P.T. Roberts, and A.J. Charlton, *Quantitative NMR spectroscopy for the rapid measurement of methylglyoxal in manuka honey.* Analytical Methods, 2010. **2**(10).

70. Garrido, R., et al., *Quantitative Proton Nuclear Magnetic Resonance evaluation and total assignment of the capsular polysaccharide Neisseria meningitidis serogroup X.* Journal of Pharmaceutical and Biomedical Analysis, 2012. **70**: p. 295-300.

71. Hasada, K., et al., *Quantitative determination of atractylon in Atractylodis Rhizoma and Atractylodis Lanceae Rhizoma by 1H-NMR spectroscopy.* J Nat Med, 2010. **64**(2): p. 161-6.

72. Aalim M. Weljie, J.N., Frank R. Jirik, and Hans J. Vogel, *Evaluating Low-Intensity Unknown Signals in Quantitative Proton NMR Mixture Analysis.* Anal. Chem., 2008. **80**: p. 8956–8965.

73. Sharma, R., et al., *A quantitative NMR protocol for the simultaneous analysis of atropine and obidoxime in parenteral injection devices.* J Pharm Biomed Anal, 2009. **49**(4): p. 1092-6.

74. Franconia, F., et al., *Quantitative MR renography using a calibrated internal signal (ERETIC).* Magnetic Resonance Imaging 2002. **20**: p. 587-592.

75. Mahon, H.P., *Calibration System for a Pulse Nuclear Induction Spectrometer.* Review of Scientific Instruments, 1969. **40**(12): p. 1644-1645.

76. Lee, D., et al., *Quantitative 19F imaging using inductively coupled reference signal injection.* Magn Reson Med, 2010. **63**(3): p. 570-3.

77. Zoelch, N., et al., *Accurate determination of brain metabolite concentrations using ERETIC as external reference.* NMR Biomed, 2017. **30**(8).

78. Jacob, D., et al., *NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics.* Metabolomics, 2017. **13**(4): p. 36.

79. Tardivel, P.J.C., et al., *ASICS: an automatic method for identification and quantification of metabolites in complex 1D 1H NMR spectra.* Metabolomics, 2017. **13**(10).

80. Simmler, C., et al., *Universal quantitative NMR analysis of complex natural samples.* Curr Opin Biotechnol, 2014. **25**: p. 51-9.

81. Letertre, M.P.M., G. Dervilly, and P. Giraudeau, *Combined Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry Approaches for Metabolomics.* Anal Chem, 2021. **93**(1): p. 500-518.

82. Cech, N.B. and C.G. Enke, *Pratical implications of some recent studies in electrospray ionization fundamentals.* Mass Spectrom Rev, 2001. **20**: p. 362-387.

83. So, P.K., B. Hu, and Z.P. Yao, *Mass spectrometry: towards in vivo analysis of biological systems.* Mol Biosyst, 2013. **9**(5): p. 915-29.

84. Eliuk, S. and A. Makarov, *Evolution of Orbitrap Mass Spectrometry Instrumentation.* Annu Rev Anal Chem (Palo Alto Calif), 2015. **8**: p. 61-80.

85. Makarov, A., *Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis.* Anal. Chem., 2000. **72**: p. 1156-1162.

86. Hardman, M. and A.A. Makarov, *Interfacing the Orbitrap Mass Analyzer to an Electrospray Ion Source.* Anal. Chem., 2003. **75**: p. 1699-1705.

87. Hu, Q., et al., *The Orbitrap: a new mass spectrometer.* J Mass Spectrom, 2005. **40**(4): p. 430-43.

88. Makarov, A., et al., *Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer.* Anal. Chem., 2006. **78**: p. 2113-2120.

89. Habchi, B., et al., *Potential of dynamically harmonized Fourier transform ion cyclotron resonance cell for high-throughput metabolomics fingerprinting: control of data quality.* Anal Bioanal Chem, 2018. **410**: p. 483-490.

90. Himmelsbach, M., *10 years of MS instrumental developments--impact on LC-MS/MS in clinical chemistry.* J Chromatogr B Analyt Technol Biomed Life Sci, 2012. **883-884**: p. 3-17.

91. Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, *An introduction to quadrupole-time-of-flight mass spectrometry.* J Mass Spectrom, 2001. **36**(8): p. 849-65.

92. Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer.* Mol Cell Proteomics, 2011. **10**(9): p. M111 011015.

93. La Barbera, G., et al., *Liquid chromatography-high resolution mass spectrometry for the analysis of phytochemicals in vegetal-derived food and beverages.* Food Res Int, 2017. **100**(Pt 1): p. 28-52.

94. Benton, H.P., et al., *Autonomous metabolomics for rapid metabolite identification in global profiling.* Anal Chem, 2015. **87**(2): p. 884-91.

95. Mullard, G., et al., *A new strategy for MS/MS data acquisition applying multiple data dependent experiments on Orbitrap mass spectrometers in non-targeted metabolomic applications.* Metabolomics, 2014. **11**(5): p. 1068-1080.

96. Hopfgartner, G., D. Tonoli, and E. Varesio, *High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices.* Anal Bioanal Chem, 2012. **402**(8): p. 2587-96.

97. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis.* Mol Cell Proteomics, 2012. **11**(6): p. O111 016717.

98. Zhang, Y., et al., *The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition.* J Proteome Res, 2015. **14**(10): p. 4359-71.

99. Colin, A.S., et al., *XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification.* Anal. Chem, 2006. **78**: p. 779-787.

100. Smith, C.A., et al., *XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification.* Anal. Chem. 2006, 2006. **78**: p. 779-787.

101. Giacomoni, F., et al., *Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics* Bioinformatics Advance, 2014.

102. Tsugawa, H., et al., *MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis.* Nat Methods, 2015. **12**(6): p. 523-6.

103. Rathahao-Paris, E., et al., *High resolution mass spectrometry for structural identification of metabolites in metabolomics.* Metabolomics, 2015. **12**(1).

104. Want, E.J., et al., *Global metabolic profiling procedures for urine using UPLC-MS.* Nat Protoc, 2010. **5**(6): p. 1005-18.

105. Gika, H.G., et al., *Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples.* J Chromatogr B Analyt Technol Biomed Life Sci, 2008. **871**(2): p. 299-305.

106. Helen, G.G., et al., *Within-Day Reproducibility of an HPLC-MS-Based Method for Metabonomic Analysis: Application to Human Urine.* Journal of Proteome Research 2007. **6**: p. 3291-3303.

107. Sangster, T., et al., *A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis.* Analyst, 2006. **131**(10): p. 1075-8.

108. Hughes, G., et al., *MSPrep--summarization, normalization and diagnostics for processing of mass spectrometry-based metabolomic data.* Bioinformatics, 2014. **30**(1): p. 133-4.

109. Xiaotao, S., et al., *Normalization and integration of large-scale metabolomics data using support vector regression.* Metabolomics, 2016. **12**(5).

110. Wang, W., et al., *Quantification of Proteins and Metabolites by Mass Spectrometry without Isotopic Labeling or Spiked Standards.* Anal. Chem, 2003. **75**: p. 4818-4826.

111. Dunn, W.B., et al., *Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry.* Nat Protoc, 2011. **6**(7): p. 1060-83.

112. Redestig, H., et al., *Compensation for Systematic Cross-Contribution Improves Normalization of Mass Spectrometry Based Metabolomics Data.* Anal. Chem, 2009. **81**: p. 7974–7980.

113. Scalbert, A., et al., *The food metabolome: a window over dietary exposure.* Am J Clin Nutr, 2014. **99**(6): p. 1286-308.

114. Lenz, E.M., et al., *A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects.* Journal of Pharmaceutical and Biomedical Analysis, 2003. **33**(5): p. 1103-1115.

115. Zhao, L., et al., *High Throughput and Quantitative Measurement of Microbial Metabolome by Gas Chromatography/Mass Spectrometry Using Automated Alkyl Chloroformate Derivatization.* Anal Chem, 2017. **89**(10): p. 5565-5577.

116. Alan L. Harvey, R. Edrada-Ebe, and R.J. Quinn, *The re-emergence of natural products for drug discovery in the genomics era.* Nature Reviews Drug Discovery, 2015.

117. Diémé, B., et al., *Metabolomics Study of Urine in Autism Spectrum Disorders Using a Multiplatform Analytical Methodology.* J. Proteome Res. , 2015. **14**: p. 5273-5282.

118. Gebretsadik, T., et al., *LC-NMR for Natural Products Analysis: A Journey from an Academic Curiosity to a Robust Analytical Tool.* Sci 2019. **1,31**.

119. Serkova, N.J., et al., *Early Detection of Graft Failure Using the Blood Metabolic Profile of a Liver Recipient.* Transplantation, 2007. **83**(4): p. 517-521.

120. David, S.W., et al., *Magnetic Resonance Diagnostics : A new technology for high-throughput clinical diagnostics.* Clinical Chemistry 47, 2001. **10**: p. 1918.

121. Djukovic, D., D. Raftery, and N. Gowda, *Chapitre 16-Mass spectrometry and NMR spectroscopy based quantitative metabolomics*, in *Proteomic and Metabolomic Approaches to Biomarker Discovery*. 2020. p. 289-311.

122. Djukovic, D., G.A. Nagana Gowda, and D. Raftery, *Mass Spectrometry and NMR Spectroscopy–Based Quantitative Metabolomics*, in *Proteomic and Metabolomic Approaches to Biomarker Discovery*. 2013. p. 279-297.

123. Marchand, J., et al., *Multidimensional NMR approaches towards highly resolved, sensitive and high-throughput quantitative metabolomics.* Curr Opin Biotechnol, 2017. **43**: p. 49-55.

124. Zhao, S. and L. Li, *Chemical derivatization in LC-MS-based metabolomics study.* TrAC Trends in Analytical Chemistry, 2020. **131**: p. 115988.

125. Nagana Gowda, G.A., et al., *NMR-Guided Mass Spectrometry for Absolute Quantitation of Human Blood Metabolites.* Anal Chem, 2018. **90**(3): p. 2001-2009.

126. Xi, X., et al., *Ultra-performance liquid chromatography-quadrupole-time of flight mass spectrometry MS(E)-based untargeted milk metabolomics in dairy cows with subclinical or clinical mastitis.* J Dairy Sci, 2017. **100**(6): p. 4884-4896.

127. Dupuy, C., et al., *SHR/NCrl rats as a model of ADHD can be discriminated from controls based on their brain, blood, or urine metabolomes.* Transl Psychiatry, 2021. **11**(1): p. 235.

128. Zhou, J. and Y. Yin, *Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry.* Analyst, 2016. **141**(23): p. 6362-6373.

129. Nilsson, L.B. and G. Eklund, *Direct quantification in bioanalytical LC-MS/MS using internal calibration via analyte/stable isotope ratio.* J Pharm Biomed Anal, 2007. **43**(3): p. 1094-9.

130. Meija, J. and Z. Mester, *Paradigms in isotope dilution mass spectrometry for elemental speciation analysis.* Analytica chimica acta, 2007. **607**: p. 115-125.

131. Rousseau, K., *Développement de nouveaux outils pour l'analyse métabolomique par spectrométrie de masse haute résolution : de l'acquisition de spectres MS/MS pour l'identification large-spectre de métabolites au marquage isotopique pour la quantification*. septembre 2020, Université Paris-Saclay: Paris.

132. Zheng, S.J., et al., *Hydrogen-Deuterium Scrambling Based on Chemical Isotope Labeling Coupled with LC-MS: Application to Amine Metabolite Identification in Untargeted Metabolomics.* Anal Chem, 2020. **92**(2): p. 2043-2051.

133. Taglang, C., et al., *Enantiospecific C-H Activation Using Ruthenium Nanocatalysts.* Angew Chem Int Ed Engl, 2015. **54**(36): p. 10474-7.

134. Fujiwara, Y., et al., *Method for regio-, chemo- and stereoselective deuterium labeling of sugars based on ruthenium-catalyzed C-H bond activation.* Chem Commun (Camb), 2010. **46**(27): p. 4977-9.

135. Heux, S., et al., *Recent advances in high-throughput (13)C-fluxomics.* Curr Opin Biotechnol, 2017. **43**: p. 104-109.

136. Fernandez, F.G.A., et al., *Cost-effective production of 13C, 15N stable isotope-labelled biomass from phototrophic microalgae for various biotechnological applications.* Biomol Eng, 2005. **22**(5-6): p. 193-200.

137. Bueschl, C., et al., *A novel stable isotope labelling assisted workflow for improved untargeted LC-HRMS based metabolomics research.* Metabolomics, 2014. **10**(4): p. 754-769.

138. Patrick Giavalisco, J.H., Jan Lisec, Alvaro Cuadros Inostroza, Gareth Catchpole, and Lothar Willmitzer, *High-Resolution Direct Infusion-Based Mass Spectrometry in Combination with Whole 13C Metabolome Isotope Labeling Allows Unambiguous Assignment of Chemical Sum Formulas.* Anal. Chem, 2008. **80**: p. 9417–9425.

139. Mayers, M.D., et al., *Quantitative Metaproteomics and Activity-Based Probe Enrichment Reveals Significant Alterations in Protein Expression from a Mouse Model of Inflammatory Bowel Disease.* J Proteome Res, 2017. **16**(2): p. 1014-1026.

140. Heuillet, M., et al., *Methodology for the Validation of Isotopic Analyses by Mass Spectrometry in Stable-Isotope Labeling Experiments.* Anal Chem, 2018. **90**(3): p. 1852-1860.

141. Blasco, H., et al., *The specific metabolome profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytosine metabolism.* Sci Rep, 2020. **10**(1): p. 16824.

142. Higashi, T. and S. Ogawa, *Chemical derivatization for enhancing sensitivity during LC/ESI-MS/MS quantification of steroids in biological samples: a review.* J Steroid Biochem Mol Biol, 2016. **162**: p. 57-69.

143. Holmes, E., I.D. Wilson, and J.K. Nicholson, *Metabolic phenotyping in health and disease.* Cell, 2008. **134**(5): p. 714-7.

144. Collino, S., F.P. Martin, and S. Rezzi, *Clinical metabolomics paves the way towards future healthcare strategies.* Br J Clin Pharmacol, 2013. **75**(3): p. 619-29.

145. Kell, D.B. and R. Goodacre, *Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery.* Drug Discov Today, 2014. **19**(2): p. 171-82.

146. Li, S., et al., *The Use of "Omics" in Lactation Research in Dairy Cows.* Int J Mol Sci, 2017. **18**(5).

147. Sun, H.Z., et al., *Metabolomics of four biofluids from dairy cows: potential biomarkers for milk production and quality.* J Proteome Res, 2015. **14**(2): p. 1287-98.

148. Pallares-Mendez, R., et al., *Metabolomics in diabetes, a review.* Ann Med, 2016. **48**(1-2): p. 89-102.

149. Jansson, J., et al., *Metabolomics reveals metabolic biomarkers of Crohn's disease.* PLoS One, 2009. **4**(7): p. e6386.

150. Orozco, J.S., et al., *Metabolomics analysis of children with autism, idiopathic-developmental delays, and Down syndrome.* Transl Psychiatry, 2019. **9**(1): p. 243.

151. Wu, Y. and L. Li, *Sample normalization methods in quantitative metabolomics.* J Chromatogr A, 2016. **1430**: p. 80-95.

152. Pezzatti, J., et al., *Implementation of liquid chromatography-high resolution mass spectrometry methods for untargeted metabolomic analyses of biological samples: A tutorial.* Anal Chim Acta, 2020. **1105**: p. 28-44.

153. Dunn, W.B., N.J. Bailey, and H.E. Johnson, *Measuring the metabolome: current analytical technologies.* Analyst, 2005. **130**(5): p. 606-25.

154. Beckonert, O., et al., *Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts.* Nat Protoc, 2007. **2**(11): p. 2692-703.

155. Paudyal, S., et al., *Rumination time and monitoring of health disorders during early lactation.* Animal, 2018. **12**(7): p. 1484-1492.

156. Goldansaz, S.A., et al., *Livestock metabolomics and the livestock metabolome: A systematic review.* PLoS One, 2017. **12**(5): p. e0177675.

157. Xu, W., et al., *Relationship between energy balance and metabolic profiles in plasma and milk of dairy cows in early lactation.* J Dairy Sci, 2020. **103**(5): p. 4795-4805.

158. Zhang, G., et al., *A Multi-Platform Metabolomics Approach Identifies Urinary Metabolite Signatures That Differentiate Ketotic From Healthy Dairy Cows.* Front Vet Sci, 2021. **8**: p. 595983.

159. Gabli, Z., et al., *Prevalence of mastitis in dairy goat farms in Eastern Algeria.* Vet World, 2019. **12**(10): p. 1563-1572.

160. Hussein, H.A., et al., *Study on prevalence and bacterial etiology of mastitis, and effects of subclinical mastitis and stage of lactation on SCC in dairy goats in Egypt.* Trop Anim Health Prod, 2020. **52**(6): p. 3091-3097.

161. Persson, Y. and I. Olofsson, *Direct and indirect measurement of somatic cell count as indicator of intramammary infection in dairy goats.* Acta Veterinaria Scandinavica, 2011. **53:15**.

162. He, Y.-L., et al., *Research Progress in Analysis of Small Molecule Metabolites in Bio-matrices by Stable Isotope Coded Derivatization Combining with Liquid Chromatography–tandem Mass Spectrometry.* Chinese Journal of Analytical Chemistry, 2017. **45**(7): p. 1066-1077.

163. Zhao, S., et al., *Metabolomic Coverage of Chemical-Group-Submetabolome Analysis: Group Classification and Four-Channel Chemical Isotope Labeling LC-MS.* Anal Chem, 2019. **91**(18): p. 12108-12115.

164. Khodadoust, S. and M. Ghaedi, *Optimization of dispersive liquid-liquid microextraction with central composite design for preconcentration of chlordiazepoxide drug and its determination by HPLC-UV.* J Sep Sci, 2013. **36**(11): p. 1734-42.

165. E. Johnson, A., Abu-Shumays, and S.R. Abboi, *USE OF FLUORESCENCE DETECTION IN HIGH-PERFORMANCE LIQUID. CHROMATOGRAPHY.* Journd of Chronrarography, 1977. **134**: p. 107-119.

166. Zhao, S., X. Luo, and L. Li, *Chemical Isotope Labeling LC-MS for High Coverage and Quantitative Profiling of the Hydroxyl Submetabolome in Metabolomics.* Anal Chem, 2016. **88**(21): p. 10617-10623.

167. Zhao, S., et al., *Development of High-Performance Chemical Isotope Labeling LC-MS for Profiling the Carbonyl Submetabolome.* Anal Chem, 2017. **89**(12): p. 6758-6765.

168. Lkhagva, A., et al., *Comparative study of five different amine-derivatization methods for metabolite analyses by liquid chromatography-tandem mass spectrometry.* J Chromatogr A, 2020. **1610**: p. 460536.

169. Bawazeer, S., et al., *A method for the analysis of sugars in biological systems using reductive amination in combination with hydrophilic interaction chromatography and high resolution mass spectrometry.* Talanta, 2017. **166**: p. 75-80.

170. Han, W. and L. Li, *Chemical Isotope Labeling LC-MS for Human Blood Metabolome Analysis*, in *Clinical Metabolomics*, M. Giera, Editor. 2018, Springer New York: New York, NY. p. 213-225.

171. Comini, M.A., *Measurement and meaning of cellular thiol:disufhide redox status.* Free Radic Res, 2016. **50**(2): p. 246-71.

172. Zhao, S. and L. Li, *Chemical isotope labeling LC-MS for Metabolomics.* Cancer Metabolomics, Advances in Experimental Medicine and Biology 1280, 2021.

173. Guo, K. and L. Li, *High-Performance Isotope Labeling for Profiling Carboxylic Acid-Containing Metabolites in Biofluids by Mass Spectrometry.* Anal. Chem., 2010. **82**: p. 8789–8793.

174. Fujiwara, T., et al., *Liquid-Chromatographic Methods for Carboxylic Acids in Biological Samples.* Molecules, 2020. **25**(21).

175. Baghdady, Y.Z. and K.A. Schug, *Review of in situ derivatization techniques for enhanced bioanalysis using liquid chromatography with mass spectrometry.* J Sep Sci, 2016. **39**(1): p. 102-14.

176. He, Y., et al., *Profiling of carboxyl-containing metabolites in smokers and non-smokers by stable isotope labeling combined with LC-MS/MS.* Anal Biochem, 2019. **569**: p. 1-9.

177. Zhu, Q.-F., et al., *Analysis of cytochrome P450 metabolites of arachidonic acid by stable isotope probe labeling coupled with ultra high-performance liquid chromatography/mass spectrometry.* Journal of Chromatography A, 2015. **1410**: p. 154-163.

178. Ag, B., *Coupling Reagents.* p. 20.

179. Albericio, F., et al., *Use of Onium Salt-Based Coupling Reagents in Peptide Synthesis.* American Chemical Society, 1998. **63**: p. 9678–9683.

180. Patel, M.K.a.D.V., *"BOP" AS A REAGENT FOR MILD AND EFFICIENT PREPARATION OF ESTERS.* Tefrahedmn Lmen. , 1994. **35**: p. 5643-5606.

181. Carpino, L.A. and A. El-Faham, *Effect of Tertiary Biases on O-Benzotriazolyluronium Salt-Induced Peptide Segment Coupling.* The Journal of Organic Chemistry, 1994. **59**.

182. Castro, B., et al., *Reactif de couplage peptifique IV [II- l'hexafluorophosphate de benzotriazolyl N-oxytrisoimethylaminophosphonium].* Tetrahedron Lett, 1975: p. 1219–1222.

183. Leng, J., et al., *Direct infusion electrospray ionization-ion mobility-mass spectrometry for comparative profiling of fatty acids based on stable isotope labeling.* Anal Chim Acta, 2015. **887**: p. 148-154.

184. Al-Warhi, T.I., H.M.A. Al-Hazimi, and A. El-Faham, *Recent development in peptide coupling reagents.* Journal of Saudi Chemical Society, 2012. **16**(2): p. 97-116.

185. Hudson, D., *Methodological Implications of Simultaneous Solid-Phase Peptide Synthesis. 1. Comparison of Different Coupling Procedures.* J. Org. Chem., 1988. **53**: p. 617-624.

186. Albericio, F., et al., *On the Use of PyAOP, a phosphonium salt derived from HOAt, in Solid-Phase Peptide Synthesis.* Tetrahedron Letters, 1997. **38**: p. 4853-4856.

187. Dourtoglou, V., J.-C. Ziegler, and B. Gross, *L'hexafluorophosphate de 0-benzotriazolyl-N,N-tetramethyluronium : un réactif de couplage peptidique nouveau et efficace.* Tetrahedron Letter, 1978. **15**: p. 1269-1272.

188. Carpino, L.A., *1-Hydroxy-7-azabenzotriazole. An efficient peptide coupling additive.* Journal of the American Chemical Society, 1993. **115**(10): p. 4397-4398.

189. Ehrlich, A., et al., *Cyclization of all - l -Pentapeptides by Means of 1-Hydroxy-7-azabenzotriazole-Derived Uronium and Phosphonium Reagents.* The Journal of Organic Chemistry, 1996. **61**(25): p. 8831-8838.

190. Kamiński, Z.J., et al., *N -Triazinylammonium Tetrafluoroborates. A New Generation of Efficient Coupling Reagents Useful for Peptide Synthesis.* Journal of the American Chemical Society, 2005. **127**(48): p. 16912-16920.

191. Mukaiyama, T., *New Synthetic Reactions Based on the Onium Salts of Aza-Arenes.* Angew. Chem. Inf. Ed. Engl., 1979. **18**: p. 707-721.

192. Sheehan, J.C. and G.P. Hess, *A New Method of Forming Peptide Bonds.* Journal of the American Chemical Society, 1955. **77**(4): p. 1067-1068.

193. Sheehan, J., P. Cruickshank, and G. Boshart, *Notes- A Convenient Synthesis of Water-Soluble Carbodiimides.* 1961. **26**: p. 4.

194. Kiinig, W. and R. Geiger, *Eine neue Methode zur Synthese von Peptiden: Aktivierung der Carboxylgruppe mit Dicyclohexylcarbodiirnid unter Zusatz von 1-Hydrox y-benzotriazolen.* Chcm. Ber. , 1970. **103**: p. 788-798.

195. Carpino, L.A., J. Xia, and A. El-Faham, *3-Hydroxy-4-oxo-3,4-dihydro-5-azabenzo-1,2,3-triazene.* The Journal of Organic Chemistry, 2004. **69**(1): p. 54-61.

196. Subiros-Funosas, R., et al., *Oxyma: an efficient additive for peptide synthesis to replace the benzotriazole-based HOBt and HOAt with a lower risk of explosion.* Chemistry, 2009. **15**(37): p. 9394-403.

197. Awad, H., et al., *Development of a new quantification method for organic acids in urine as potential biomarkers for respiratory illness.* Journal of Chromatography B, 2019. **1122-1123**: p. 29-38.

198. Peng, J. and L. Li, *Liquid–liquid extraction combined with differential isotope dimethylaminophenacyl labeling for improved metabolomic profiling of organic acids.* Analytica Chimica Acta, 2013. **803**: p. 97-105.

199. Willacey, C.C.W., et al., *LC–MS/MS analysis of the central energy and carbon metabolites in biological samples following derivatization by dimethylaminophenacyl bromide.* Journal of Chromatography A, 2019. **1608**: p. 460413.

200. Sentellas, S., O. Nunez, and J. Saurina, *Recent Advances in the Determination of Biogenic Amines in Food Samples by (U)HPLC.* J Agric Food Chem, 2016. **64**(11): p. 7667-7678.

201. Catalan, U., et al., *In vitro Metabolomic Approaches to Investigating the Potential Biological Effects of Phenolic Compounds: An Update.* Genomics Proteomics Bioinformatics, 2017. **15**(4): p. 236-245.

202. Betta, F.D., et al., *Phenolic Compounds Determined by LC-MS/MS and In Vitro Antioxidant Capacity of Brazilian Fruits in Two Edible Ripening Stages.* Plant Foods Hum Nutr, 2018. **73**(4): p. 302-307.

203. Guo, K. and L. Li, *Differential 12C-/13C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome.* Anal. Chem, 2009. **81**: p. 3919-3932.

204. Zheng, J., R.A. Dixon, and L. Li, *Development of Isotope Labeling LC–MS for Human Salivary Metabolomics and Application to Profiling Metabolome Changes Associated with Mild Cognitive Impairment.* Analytical Chemistry, 2012. **84**(24): p. 10802-10811.

205. Su, X., et al., *Dansylation isotope labeling liquid chromatography mass spectrometry for parallel profiling of human urinary and fecal submetabolomes.* Analytica Chimica Acta, 2016. **903**: p. 100-109.

206. Xu, W., et al., *Development of High-Performance Chemical Isotope Labeling LC–MS for Profiling the Human Fecal Metabolome.* Analytical Chemistry, 2015. **87**(2): p. 829-836.

207. Zhou, R., K. Guo, and L. Li, *5-Diethylamino-naphthalene-1-sulfonyl Chloride (DensCl): A Novel Triplex Isotope Labeling Reagent for Quantitative Metabolome Analysis by Liquid Chromatography Mass Spectrometry.* Analytical Chemistry, 2013. **85**(23): p. 11532-11539.

208. Song, C., et al., *Accurate Determination of Amino Acids in Serum Samples by Liquid Chromatography–Tandem Mass Spectrometry Using a Stable Isotope Labeling Strategy.* Journal of Chromatographic Science, 2015. **53**(9): p. 1536-1541.

209. Peng, J., et al., *Development of Isotope Labeling Liquid Chromatography Mass Spectrometry for Mouse Urine Metabolomics: Quantitative Metabolomic Study of Transgenic Mice Related to Alzheimer's Disease.* Journal of Proteome Research, 2014. **13**(10): p. 4457-4469.

210. Zhendong, L., T. Jaspaul, and L. Liang, *Nanoflow LC–MS for High-Performance Chemical Isotope Labeling Quantitative Metabolomics.* Analytical Chemistry, 2015. **87**(22): p. 11468-11474.

211. Sommer, U., T. Koal, and A. Peck, *Standardized Targeted Metabolomics Using the BIOCRATES MxP Quant 500 Kit on the ACQUITY UPLC I-Class PLUS and Xevo TQ-XS Mass Spectrometer.* Waters corporation, 2020.

212. Sigrist, H. and P. Zahler, *Characterization of phenylisothiocyanate as a hydrophobic membrane label.* FEBS Letters, 1978. **95**(1): p. 116-120.

213. Lavi, L.E., et al., *Sensitive analysis of asparagine and glutamine in physiological fluids and cells by precolumn derivatization with phenylisothiocyanate and reversed-phase high-performance liquid chromatography.* Journal of Chromatography B: Biomedical Sciences and Applications, 1986. **377**: p. 155-163.

214. Battaglia, A., et al., *High-performance liquid chromatographic analysis of physiological amino acids in human brain tumors by pre-column derivatization with phenylisothiocyanate.* Journal of Chromatography B, 1999. **730**: p. 81-93.

215. Fekkes, D., *State-of-the-art of high-performance liquid chromatographic analysis of amino acids in physiological samples.* Journal of Chromatography B, 1996. **682**: p. 3 22.

216. Hampel, D., et al., *Validation and Application of Biocrates AbsoluteIDQ® p180 Targeted Metabolomics Kit Using Human Milk.* 2019: p. 10.

217. Liem-Nguyen, V., et al., *Determination of picomolar concentrations of thiol compounds in natural waters and biological samples by tandem mass spectrometry with online preconcentration and isotope-labeling derivatization.* Anal Chim Acta, 2019. **1067**: p. 71-78.

218. Huang, Y.-Q., et al., *Use of isotope differential derivatization for simultaneous determination of thiols and oxidized thiols by liquid chromatography tandem mass spectrometry.* Analytical Biochemistry, 2011. **416**(2): p. 159-166.

219. Guo, N., et al., *Profiling of carbonyl compounds in serum by stable isotope labeling - Double precursor ion scan - Mass spectrometry analysis.* Anal Chim Acta, 2017. **967**: p. 42-51.

220. Yu, L., et al., *Profiling of aldehyde-containing compounds by stable isotope labelling-assisted mass spectrometry analysis.* Analyst, 2015. **140**(15): p. 5276-86.

221. Kishikawa, N., M.H. El-Maghrabey, and N. Kuroda, *Chromatographic methods and sample pretreatment techniques for aldehydes determination in biological, food, and environmental samples.* J Pharm Biomed Anal, 2019. **175**: p. 112782.

222. Fathy Bakr Ali, M., et al., *Chromatographic determination of aliphatic aldehydes in human serum after pre-column derivatization using 2,2'-furil, a novel fluorogenic reagent.* J Chromatogr A, 2013. **1300**: p. 199-203.

223. Jia, S., et al., *Chemical Isotope Labeling Exposome (CIL-EXPOSOME): One High-Throughput Platform for Human Urinary Global Exposome Characterization.* Environ Sci Technol, 2019. **53**(9): p. 5445-5453.

224. Willacey, C.C.W., et al., *Metabolic profiling of material-limited cell samples by dimethylaminophenacyl bromide derivatization with UPLC-MS/MS analysis.* Microchemical Journal, 2020. **159**.

225. Xian, L., et al., *High-Performance Chemical Isotope Labeling Liquid Chromatography–Mass Spectrometry for Profiling the Metabolomic Reprogramming Elicited by Ammonium Limitation in Yeast.* Journal of Proteome Research, 2016. **15**(5): p. 1602-1612.

226. Plewa, S., et al., *LC-MS/MS based targeted metabolomics method for analysis of serum and cerebrospinal fluid.* Journal of Medical Science, 2019. **88**(1): p. 12-20.

227. Jeffrey, J.A., et al., *Synthesis and Application of 13C Enriched phenylisothiocyanate as a 13C NMR protein probe.* Biochemical and Biophysical research communications, 1987. **142**: p. 1064-1071.

Cécile MARTIAS

Phénotypage métabolique multi-matrices biologiques et multi-plateformes analytiques

Résumé

L'étude de petites molécules permet une meilleure compréhension des systèmes biologiques. Leur détection et quantification permettent de réaliser un diagnostic (précoce ou prédictif) et de déterminer une pathologie notamment chez l'Homme et l'animal. De nos jours, il n'existe pas de consensus quant à la détection de ces molécules. L'objectif de ce projet est réaliser un développement méthodologique standardisé afin de détecter et quantifier le plus grand nombre de petites molécules chez l'humain et le caprin. La première étude a permis la visualisation de ces molécules sous forme de carte, chez l'Homme, et d'atlas chez le caprin. Pour se faire, les échantillons les plus communément utilisés sont sélectionnés, sang, urine, salive et fèces chez l'Homme et sang, lait et fèces chez le caprin. Le deuxième objectif consiste au développement du marquage de ces petites molécules afin de les visualiser et de les quantifier de manière robuste.

Mots clés : Métabolomique ; échantillon biologique ; marquage des fonctions chimiques

Abstract

The studies of small molecules allow a better understanding of the biological systems. The process to detect and quantify them support the achievement of the diagnosis (early or predictive). It also allows to determine a pathology, particularly on humans and animals. For the time being, there is no consensus on the detection of these molecules. The objective of this project is to achieve the development of a standardized methodological approach, to detect and quantify the largest number of small molecules in humans and in goats samples. The first study performed permitted us to visualize these molecules in the form of a map for the humans and in the form of an atlas for the goats. To achieve this, the most commonly selected samples were blood, urine, saliva and faeces for humans. For goats, we selected blood, milk and faeces. The second objective is to develop the labelling of these small molecules in order to visualize and quantify them in a robust way.

Keywords : Metabolomic ; biological sample; chemical derivatization